CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 126, No. 6 December 2009

JASA EXPRESS LETTERS

LETTERS TO THE EDITOR

GENERAL LINEAR ACOUSTICS [20]

NONLINEAR ACOUSTICS [25]

(Continued)

CONTENTS—*Continued from preceding page*

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

UNDERWATER SOUND [30]

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

TRANSDUCTION [38]

NOISE: ITS EFFECTS AND CONTROL [50]

ARCHITECTURAL ACOUSTICS [55]

(Continued)

CONTENTS—*Continued from preceding page*

ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION [58]

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

PSYCHOLOGICAL ACOUSTICS [66]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Exploiting forward scattering for detecting submerged proud/half-buried unexploded ordnance

J. A. Bucaro[a)]
Excet, Inc., Springfield, Virginia 22151
bucaro@pa.nrl.navy.mil

B. H. Houston, H. Simpson, and L. R. Dragonette
The Naval Research Laboratory, Washington, DC 20375
brian.houston@nrl.navy.mil, harry.simpson@nrl.navy.mil, dragonet@pa.nrl.navy.mil

L. Kraus and T. Yoder
Global Strategies Group (North America), Crofton, Maryland 20815
larry.kraus.ctr@nrl.navy.mil, yoder@pa.nrl.navy.mil

Abstract: Laboratory underwater bistatic scattering measurements are reported for free, proud, and half-buried unexploded ordnances for 0° and 90° source angles. Forward echoes are larger than backscattered returns, and half burial significantly decreases the latter but not the former. Results agree with analytic predictions borrowed from radar. The forward echo and source signal are separated by measurements made with and without the target, a method not possible in a target search. For this, a method is described that uses knowledge of the source location and the hyperbolic character in time-cross range of the signals received at points along a line.

PACS numbers: 43.20.Fn, 43.30.Vh [AN]
Date Received: July 7, 2009 **Date Accepted:** September 29, 2009

1. Introduction

Interest in acoustic techniques to detect and classify underwater unexploded ordnance (UXO) on or in the sediment has grown steadily as the danger posed to the public by such objects has become more apparent. Generally, experimental studies of scattering from proud or buried targets[1–4] have focused on monostatic scattering wherein the source and receiver are co-located. This letter presents results related to bistatic scattering and, in particular, its forward scattered component. We would like to explore whether forward scattering provides some advantage relative to backscattering especially regarding partially buried targets probed below the critical angle of the sediment/water interface.

2. Forward scattering characteristics

Many UXOs are more or less cylindrical in shape including the particular target whose bistatic target strength (TS) we measure here. TS is defined in the usual sense[5] as 10 $\log_{10}$ of the ratio of the intensity returned by the target in some direction, at a distance of 1 m from its acoustic center, to the incident intensity from a distant source (plane wave). Apart from elastic effects, we can estimate what one would expect for the forward scattered TS of a finite cylinder using the analysis of Ross[6] as applied to the radar cross-section of a perfectly conducting cylinder. As discussed by Bowman[7] for an electromagnetic wave incident normally on an infinite conducting cylinder, the problem reduces to the solution of a scalar wave equation for either horizontal or vertical polarization (parallel or perpendicular to the cylinder axis) of the same form as that for the velocity potential of an acoustic

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

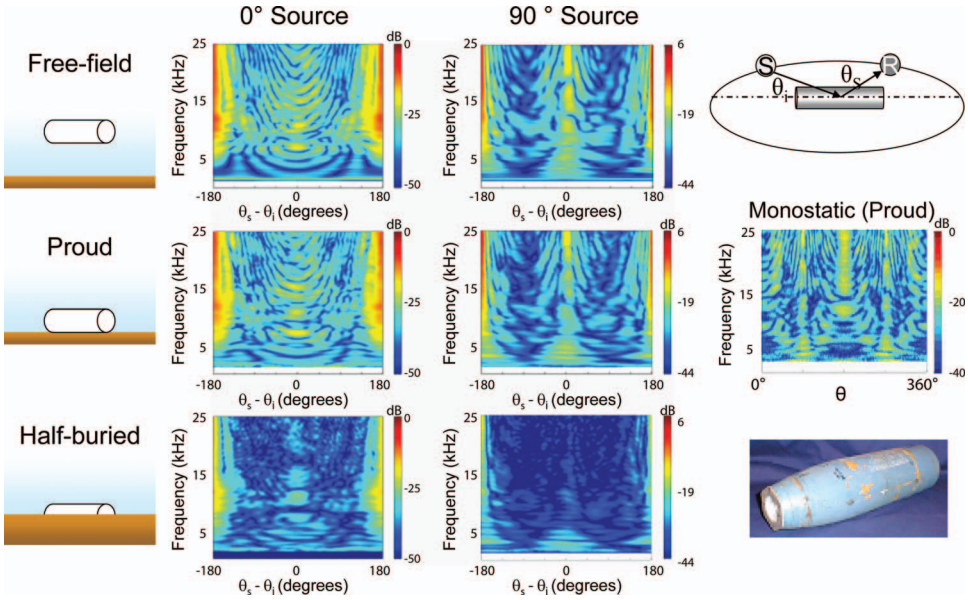


Fig. 1. Bistatic TS displayed in dB as a color map measured for the 5 in. rocket versus frequency (2–25 kHz) and the source angle minus receiver angle. 0° and 90° source angles are on the left and right, respectively. Free-field (upper row), proud (middle row), and half-buried (lower row). Monostatic measurements (Ref. 2) for the proud target (middle right).

wave with soft or rigid boundary conditions, respectively. For this case, in the short wavelength limit, the forward scattered TS depends in large part on the target's projected area with a correction term that takes into account the actual surfaces illuminated by the sound wave.

For the beam incidence (90°) case, Ross'[6] expression for forward scattering with wavenumber k gives

$$\mathrm{TS} = 10 \log \left| \sqrt{\sigma^{\mathrm{PO}}(90°)}\, W \exp\left(\frac{j\pi}{2}\right) + \sqrt{\sigma^{\mathrm{side}}} \exp\left(\frac{j3\pi}{4}\right) \right|^2, \tag{1}$$

where $\sigma^{\mathrm{PO}}(90°) = 4(aL/\lambda)^2$, $W = W_{hh} \sim 1 + 0.498(ka)^{-2/3} - 0.011(ka)^{-4/3}$ with $\sigma^{\mathrm{side}} = 0$ for horizontal polarization, $W = W_{vv} \sim 1 - 0.432(ka)^{-2/3} - 0.214(ka)^{-4/3}$ with $\sigma^{\mathrm{side}} = 7/\pi(a^3/\lambda)$ for vertical polarization, λ is the acoustic wavelength, $k = 2\pi/\lambda$, and L and a are the cylinder length and radius. The unity term in the expression for W is the physical optics result while the terms in inverse powers of ka account for the effect of the curvature near the shadow boundary. The last term gives the scattering from the ends.

For 0° incidence, Ross' result gives

$$\mathrm{TS} = 10 \log \left| [\sigma^{\mathrm{PO}}(0°)]^{1/2} \exp(j\pi/2) + [\sigma^{\mathrm{CS}}(0°)]^{1/2} \exp(j3\pi/4) \right|^2, \tag{2}$$

where $\sigma^{\mathrm{PO}}(0°) = \pi^2(a^2/\lambda)^2$ and $\sigma^{\mathrm{CS}}(0°) = 0.3(\pi aL)(a/\lambda)$. The first term is the physical optics result and the second term takes into account the contribution from the long, curved surface of the cylinder.

3. Experimental measurements and the case for detection of buried targets by forward scattering

We have carried out bistatic acoustic scattering measurements on ordnance whose monostatic free-field target scattering has already been reported.[2] The particular target studied here is a 5 in. rocket warhead filled with a polymer material (see photo in Fig. 1). Measurements were carried out in both free-field and sediment pool facilities with the target placed proud of, and

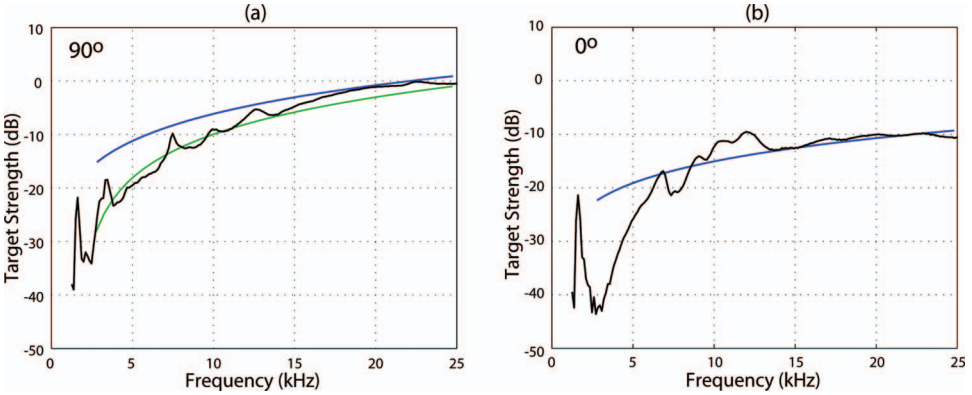


Fig. 2. Measured free-field forward scattered TS versus frequency in black. (a) 90° source; blue/green soft/rigid boundary condition theory, respectively; (b) 0° source; blue theory.

then half buried in, the sediment. In the sediment facility, the source and receiver were 10 cm above the sediment surface; and for source-to-target and receiver-to-target distances used here (2.7 and 2.0 m), the sound strikes the sediment at an angle much smaller than the critical angle ($\theta_C \sim 27°$).

The bistatic measurement process we employed is similar to that reported in Ref. 2 for the monostatic case, which used a near-field cylindrical source array and a small "point' receiver rotating about the target. The data were collected and processed to recover complex scattering cross-section expressible as TS. The incident acoustic pressure in the form of a broadband pulse, the pool clutter (background) pressure, and the scattered pressure were measured in the following way. First, before positioning the target, the source was excited and the incident pressure measured at the location corresponding to the target center. Second the source was excited and the non-target pressure field measured as a function of θ at each receiver position to be used in the scattering experiment. This measurement contains scattering from pool clutter and, in the forward scattering plane, the incident pressure field. Lastly, the target was inserted and the scattered pressure field as a function of scattering angle θ_s was measured.

The TS was obtained by subtracting measurements made with and without the target. This process removes energy associated with the incident wave for bistatic angles in the forward plane as well as from spurious reflections from the finite-sized pool. With the non-target data file removed from the scattered signal, the range-normalized and incident pressure-normalized parameter $X(f, \theta)$ is formed in terms of the scattered signal $P_{\text{scat}}(f, \theta)$ and the incident field measured at the target center $P_{\text{inc}}(f)$ as follows:

$$X(f,\theta) = \frac{P_{\text{scat}}(f,\theta)}{P_{\text{inc}}(f)} \frac{R_{\text{TR}}}{e^{ikR_{\text{TR}}}}, \tag{3}$$

where f is frequency and R_{TR} is the distance from the target center to the receiver. Here the R dependent factor removes the effects of range from the pressure ratio term. The scattering data were measured at a range (2 m) that is in the near-field for some target aspects and in the far-field for others. Since bistatic data can be readily projected to the far-field, we performed this projection on all the echo measurements. TS values are then defined and displayed as $10 \log_{10}(|X(f, \theta)|^2)$. Measured bistatic TS versus frequency and $\Delta\theta = \theta_S - \theta_i$ is shown in Fig. 1 for $\theta_S = 0°$ and 90° for the free field, proud, and half-buried cases. Also shown for comparison is the monostatic TS reported by Bucaro *et al.*[2]

We show in Fig. 2(a) the result of Eq. (1) compared to the measured free-field data for the beam incidence (90°) case taking the rocket as a circular cylinder of length 18 in. and radius 2.5 in. As can be seen, the two agree well over almost the entire frequency band. Overall, the data are closer to the rigid result than to the free case, which is not surprising given that the

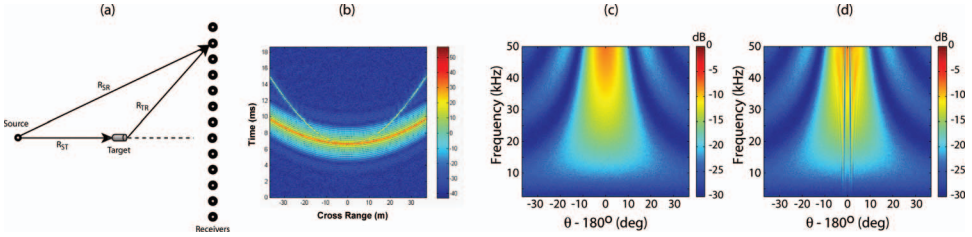


Fig. 3. (a) Hypothetical long range forward scattering measurement range: R_{ST}, R_{SR}, and R_{TR} are the source-to-target distance, the distance from the source to a particular receiver, and the distance from the target to a particular receiver, respectively; (b) time-cross range plots in dB for the direct source 150 m from the center receiver (heavy line) and for the forward scattered signal ×30 with TS given by Eq. (4); (c) scattering TS in dB versus frequency and angle in the forward scattered sector for the scattering response given by Eq. (4); and (d) that extracted from the numerical results shown in (b).

target is a relatively thick-walled structure (the cylindrical wall is ~0.5 in. steel). We note that the sharp peak at the lowest frequencies (which is even more pronounced in the 0° case) is due to a resonance of the air-filled back compartment (the end wall thickness is ~0.1 in.) as confirmed by finite element calculations we carried out indicating that at these low frequencies the structure is not totally rigid. As shown in Fig. 2(b), for end-on incidence (0°), the agreement is good over most of the band, but Eq. (2) over-predicts TS at the lower frequencies. The analytic result has the same frequency dependence as that computed numerically using a *T*-matrix approach[8] for a rigid, hemispherically capped cylinder of about the same length for equivalent frequencies up to about 8 kHz. The latter has a TS level of about 4 dB higher, and this is presumably due to additional scattering from the curved versus flat ends. Why Eq. (2) is higher than the measured result at low frequencies is unclear.

Consideration of these results leads to the following observations. (1) The forward scattered echo has TS levels that are significantly higher than any other return, (2) half burial of the target significantly reduces the backscattered TS levels but not the forward scattered levels, and (3) the leading forward scattering terms for both 0° and 90° incidences are $\sigma^{PO}(90°) = 4(aL/\lambda)^2$ and $\sigma^{PO}(0°) = \pi^2(a^2/\lambda)^2$, which are just the projected areas. We therefore expect that the forward scattered TS for other aspects or other targets would be of order of the projected areas with a 6 dB per octave increase with frequency.

We conclude that for the $\theta_i < \theta_C$ case, half-buried targets would be difficult to detect in backscattering. The question then becomes as follows: In order to exploit the higher TS levels for the forward scattered echo upon burial, is it possible to extract the forward scattered echo from the strong overlapping incident field *without* having to remove the target as is done in the laboratory measurement?

4. Practical extraction of forward scattered signal

In the laboratory measurements presented here, we have been able to obtain an accurate measure of the forward scattered TS versus frequency and angle by precise mapping of the incident field (which can be one or more orders of magnitude larger than the scattered pressure) at the receiver locations *prior to target insertion*. This is of course not possible in an actual search in the environment for proud and buried targets. Field approaches, which attempt to extract the incident field, include, for example, mode filtering in a water channel[9] and apex shifted Radon transforms[10] as applied to ground penetrating radar. The former requires long vertical arrays that are not practical for our application. We illustrate here a source estimation technique related to the latter,[10] which uses knowledge of the source location and directivity.

Figure 3(a) describes a possible source-receiver-target geometry for long range (below critical angle) detection of proud or buried UXO. These receiver positions might be realized, for example, using an autonomous underwater vehicle (AUV) moving in a straight line. The location of the source relative to the *i*th receiver [and therefore the corresponding distance $R_{SR}(i)$ to the

*i*th receiver] is assumed known *a priori* apart from random fluctuations but not R_{ST} or $R_{TR}(i)$, the source-to-target and target to *i*th receiver distances. In Fig. 3(b) we show the time versus cross range plots calculated for signals at the various receiver positions due to a broadband (2–50 kHz) point source (monopole) at $R_{SR}=150$ m from the closest receiver together with a forward scattered signal (×30) from an "unknown" target position (100 m from the source). For the latter we use a TS given by $10 \log \sigma^{PO}(0°)=10 \log \pi^2(a^2/\lambda)^2$ [the leading term in Eq. (2) for the 0° source case] since 0° is the aspect with the lowest forward scattered level, and we take a that for the 5 in. rocket. The scattering angle dependence (the forward lobe structure) associated with the physical optics area term above is assumed to be that for the far-field diffraction from an aperture of radius a leading to a TS (k, θ) given by

$$\mathrm{TS}(k,\theta) = 10 \log[a^2/4(ka)^2(\sin^2(ka \sin \theta))/(ka \sin \theta)^2], \quad (4)$$

where the frequency can be made explicit using $k=2\pi f/C$ with C the sound speed. We note that the associated scattered pressure we used has an $\exp(j\pi/2)$ phase term [see Eq. (2)]. We have also added −30 dB of random Gaussian noise and a random phase term $\exp(j2\pi fC^{-1}\Delta R(i))$ to the *i*th receiver signal with $0 \le 2\pi fC^{-1}\Delta R(i) \le 4\times10^{-4}\times f$ to simulate random deviations (up to 0.1 m) of the AUV path from a straight line.

As can be seen in Fig. 3(b), the source and target scattering signals map into two hyperbolas weakly modulated by the random phase term and noise since both free space Green's functions are of the form $R^{-1}\exp(j2\pi fC^{-1}R)$. Our proposed technique is as follows. (1) There is an angular region (here beyond $\sim\pm20°$ from forward) with no overlap of source and echo time signals, and here the forward scattered component is directly accessible. (2) In the overlap region, we would like to subtract the "known" source signal at each receiver from the total signal using the appropriate Green's function leaving the desired scattered pressure. However, the source signal is not known precisely due to the unknown random deviations of the AUV [the $\Delta R(i)$] from a straight line. We can, however, estimate each receiver position deviation by finding the local minimum for the difference between the received signal (the sum of the source, scattered, and noise signals) and an estimate of the source signal based on the known Green's function. Good estimates of the receiver deviations are so obtained because the source signal is much stronger than the other components of the received signal.

Carrying out this procedure on the data in Fig. 3(b) produces the TS frequency/angle map shown in Fig. 3(d) compared to that given directly by Eq. (4) in Fig. 3(c), both with the added noise. Apart from the obvious artifact (the vertical line structure), the process recovers TS (f, θ). The line structure corresponds to loci of receiver positions and frequencies where the path difference $R_{ST}+R_{TR}-R_{SR}=\pi/2$, which when added to the $\pi/2$ in the TS phase term [see Eq. (2)] leaves the echo with no quadrature component, i.e., π out of phase with the source signal. As a consequence, our simple fitting procedure for determining the ΔR incorrectly eliminates the scattered signal at these particular (f, θ) points. Although not shown here, the extracted results and Eq. (4) at exactly 180° agree very well (±0.3 dB) over the entire band.

5. Concluding remarks

Bistatic scattering measurements at vertical angles well below the critical angle made on a 5 in. diameter UXO in the free-field, proud, and half buried in sediment indicate that for these conditions exploitation of forward scattering may provide a detectable signal with levels higher than that for backscattering and one that persists under partial burial. We illustrate for a point source in an infinite medium a technique that might be employed in a relatively simple environment to extract the forward scattered component from the much stronger incident field. Work is underway to develop more sophisticated approaches and to demonstrate them in more complex environments.

Acknowledgment

This work was supported by the SERDP Program and ONR.

References and links

[1]H. J. Simpson, B. H. Houston, and R. Lim, "Laboratory measurements of sound scattering from a buried sphere above and below the critical angle," J. Acoust. Soc. Am. **113**, 39–42 (2003).
[2]J. A. Bucaro, B. H. Houston, M. Saniga, L. R. Dragonette, T. Yoder, S. Dey, L. Kraus, and L. Carin, "Broadband acoustic scattering measurements of underwater unexploded ordnance (UXO)," J. Acoust. Soc. Am. **123**, 738–746 (2008).
[3]J.L. Lopes, C.L. Nesbit, R. Lim, D. Tang, K.L. Williams, and E.I. Thorsos, "Shallow grazing angle sonar detection of targets buried under a rippled sand interface," Proceedings of the Oceans 2002, Biloxi, MS, 28 October–1 November 2002.
[4]A. Tesei, J. A. Fawcett, and R. Lim, "Physics-based detection of man-made elastic objects buried in high-density-clutter areas of saturated sediments," Appl. Acoust. **69**, 422–437 (2008).
[5]R. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983), pp. 291–327.
[6]R. A. Ross, "Forward scattering from a finite, circular cylinder," Progress in Electromagnetics Research C, **2**, 207–215 (2008).
[7]J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (Hemisphere, New York, 1987), pp. 89–91.
[8]A. Sarkissian, C. F. Gaumond, and L. R. Dragonette, "T-matrix implementation of forward scattering from rigid structures," J. Acoust. Soc. Am. **94**, 3448–3453 (1993).
[9]A. Sarkissian, "Extraction of a target scattering response from measurements made over long ranges in shallow water," J. Acoust. Soc. Am. **102**, 825–832 (1997).
[10]F. Yong, Z. Zheng-ou, and X. Jia-li, "Clutter reduction based on apex shifted radon transform in sub-surface forward-looking ground penetrating radar," *International Conference on Radar CIE'06* (INSPEC, 2006), pp. 1–3, Paper No. 9431360.

Masking with interaurally "double-delayed" stimuli: The range of internal delays in the human brain

Torsten Marquardt and David McAlpine
UCL Ear Institute, 332 Gray's Inn Road, London WC1X 8EE, United Kingdom
t.marquardt@ucl.ac.uk, d.mcalpine@ucl.ac.uk

Abstract: Is binaural processing in humans different to that of other mammals? While psychophysical data suggest that the range of internal delays necessary for processing interaural time differences is at least ±3 ms, physiological data from small mammals indicate a more limited range. This study demonstrates that binaural detection is impeded by reduced interaural coherence in auditory channels remote from the signal frequency, in accordance with the wider critical bandwidths reported for binaural processing. This explains previous psychophysical data without requiring long internal delays. The current psychophysical data support the view that human binaural processing is similar to that of other mammals.

PACS numbers: 43.66.Pn, 43.66.Dc, 43.66.Ba [QJF]
Date Received: August 12, 2009 **Date Accepted:** September 29, 2009

1. Introduction

To account for human sensitivity to interaural time differences (ITDs), Jeffress[1] envisaged an array of binaural coincidence detectors, each maximally activated only when the "external" interaural delay was compensated by an equal and opposite "internal" interaural delay (arising from a difference in the axonal conduction delay from each ear). Although physiological investigations confirm the existence of neurons that respond as such binaural coincidence detectors, the range of the compensatory internal delays is open to question. In particular, electrophysiological recordings in a range of mammalian species suggest the existence of a "π-limit," with no binaural coincidence detectors tuned to ITDs longer than half a period (relative to the center frequency of the auditory channel under consideration).[2] This has been difficult to reconcile with psychophysical studies purporting to show evidence of internal delays of at least several times this magnitude. Consequently, the extent to which the π-limit might be extrapolated to the human brain is unclear. van der Heijden and Trahiotis[3] addressed the question of the existence of long internal delays in humans, by employing a binaural detection paradigm in which they measured the detection threshold for a diotic tone that was masked by a "double-delayed" noise (DDN). DDN is the sum of two independent binaural noises with ITD of equal magnitude but opposite sign. In contrast, a "single-delayed" noise (SDN) is generated by only one such noise source with either positive or negative ITD—a masker that is routinely employed in measurements of binaural-masking-level-difference (BMLD). Figures 1(a) and 1(b) show cross-correlograms of SDN and DDN stimuli. Since DDN is the sum of two SDN sources, the normalized cross-correlogram for the DDN is the average of those of its two SDN components. This is illustrated in Fig. 1(c) for the frequency channel containing the target tone (500 Hz). The data of van der Heijden and Trahiotis[3] [a sub-set of which is reproduced here in Fig. 1(d)] showed that thresholds for SDN maskers were lower than for DDN maskers. Their model simulations suggested that internal delays as large as 3 ms must exist, and are employed under the SDN masker conditions to profit from the deeper modulation in this region of the cross-correlation function [Fig. 1(c)].

The lack of physiological evidence for such large internal-delay magnitudes calls for alternative explanations for van der Heijden and Trahiotis' data.[3] For the 500-Hz target em-

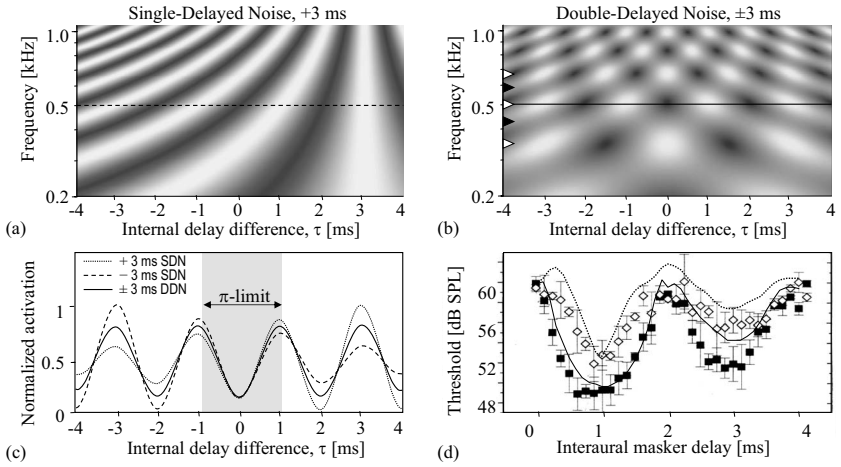


Fig. 1. Rescaled cross-correlograms for SDN (a) and DDN (b) maskers with ITD of 3 ms magnitude. Their cross-correlation values are computed using Eq. (1) (here with $\tau = \pi\{-4\ \text{ms}, \ldots, 4\ \text{ms}\}$), and are rescaled to express normalized neural activation: The activity of a neural coincidence detector is one (white), when the signals arriving from the left and right ears are fully correlated, and zero (black), when the arriving signals have a correlation of −1. The frequency channels of the target tone (500 Hz) are indicated by horizontal lines. Solid triangles and open triangles in (b) mark examples of frequency channels of zero and high IACs, respectively. (c) Normalized activation within the target-tone channel to SDN and DDN. The physiologically observed range of internal delays is shown gray (π-limit). (d) Detection thresholds for a diotic 500-Hz tone with SDN (solid squares) and DDN (open diamonds) maskers obtained by van der Heijden and Trahiotis (Ref. 3). Thresholds predicted by our across-frequency convergence model are plotted as solid (SDN) and dotted (DDN) lines.

ployed, the π-limit is ±1 ms and the regions of deeper modulation that presumably lowered thresholds for the SDN maskers fall outside this range of internal delays [gray area in Fig. 1(c)]. However, while the π-limited cross-correlation functions to DDN and SDN are almost identical *within* the target-frequency channel, large differences are apparent *across* frequency bands [Figs. 1(a) and 1(b)]. For SDN, the cross-correlogram is strongly modulated over the whole frequency range, indicating high interaural coherence (IAC) (the maximum cross-correlation within a frequency channel) across frequency. For DDN, however, the modulation is only pronounced in certain frequency bands. Frequency bands alternate between high IAC (e.g., at 333, 500, and 667 Hz) and zero IAC (e.g., at 417 and 583 Hz). Here, we demonstrate that these differences in the modulation of off-frequency channels provide an alternative explanation for the observed threshold differences between the two masking conditions; an explanation that does not require long, physiologically unrealistic, internal delays.

2. Psychophysical experiment

To investigate the influence of IAC in off-frequency channels on binaural detection, we measured masked tone thresholds in either DDN or SDN, each flanked by noise bands of different interaural configurations (similar to the paradigm applied by Sondhi and Guttman[4]). The addition of the flanking bands modifies the IAC in channels tuned to frequencies around the transitions between inner and flanking bands. We predict that detection of the target tone will be impaired by these flanking bands in a manner that depends on the IAC at these transitions and the proximity of these transitions to the target-frequency.

The interaural timing configurations of the four masker types employed in our experiment are illustrated in Fig. 2(a). The three frequency bands (one inner band and two flanking bands) had equal power density (45.5 dB sound pressure spectrum level), and spanned the total range 50–950 Hz. The inner band, consisting of either SDN or DDN with an ITD magnitude of 1 ms (0.5 cycles at 500 Hz), was positioned symmetrically around the frequency of the target tone (500 Hz). The upper and lower flanking bands were unilaterally phase delayed such that

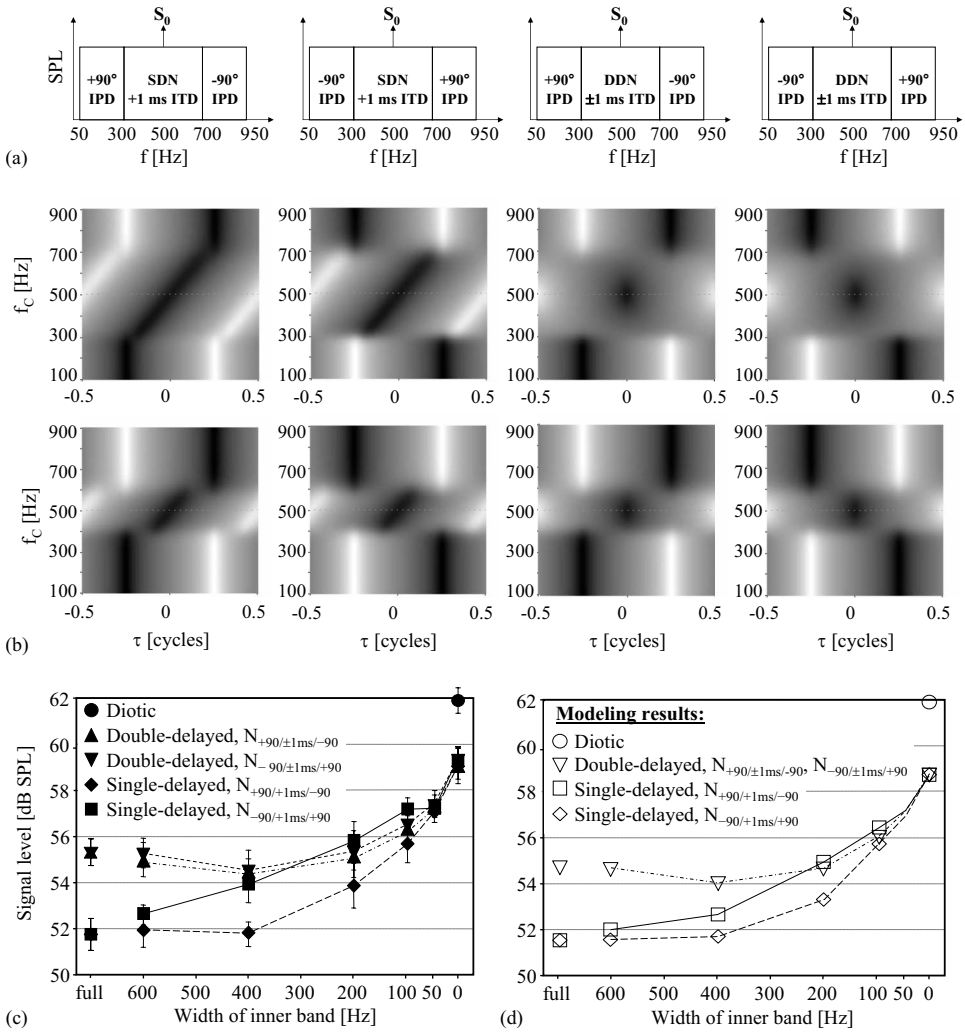


Fig. 2. (a) Long-term average spectra of the four masker configurations: $N_{+90°/+1\ ms/-90°}$, $N_{-90°/+1\ ms/+90°}$, $N_{+90°/\pm1\ ms/-90°}$, and $N_{-90°/\pm1\ ms/+90°}$ (left to right column). Labels inside the sub-bands denote their interaural timing configurations. Below each spectrum, in panel (b), are rescaled cross-correlograms of each of these maskers with 400-Hz (upper row) and 200-Hz (lower row) inner bandwidths. In accordance with the π-limit, the range of internal-delay differences (τ) is limited to ±0.5 cycles and expressed in phase. The rescaled cross-correlation values are computed using Eq. (1) (for rescaling, see caption of Fig. 1). (c) Average detection threshold as a function of the width of the inner band for the four masker types (see legend). Error bars indicate across-listeners standard errors. (d) Thresholds predicted by our across-frequency convergence model.

each band had an interaural *phase* delay (IPD) of 90°, but of opposite sign to the other. This produced a total of four masker configurations: $N_{+90°/+1\ ms/-90°}$, $N_{-90°/+1\ ms/+90°}$, $N_{+90°/\pm1\ ms/-90°}$, and $N_{-90°/\pm1\ ms/+90°}$. The masker noises were generated in the frequency domain from spectral components with 1-Hz-spacing and Rayleigh-distributed amplitude. The phases of the spectral components of the left ear stimulus were uniformly distributed, and the right ear phases were assigned relative to the left ear phases to produce the spectrum of ITDs illustrated in Fig. 2(a). For maskers containing DDN, two independent SDNs of half power were added: one with +1 ms ITD and one with −1 ms ITD. After inverse Fourier transformation, the 1-s long masker was truncated to 300 ms (including 10-ms cosine ramps).

As in the study by Sondhi and Guttman,[4] pure-tone thresholds were measured as a function of the width of the inner masker band (0, 50, 100, 200, 400, 600, and 900 Hz). A two-interval, two-alternative forced-choice task was employed with an inter-stimulus interval of 300 ms. For each stimulus interval, a new masker noise was generated. A diotic target tone (500 Hz, 280-ms long including 10-ms cosine ramps) was computed in the time domain and added (temporally centered) to the masker in one of the two intervals (chosen randomly on each trial with equal probability). This trial sequence was then converted to electric analog signals by a 16-bit soundcard (22 050 Hz sampling rate), and presented acoustically via earphones (Beyerdynamic DT 48A with supra-aural cushions). The subject was directed to identify the interval that contained the target tone. Correct-answer feedback was given after each response. The target-tone level, initially set well above the expected thresholds, was varied adaptively using a "3down-1up" procedure to estimate the 79.4% correct threshold. The initial step size of the adaptive track was 4 dB, and was reduced to 2 dB and then to 1 dB following two reversals at each of the former step sizes. A track was terminated after 12 reversals using 1-dB steps, and the threshold for that track calculated as the average target-tone level at the last ten reversals.

Four subjects (aged 20–33; trained until threshold stabilized) were tested under all masker conditions plus one diotic condition. Each subject was tested four times, and each time all conditions were tested in a different random order. The median formed the subject's threshold. Figure 2(c) shows the average thresholds across all subjects. For all four masker types, the thresholds increase as the flanking bands close in on the target-frequency. Under the $N_{-90°/+1\ ms/+90°}$ masker conditions, the threshold clearly increased when the flanking bands were still 200 Hz from the target (i.e., with inner bandwidth of 400 Hz). In contrast, for the DDN and the $N_{+90°/+1\ ms/-90°}$ conditions, thresholds did not increase until the flanking bands were approximately 100 Hz from the target. Although SDN maskers with large inner bandwidths ($>$400 Hz) produced lower thresholds than the DDN maskers, this advantage was not so clear for maskers with lower inner bandwidths, especially with the SDN masker, $N_{-90°/+1\ ms/+90°}$.

The data can be understood by examining the cross-correlograms in Fig. 2(b). Within frequency channels, where spectral components from neighboring bands with different interaural timings merge, the IAC and consequently the modulation in activation (along the τ-dimension) are reduced. In the case of the $N_{+90°/+1\ ms/-90}$ masker, the IPD difference between bands at the transition remained small such that, for an inner bandwidth of 400 Hz, IAC in the transition channels was hardly reduced and detection threshold remained low. In contrast, for the $N_{-90°/+1\ ms/+90°}$ configuration with 400-Hz inner bandwidth, the difference in IPD was almost 180°, leading to near zero IAC at both transitions, and presumably causing the substantial threshold increase. Thus, reduced IAC appears detrimental to binaural detection, even outside the target's frequency channel (which has a critical bandwidth of just 78 Hz).[5]

Of particular interest are the DDN maskers $N_{+90°/\pm1\ ms/-90°}$ and $N_{-90°/\pm1\ ms/+90°}$. Due to the symmetry of DDN around zero IPD, these two masker types produced almost identical results. DDN has inherently low IAC at certain frequencies outside the target channel [Fig. 1(b)]. As long as the flanking noise did not enter the frequency channels of high IAC in the proximity of the target, its addition had little effect on the already elevated thresholds produced by the DDN masker alone. Thus, we suggest that the elevated thresholds observed under DDN masker conditions, in both our data and those of van der Heijden and Trahiotis,[3] are the result of the off-frequency bands of low IAC inherent in the DDN masker.

3. The frequency convergence model

To demonstrate how IAC in the off-frequency channels can quantitatively account for both our data and those of van der Heijden and Trahiotis,[3] we introduce a simple phenomenological model based on binaural cross-correlation *followed* by across-frequency convergence. The activation of the binaural coincidence detector neurons is modeled by a rescaled cross-correlogram using 81 frequency channels (f_c=$\{100, 110, \ldots, 900$ Hz$\}$) with 201 internal-delay channels per frequency channel. The range of internal delays was restricted to the π-limit as suggested by physiological data (τ=$\{-0.5/f_c, \ldots, 0.5/f_c\}$).

$$\mathrm{CC}(f_c,\tau)=\sum_{f=1\ \mathrm{Hz}}^{2\ \mathrm{kHz}} H_{\mathrm{st}}{}^2(f)H_{\mathrm{AF}}{}^2(f_c,f)\left\{\eta\frac{1}{2}\cos[\mathrm{IPD}_{\mathrm{st}}(f)-2\pi\tau f]+\frac{1}{2}\right\}. \quad (1)$$

The computation of the cross-correlation function within a frequency channel f_c in Eq. (1) is based on the sum of sinusoidal cross-correlation functions of individual spectral components, which are power-weighted by the square of the cochlear filter shape (H_{AF}). (The cross-correlation function is the inverse Fourier transform of the cross-power spectrum.) Although white noise spectra, and consequently cross-correlograms, vary from one stimulus sample to the next, the long-term average of their magnitude spectra is "flat." Therefore, we defined the masker magnitude spectrum in our model as $H_{\mathrm{st}}(50-950\ \mathrm{Hz})=1$, so that the activity of a neural coincidence detector, $\mathrm{CC}(f_c,\tau)$, constitutes a long-term average, and is normalized such that its activity is one, when the signals arriving from the left and right ears are fully correlated, and zero, when the arriving signals have a correlation of –1 (for $\eta=1$). The parameter η modeled the effect of internal noise in the (monaural) auditory periphery, which reduces IAC and, consequently, the modulation of the cross-correlation function (for $\eta<1$). H_{AF} had the shape of a fourth-order gammatone filter with unit power transmission, and equivalent rectangular bandwidth (ERB) as estimated monaurally.[5] $\mathrm{IPD}_{\mathrm{st}}(f)$ describes the frequency dependence of the stimulus IPD. Input spectra were specified with 1-Hz frequency resolution.

Rescaled cross-correlograms were calculated separately for the signal alone ($\mathrm{CC}_{\mathrm{signal}}$) and the noise alone ($\mathrm{CC}_{\mathrm{noise}}$). Element-wise division of $\mathrm{CC}_{\mathrm{signal}}$ by $\mathrm{CC}_{\mathrm{noise}}$ produces the local signal-to-noise ratio (SNR) at each coincidence-detecting neuron. In each frequency channel f_c, the neuron with the maximum SNR was then identified [$\tau_{\max}(f_c)$]. The activations of these neurons are then summed across frequency channels in a weighted manner, before the global SNR was calculated [Eq. (2)]. This ensured that also activations beyond the target-frequency channel influenced signal detection ("across-frequency convergence").

$$\mathrm{SNR}_{\mathrm{global}}=\frac{\sum_{f_C}\omega(f_c)\mathrm{CC}_{\mathrm{signal}}[f_c,\tau_{\max}(f_c)]}{\sum_{f_C}\omega(f_c)\mathrm{CC}_{\mathrm{noise}}[f_c,\tau_{\max}(f_c)]}. \quad (2)$$

The $\mathrm{SNR}_{\mathrm{global}}$ [Eq. (2)], calculated for a diotic stimulus condition, normalized by the $\mathrm{SNR}_{\mathrm{global}}$ calculated for the diotic stimulus, gives the BMLD estimate of our model. (Because the model predicts only BMLD, note that simulated masked thresholds, as plotted in the figures, are the BMLD relative to the experimental masked threshold for the diotic condition.)

The impact of IAC in off-frequency channels on the BMLD can be easily understood when considering that signal activation in these channels is almost negligible. Therefore, $\tau_{\max}(f_c)$ is, here, essentially the internal delay producing the lowest masker activation. Because the cross-correlation function in frequency channels with lower IAC is less modulated, the activation minimum at $\tau_{\max}(f_c)$ and, consequently, the noise contribution to the global SNR from such channels are larger compared to an off-frequency frequency channel with a higher IAC.

The only free parameters of our model are the spectral weighting function $\omega(f_c)$ and the internal noise parameter η, which were manually adjusted to produce the closest agreement between the model and our data. Figure 2(d) shows the threshold functions predicted by our model. They were obtained using $\eta=0.93$, and a two-box weighting comprised of a low-weight convergence over a range larger than ±3 ERB [$\omega(220\ \mathrm{Hz}<f_c<780\ \mathrm{Hz})=0.02$], and a higher-weight convergence within approximately ±1 ERB of the target channel [$\omega(410\ \mathrm{Hz}<f_c<590\ \mathrm{Hz})=0.3$, $\omega(500\ \mathrm{Hz})=1$]. We felt that this simple two-box description illustrates best the dual character of the across-frequency convergence. It reproduces the features of our psychophysical results surprisingly well. Attempts to use more sophisticated weighting functions did not yield significantly better fits.

Our model (with unaltered parameters) also provides a reasonable account of the data of van der Heijden and Trahiotis[3] [Fig. 1(d)], in that it captures the general shape of the func-

tions, e.g., the wider troughs with SDN maskers and, most importantly, the higher thresholds with DDN maskers. It falls short, however, in two aspects: First, our model predicts slightly stronger damping of the oscillations than the psychophysical data of van der Heijden and Trahiotis,[3] although we ameliorated this side-effect of across-frequency convergence largely by introducing it after, rather than before, the stage of binaural cross-correlation. Note that the latter would be equivalent to a simple widening of the 500-Hz target channel. Such a π-limited model cannot reproduce the BMLD difference between SDN and DDN at 1 and 3 ms ITDs, because the best internal delay τ_{max} is then zero, where SDN and DDN produce identical cross-correlation values [see Fig. 1(c)]. Second, our binaural model does not reduce the impact of off-frequency channels near diotic conditions (around 0, 2, and 4 ms masker ITDs), when the critical bandwidth is expected to decrease to that found for monaural listening.[5] Since our purely binaural model does not incorporate a monaural detector path, those simulated thresholds are somewhat elevated.

The model of van der Heijden and Trahiotis,[3] on the other hand, cannot account for our data since it only considers the target-frequency channel, having a 3-dB-bandwidth of just 90 Hz. The long internal delays proposed in their model will always provide an advantage under SDN masker conditions (as long as the flanking noise is outside the target-frequency channel). Our experiment, however, showed similar thresholds for SDN and DDN masker conditions ($N_{-90°/+1\ ms/+90°}$ and $N_{-90°/\pm1\ ms/+90}$), even when the inner bands were as wide as 400 Hz.

4. Conclusion

Our findings suggest that binaural detection is influenced by IAC outside the target-frequency channel, which is in line with the phenomenon of wider binaural, compared to monaural, critical bandwidths.[4] Convergence of binaural neurons across different frequency bands has been observed physiologically,[6] although the questions remain as to how exactly this is implemented, and what function it might serve. Additionally, our findings provide an explanation for the observed difference in thresholds for DDN and SDN maskers that does not require long internal delays. Together with a recent functional imaging study,[7] our data further support the view that the human binaural system is similar to that of other mammals.

Acknowledgments

Data were collected during a Bogue Visiting Fellowship to TM at Dalhousie University (NS, Canada), hosted by Dennis Phillips and Susan Boehnke. We thank her and Isabel Dean for critical reading of the manuscript, and John Agapiou for many helpful suggestions.

References and links

[1]L. A. Jeffress, "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–49 (1948).
[2]P. X. Joris and T. C. T. Yin, "A matter of time: Interaural delays in binaural processing," Trends Neurosci. **30**, 70–78 (2007).
[3]M. van der Heijden and C. Trahiotis, "Masking with interaurally delayed stimuli: The use of 'internal' delays in binaural detection," J. Acoust. Soc. Am. **105**, 388–399 (1999).
[4]M. M. Sondhi and N. Guttman, "Width of the spectrum effective in the binaural release of masking," J. Acoust. Soc. Am. **40**, 600–606 (1966).
[5]B. G. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138 (1990).
[6]D. McAlpine, D. Jiang, T. M. Shackleton, and A. R. Palmer, "Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus," J. Neurosci. **18**, 6026–6039 (1998).
[7]S. K. Thompson, K. von Kriegstein, A. Deane-Pratt, T. Marquardt, R. Deichmann, T. D. Griffiths, and D. McAlpine, "Representation of interaural time delay in the human auditory midbrain," Nat. Neurosci. **9**, 1096–1098 (2006).

Accuracy of the deterministic travel time retrieval from cross-correlations of non-diffuse ambient noise[a)]

Oleg A. Godin
CIRES, University of Colorado, and NOAA/Earth System Research Laboratory, DSRC, Mail Code R/PSD, 325 Broadway, Boulder, Colorado 80305-3328
oleg.godin@noaa.gov

Abstract: Measurements of long-range cross-correlations of ambient noise underlie acoustic noise interferometry, a promising technique for passive remote sensing of the environment. Previously established simple, exact relations between deterministic Green's functions and the cross-correlation function of perfectly diffuse noise do not necessarily hold for noise fields in the ocean and atmosphere. Here, the method of a stationary phase is applied to study the information content of the cross-correlation function of non-diffuse noise and to quantify the accuracy of passive measurements of the acoustic travel times.

PACS numbers: 43.30.Pc, 43.20.Bi, 43.30.Nb, 43.60.Rw [AN]
Date Received: August 21, 2009 **Date Accepted:** October 6, 2009

1. Introduction

A two-point correlation function of a perfectly diffuse noise field contains all the information about the environment that can be obtained using transceivers placed at the two points.[1–4] The emergence of exact deterministic Green's functions (GFs) from diffuse noise has been demonstrated theoretically[4–10] and in laboratory experiments[5,6] for fluid, solid, and fluid-solid systems, including moving media (see also Refs. 3 and 11). Existence of a simple, algebraic relationship between deterministic GFs and the cross-correlation function of perfectly diffuse noise is fascinating and thought-provoking; it provides deep physical insights into the nature of random wave fields. However, these results cannot be directly applied to passive remote sensing of the environment as long as neither seismic noise nor ambient acoustic noise in the ocean and atmosphere is perfectly diffuse, except at very high frequencies where noise of thermal origin[5,6,10,12] dominates.

Asymptotic techniques allow one to relate parameters of the noise cross-correlation function and the deterministic GFs without making the assumption that the noise is perfectly diffuse. The first approximation of the stationary phase method[13] was used by Snieder,[14,15] Sabra *et al.*,[16] Godin,[4] Brooks and Gerstoft,[17] and Garnier and Papanicolaou[18] to study long-range correlations of high-frequency waves generated by distributions of random sources, the density of which is either constant or slowly varying in space. It was found that in inhomogeneous, moving or motionless media, deterministic eikonals (or travel times) along all eigenrays connecting any two points can be retrieved from the cross-correlation function of noise measured at the two points.[4] However, it remains an open question: what accuracy of deterministic travel time retrieval and, hence, of sound speed and flow velocity inversions, can be achieved with passive measurements. To address this question it is natural to analyze higher-order terms in the asymptotic expansion of the noise cross-correlation function obtained by the stationary phase method.

[a)]Parts of this work have been previously reported at the 157th ASA Meeting (Portland, OR, May 2009) and the 12th L. M. Brekhovskikh Conference on Ocean Acoustics (Moscow, Russia, June 2009).

In this paper, we consider high-frequency sound fields in moving and motionless fluids generated by random sources distributed on a curve or on a surface. We use the method developed in Ref. 4 to determine what environmental information can be extracted from the noise cross-correlation without having detailed knowledge about the noise field sources and properties and, in particular, to quantify the impact of non-uniformity of the spatial distribution of noise sources on the accuracy of passive measurements of acoustic travel times.

2. Cross-correlation function of noise

Consider acoustic waves in an inhomogeneous moving fluid with sound speed $c(\mathbf{x})$, mass density $\rho(\mathbf{x})$, and flow velocity $\mathbf{u}(\mathbf{x})$. As in Refs. 4 and 10, define the GF in the time domain, $G(\mathbf{x},\mathbf{y},t)$, as the acoustic pressure p at $\mathbf{x}$ due to a point source of volume velocity with density $B(\mathbf{x},t)=\delta(\mathbf{x}-\mathbf{y})\delta'(t)$, where δ is the delta function and t is time. The frequency spectrum $G(\mathbf{x},\mathbf{y},\omega)$ of the time-domain GF $G(\mathbf{x},\mathbf{y},t)$ has the meaning of the continuous wave (cw) Green's function. Time dependence $\exp(-i\omega t)$ of cw fields is assumed and suppressed. The GF $\tilde{G}(\mathbf{x},\mathbf{y},\omega)$ in a medium with reversed flow, i.e., with parameters $c(\mathbf{x})$, $\rho(\mathbf{x})$, and $-\mathbf{u}(\mathbf{x})$, is generally different from the GF $G(\mathbf{x},\mathbf{y},\omega)$ in the original medium, but a simple, reciprocity-type symmetry relation[19] holds: $G(\mathbf{x},\mathbf{y},\omega)=\tilde{G}(\mathbf{y},\mathbf{x},\omega)$. At high frequencies, when the ray approximation becomes applicable,

$$G(\mathbf{x},\mathbf{y},\omega)=\sum_n A_n(\mathbf{x},\mathbf{y},\omega)\exp[i\omega\varphi_n(\mathbf{x},\mathbf{y})],\quad A_n(\mathbf{x},\mathbf{y},\omega)=A_n^{(0)}(\mathbf{x},\mathbf{y})+\frac{i}{\omega}A_n^{(1)}(\mathbf{x},\mathbf{y})+\ \ldots\ , \tag{1}$$

where A_n and φ_n are, respectively, slowly varying complex amplitudes and rapidly varying, real-valued eikonals of individual ray components of the field.[13] The number of summands in the right-hand side of Eq. (1) depends on the environment and possibly on the source and receiver positions.

Let an acoustic field be generated by random, delta-correlated cw sources of volume velocity with density B:

$$\langle B(\mathbf{x})\rangle=0,\quad \langle B(\mathbf{x}_1)B^*(\mathbf{x}_2)\rangle=Q(\mathbf{x}_1)\delta_S(\mathbf{x}_1-\mathbf{x}_2). \tag{2}$$

Here and below, the asterisk and angular brackets designate complex conjugation and the average over the statistical ensemble of random sources. The sources are located either on a curve $\mathbf{x}=\mathbf{X}(s)$, where s is the arc length of the curve, or on a surface $\mathbf{x}=\mathbf{Y}(\xi^{(1)},\xi^{(2)})$, where $\xi^{(1,2)}$ are curvilinear coordinates on the surface. For sources on a curve, $\mathbf{x}_j=\mathbf{X}(s_j)$, $j=1,2$ and $\delta_S(\mathbf{x}_1-\mathbf{x}_2)\equiv\delta(s_1-s_2)$, while for sources on a surface $\mathbf{x}_j=\mathbf{Y}(\xi_j^{(1)},\xi_j^{(2)})$ and $\delta_S(\mathbf{x}_1-\mathbf{x}_2)\equiv\delta(\xi_1^{(1)}-\xi_2^{(1)})\delta(\xi_1^{(2)}-\xi_2^{(2)})$ in Eq. (2). Acoustic pressure generated by the random sources has zero mean: $\langle p(\mathbf{x})\rangle=0$, while its cross-correlation function equals[4,10]

$$C(\mathbf{x}_1,\mathbf{x}_2)\equiv\langle p(\mathbf{x}_1)p^*(\mathbf{x}_2)\rangle=\omega^{-2}\int_S Q(\mathbf{y})G(\mathbf{x}_1,\mathbf{y},\omega)G^*(\mathbf{x}_2,\mathbf{y},\omega)dS, \tag{3}$$

where $dS=ds$, $\mathbf{y}=\mathbf{X}(s)$ and $dS=d\xi^{(1)}d\xi^{(2)}$, $\mathbf{y}=\mathbf{Y}(\xi^{(1)},\xi^{(2)})$ for the sources located on a curve and on a surface, respectively.

3. Two-dimensional (2D) problem

Let the curve $\mathbf{x}=\mathbf{X}(s)$, where sound sources are located, be the boundary $\partial\Omega$ of a finite domain Ω on the plane $x_3=0$ and all receivers be located in the same plane either inside or outside Ω. Physically, this 2D problem represents random fields due to an interface wave or an adiabatic normal mode in a horizontally inhomogeneous waveguide.

From Eqs. (1) and (3), we have

$$C(\mathbf{x}_1,\mathbf{x}_2)=\omega^{-2}\sum_{n,m}\int_S Q(\mathbf{y})A_n(\mathbf{x}_1,\mathbf{y},\omega)A_m^*(\mathbf{x}_2,\mathbf{y},\omega)\exp[i\omega\varphi_n(\mathbf{x}_1,\mathbf{y})-i\omega\varphi_m(\mathbf{x}_2,\mathbf{y})]dS. \quad (4)$$

The integrand in Eq. (4) contains a rapidly oscillating exponential, and high-frequency asymptotic expansion of the integral is dominated by contributions of the stationary points of the exponent.[13] By taking into account that only one ray can pass through a given point in a given direction, it can be demonstrated that only diagonal terms ($n=m$) in the ray sum in the right-hand side of Eq. (4) have stationary points. There is a one-to-one correspondence between the stationary points on $\partial\Omega$ and the intersections between $\partial\Omega$ and rays passing through both receivers at $\mathbf{x}_1$ and $\mathbf{x}_2$: every stationary point lies on an extension of a ray, which passes through both $\mathbf{x}_1$ and $\mathbf{x}_2$, in a direction opposite of the direction of sound propagation, and every intersection of $\partial\Omega$ and such a ray extension is a stationary point.[4]

Consider a stationary point $\mathbf{y}_s\in\partial\Omega$ of the n-th diagonal term in the sum in the right-hand side of Eq. (4). For definiteness, let a ray leaving $\partial\Omega$ at $\mathbf{y}_s$ go first through $\mathbf{x}_2$ and then through $\mathbf{x}_1$. Applying the method of stationary phase[13] to evaluate the integral in the right-hand side of Eq. (4), after some algebra one finds the contribution of the stationary point to the noise cross-correlation function $C(\mathbf{x}_1,\mathbf{x}_2)$:

$$C^{(n)}(\mathbf{x}_1,\mathbf{x}_2)=\alpha_n A_n^{(0)}(\mathbf{x}_1,\mathbf{x}_2,\omega)\exp[i\omega\varphi_n(\mathbf{x}_1,\mathbf{x}_2)+i\beta_n/\omega][1+O(\omega^{-2})],$$

$$\alpha_n=-\rho cQ/2q_n|_{\mathbf{y}=\mathbf{y}_s}, \quad (5)$$

$$\beta_n=\left[\frac{F''}{2F\Phi''}+\frac{5(\Phi''')^2}{24(\Phi'')^3}-\frac{F'\Phi'''}{2F(\Phi'')^2}-\frac{\Phi^{(\mathrm{IV})}}{8(\Phi'')^2}+\frac{A_n^{(1)}(\mathbf{x}_1,\mathbf{y})}{A_n^{(0)}(\mathbf{x}_1,\mathbf{y})}-\frac{A_n^{(1)*}(\mathbf{x}_2,\mathbf{y})}{A_n^{(0)*}(\mathbf{x}_2,\mathbf{y})}\right]_{\mathbf{y}=\mathbf{y}_s},$$

$$\Phi(s)=\varphi_n(\mathbf{x}_1,\mathbf{y})-\varphi_n(\mathbf{x}_2,\mathbf{y}),\quad F(s)=A_n^{(0)}(\mathbf{x}_1,\mathbf{y})A_n^{(0)*}(\mathbf{x}_2,\mathbf{y})Q(y),\quad \mathbf{y}=\mathbf{X}(s). \quad (6)$$

Here, primes designate differentiation with respect to s, q_n is the normal to $\partial\Omega$ component of the vector $(c\mathbf{h}_n+h_n\mathbf{u})/c^2h_n^2$, and $\mathbf{h}_n=-\partial\varphi_n(\mathbf{x}_1,\mathbf{y})/\partial\mathbf{y}$. Calculation of the leading-order term in the asymptotic expansion (5) in the inverse powers of frequency is discussed in Ref. 4. Equation (6) for the first-order correction follows from Eqs. (A.1.9) and (A.1.12) in Ref. 13.

The component $C^{(n)}(\mathbf{x}_1,\mathbf{x}_2)$ (5) of the noise cross-correlation function has the same eikonal as the n-th term in the sum (1) for the deterministic GF $G(\mathbf{x}_1,\mathbf{x}_2,\omega)$ and differs from the latter by the additional factor

$$\gamma_n=\alpha_n\exp\{i\omega^{-1}[\beta_n-A_n^{(1)}(\mathbf{x}_1,\mathbf{x}_2)/A_n^{(0)}(\mathbf{x}_1,\mathbf{x}_2)]\}[1+O(\omega^{-2})]. \quad (7)$$

The shading α_n (5) of the amplitudes of ray arrivals is determined by variations in the source density as well as by environmental parameters in the vicinity of $\partial\Omega$. The values that α_n (5) takes at different n can vary significantly. On the other hand, phase differences are proportional to ω^{-1} and are small for high-frequency waves. We emphasize that possible $O(1)$ phase shifts in individual ray components of the deterministic GF, which occur when the ray touches a caustic,[13] are contained in $A_n^{(0)}(\mathbf{x}_1,\mathbf{x}_2)$ and, according to Eqs. (5) and (7), are reproduced faithfully by the noise cross-correlation function.

Quite similar to the derivation of Eqs. (5)–(7) outlined above, one can show that the sum of contributions of the stationary points on $\partial\Omega$, for which the corresponding rays go first through $\mathbf{x}_1$ and then through $\mathbf{x}_2$, approximates $G^*(\mathbf{x}_2,\mathbf{x}_1,\omega)=\tilde{G}^*(\mathbf{x}_1,\mathbf{x}_2,\omega)$. Consequently, the noise cross-correlation function $C(\mathbf{x}_1,\mathbf{x}_2)$ approximates $G(\mathbf{x}_1,\mathbf{x}_2,\omega)+G^*(\mathbf{x}_2,\mathbf{x}_1,\omega)$.

4. Three-dimensional problem

Let sound sources be distributed on a surface $\mathbf{x}=\mathbf{Y}(\xi^{(1)},\xi^{(2)})$, which coincides with the boundary $\partial\Omega$ of a finite volume Ω, and all receivers be located either inside or outside Ω. As in the 2D

problem, the high-frequency asymptotics of the surface integral in Eq. (4) for the noise cross-correlation function are determined by contributions of the vicinities of stationary points of the rapidly oscillating exponential in the integrand. Physically, the stationary points on $\partial\Omega$ are origination points of the rays which go through both receivers.[4] Application of the two-dimensional stationary phase method[13,20] again leads to Eqs. (5) and (7), where now

$$\beta_n = \left\{ \frac{\hat{L}F}{F} - \frac{1}{2}\hat{L}^2\Phi - \frac{1}{F}\hat{M}[F,\hat{L}\Phi] + \frac{1}{12}\hat{L}^2\hat{M}[\Psi,\Psi] + \frac{1}{3}\hat{M}[\hat{L}\Phi,\hat{L}\Phi] + \frac{A_n^{(1)}(\mathbf{x}_1,\mathbf{y})}{A_n^{(0)}(\mathbf{x}_1,\mathbf{y})} - \frac{A_n^{(1)*}(\mathbf{x}_2,\mathbf{y})}{A_n^{(0)*}(\mathbf{x}_2,\mathbf{y})} \right\}_{\mathbf{y}=\mathbf{y}_s}, \quad F(\xi^{(1)},\xi^{(2)}) = A_n^{(0)}(\mathbf{x}_1,\mathbf{y})A_n^{(0)*}(\mathbf{x}_2,\mathbf{y})Q(\mathbf{y}),$$

$$\Phi(\xi^{(1)},\xi^{(2)}) = \varphi_n(\mathbf{x}_1,\mathbf{y}) - \varphi_n(\mathbf{x}_2,\mathbf{y}), \quad \mathbf{y} = \mathbf{Y}(\xi^{(1)},\xi^{(2)}), \quad \mathbf{y}_s = \mathbf{Y}(\xi_s^{(1)},\xi_s^{(2)}),$$

$$\Psi(\xi^{(1)},\xi^{(2)}) = \Phi(\xi^{(1)},\xi^{(2)}) - \sum_{j,k=1}^{2} \partial^2\Phi/\partial\xi^{(j)}\,\partial\,\xi^{(k)}|_{\mathbf{y}=\mathbf{y}_s}(\xi^{(1)}-\xi_s^{(1)})(\xi^{(2)}-\xi_s^{(2)}). \tag{8}$$

Differential operators $\hat{L}$ and $\hat{M}$ in Eq. (8) are defined as follows:

$$\hat{L}F = \frac{1}{2}\sum_{j,k=1}^{2} H_{jk}\frac{\partial^2 F}{\partial\xi^{(j)}\,\partial\,\xi^{(k)}}, \quad \hat{M}[f,F] = \sum_{j,k=1}^{2} H_{jk}\frac{\partial f}{\partial\xi^{(j)}}\frac{\partial F}{\partial\xi^{(k)}}, \quad \{H_{jk}\} = \left\| \left.\frac{\partial^2\Phi}{\partial\xi^{(j)}\,\partial\,\xi^{(k)}}\right|_{\mathbf{y}=\mathbf{y}_s} \right\|^{-1}. \tag{9}$$

Compared to the results of Ref. 4, Eqs. (5) and (8) provide higher accuracy by calculating the first-order correction in addition to the leading-order term of the high-frequency asymptotic expansion of the noise cross-correlation function.

The first-order correction β_n (8) depends on the derivatives (up to the fourth order) of the eikonal, on diffraction corrections $A_n^{(1)}$ to the zero-order ray amplitudes $A_n^{(0)}$, and on the first- and second-order derivatives of $A_n^{(0)}$ and the source density Q along $\partial\Omega$. When the ray amplitudes $A_n^{(0)}$ and $A_n^{(1)}$ or at least the quotients $A_n^{(0)}(\mathbf{x}_2,\mathbf{y})/A_n^{(0)}(\mathbf{x}_1,\mathbf{y})$ and $A_n^{(1)}(\mathbf{x},\mathbf{y})/A_n^{(0)}(\mathbf{x},\mathbf{y})$ are real-valued, $\mathrm{Im}\,\beta_n=0$, and β_n describes a correction to the phase only. Note that $|\beta_n/\omega| \ll 1$ as long as the stationary phase method is applicable.[13]

In a gradually varying inhomogeneous fluid with $m \equiv |\mathbf{u}|/c < 1$, the representative spatial scales of the ray amplitude $A_n^{(0)}(\mathbf{x},\mathbf{y})$ and acoustic wave vector $\mathbf{k}=\omega\partial\varphi(\mathbf{x},\mathbf{y})/\partial\mathbf{x}$ variation are $\min(|\mathbf{x}-\mathbf{y}|,l)$ and $A_n^{(1)}(\mathbf{x},\mathbf{y})/A_n^{(0)}(\mathbf{x},\mathbf{y}) = O(c_0/\min(|\mathbf{x}-\mathbf{y}|,l))$, where c_0 is a representative value of the sound speed and l is a representative spatial scale of ρ, c, and $\mathbf{u}$ inhomogeneities. Let b be the representative spatial scale of the variation in the source density Q along $\partial\Omega$, and $\tau_{1,2}$ be the representative travel times from the stationary point to receivers at $\mathbf{x}_{1,2}$. We assume that $c_0\tau_{1,2}$ are either of the same order of magnitude or large compared to $|\mathbf{x}_2-\mathbf{x}_1|$. Taking into account that the function Φ (8) and all its derivatives with respect to $\xi^{(1,2)}$ are zero when $\mathbf{x}_1=\mathbf{x}_2$, for the quantities H_{jk} in Eq. (9) we have $H_{jk}=O(c_0^3\tau^2/|\mathbf{x}_2-\mathbf{x}_1|)$. Here $\tau=\min(\tau_1,\tau_2,l/c_0)$. Using the definitions (9) of the differential operators $\hat{L}$ and $\hat{M}$ to evaluate respective terms in the right-hand side of Eq. (8), we obtain the following order-of-magnitude estimate $\beta_n \sim c_0(c_0\tau/b+1)^2/|\mathbf{x}_2-\mathbf{x}_1|$. If the phase of $C^{(n)}(\mathbf{x}_1,\mathbf{x}_2)$ (5) is used to measure the eikonal $\varphi_n(\mathbf{x}_1,\mathbf{x}_2)$, the relative error is inversely proportional to the square of $\omega\varphi_n(\mathbf{x}_1,\mathbf{x}_2)$. When the source density varies slowly ($b > c_0\tau$) and $|\mathbf{x}_2-\mathbf{x}_1| \sim l$, the phase correction β_n/ω is rather small and generally is of the same order of magnitude as the difference between the eikonal and the phase of the GF $G(\mathbf{x}_1,\mathbf{x}_2,\omega)$. In the case of faster varying Q or a distant noise source, where $c_0\tau \gg b$, the phase error is much larger and is inversely proportional to b^2. An inspection of Eq. (6) shows that the above estimates and qualitative conclusions remain valid in the 2D problem.

Eikonal non-reciprocity $|\varphi_n(\mathbf{x}_2,\mathbf{x}_1)-\varphi_n(\mathbf{x}_1,\mathbf{x}_2)|\sim mc_0^{-1}|\mathbf{x}_2-\mathbf{x}_1|$ arises due to fluid motion and can be used to determine the flow velocity.[13,19] In geophysical applications, the flow velocity is typically much smaller than the sound speed ($m\ll 1$), and precise measurements of the eikonals are required to characterize the flow acoustically. Using the above estimates of the phase correction β_n/ω (8), we obtain the condition $\omega|\mathbf{x}_2-\mathbf{x}_1|/c_0\gg m^{-1/2}(1+c_0\tau/b)$ under which measurements of the noise cross-correlations can be used for the flow velocity tomography.

5. A simple example

To illustrate general results, consider an acoustic field in a homogeneous, quiescent ($\mathbf{u}=0$) fluid generated by random sound sources distributed on a plane $x_3=0$. Receivers are located at points $\mathbf{x}_j=(x_{j,1},x_{j,2},x_{j,3})$, $j=1,2$ in the half-space $x_3>0$. Let $x_{2,3}>x_{1,3}$ for definiteness. We choose Cartesian coordinates $x_{1,2}$ to serve as coordinates $\xi^{(1,2)}$ in the plane $x_3=0$. In this problem, the cw GF is[13]

$$G(\mathbf{x},\mathbf{y},\omega)=i\rho\omega^3\exp(ik|\mathbf{x}-\mathbf{y}|)/4\pi|\mathbf{x}-\mathbf{y}|,\quad k=\omega/c, \tag{10}$$

and Eq. (3) for the noise cross-correlation function becomes

$$C(\mathbf{x}_1,\mathbf{x}_2)=\left(\frac{\rho\omega^2}{4\pi}\right)^2\int_{y_3=0}\int\exp\left[ik(|\mathbf{x}_1-\mathbf{y}|-|\mathbf{x}_2-\mathbf{y}|)\right]\frac{Q(\mathbf{y})dy_1dy_2}{|\mathbf{x}_1-\mathbf{y}||\mathbf{x}_2-\mathbf{y}|}. \tag{11}$$

In homogeneous, unbounded fluids, the sums over rays in Eqs. (1) and (4) contain a single summand, and higher-order ray amplitudes $A_n^{(j)}$, $j=1,2,\ldots$ equal zero.

The only stationary point of the exponent in the integrand in Eq. (11) is located at $\mathbf{y}_s=(x_{1,3}-x_{2,3})^{-1}(x_{1,3}\mathbf{x}_2-x_{2,3}\mathbf{x}_1)$. It lies at the intersection of the plane $y_3=0$ with the straight line through the receivers. Asymptotic expansion of the integral in the right-hand side of Eq. (11) can be obtained directly, using neither the reasoning[4] employed to evaluate the leading-order term nor general expressions[20] for the first-order correction; e.g., the double integral (11) can be evaluated by applying well-known stationary phase results for single integrals[13] to iterated integrals over y_1 and y_2, leading to

$$C(\mathbf{x}_1,\mathbf{x}_2)=G^*(\mathbf{x}_2,\mathbf{x}_1,\omega)\frac{\rho c|\mathbf{x}_1-\mathbf{x}_2|}{2(x_{1,3}-x_{2,3})}Q(\mathbf{y}_s)\exp\left[\frac{i}{\omega}\beta(\mathbf{x}_1,\mathbf{x}_2)\right]\left[1+O\left(\frac{1}{\omega^2}\right)\right], \tag{12}$$

where

$$\beta=\frac{c|\mathbf{x}_2-\mathbf{x}_1|}{(x_{2,3}-x_{1,3})^2}\left\{1-(x_{1,3}+x_{2,3})\frac{(\mathbf{x}_2-\mathbf{x}_1)\cdot\nabla Q}{(x_{2,3}-x_{1,3})Q}+\frac{x_{1,3}x_{2,3}}{2}\left[\frac{\Delta Q}{Q}+\frac{[(\mathbf{x}_2-\mathbf{x}_1)\cdot\nabla]^2Q}{(x_{2,3}-x_{1,3})^2Q}\right]\right\}\Bigg|_{\mathbf{y}=\mathbf{y}_s}. \tag{13}$$

When $x_{2,3}<x_{1,3}$, $G^*(\mathbf{x}_2,\mathbf{x}_1,\omega)$ in Eq. (12) is replaced by $G(\mathbf{x}_1,\mathbf{x}_2,\omega)$. In the example considered, the noise correlation function approximates only one term of the sum $G(\mathbf{x}_1,\mathbf{x}_2,\omega)+G^*(\mathbf{x}_2,\mathbf{x}_1,\omega)$ because the insonification is one-sided. An inspection shows that the results (12) and (13) agree with the general equations (5), (7), and (8).

When noise cross-correlation is used to measure the phase of the deterministic GF between points $\mathbf{x}_1$ and $\mathbf{x}_2$, the phase error β/ω (13) depends on the positions of the receivers and on the first and second derivatives of the noise density Q at the stationary point $\mathbf{y}_s$. When $x_{1,3}\neq x_{2,3}$, β (13) is always finite and can be zero at some points. Assuming that distances from $\mathbf{x}_1$ and $\mathbf{x}_2$ to the plane $x_3=0$ are not too close to each other [namely, that $|\mathbf{x}_2-\mathbf{x}_1|/|x_{2,3}-x_{1,3}|=O(1)$], a simple order-of-magnitude estimate $\beta/\omega\sim(k|\mathbf{x}_2-\mathbf{x}_1|)^{-1}(1+x_{1,3}x_{2,3}/b^2)$ follows from Eq. (13), where b is a representative spatial scale of the function Q variation. Hence, the relative error of the passive phase measurement is inversely proportional to the squares of the wave frequency and the distance between the receivers. The phase error becomes sensitive to the rate of Q variation when $b<x_{j,3}$. Then β is proportional to $x_{1,3}x_{2,3}/b^2$ or the solid angle formed by a

square with side b in the vicinity of $\mathbf{y}_s$ as seen from the receivers. The fact that the phase error increases when the receivers move further away from distant noise sources may be unexpected but is easy to understand. The vicinity $\partial\Omega_s$ of the stationary point $\mathbf{y}_s$, which provides the main contribution to the integral in Eq. (11), is the region where the exponential in the integrand deviates from its value at $\mathbf{y}_s$ by $O(1)$.[13] Spatial dimensions of $\partial\Omega_s$ are $d_s \sim [x_{1,3}x_{2,3}/k|\mathbf{x}_2-\mathbf{x}_1|]^{1/2}$ and *increase* with $x_{j,3}$. Terms in the integrand, which are linear in $\mathbf{y}-\mathbf{y}_s$, cancel out and do not contribute to the integral. At large $x_{j,3}$, the leading-order correction is due to quadratic terms in the Taylor expansion of $Q(\mathbf{y})$, which are $O(d_s^2/b^2)=O(x_{1,3}x_{2,3}/k|\mathbf{x}_2-\mathbf{x}_1|b^2)$ in $\partial\Omega_s$, in agreement with the estimate of β/ω obtained above from Eq. (13).

6. Discussion

Equations (5), (7), and (12) show that, unlike the amplitude information contained in the deterministic GF $G(\mathbf{x}_2,\mathbf{x}_1,\omega)$, the phase information or the travel times between $\mathbf{x}_1$ and $\mathbf{x}_2$ along various eigenrays can be retrieved from the noise cross-correlation function without any detailed knowledge about properties or location of the noise sources. Passive measurements of the travel times require that the random source density varies gradually in space or, equivalently, noise directivity in each point is a gradual function of direction.

We have quantified errors (6), (8), and (13) in the eikonal estimates retrieved from the noise cross-correlation function. These errors stem from non-uniformity of the random sources distribution or, in other words, from the noise field not being perfectly diffuse. Our results refer to the eikonal estimates, which are obtained with sufficiently large noise averaging time. Additional errors may occur when the averaging time is not sufficient to ensure that time averages approximate statistical mean values (see Ref. 11).

Our analysis relies on the second approximation of the stationary phase method. It reveals the dependence of the phase (and, consequently, travel time) errors on the geometry of the problem and the rate of variation in the noise sources' density. Under conditions of validity of the stationary phase method,[13] the phase correction $\beta_n/\omega \ll 1$. Estimates of β_n/ω can be used to determine the feasibility of the passive measurements of acoustic travel times in a specific setting. When the phase errors calculated using Eq. (6), Eq. (8), or Eq. (13) are of the order of unity or larger, the stationary-point arguments do not apply. Then, peaks in the noise cross-correlation function in the time domain do not necessarily correspond to the travel times between the receivers, and the noise cross-correlations cannot be used to characterize the environment without detailed knowledge of the noise sources.

Acknowledgments

This work was supported in part by the Office of Naval Research through Grant No. N00014-08-1-0100. I thank M. Charnotskii, V. G. Irisov, K. G. Sabra, and R. I. Weaver for stimulating discussions.

References and links

[1]S. M. Rytov, "On thermal fluctuations in distributed systems," Dokl. Akad. Nauk SSSR **110**, 371 (1956).
[2]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[3]R. L. Weaver and O. I. Lobkis, "Diffuse fields in ultrasonics and seismology," Geophysics **71**, SI5–SI9 (2006).
[4]O. A. Godin, "Recovering the acoustic Green's function from ambient noise cross-correlation in an inhomogeneous moving medium," Phys. Rev. Lett. **97**, 054301 (2006).
[5]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[6]R. L. Weaver and O. I. Lobkis, "Elastic wave thermal fluctuations, ultrasonic waveforms by correlation of thermal phonons," J. Acoust. Soc. Am. **113**, 2611–2621 (2003).
[7]A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from field-field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[8]K. Wapenaar, "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301 (2004).
[9]R. L. Weaver and O. I. Lobkis, "Diffuse fields in open systems and the emergence of the Green's function," J. Acoust. Soc. Am. **116**, 2731–2734 (2004).

[10]O. A. Godin, "Retrieval of Green's functions of elastic waves from thermal fluctuations of fluid-solid systems," J. Acoust. Soc. Am. **125**, 1960–1970 (2009).
[11]P. Gouédard, L. Stehly, F. Brenguier, M. Campillo, Y. Colin de Verdière, E. Larose, L. Margerin, P. Roux, F. J. Sánchez-Sesma, N. M. Shapiro, and R. L. Weaver, "Cross-correlation of random fields: Mathematical approach and applications," Geophys. Prospect. **56**, 375–393 (2008).
[12]R. H. Mellen, "The thermal-noise limit in the detection of underwater acoustic signals," J. Acoust. Soc. Am. **24**, 478–480 (1952).
[13]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media. 2: Point Sources and Bounded Beams*, 2nd extended ed. (Springer, Berlin, 1999).
[14]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[15]R. Snieder, "Extracting the Green's function of attenuating heterogeneous acoustic media from uncorrelated waves," J. Acoust. Soc. Am. **121**, 2637–2643 (2007).
[16]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[17]L. A. Brooks and P. Gerstoft, "Ocean acoustic interferometry," J. Acoust. Soc. Am. **121**, 3377–3385 (2007).
[18]J. Garnier and G. Papanicolaou, "Passive sensor imaging using cross correlations of noisy signals in a scattering medium," SIAM J. Imaging Sci. **2**, 396–437 (2009).
[19]O. A. Godin, "Reciprocity and energy theorems for waves in a compressible inhomogeneous moving fluid," Wave Motion **25**, 143–167 (1997).
[20]M. V. Fedoryuk, *Asymptotic Expansions: Integrals and Series* (Nauka, Moscow, 1987), p. 189 (in Russian).

Acoustic measurements of clay-size particles

Wayne O. Carpenter, Jr.[a] **and James P. Chambers**
National Center for Physical Acoustics, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677-1848
wocarpen@olemiss.edu, chambers@olemiss.edu

Daniel G. Wren, Roger A. Kuhnle, and Jeffrey A. Diers
USDA-ARS-NSL, P.O. Box 1157, Oxford, Mississippi 38655
daniel.wren@ars.usda.gov, roger.kuhnle@ars.usda.gov, jeffrey.diers@ars.usda.gov

Abstract: Knowledge of sediment concentration is important in the study of streams and rivers. The work presented explores using high frequency (20 MHz) acoustic signal attenuation to measure the concentration of fine sediment particles ($0.2-5.0$ μm) in a fluvial environment. A small laboratory tank with a pitch-catch transducer configuration measured a 35 dB change in signal level over a wide range of kaolinite and bentonite concentrations (1–14 g/l) over a range of distances (180–357 mm). The data suggest that a fixed distance of 180 mm between the transducers will be capable of measuring the entire range of concentrations.

PACS numbers: 43.30.Ma, 43.35.Bf, 43.35.Yb [TM]
Date Received: August 18, 2009 **Date Accepted:** October 5, 2009

1. Introduction

Suspended sediments are a global-scale pollutant whose yield has been estimated at 20×10^9 tons/year.[1] In many streams, the majority of sediment moves during flood events caused by a few large storms per year.[2] These flood events are unpredictable and frequently occur at night, making the collection of physical sediment samples difficult and sometimes dangerous. Manual techniques yield samples that are widely spaced in time, small in number, and flow intrusive. To improve the spatial and temporal densities of suspended-sediment data, the continuing development of automated measurement systems is essential. Ultrasonic measurement systems can detect particles with a high degree of both spatial and temporal resolutions, making them ideal for addressing the needs of those who rely on sediment data.[3,4]

Acoustic technology has great potential for improving the current state of suspended-sediment measurement technology; it can be relatively inexpensive, lends itself well to remote deployment, and is non-intrusive.[5] Most acoustic systems have targeted sand sized particles ($62-2000$ μm) due to their heterogeneous distribution with depth.[4] However, a large portion of the sediment load in a stream may be <62 μm, a size range that is well-distributed throughout the cross-section.[6,7] These fine sediment particle concentrations have been observed to be as high as 56% by weight.[8,9]

This paper describes experiments aimed at the development of a device that will use measurements of acoustic signal attenuation during propagation through water containing clay particles to determine the particle concentration. Two clays, kaolinite and bentonite, were chosen for several reasons. The mechanical failures in rocks and unconsolidated materials frequently involve clay-rich lithologies.[10] In the sea floor, clays represent 40%–80% of silicate minerals in pelagic sediments.[11] There is also a strong effect of kaolinite content on the acoustic properties in sedimentary rocks and unconsolidated mixtures.[12,13] Bentonite at 0.2 μm and kaolinite at 5 μm represent a wide range of clay particle sizes. These two clays are also readily available from commercial sources.

[a]Author to whom correspondence should be addressed.

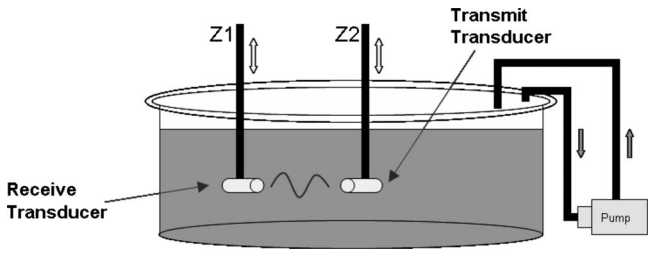


Fig. 1. Recirculation tank

Previous work on acoustic attenuation by fine particles focused on fixed pitch/catch configuration distances. Greenwood *et al.*[14] measured ultrasonic attenuation in a kaolin-water slurry (1 μm particles) from 0–0.2 volume fraction and utilized a frequency range 0.5–3 MHz at a fixed distance of approximately 0.1016 m. At each frequency, they found that attenuation was a linear function of the volume fraction and had a linear dependence on frequency within the 0.5–3 MHz range. To measure kaolinite/water suspensions in near 40% solid-volume, Green and Esquivel-Sirvent[10] used a bench-top 1 liter suspension chamber with two ultrasonic transducers (3.5 and 7 MHz) and a micrometer stage allowing adjustments in transmit/receive path length. The effect of the increasing frequency produced only a moderate increase in attenuation and a slight increase in the concentration of maximum absorption. The present work focuses on suspended-sediment concentrations much smaller (1–14 g/l) than some of the aforementioned studies. These lower concentrations, which encompass what one would find in ephemeral streams,[12] made it possible to neglect multiple scattering since the concentrations were on the order of 10 g/l.[15]

2. Methods and equipment

Experimental data were collected in a recirculation tank at the National Center for Physical Acoustics (NCPA) in collaboration with the USDA-ARS-National Sedimentation Laboratory. A 110 gal (416.9 l) cylindrical tank was used to recirculate and suspend the fine sediment particles, as illustrated in Fig. 1. The dimensions of the tank were 126 cm long, 85 cm wide, and 51.5 cm deep. The water and fine sediment particles were recirculated using a Weg $\frac{1}{2}$ hp 220 V centrifugal pump. The tank had extruded aluminum rails mounted on its top to allow three-dimensional alignment of the transducers. A point gauge was mounted to the rails to verify that the same water level was used during the measurements.

Two 20 MHz immersion transducers were used in a pitch-catch configuration to send and receive acoustic signals. The transducers were placed 4.5 cm under the water surface and separated by 2.8–36.0 cm in 2.54 cm increments. At each range setting, the transducers were aligned by adjusting their position until the maximum signal amplitude was achieved, thus ensuring that the active elements were aligned with one another. The data were collected and averaged for 1000 bursts per range setting. Each experiment was repeated three times and the results were averaged. Averaging was used to remove statistical variations due to the random relative motion of the scatterers.[16]

The acoustic data collection system consisted of off-the-shelf components for transmitting and acquiring acoustic data. A custom written LABVIEW program was used to operate the instruments. A Hewlett Packard 3314A function generator transmitted the acoustic signal. The signal was set to 10 V_{peak}, 100 cycles, with a 10 ms delay between bursts to allow the reverberant signal to completely dissipate. The signal was sent and received with NDT Systems IBHG202 20 MHz immersion transducers with a $\frac{1}{4}$ in. (3 mm diameter) active element. The near-field length, N, for these transducers is given as 136 mm and the half angle beam width is 0.365°. Thus, we expect to be well in the far field at a distance of 230 mm and to have no multipath effects from the surface. The received acoustic signal was amplified by an Olympus 5682 500 KHz–25 MHz preamplifier,

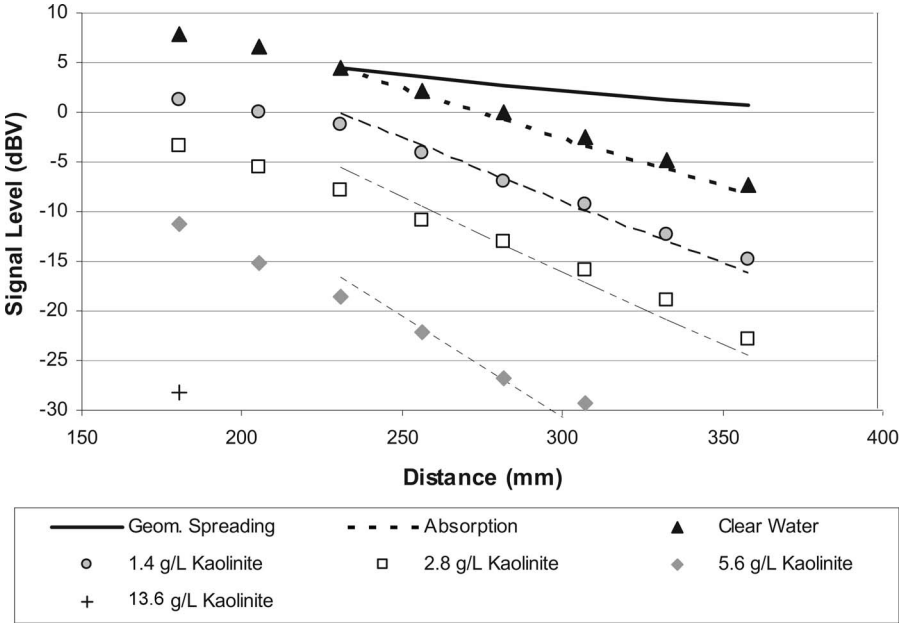


Fig. 2. Signal level (dBV) for kaolinite concentrations (1.4–13.6 g/l).

and then captured with a National Instruments two-channel, 8-bit, 1 GS/s per channel oscilloscope card (NI PXI-PCI-5152).

Kaolinite and bentonite clays with concentrations of 1–14 g/l in tap water (70–72 °F) were used in the experiments. The kaolinite particles had a particle diameter range 2–5 μm. The bentonite particles had a particle diameter of 0.2–1 μm. The kaolinite particles stayed well-separated due to the recirculation pump, but the bentonite particles tended to flocculate. Therefore, prior to adding the bentonite to the circulation tank, sodium hexametaphosphate, a deflocculant, was added in a 1:100 ratio and allowed to completely dissolve.

Duplicate 1-l pump samples were taken in pre-tared flasks at each concentration. Samples of the fluid/sediment mixture were used to verify that a uniform distribution of particles in the tank was achieved. The physical samples were obtained using a vacuum pump connected by a hose to a small bronze tube near the receive transducer. Concentrations were determined by weighing each fluid/sediment sample, decanting the sample, and placing the sediment into pre-tared pans that were oven dried at 60 °C for approximately 48 h. Then, the flasks were re-weighed to obtain particle mass and sediment concentration.

3. Theoretical model

In the following expression, the received voltage, in the far field, as a function of range and concentration is determined by both spherical spreading and absorption due to either pure water or the water-sediment mixture.[17]

$$V = \frac{V_0 R_0}{R} e^{-\alpha(R-R_0)}. \tag{1}$$

The reference voltage, V_0, is recorded at the reference position, R_0. The reference data point V_0R_0 must be selected outside of the near-field of the transducer. The measured voltage, V, corresponds to the position, R. The total sound attenuation coefficient, α, is a function of water temperature and suspended-particle concentration. It should be noted that the purpose of this work is not necessarily to determine the attenuation of the sediment from scattering theory,[17] but rather to determine if the presence of suspended particles in concentrations that commonly occur in streams will cause enough signal attenuation to be usable for field measurements.

Table 1. Kaolinite concentrations and calculated absorption coefficients.

Concentration (g/l)	Absorption (Np/m)
1.4	11.2
2.8	13.7
5.6	19.6

4. Experimental findings

Figure 2 shows the results for signal loss in the range between 180 and 360 mm for both pure water and kaolinite mixtures. The geometrical spreading curve shows the 6 dB loss per doubling of distance (6 dB/DD) associated with simple spherical spreading while the data show the significant contribution of attenuation to signal loss from both tap water and suspended-sediment mixtures. For the purpose of prediction and evaluation of an absorption coefficient, the data at a range of 230 mm were used as the start of the far field as the data closer than this show the residual effect of the near field. The absorption coefficient for the water used in these experiments was found to be 8.3 Np/m, which is comparable to the 8.7 Np/m one can calculate for distilled water. Any discrepancy, perhaps due to specific water chemistry, would be present for both the reference clear water and water-sediment mixtures both in the laboratory experiment and ultimately field measurements. Absorption coefficients were estimated for concentrations of kaolinite ranging from 1.4–5.6 g/l using a best fit to the data with Eq. (1) and are shown in Table 1. Significant attenuation for the 13.6 g/l concentration was observed at 180 mm, but past this range, the data fell into the noise and was unusable to estimate the absorption coefficient.

Figure 3 shows the signal level changes for bentonite with concentrations ranging from 1.7–13.4 g/l. Figure 3 also shows the predicted signal levels for geometrical spreading and absorption in clear and sediment laden water at pre-determined distances for data past 230 mm. Similar to the kaolinite measurements, as the concentration increases, absorption increases and the signal level decreases. However, the signal levels for the bentonite mixtures are higher and the absorption is lower than in the kaolinite mixtures. Kaolinite particles are larger than bento-

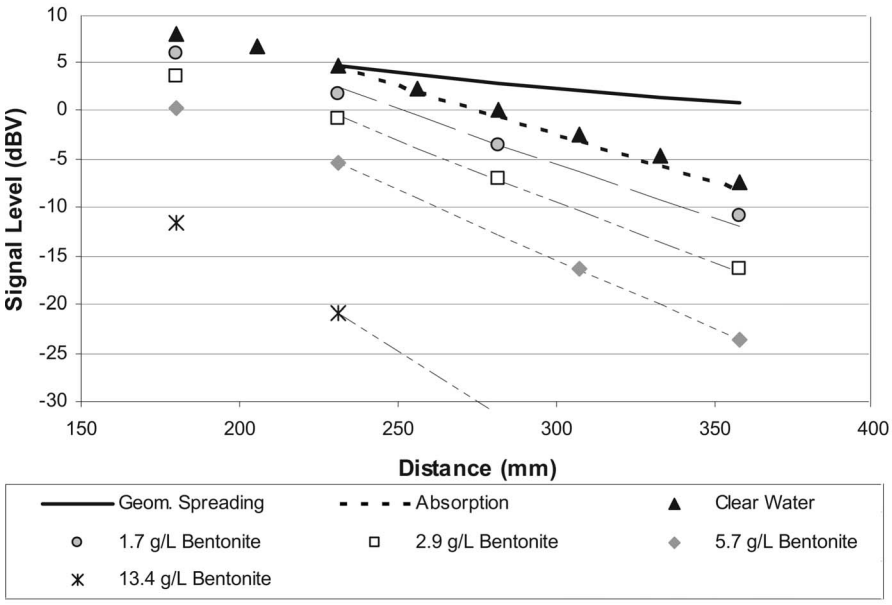


Fig. 3. Signal level (dB/V) for bentonite concentrations (1.7–13.4 g/l) at pre-determined distances.

Table 2. Bentonite concentrations and calculated absorption coefficients.

Concentration (g/l)	Absorption (Np/m)
1.7	9.5
2.9	11.2
5.7	13.0
13.4	20.6

nite, and should be expected to scatter and attenuate more signal than a similar concentration of bentonite particles. The calculated attenuations for the bentonite mixtures are shown in Table 2.

While we have performed calculations to investigate the attenuation beginning at 230 mm and beyond, the data at 180 mm have significance in that we are able to investigate the full range of sediment concentrations considered. At distances greater than 180 mm, the attenuation caused by high concentrations is greater than the dynamic range of our instrumentation and the signal is lost in the noise. Conversely, at ranges closer than 180 mm (data not presented), it was difficult to distinguish the attenuation of the lowest sediment concentrations from that of clear water. Thus, Fig. 4 shows signal level data (relative to clear water) for various kaolinite and bentonite concentrations at 180 mm. The signal level loss due to water attenuation (approximately 8 dBV) has been subtracted from these data to show the effect of increasing concentration. As indicated in the plot, an increase in suspended-sediment concentration resulted in a corresponding decrease in signal level. An increase in particle size diameter (i.e., kaolinite particles are larger than bentonite particles) also resulted in a decrease in signal level. The 20 MHz acoustic system was able to measure signal loss over the entire range of concentrations considered at the 180 mm range.

The data in Fig. 4 could be used for the inverse problem of interest. That is, one could use the signal level loss to calculate sediment concentration. However, the particle size will either need to be known or estimated *a priori*. Furthermore, if the particle size has a wide distribution, unlike the narrow range used here, interpolation between the results or expansion

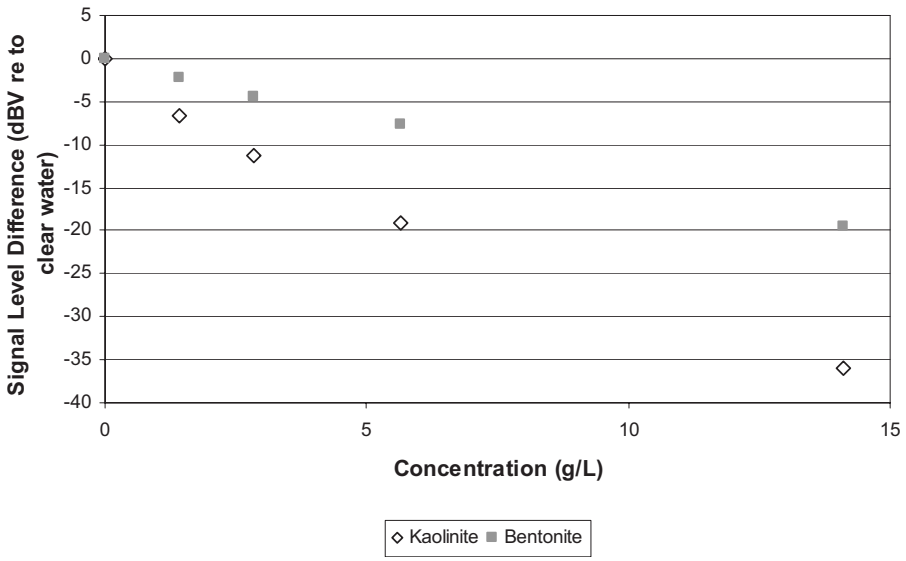


Fig. 4. Signal level difference (dB/V) for increasing kaolinite and bentonite concentrations relative to clear water at a distance of 180 mm.

of the attenuation database to account for a greater range of sizes would be necessary. Alternately, multiple frequencies or a combination of forward and backscatter data could be used to simultaneously estimate particle size and concentration. These topics are part of ongoing research and are not covered here.

5. Conclusion

The feasibility of measuring particle concentration using the attenuation of high frequency acoustic waves traversing a suspension of fine particles in water was investigated in a series of laboratory experiments. The data showed that as concentration increased, attenuation increased and signal level decreased over a distance and using a frequency that are physically reasonable for field measurements. The use of a 20 MHz pitch/catch transducer arrangement to measure the concentration of 5 μm particles, or smaller, suspended in water appears to be practical at a range of distances from 180 to 357 mm. The tested concentration ranges were 1.4–13.6 g/l for kaolinite and 1.7–13.4 g/l for bentonite. This work suggests that fixed distance measurements of approximately 180 mm can be used to evaluate the broad range of concentrations expected in the field.

Acknowledgment

This work was funded by the Federal Interagency Sedimentation Project (FISP).

References and links

[1]J. N. Holeman, "The sediment yield of major rivers of the world," Water Resour. Res. **4**, 737–747 (1968).
[2]M. E. Nelson and P. C. Benedict, "Measurement and analysis of suspended loads in streams," Trans. Am. Soc. Civ. Eng., Proceedings, No. 31, Paper No. 2450, 891–918 (1950).
[3]C. Shen and U. Lemmin, "Ultrasonic measurements of suspended sediments: A concentration profiling system with attenuation compensation," Meas. Sci. Technol. **7**, 1191–1194 (1996).
[4]P. D. Thorne, G. P. Holdaway, and P. J. Hardcastle, "Constraining acoustic backscatter estimates of suspended sediment concentration profiles using the bed echo," J. Acoust. Soc. Am. **98**(4), 2280–2288 (1995).
[5]D. G. Wren, B. D. Barkdoll, R. A. Kuhnle, and R. W. Derrow, "Field techniques for suspended-sediment measurement," J. Hydrol. Eng. **126**, 97–104 (2000).
[6]R. A. Kuhnle, S. J. Bennett, C. V. Alonso, R. L. Bingner, and E. Langendoen, "Sediment transport processes in agricultural watersheds," Int. J. Sediment Res. **15**, 182–197 (2000).
[7]R. A. Kuhnle, D. G. Wren, and J. P. Chambers, "Prediction of the grain size of suspended sediment: Implications for calculating suspended sediment concentrations using single frequency acoustic backscatter," Int. J. Sediment Res. **22**, 1–15 (2007).
[8]E. W. Lane, "Notes on limit of sediment concentration," J. Sediment Petrol. **10**, 95–96 (1940).
[9]C. T. Yang, F. J. M. Simoes, "Wash load and bed-material load transport in the Yellow River," J. Hydrol. Eng. **131**(5), 415–416 (2005).
[10]D. H. Green and R. Esquivel-Sirvent, "Acoustic behavior at the fluid/solid transition of kaolinite suspensions," J. Geophys. **64**(1), 89 (1999).
[11]D. Eisma, *Suspended Matter in the Aquatic Environments* (Springer-Verlag, Berlin, 1993).
[12]B. Kowallis, L. E. A. Jones, and H. F. Wang, "Velocity-porosity-clay content: Systematics of poorly consolidated sandstones," J. Geophys. **89**, 10355–10364 (1984).
[13]D. Marion, A. Nur, H. Yin, and D. Han, "Compressional velocity and porosity in sand-clay mixtures," J. Geophys. **57**, 554–563 (1992).
[14]M. S. Greenwood, J. L. Mai, and M. S. Good, "Attenuation measurements of ultrasound in a kaolin-water slurry: A linear dependence upon frequency," J. Acoust. Soc. Am. **94**(2), 908–916 (1993).
[15]J. Y. Sheng and A. E. Hay, "An examination of the spherical scatterer approximation in aqueous suspensions of sand," J. Acoust. Soc. Am. **83**, 598–610 (1988).
[16]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Saunders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), p. 161.
[17]A. E. Hay, "Sound scattering from a particle-laden, turbulent jet," J. Acoust. Soc. Am. **90**(4), 2055–2074 (1991).

 [DOI: 10.1121/1.3258062] Published Online 16 November 2009

The intelligibility of pointillistic speech

Gerald Kidd, Jr., Timothy M. Streeter, Antje Ihlefeld, Ross K. Maddox, and Christine R. Mason
Hearing Research Center, Boston University, Boston, Massachusetts 02215
gkidd@bu.edu, timstr@cns.bu.edu, antje1@gmail.com, rkmaddox@bu.edu, cmason@bu.edu

Abstract: A form of processed speech is described that is highly discriminable in a closed-set identification format. The processing renders speech into a set of sinusoidal pulses played synchronously across frequency. The processing and results from several experiments are described. The number and width of frequency analysis channels and tone-pulse duration were variables. In one condition, various proportions of the tones were randomly removed. The processed speech was remarkably resilient to these manipulations. This type of speech may be useful for examining multitalker listening situations in which a high degree of stimulus control is required.

PACS numbers: 43.66.Mk, 43.72.Gy, 43.72.Ja [QJF]
Date Received: August 29, 2009 **Date Accepted:** October 9, 2009

1. Introduction

It has been demonstrated in a variety of ways that human speech is remarkably resilient and can convey meaning even under conditions of extreme distortion. For example, early work on "infinitely" peak-clipped speech revealed that the speech retained a high degree of intelligibility (e.g., Licklider and Pollack, 1948). It was also found that speech could be interrupted frequently and yet still be understood (Miller and Licklider, 1950). Furthermore, it has long been known that the information in speech is distributed across a wide range of frequencies and conveys meaning through the variation over time within these different frequency channels. Limiting the information to a subset of channels can provide some (highly predictable) degree of intelligibility (e.g., French and Steinberg, 1947). More recently interest in the essential aspects of speech has increased due to the development of cochlear implants. Shannon *et al.* (1995) described a means for simulating cochlear implant processing and demonstrated that such speech could be highly intelligible. In their procedure, the amplitude envelopes of several bands of speech are extracted and used to modulate noiseband carriers limiting the speech cues primarily to those conveyed by the envelopes. This type of "vocoded" speech has a long history (Dudley, 1939) and has been utilized and modified in various ways by many recent investigators (e.g., Dorman *et al.*, 1997; Loizou *et al.*, 1999; Arbogast *et al.*, 2002; Qin and Oxenham, 2003; Yang and Fu, 2005; Brungart *et al.*, 2005; Nie *et al.*, 2005; Poissant *et al.*, 2006; Throckmorton *et al.*, 2006; Stickney *et al.*, 2007; Whitmal *et al.*, 2007; Souza and Rosen, 2009).

A method of representing speech is presented here which combines some features of previous methods in addition to more severely quantizing the information. This speech is referred to as "pointillistic speech" because in the limit the speech signal is reduced to a time-frequency matrix of points with each point (or "element") consisting of a brief pulsed "pure" tone represented by only two values (its frequency and amplitude). Using this technique, speech identification results are reported providing a parametric examination of the effects of manipulating variables in the processing (number of time elements, number of frequency analysis channels, and proportion of resulting matrix removed).

2. Methods

2.1 The processing algorithm

The speech was filtered into 4, 8, or 16 contiguous frequency bands spaced logarithmically with a total range of 267–10667 Hz. Within each analysis band, the Hilbert magnitude and phase were

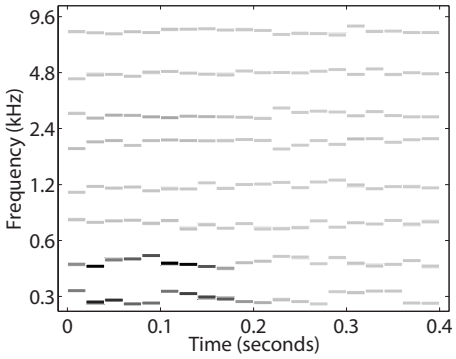


Fig. 1. A schematic illustration of the result of the processing in which the speech signal is replaced by a matrix of tone bursts. In this example there were eight tones per time window that were each 20 ms.

computed yielding functions describing the variation in the amplitude envelopes and instantaneous phases over time. The average values of the envelope and frequency were computed in each band for discrete contiguous time windows of 10, 20, 40, and 80 ms. The envelope value was calculated by squaring the mean of the absolute value of the Hilbert envelope and the frequency value was determined by taking the first time-derivative of the instantaneous phase function averaged over all positive frequencies. This yielded two numbers representing the stimulus at each time-frequency point. Then, a 0°-phase sinusoid of that amplitude and frequency at the total duration of the time window (including a 3-ms rise-decay) was created. The resulting sinusoidal elements were concatenated in time and summed across frequency producing sets of temporally contiguous, non-overlapping, synchronously gated tones. Figure 1 illustrates the result of this process for a single word under one of the conditions (8 tones, 20 ms) used in Exp. 1. The resulting speech has some of the envelope and fine structure information preserved, as in cochlear implant simulation (vocoded) speech (e.g., Nie *et al.*, 2005; Throckmorton *et al.*, 2006; Stickney *et al.* 2007), combined with more quantized envelope time segments (e.g., Loizou *et al.*, 1999; Brungart *et al.*, 2007; Li and Loizou, 2008).

2.2 Listeners

A total of 21 paid listeners (ages 19–31) participated in this study with 11 listeners participating in Exp. 1, 6 listeners participating in Exp. 2, and 6 listeners participating in Exp. 3. Two of the listeners participated in both Exps. 1 and 3 but none had participated in previous experiments in this laboratory.

2.3 Stimuli

The speech was from a laboratory-designed monosyllabic corpus (Kidd *et al.*, 2008) consisting of eight tokens from each of five categories: ⟨*subject*⟩ ⟨*verb*⟩ ⟨*number*⟩ ⟨*adjective*⟩ ⟨*object*⟩. In most conditions, for each trial, five words (one from each category without replacement) from one randomly chosen talker (of eight males and eight females) were concatenated in syntactically correct order. For one condition in Exp. 3, five of the entire 40 words were randomly chosen and concatenated on each trial producing sequences that were very unlikely to be syntactically correct.

2.4 Procedures

Listeners were seated in a sound-treated IAC chamber wearing Sennheiser HD280 Pro earphones. Stimuli were presented diotically through Tucker-Davis Technologies hardware at 60 dB sound pressure level. Listeners were instructed to report all five keywords by clicking on a response graphical user interface (GUI). The GUI had a button for each keyword organized in columns according to the word categories and within each column sorted in alphabetical (or in the case of numbers, numerical) order. Each keyword was scored individually and the listeners received no feedback. In Exps. 1 and 2, each listener was tested in 12 conditions: 4 time windows by 3 number-

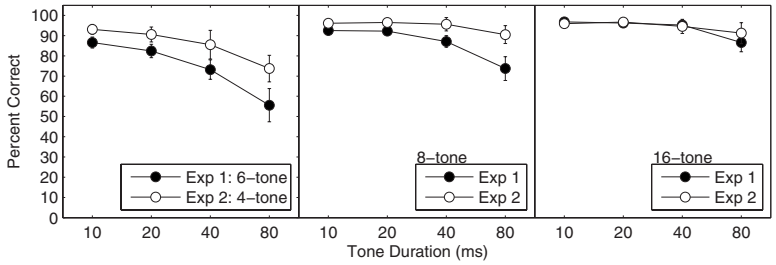


Fig. 2. Group mean word identification from Exps. 1 and 2. Error bars are ±1 standard deviation.

of-analysis-bands cases. All 12 conditions were presented three times in random order in each block of five-word trials and each listener completed 12 blocks. In Exp. 1, the tones for every band, every other band, or every third band of the 16-band case (starting with the lowest frequency) were used yielding 6, 8, or 16 simultaneous tones in each time window. For Exp. 2, the 16-tone case was identical and the 4- and 8-tone cases (in which the original analysis bands were wider) were also tested. In Exp. 3, only one combination of time windows and number of tones was tested (16 tones, 10 ms), while the number of elements in the signal was manipulated by randomly removing various proportions (0, 0.5, 0.66, 0.75, and 0.875) of the time-frequency bins representing the signal on a per-word per-trial basis. All proportions were tested for both syntactically correct and random order five-word utterances. Each block consisted of 15 trials for each of the five proportions in one of the word order conditions. Eight blocks were completed by each listener. Each experiment lasted about 2 hours.

3. Results

3.1 Experiments 1 and 2: Effect of number of analysis bands and duration of time windows

The results of Exps. 1 and 2 are shown in Fig. 2. Pointillistic speech, at least as assessed in this closed-set paradigm, is highly identifiable under certain combinations of stimulus parameters. Predictably, for any time window, performance was best for the highest number of tones and declined as that number decreased. Furthermore, for any given number of tones, performance generally decreased as the tone duration increased beyond 20 ms. The highest group-mean scores overall were obtained for the 16-tone signal at 10, 20, or 40 ms tone durations where the group mean values obtained were approximately 95% correct or better. However, even for the conditions with fewer tones, performance was better than 85% correct for the 10 ms duration. The poorest performance was for the fewest tones and longest time segments which provided the coarsest representation. However, even in that condition listeners achieved about 55% correct identification when 6 of the 16 analysis bands were included and almost 74% correct when the information was extracted from four wide bands covering the entire range. The variability across listeners was generally rather small but increased for the conditions under which performance was relatively poor. The 16-tone case allows a comparison of the two groups of listeners. Overall the results are remarkably similar with the 6 listeners in Exp. 2 performing slightly better than the 11 in Exp. 1 for the longest duration. In general, performance is better when the information comes from a wider analysis band (Exp. 2) than when narrow bands are used but some are excluded. In the left panel this is true even when the wide band analysis is only four tones as compared to six tones from every third band in the original set of 16.

3.2 Experiment 3: Effect of random removal of elements

The results are shown in Fig. 3 for the syntactically correct and random word order conditions. Rather remarkably the speech was highly robust with respect to this type of distortion. When the target words were presented in correct syntactic order, removing 2/3 of the elements only reduced identification performance to 92% correct. For comparison, the asterisks are results from Exp. 1 for the 10-ms six-tone and eight-tone cases. These correspond to proportions removed of

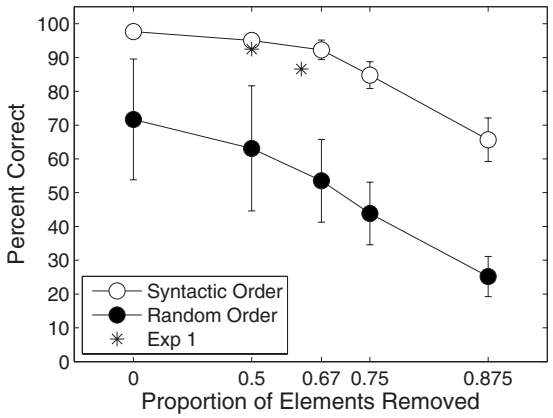


Fig. 3. Group mean word identification from Exp. 3. Error bars are ±1 standard deviation.

0.625 and 0.5, respectively. As can be seen, performance is somewhat worse when the removed elements are from fixed frequency regions across all time. For words in random order, the effect of removing elements was greater, with mean performance of about 52% correct when only 1/3 of the elements remain dropping to about 25% correct when only 1/8 of the elements remain. The difference between syntactic and random order doubtless reflects the advantage of appropriate contextual order as well as some inherent differences in chance performance related to guessing or the ability to match the distorted token to 1 of 8 versus 1 of 40 alternatives. There were also larger differences across listeners for the random word order conditions. However, it is clear that these words may be discriminable in a closed-set identification paradigm when the spectrotemporal representation is quite sparse.

4. Discussion

There is a vast literature and long history describing various methods of distorting or recoding the speech stimulus and evaluating those effects on intelligibility. While a comprehensive review of that literature is beyond the scope of this letter, there are a few methods that are sufficiently similar to the current method as to deserve comment. As noted above, there is considerable interest in the essential aspects of speech that must be preserved in the processing provided by cochlear implants. Shannon *et al.* (1995) demonstrated that speech-envelope-modulated noise could be quite intelligible if a sufficient number of frequency bands were used. Brungart *et al.* (2005) compared three variations in speech processing and found that each method provided impressively high intelligibility, at least with a sufficient number of bands and high signal-to-noise ratios in closed-set tasks. The methods they compared were noise band or tonal carriers modulated by the speech envelopes for fixed frequency bands, and "sinewave speech" tracking formants like that used by Remez *et al.* (1981). Not surprisingly, they reported that intelligibility declined with decreasing numbers of frequency bands for all of the methods. In the current study, reducing the resolution of the speech (smaller number or longer duration of tones) degraded performance, likely due to an inadequate representation of important contrasts signaling phoneme identity. In Exp. 2, the number of contiguous bands from which the pointillistic speech was derived was varied and compared to speech processed into a subset of narrower analysis bands. Generally, the wider bands yielded better identification than the narrower bands. This finding may be related to that reported recently by Souza and Rosen (2009) who found better performance in speech recognition of sine-carrier speech for higher envelope cutoff frequencies. The processing method described here is similar in its initial stages to vocoded speech in which the carriers are pure tones modulated by narrow-band-derived envelope functions (see Arbogast *et al.*, 2002). One difference is that both frequency and amplitude are represented separately in the elements comprising the signal. However, this is not unique, either. Several investigators have demonstrated that adding slowly varying frequency modulation to

individual frequency channels previously conveying only amplitude envelope information provided speech intelligibility advantages especially in multiple talker situations (e.g., Nie *et al.*, 2005; Stickney *et al.*, 2007). This procedure is similar to the current method in that it preserves some of both the envelope and frequency information in the channels. The primary difference between that technique and the one presented here is that the within-channel variations in envelope and frequency were continuous throughout the stimulus rather than being severely quantized and replaced by equal-phase tone segments. Also similar is the procedure used by Throckmorton *et al.* (2006) in which the effect of limiting the carrier frequency to one of a small set of discrete frequencies within each channel for 2-ms time windows was investigated. Of course all digital representations of speech are quantized although the sampling rates used typically give resolution on the order of tens of microseconds. Longer time scale quantization has been used in cochlear implant coding methods and in simulated cochlear implant speech on the order of 4 ms in which the resolution in amplitude of the pure-tone carriers was varied in discrete segments or steps (Loizou *et al.*, 1999). Brungart *et al.* (2007) demonstrated that a uniform broadband noise filtered into the time-frequency regions (1/3-octave bands in 7.8 ms segments) in which the speech that it mimicked had 90% of its energy could also convey meaning. The current findings indicate that even longer discrete time segments containing both frequency and amplitude information may also result in good intelligibility under certain conditions. While it is difficult to compare performance across studies due to differences in processing as well as speech materials, Loizou *et al.* (1999) found that listeners could achieve 90% correct or better with five or more channels using open set sentence identification when the duration of each tone was 4 ms and the frequency of the sinusoidal carriers did not vary.

The nature of the speech stimulus in the current study is also similar to speech processed using a variation in the ideal binary mask for "ideal time-frequency segregation" (e.g., Brungart *et al.*, 2006, 2007; Li and Loizou, 2007, 2008; Kjems *et al.*, 2009). Analogous to the findings in Exp. 3, these authors noted that high proportions of the time-frequency bins in their approach could be removed or masked while intelligibility was retained. The similarity between the different approaches lies in the rendering of the speech stimulus into a matrix of discrete spectrotemporal units that may be analyzed or manipulated independently. However, the current method requires very little stored information to reconstruct the stimulus, only two values per matrix element. Unlike here, the ideal time-frequency algorithm usually preserves the original speech stimulus in those time-frequency units (although the Li and Loizou (2008) study also applied the binary mask technique to vocoded stimuli) and is used to separate a target source from an overlapping speech masking source using *a priori* knowledge of each prior to the mixture. It is possible, and we are currently exploring this issue, to represent two distinct sources using acoustically non-overlapping sets of tones in the current procedure. The results of Exp. 3, which demonstrated that high identification scores may be obtained in a closed-set format even when one-half or more of the elements are randomly (compared to "ideally") removed, suggest that such a multisource approach is feasible. Thus two, or perhaps even three, intelligible speech sources could be represented with mutually exclusive tonal elements. Presumably, relative level would be the primary cue in determining the most salient source although this issue awaits future investigation. Also of interest is the relative proportion of energetic and informational masking for various methods. The two techniques share many similarities and we plan future work to compare and contrast the two approaches.

Acknowledgments

This work was supported by grants from NIH/NIDCD and AFOSR.

References and links

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., and Kidd, G., Jr. (**2005**). "Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**, 292–304.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-

on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.
Brungart, D., Iyer, N., and Simpson, B. (**2007**). "Speech perception from a crudely quantized spectrogram: A figure-ground analogy (A)," J. Acoust. Soc. Am. **121**, 3186.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.
Dudley, H. (**1939**). "The vocoder," Bell Lab. Rec. **18**, 122–126.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Kidd, G., Jr., Best, V., and Mason, C. R. (**2008**). "Listening to every other word: Examining the strength of linkage variables in forming streams of speech," J. Acoust. Soc. Am. **124**, 3793–3802.
Kjems, U., Boldt, J., Pedersen, M., Lunner, T., and Wang, D. (**2009**). "Role of mask pattern in intelligibility of ideal binary-masked noisy speech," J. Acoust. Soc. Am. **126**, 1415–1426.
Licklider, J. C. R., and Pollack, I. (**1948**). "Effects of differentiation, integration and infinite peak clipping on the intelligibility of speech," J. Acoust. Soc. Am. **20**, 42–51.
Li, N., and Loizou, P. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**, 1165–1172.
Li, N., and Loizou, P. (**2008**). "Effect of spectral resolution on the intelligibility of ideal binary masked speech," J. Acoust. Soc. Am. **123**, EL59–EL64.
Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–2103.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
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.
Poissant, S. F., Whitmal, N. A., III, and Freyman, R. L. (**2006**). "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," J. Acoust. Soc. Am. **119**, 1606–1615.
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.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Souza, P., and Rosen, S. (**2009**). "Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech," J. Acoust. Soc. Am. **126**, 792–805.
Stickney, G. S., Assman, P. F., Chang, J., and Zhang, F.-G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.
Throckmorton, C. S., Kucukoglu, M. S., Remus, J. J., and Collins, L. M. (**2006**). "Acoustic model investigation of a multiple carrier frequency algorithm for encoding fine frequency structure: Implications for cochlear implants," Hear. Res. **218**, 30–42.
Whitmal, N. A., III, Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.
Yang, L., and Fu, Q.-J. (**2005**). "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," J. Acoust. Soc. Am. **117**, 1001–1004.

LETTERS TO THE EDITOR

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

A special relation between Young's modulus, Rayleigh-wave velocity, and Poisson's ratio (L)

Peter G. Malischewsky[a] and Tran Thanh Tuan
Institute for Geosciences, Friedrich-Schiller University Jena, Burgweg 11, D-07749 Jena, Germany

(Received 9 March 2009; revised 17 August 2009; accepted 1 September 2009)

Bayón *et al.* [(2005). J. Acoust. Soc. Am. **117**, 3469–3477] described a method for the determination of Young's modulus by measuring the Rayleigh-wave velocity and the ellipticity of Rayleigh waves, and found a peculiar almost linear relation between a non-dimensional quantity connecting Young's modulus, Rayleigh-wave velocity and density, and Poisson's ratio. The analytical reason for this special behavior remained unclear. It is demonstrated here that this behavior is a simple consequence of the mathematical form of the Rayleigh-wave velocity as a function of Poisson's ratio. The consequences for auxetic materials (those materials for which Poisson's ratio is negative) are discussed, as well as the determination of the shear and bulk moduli. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3243464]

PACS number(s): 43.20.Jr, 43.20.Bi [SFW] Pages: 2851–2853

I. INTRODUCTION

This paper investigates a special relation between Young's modulus E, Rayleigh-wave velocity v_R, and Poisson's ratio ν. Bayón *et al.* (2005) observed in determining the elastic behavior of a material by measuring the properties of ultrasonic Rayleigh waves (velocity and ellipticity) propagating over one face of the material a strange connection between two quantities that seemed to be totally independent, namely, the quantities ν and $B \equiv E/(\rho v_R^2)$, which are considered as functions of the reciprocal ellipticity a. We use here the same notation as in Bayón *et al.*, 2005: ν is Poisson's ratio, E is Young's modulus, ρ is density, and v_R is the Rayleigh-wave velocity. They concluded that a linear relationship between ν and B seems to exist and it was approximately deduced as

$$B(\nu) = 2.618 + 1.332\nu. \tag{1}$$

However, an analytical explanation for this peculiar relation has not been found.

II. EXPLANATION FOR THE LINEARITY OF *B*

Let the starting-point be the following well-known formula connecting E, ρ, ν, and shear-wave velocity v_S:

$$E = 2(1+\nu)\rho v_S^2 = 2(1+\nu)\rho v_R^2/\eta, \tag{2}$$

where we have introduced in accordance with Bayón *et al.* (2005) the squared ratio of v_R to v_S as follows:

$$\eta = (v_R/v_S)^2. \tag{3}$$

By using Malischewsky's formula (see Malischewsky Auning, 2004) and the squared ratio of shear-wave velocity to velocity of longitudinal waves v_P denoted by γ,

$$\gamma = \left(\frac{v_S}{v_P}\right)^2 = \frac{1-2\nu}{2(1-\nu)}, \tag{4}$$

we obtain η as

$$\eta = \frac{2}{3}\left[4 - \sqrt[3]{h_3(\gamma)} + \frac{2(1-6\gamma)}{\sqrt[3]{h_3(\gamma)}}\right], \tag{5}$$

with the auxiliary functions

$$h_1(\gamma) = 3\sqrt{33 - 186\gamma + 321\gamma^2 - 192\gamma^3},$$

$$h_3(\gamma) = 17 - 45\gamma + h_1(\gamma), \tag{6}$$

where the main values of the cubic roots are to be used; i.e., the cubic root of a negative number is complex. It should be noted that there are a lot of different approximations for the Rayleigh wave velocity such as the well-known approximations of Achenbach (1973) and Bergmann (1948), which were discussed and improved by Vinh and Malischewsky (2007). It is clear that $\eta = \eta(\nu)$ is a non-linear function of ν and $B = B(\nu)$ for the present, too. We expand now the interval of considered Poisson ratios from -1 to 0.5 because of two reasons: The auxetic range of negative Poisson ratios becomes more and more interesting in material sciences (see,

[a]Author to whom correspondence should be addressed. Electronic mail: p.mali@uni-jena.de

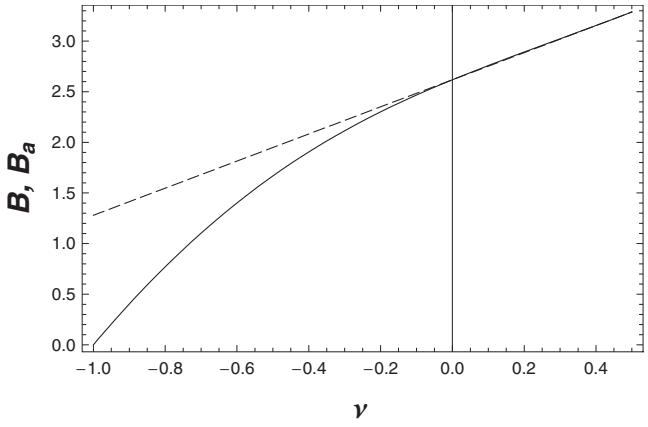


FIG. 1. Exact function B (solid) and its linear approximation (dashed) as a function of Poisson's ratio ν from Eqs. (2)–(6).

e.g., Yang *et al.*, 2004) and, second, the special behavior of B in the interval [0,0.5] becomes more obvious. If we plot the function $B \equiv E/(\rho v_R^2)$ from Eq. (2) by using Eqs. (4) and (5),

$$B = 2(1+\nu)/\eta, \tag{7}$$

and approximation (1), which is nearly identical to the result in Eq. (16), into Fig. 1, it can be realized that B is a quasi-linear function of ν in the interval [0,0.5], which becomes also obvious by calculating the curvature as a function of ν. This behavior becomes understandable by carrying out a Taylor expansion of function $1/\eta(\nu)$ up to the quadratic term $(1.309-0.59\nu+0.37\nu^2)$. The higher terms will not bring new information in the interval $\nu \in [0,0.5]$ under consideration. Multiplication of this Taylor expansion with $2(1+\nu)$ yields a cubic equation, which is a fairly good approximation of $B(\nu)$ and has the following peculiarity: The quadratic and cubic terms cancel approximately with the consequence that the linear term is the leading term. That is the analytical reason for this behavior of $B(\nu)$. By using a cubic test function with arbitrary quadratic and cubic coefficients, we observe that this quasi-linear behavior of $B(\nu)$ corresponds just with that region of the cubic function, where it has a point of inflection within the interval $\nu \in [0,0.5]$, in our case nearby $\nu=0.3$. This is intuitively clear. There is no point of inflection for $\nu \in [-1,0]$.

By making the linear ansatz $B_a = c + d\nu$ with unknown constants we obtain from Eq. (7)

$$\eta = \frac{2(1+\nu)}{c+d\nu}, \tag{8}$$

and by taking into account that this η has to satisfy Rayleigh's equation for every ν, i.e.,

$$\Delta(\nu,c,d) = (2-\eta)^4 - 16(1-\eta)\left[1 - \frac{1-2\nu}{2(1-\nu)}\eta\right] = 0. \tag{9}$$

After some simple algebraic manipulations, Eq. (9) becomes

$$F(c) + \nu(\nu+1)G(c,d,\nu) = 0, \quad \forall\ \nu, \tag{10}$$

where

$$F(c) = (c-1)(c^2-3c+1) \tag{11}$$

and $G(c,d,\nu)$ is a polynomial of third order with respect to d. For $\nu=0$ Eq. (10) becomes $F(c)=0$, and we are able to determine c from the equation

$$F(c) = (c-1)(c^2-3c+1) = 0. \tag{12}$$

It is clear from Fig. 1 that the only suitable solution is

$$c = \frac{3+\sqrt{5}}{2} \approx 2.61803. \tag{13}$$

This remarkable number is the exact value of B for the Poisson ratio 0, which could be also obtained from the exact formula (5) by some arithmetic transformations in a less elegant manner. A by-product of Eq. (13) is an analytical representation of the Rayleigh-wave velocity for the Poisson ratio 0 as follows:

$$v_R(0)/v_S = 2/\sqrt{3+\sqrt{5}} \approx 0.874. \tag{14}$$

The other unknown constant d could be found by solving the implicit function $G(2.61803,d,\nu)=0$ and it is a function of ν. Because we have already realized that B is almost a straight line for $\nu \in [0,0.5]$, we can determine the constant d simply in such a manner that the approximation is also exact at the other end of the interval

$$B(0.5) = 3.28723 = \tfrac{1}{2}(3+\sqrt{5}) + 0.5d \Rightarrow d = 1.3384. \tag{15}$$

So we have obtained the following approximation of B in the interval [0, 0.5]:

$$B_a = 2.61803 + 1.3384\nu, \tag{16}$$

which is only slightly different from approximation (1) of Bayón *et al.* (2005), but the relative mean deviation from the exact value is 1.4-fold less than that of Eq. (1). We could obtain a yet better approximation than Eq. (16) by freely varying c and d. But this is not necessary because the error of Eqs. (1) and (16) is small enough (equal or less than 0.2% over the whole interval). We are also able to use a polynomial of third order to get a very good approximation within the whole interval $\nu \in [-1,0.5]$. The polynomial is obtained by carrying out a Taylor expansion of B at a certain ν-value, which is determined by the method of Vinh and Malischewsky (2006). The result is

$$\bar{B}_a = 2.618 + 1.446\nu - 0.521\nu^2 + 0.682\nu^3, \tag{17}$$

with a relative deviation less than 1.2% within the interval [−0.8, 0.5].

Finally let us present B and Poisson's ratio ν as in Bayón *et al.*, 2005 as a function of the reciprocal ellipticity a

$$a = \frac{2-\eta}{2\sqrt{1-\eta}} \tag{18}$$

in Fig. 2. We see the same picture as in Fig. 1 of Bayón *et al.*, 2005 for positive Poisson ratios, but the two curves separate from each other for negative Poisson ratios. Relation (16) is no longer valid in this range but this is no obstacle for the application of the method: ν and B and E, respectively, are determined for constant a from the corresponding two curves without difficulty.

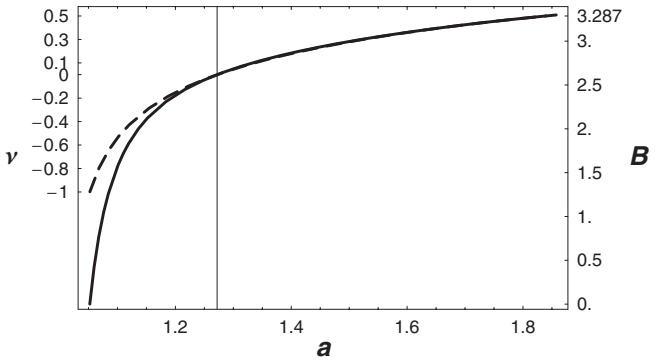


FIG. 2. Poisson's ratio ν (dashed, left-hand scale) and B (full, right-hand scale) as functions of the reciprocal ellipticity a. Regions of negative and positive Poisson ratios, respectively, are separated by a thin vertical line.

III. DETERMINATION OF THE SHEAR AND BULK MODULI

The determination of the shear modulus G and bulk modulus K, respectively, is simple when E and ν are known. However, in order to demonstrate better the contrast in determining these moduli by the method described here, we define similar functions analogously to B for G and K, namely,

$$B_1 = \frac{G}{\rho v_R^2} = 1/\eta, \quad B_2 = \frac{K}{\rho v_R^2} = \frac{2(1+\nu)}{3(1-2\nu)\eta}. \tag{19}$$

The functions B, B_1, B_2, and ν are drawn in Fig. 3 in dependence on a. Only the function B can be shifted down by linear transformation (16) for positive Poisson ratios in order to become almost identical with the function $\nu(a)$. For the other functions such a procedure is not possible, but this is no limitation for the applicability of the method described here. Note the points of intersections in Fig. 3, where certain moduli have identical numerical values for certain Poisson ratios.

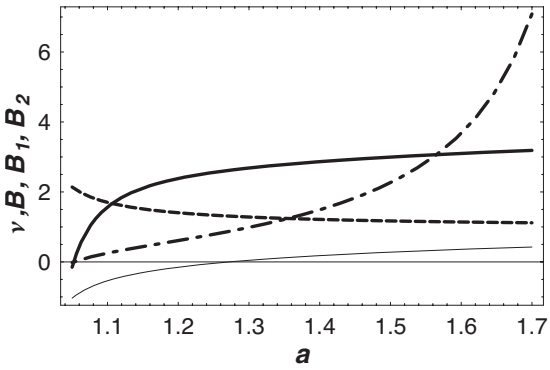


FIG. 3. The functions ν (thin), B (full, thick), B_1 (dashed), and B_2 (dashed-dotted) in dependence on a.

IV. CONCLUSIONS

The quasi-linearity of the function $B(\nu)$ for $\nu > 0$ is an interesting by-product in determining Young's modulus from the measurement of the Rayleigh-wave velocity and its ellipticity. It is a consequence of the special form of the function $1/\eta(\nu)$ in connection with its multiplication with the linear function $(1+\nu)$. This quasi-linearity allows a convenient and simple graphical determination of both quantities ν and E from one curve alone. It breaks down in the auxetic range $\nu < 0$, but this is not a limitation of the applicability of the method. We have only to use two curves in that range. The same is true for the whole range of Poisson ratios if we would determine the moduli G and K in a similar manner.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. MA 1520/6-2. Further, P.G.M. gratefully acknowledges the support of Bundesministerium für Bildung und Forschung (BMBF) in the framework of the joint project "WTZ Germany-Israel: System Earth" under Grant No. 03F0448A.

Achenbach, J. D. (**1973**). *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam).

Bayón, A., Gascón, F., and Nieves, F. J. (**2005**). "Estimation of dynamic elastic constants from the amplitude and velocity of Rayleigh waves," J. Acoust. Soc. Am. **117**, 3469–3477.

Bergmann, L. (**1948**). *Ultrasonics and Their Scientific and Technical Application* (Wiley, New York).

Malischewsky Auning, P. G. (**2004**). "A note on Rayleigh-wave velocities as a function of the material parameters," Geofis. Int. **43**, 507–509.

Vinh, P. C., and Malischewsky, P. G. (**2006**). "Explanation for Malischewsky's approximate expression for the Rayleigh wave velocity," Ultrasonics **45**, 77–81.

Vinh, P. C., and Malischewsky, P. G. (**2007**). "An improved approximation of Bergmann's form for the Rayleigh wave velocity," Ultrasonics **47**, 49–54.

Yang, W., Li, Z.-M., Shi, W., Xie, B.-H., and Yang, M.-B. (**2004**). "On auxetic materials," J. Mater. Sci. **39**, 3269–3279.

Authors' Response to Strasberg's "Comment on 'Measurement of the frequency dependence of the ultrasonic parametric threshold amplitude for a fluid-filled cavity' " [J. Acoust. Soc. Am. 125, 1857 (2009)] (L)

A. Teklu
Department of Physics and Astronomy, College of Charleston, Charleston, South Carolina 29424

Michael S. McPherson, M. A. Breazeale, and Roger D. Hasse
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

Nico F. Declercq
Georgia Tech Lorraine, Laboratory for Ultrasonic Nondestructive Evaluation, 2 rue Marconi, 57070 Metz-Technopole, France

(Received 27 June 2009; accepted 15 September 2009)

This letter is a response to Strasberg's recent paper, "Comment on 'Measurement of the frequency dependence of the ultrasonic parametric threshold amplitude for a fluid-filled cavity.'" The authors dispute the conclusions of Strasberg regarding the effect observed by Teklu *et al.* [J. Acoust. Soc. Am. **120**, 657–660 (2006)] published previously in JASA.
 [DOI: 10.1121/1.3243877]

PACS number(s): 43.25.Gf [ROC] Pages: 2854–2855

We would like to thank Strasberg for bringing attention to the frequency dependence of acoustical attenuation. At first we thought parametrically excited fractional harmonics would be a way to measure this attenuation.[1] We were aware of the dependence pointed out by Strasberg at the time of our 2006 paper, but rejected the situation described by Strasberg in favor of a more complete explanation based on the solutions of Mathieu's equation, which is appropriate for our system. The behavior of the solutions is depicted in the stability chart for Mathieu functions published by McLachlan.[2] This stability chart is unusual in that the coordinates are not independent, but if properly used, it accounts for the dependence of the threshold amplitude on cavity length. The curve beginning at $a=1$ corresponds to data taken by Adler in 1970; that beginning at $a=4$ corresponds to the 2006 data.[3,4] Subharmonics occur when the threshold line between unstable and stable regions labeled in the curve is crossed. In the experiments we described, these curves were crossed near $q=0$. The presence of such parametrically excited subharmonics is demonstrated optically by the appearance of extra diffraction orders between the ordinary diffraction orders when the transducer drive voltage is above a threshold value. It has been stated in the 2006 paper that parametric resonance is observed only if the energy supplied to the system reaches a critical value large enough to overcome attenuation in the system, mathematically interpreted as Eq. (12) in the 2006 paper. When the cavity length is large, the threshold amplitude is influenced by nonlinearity of the medium, acoustic attenuation, beam spreading, or detuning, or any combinations of these, and results in the decrease in the threshold amplitude of subharmonics with the increase in the drive frequency as shown in Fig. 3 of the 2006 paper. However, if the cavity length is small, energy losses are dominated by attenuation and result in the frequency dependence of the threshold amplitude shown in Fig. 4 of the 2006 paper. It is worth mentioning that a linear system is sufficient to excite parametric resonance.

Strasberg's objection to the lack of the $\frac{1}{2}$ harmonic subfrequency in the spectrum can be explained by the existence of frequency doublets, which appear around the half frequency. The fact that doublets were observed has been explained by Eller.[5] The doublets occur because two coupled parametric oscillators are excited. Presumably the two surfaces of the reflector are coupled to produce the doublets; however, this hypothesis has not been conclusively confirmed.

The thresholds calculated by Strasberg evidently were made by assuming that the signal outputs of the transducers used in the 1970 and 2006 papers are the same, or at least of the same order of magnitude. However, in the 2006 paper, a sinusoidal wave from a function generator via a power amplifier with impedance matching network was applied to *X*-cut quartz transducers. We are not quite sure how power was supplied to the transducer in the 1970 paper. Therefore, unless we exactly measure the power output signal of the transducers at subharmonic frequencies, we cannot compare the vibration amplitudes of the transducers used in the two different experiments.

We appreciate the diligence of Strasberg. However, we do not agree with his conclusion that the spectra may not be caused by acoustical parametric oscillation. Acoustical parametric oscillation also is a reasonable explanation of the data of Yen.[6]

[1]L. Adler and M. A. Breazeale, "Measurement of ultrasonic attenuation from the threshold of parametrically excited fractional harmonics in a liquid," J. Phys. D: Appl. Phys. **4**, L3–L4 (1971).
[2]N. W. McLachlan, *Theory and Application of Mathieu's Function* (Oxford University Press, Oxford, 1947), p. 40.

[3]A. Teklu, M. S. McPherson, M. A. Breazeal, R. D. Hasse, and N. F. Declercq, "Measurement of the frequency dependence of the ultrasonic parametric threshold amplitude for a fluid-filled cavity," J. Acoust. Soc. Am. **120**, 657–660 (2006).
[4]M. S. McPherson, M. A. Breazeale, and A. Teklu, "Stability zones and acoustic parametric oscillation," official publication of the 19th International Congress on Acoustics, Madrid, CD ed. (2007).
[5]A. I. Eller, "Fractional-harmonic frequency pairs in nonlinear systems," J. Acoust. Soc. Am. **53**, 758–765 (1973).
[6]N. Yen, "Experimental investigation of subharmonic generation in an acoustic interferometer," J. Acoust. Soc. Am. **57**, 1357–1362 (1975).

Broadband directive sources for acoustic discrete-time simulations (L)

José Escolano[a)]
Department of Telecommunication Engineering, University of Jaén, 23700 Linares, Spain

José J. López and Basilio Pueo
Institute for Telecommunications and Multimedia Applications, Universidad Politécnica Valencia, 46021 Valencia, Spain

(Received 19 November 2008; revised 2 October 2009; accepted 3 October 2009)

This work investigates a generalized implementation of directive sources on discrete-time methods. Results based on previous work for sinusoidal sources have been extended to accomplish real life source requirements, which have a frequency-dependent pattern and show gradual transitions between the different directivities at successive frequencies. One of the main advantages of this method lies in its capacity to be applied to any discrete-time method, since it is based on a proper combination of impulse responses radiated from discrete positions, which, in turn, will produce a particular pressure distribution at far-field. Furthermore, some considerations about mesh-dispersion effects on source modeling are addressed using some examples.
 [DOI: 10.1121/1.3257580]

PACS number(s): 43.55.Ka, 43.58.Ta [NX] Pages: 2856–2859

I. INTRODUCTION

Acoustic impulse responses can be accurately predicted in a relatively efficient way with the use of numerical approaches known as discrete-time methods. In these methods, a mesh of discretized spatial points represents the sound field at a given time step, which is updated through a linear combination of previous and surrounding discretized points. Among different discrete-time paradigms applied to acoustic problems, the most common are finite-difference time-domain[1] and digital waveguide mesh[2] (DWM).

In sound source modeling, the directivity is a fundamental feature to a final simulation result. Unfortunately, there have only been a few attempts to solve this problem for such particular discrete-time methods, as they can only consider sources as monopoles. Authors Ref. 3 proposed a method for a *wave*-DWM (W-DWM),[2] achieving source modeling with a frequency-independent directivity and for those directivity diagrams which have dipole and quadrupole shapes. However, this is not a generalized method that can be used in other discrete-time methods, since W-DWM works with characteristic (*wave*) variables.[2] Furthermore, the same authors proposed an alternative method, also for a W-DWM, which was inspired from a well-known antenna theory principle;[3] this method provides very interesting and accurate results which are also frequency dependent. However, in the results, they do not explore gradual transitions between the different directivity diagrams at consecutive frequencies and the impact of the mesh dispersion.

At about the same time, an alternative method was proposed,[4] based on a weighted combination of monopoles, in which the far-field pressure distribution is equal to that of a defined directivity. One of the main advantages of this method relies on the fact that, since it works with monopoles, it can be easily implemented in any discrete-time method. However, this method is limited to sinusoidal sources. This letter proposes an extension of the previous method, incorporating broadband directive sources with a frequency-dependent directivity.

II. METHOD

In Ref. 4 an algorithm for sinusoidal directive sources modeling has been presented. This method is similar to the monopole synthesis proposed in Refs. 5 and 6. However, real sources in room acoustics normally have a certain bandwidth and its directivity is not constant with the frequency. For that reason, it is necessary to modify the previous algorithm in order to synthesize broadband directive sources.

The previous algorithm relies on the fact that a spherical distribution of pressure at the far-field, which is produced by a source, corresponded to its directivity diagram. The scheme followed for implementing the directive source was already presented, where $\underline{\mathbf{r}}_s=[\mathbf{r}_{s_1},\ldots,\mathbf{r}_{s_N}]^T$ is a matrix that contains the position of the monopole sources (the underline indicates a matrix and T indicates transpose) and $\underline{\mathbf{r}}_r=[\mathbf{r}_{r_1},\ldots,\mathbf{r}_{r_M}]^T$ is the receivers position matrix. The receiver points are distributed in a discretized sphere. It should be noted that the pressure at positions $\underline{\mathbf{r}}_r$, $\mathbf{P}_r(\omega)=[P(\mathbf{r}_{r_1},\omega),\ldots,P(\mathbf{r}_{r_M},\omega)]^T$ is the known data, whereas the pressure at positions $\underline{\mathbf{r}}_s$, $\mathbf{P}_s(\omega)=[P(\mathbf{r}_{s_1},\omega),\ldots,P(\mathbf{r}_{s_N},\omega)]^T$ are sought through this algorithm.

Both pressure vectors are related through a plane wave Green's function matrix $\underline{\mathbf{C}}$, forming a system of equations $\mathbf{P}_r=\underline{\mathbf{C}}\mathbf{P}_s$. The resulting system of equations will be normally overdetermined, since the number of receiving positions employed is higher than the number of monopole sources in order to obtain soft directivity diagrams. The monopole

a)Author to whom correspondence should be addressed. Electronic mail: escolano@ujaen.es

weights, $\mathbf{P}_s$, can be calculated through the pseudoinverse of the matrix $\underline{\mathbf{C}}$, and particularly, by using the least mean square method;[7] however, this formulation is specific to a frequency ω. Although the relation $M>N$ is not compulsory, an undetermined system of equations can be also solved by the least mean square method.[7]

From this basis, an extension to a broadband source is presented: for a given sampling frequency f_s, let us consider a source with a certain cut-off frequency, since without loss of generality, source radiation (or signal content) properties could be assumed by having low-pass filter characteristics. The source bandwidth can be discretized to a uniform distribution of $L+1$ frequencies $\boldsymbol{\omega}=[\omega_0,\ldots,\omega_l,\ldots,\omega_L]$, where ω_L equals $2\pi f_s/2$. The number of frequencies $L+1$ can be considered as samples of the Fourier transform of the impulse response of each source.

In a similar fashion to the sinusoidal algorithm,[4] let us define a Green's function matrix for a specific frequency ω_l as $\underline{\mathbf{C}}(\omega_l)=(C_{n,m}(\omega_l))_{N\times M}$, where each element is calculated through $C_{n,m}(\omega_l)=e^{j(\omega_l/c)\|\mathbf{r}_{s_n}-\mathbf{r}_{r_m}\|}$, where a far-field condition is assumed and $\omega_L\|\mathbf{r}_{s_n}-\mathbf{r}_{r_m}\|/c\gg 1$ must be assured for all m and n values.

For each frequency, a certain directivity, in terms of the pressure at a far-field position, $\mathbf{P}_r(\omega_l)$, is defined. Through the pseudoinverse of Green's matrix $\underline{\mathbf{C}}(\omega_l)$, the amplitude and phase of the monopole sources, $\mathbf{P}_s(\omega_l)$, are solved. Once this process is repeated for every frequency in $\boldsymbol{\omega}$, it is then possible to define the following matrix, $\underline{\mathbf{P}}_s=[\mathbf{P}_s(\omega_0),\ldots,\mathbf{P}_s(\omega_L),\mathbf{P}_s(\omega_{L-1})^*,\ldots,\mathbf{P}_s(\omega_1)^*]$, where in each row, new elements have been added in order to obtain a Hermitian vector per row. Each column represents $\mathbf{P}_s$ for a given frequency ω_l and each row represents the sampled spectrum of each one of the monopole sources $\mathbf{r}_{s_n}$, $P(\mathbf{r}_{s_n},\omega)$.

The next step is to apply the inverse discrete Fourier transform[8] on each one of the rows, giving as a result a time domain matrix, $\underline{\mathbf{p}}_s$, where each row represents the impulse response of a source, $p(\mathbf{r}_{s_n},n)$ and $n\in[n_0,\ldots,n_l,\ldots,n_{2L-1}]$, represents the time steps. These impulse responses are applied to a given anechoic source $s(n)$, to be radiated with a particular directivity, using the linear convolution operator, $s(\mathbf{r}_{s_n},n)=s(n)*p(\mathbf{r}_{s_n},n)$.

The main advantage of this algorithm lies in the fact that different directivities can be defined for different frequencies and then, the design of complex sound sources is straightforward. However, for a given source distribution $\underline{\mathbf{r}}_s$, it is only possible to synthesize a limited frequency band which depends on the maximum and minimum relative distance between monopoles: the lower frequencies are required, the larger distances between the monopoles are necessary, and the highest synthesizable frequency is determined by the closest possible distance between monopoles. In the DWM case, through the classical array signal processing theory[9] and Courant condition,[2] it is possible to determine the maximum frequency $f_{\max}=f_s/(2\sqrt{N})$, where N is the number of dimensions in the problem. The low frequency problem is not really a limitation, since most of the real sources radiate omnidirectionally; in the situation where sources have a directional pattern at low frequencies, then larger distances between monopoles will be required.[4]

Regarding to the number of monopoles, it is possible to find a certain distribution of $(m+1)^2$ monopoles to describe a directivity based on spherical harmonics distribution of order m.[5]

Once the corresponding impulse responses are calculated, they can be included in the updating equations as hard (forcing $p(\mathbf{r}_s,n)$ in their corresponding monopole positions), soft (including source terms in the inhomogeneous wave equation) or transparent sources.[10] For simplicity, the following discussion uses soft sources. These soft sources may be more properly expressed as a function of the source strength q as $p(\mathbf{r}_{s_n},n)=-\rho c^2\Delta t(q(\mathbf{r}_{s_n},n)-q(\mathbf{r}_{s_n},n-1))$.

Due to the circular convolution intrinsic to the DFT, some time aliasing effects could appear on both sides of the calculated impulse response.[8] In order to avoid these effects, the impulse responses of the directivity filters should be long enough in order to assure that they tend to zero at both sides, i.e., using a sufficiently large number of points in the Fourier transform process.

Therefore, the number of frequencies $L+1$ should be increased, using an interpolation function. A window function can also be used, although this produces a loss in frequency resolution, which is not usually a limitation, since source directivity is commonly roughly defined in octaves or third-octaves.

III. EXAMPLES

This section presents several examples of the suitability of the method for broadband directivities. Simulations are obtained in a *Kirchhoff*-DWM (Ref. 2) using a sampling frequency $f_s=40$ kHz in a two-dimensional mesh conformed by 900×900 spatial samples. Let us consider a directivity pattern with an angular resolution $\Delta\theta=\pi/4$ with $\mathbf{P}_{r(\mathrm{dB})}=[0,-2,-5,-3,-4,-9,-3,-1]$, where $\mathbf{P}_{r(\mathrm{dB})}(i)$ is the directivity at the angle $\theta_i=i\pi/4$, expressed in dB. The first example consists of a constant directivity function for a particular bandwidth, where 1024 frequency samples are used for the definition of the whole spectrum. In order to synthesize this directivity, 12 sources, homogeneously distributed in a circle shape, for simplicity, are placed at a radius 14 cm, taking into account the staircase effect due to the sampling mesh. The distance between the sources determines a bandwidth 590–1500 Hz. Figure 1(a) shows and compares the broadband results obtained (solid line) with respect to the expected values (dotted line) for each of the indicated angular positions $\Delta\theta$. Results are normalized and presented with respect to the pressure level obtained at direction $\theta=0$. The results show the highest error lower than 1 dB.

For the same directivity pattern, but this time synthesized through a circle shaped source distribution situated at a radius of 5.8 cm, using again 12 point sources. The bandwidth is now between 1.47 and 3.75 kHz. Figure 1(b) shows the results in a similar way that of the previous one. However, more fluctuations appear with respect to the average value in the results due to the inherent mesh-dispersion in the DWM grid. It has been observed that these errors are more related to angular dispersion, rather than due to the frequency-dependent dispersion;[11] this could be deduced

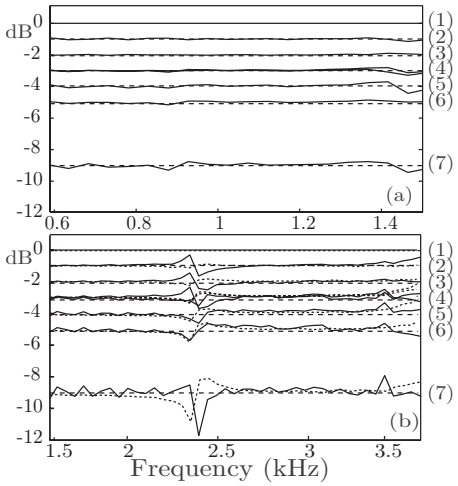


FIG. 1. Simulated broadband directivity results (solid line), compared with the expected values (dashed line), using a point sources circle configuration with radius (a) 14 cm and (b) 5.8 cm. The dotted line in (b) is obtained with the same conditions as the solid line, but using an IDWM. Each pressure value has been normalized for direction $\theta=0$ at each frequency, and the angular position $\Delta\theta$ is indicated in the right side as 0 (line 1), $7\pi/4$ (2), $\pi/4$ (3), $3\pi/4$, $6\pi/4$ (4), π (5), $2\pi/4$ (6), and $5\pi/4$ (7).

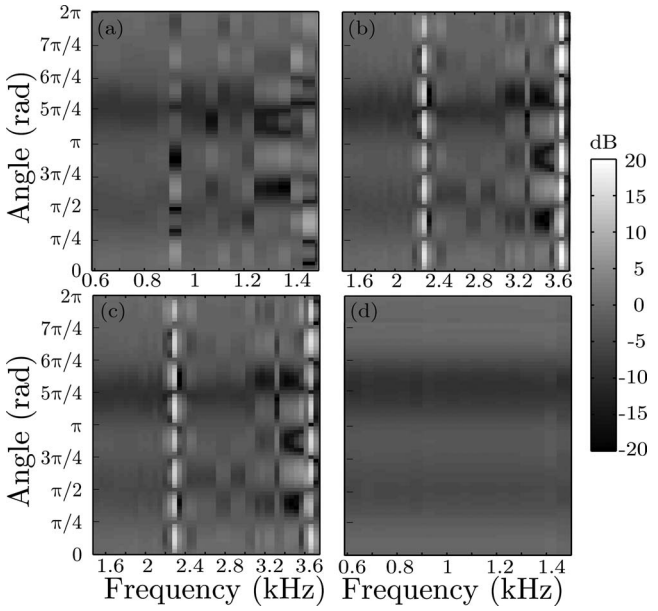


FIG. 2. (a) Same results as in Fig. 1(a), measured with an angular resolution of $\Delta\theta=\pi/36$. (b) Same results as in Fig. 1(b), measured with $\Delta\theta=\pi/36$. (c) Same results as in Fig. 1(b), measured with $\Delta\theta=\pi/36$ and using an IDWM. (d) Same results as Fig. 2(a) after a spline interpolation of the directivity pattern.

simply by assuming an isotropic dispersion: in that case, all the points situated in a circle distribution will have the same phase error but this would not affect the wave amplitude superposition. A simple solution could be the use of the interpolated[12] or triangular[2] DWM (IDWM and TDWM, respectively), since they spread the angular dispersion evenly nearby. The dotted line in Fig. 1(b) shows the same example, but using IDWM. It can be observed that the results are improved, and variations with respect to the theoretical references will decrease. Nevertheless, some errors remain, but they are due to the use of the least mean square method which calculates the pseudoinverse of $\mathbf{C}$, which is a frequency-dependent matrix.

Since the angular resolution is low, it is necessary to check the behavior in other angles, where there has been no new directivity information added. Figure 2(a) shows the same results as those of Fig. 1(a), but represented with a higher angular resolution $\Delta\theta=\pi/36$. No modifications over the previous results are made, except that more angles are measured and represented. With this low angular resolution, a considerable number of lobes and noticeable transitions between the directivity of close frequencies are observed, which are unnatural. The apparition of these lobes is related to the least mean square method instead of the propagation method, since this algorithm minimizes the squared error in the known points; but in the rest of angles, the behavior is unpredictable. Figures 2(b) and 2(c) represent the same results as in Fig. 1(b), except that more angular resolution has been measured, where Fig. 2(b) is the result with the conventional DWM method and Fig. 2(c) with the IDWM: no relevant differences in lobes' distribution could be found between the DWM [Fig. 2(b)] and the IDWM [Fig. 2(c)] simulations.

A simple idea to overcome this limitation would be to interpolate the directivity values for all $\mathbf{r}\in\mathbf{r}_r$, and then, to apply the proposed algorithm, which will now have a higher angular resolution as a result of that interpolation of data between angles for each frequency. Figure 2(d) shows the same directivity pattern as Fig. 2(a) after increasing the angular resolution through a spline interpolation, with an angular resolution $\Delta\theta=\pi/36$. Results show that the calculated directivity is now practically constant, and no important differences can be observed, proving that the proposed method is suitable for defining frequency-independent directivity sources in a DWM. Of course, this interpolation would only be necessary, if the original directivities had been measured with a poor angular resolution. This method is also applicable to other discrete-time methods, since all of them include monopoles sources.

The main advantage of this method lies in the fact that different directivities at different frequencies, according to a given frequency-dependent radiation pattern, can be defined, including soft transitions between the original data. Let us consider another example: a DWM grid with the same characteristics as the previous example, but using the following directivity pattern: at $f=500$ Hz, $\mathbf{P}_{r(\mathrm{dB})}=[0,-2,-5,-3,-4,-9,-3,-1]$; at $f=1000$ Hz, $\mathbf{P}_{r(\mathrm{dB})}=[0,-2,-9,-3,-8,-1,-12,-5]$; and finally, at $f=2000$ Hz, $\mathbf{P}_{r(\mathrm{dB})}=[0,-8,-10,-3,-0,-5,-4,-1]$; the null frequency and $f_s/2$ have been considered omnidirectional just to simplify criteria. Since the angular resolution is quite low, the directivities are interpolated using spline functions in order to achieve $\Delta\theta=\pi/36$. Furthermore, since the number of known frequencies have also fairly reduced, the rest of the frequencies are linearly interpolated using 512 frequency samples to represent the whole spectrum.

Two point source distributions are used for synthesizing this frequency-dependent directivity. Again, both distributions consisting of a circle shape source arrangement of 12 point sources homogeneously distributed, one with a radius

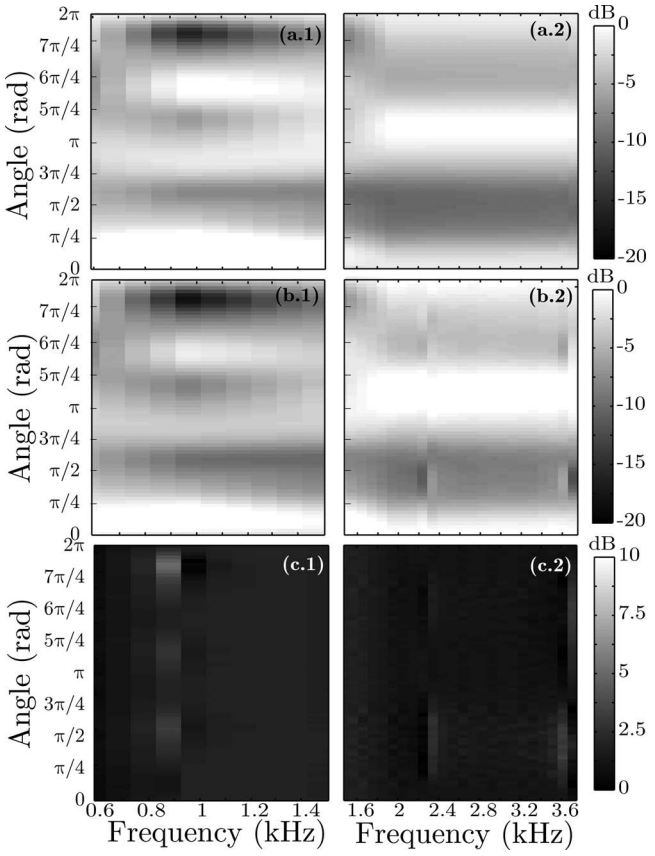


FIG. 3. Different directivity results: (a) theoretical directivity pattern obtained through a circle shape source distribution with a radius (a1) 14 cm and (a2) 5.8 cm. (b) Same directivity results obtained through the proposed method in an IDWM with a radius (b1) 14 cm and (b2) 5.8 cm. (c) Error obtained as the difference between the expected directivity and the corresponding synthesized results.

of 5.8 cm (between 1.47 and 3.75 kHz) and the other with a radius of 14 cm (between 0.59 and 1.5 kHz). Figure 3(a1) represents the expected pattern using the circle of sources with a radius 14 cm, whereas Fig. 3(b1) represents the results obtained in the simulation of the 14-cm-radius source distribution. A close agreement between both graphs can be observed. Also, Fig. 3(b2) shows the results of the 5.8-cm-radius source distribution, compared to the theoretical directivity pattern. Again, the agreement is good. Figures 3(c1) and 3(c2) presents corresponding error distribution. The maximum error is estimated approximately 4 dB, although these errors are located in very few points. The average errors are mostly negligible.

IV. CONCLUSIONS

In a previous publication,[4] a method for sinusoidal directive source modeling in discrete-time methods based on monopole synthesis was proposed. Although the achieved directivity was accurate, this method itself is not sufficient for more complete simulations where broadband sound sources are employed. From the basis of that method, an extension to broadband directivity patterns has been presented. In this improved method, the goal is to calculate a suitable set of impulse responses for a given distribution of monopoles, obtaining not only accurate results but also smooth transitions between directivities for consecutive frequencies. Furthermore, special attention has been paid to its use in isotropic dispersion meshes, such as an IDWM, which has been tested, and a reduction in the errors has been observed. This model is also valid for the remainder discrete-time methods, outperforming previously proposed methods.

ACKNOWLEDGMENTS

The authors want to thank Professor Felipe Orduña (National Autonomous University of Mexico) for his suggestions about this work. This work has been supported by the Spanish Ministry of Science and Technology (MCYT) under Project No. TEC2006-13883–C04–01 and FEDER funds.

[1]D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," J. Acoust. Soc. Am. **98**, 3302–3308 (1995).
[2]D. T. Murphy, A. Kelloniemi, J. Mullen, and S. Shelley, "Acoustic modeling using the digital waveguide mesh," IEEE Signal Process. Mag. **24**, 55–66 (2007).
[3]H. Hacıhabiboğlu, B. Günel, and A. M. Kondoz, "Time-domain simulation of directive sources in 3D digital waveguide mesh-based acoustical models," IEEE Trans. Audio, Speech, Lang. Process. **16**, 934–946 (2008).
[4]J. Escolano, J. J. López, and B. Pueo, "Directive sources in acoustic discrete-time domain simulations based on directivity diagrams," J. Acoust. Soc. Am. **121**, EL256–EL262 (2007).
[5]F. Giron, "Investigations about the directivity of sound sources," Ph.D. thesis, Ruhr Universität, Bochum, Germany (1996).
[6]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).
[7]C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Its Applications* (Wiley, New York, 1971).
[8]A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1999).
[9]D. H. Johnson and D. E. Dudgeon, *Array Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[10]J. B. Schneider, C. L. Wagner, and S. L. Broschat, "Implementation of transparent sources embedded in acoustic finite-difference time-domain grids," J. Acoust. Soc. Am. **103**, 136–142 (1998).
[11]J. Escolano, "Contributions to discrete-time methods for room acoustic simulation," Ph.D thesis, Technical University of Valencia, Valencia, Spain (2008).
[12]L. Savioja and V. Välimäki, "Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques," IEEE Trans. Speech Audio Process. **8**, 184–194 (2000).

Effects of voicing in the recognition of concurrent syllables (L)[a)]

Martin D. Vestergaard and Roy D. Patterson
Department of Physiology, Development and Neuroscience, Centre for the Neural Basis of Hearing, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 14 September 2009; accepted 5 October 2009)

This letter reports a study designed to measure the benefits of voicing in the recognition of concurrent syllables. The target and distracter syllables were either voiced or whispered, producing four combinations of vocal contrast. Results show that listeners use voicing whenever it is present either to detect a target syllable or to reject a distracter. When the predictable effects of audibility were taken into account, limited evidence remained for the harmonic cancellation mechanism thought to make rejecting distracter syllables more effective than enhancing target syllables.
 [DOI: 10.1121/1.3257582]

PACS number(s): 43.71.Bp, 43.71.An, 43.66.Ba, 43.72.Qr [JES] Pages: 2860–2863

I. INTRODUCTION

It has been shown that segregation of competing voices is facilitated if the target and distracter voices differ either in glottal pulse rate (GPR) or vocal tract length (VTL) (e.g., Darwin *et al.*, 2003). Moreover, Vestergaard *et al.* (2009) measured the interaction of GPR and VTL in concurrent syllable recognition using a paradigm that controlled temporal glimpsing and the idiosyncrasies of individual voices. They showed that, when the signal-to-noise ratio (SNR) is 0 dB, a two-semitone (ST) difference in GPR produced the same performance advantage as a 20% difference in VTL. This letter reports an extension of Vestergaard *et al.*, 2009, designed to measure the benefits of voicing to the recognition of concurrent syllables.

It has been proposed that the benefit of a pitch difference in the recognition of concurrent speech is that it helps the listener reject sounds that fit the harmonic structure of the distracting voice (cancellation theory) rather than assists the listener in selecting sounds that fit the harmonic structure of the target voice (enhancement theory). In a series of double-vowel experiments, de Cheveigné and colleagues (de Cheveigné *et al.*, 1997b, 1997a; de Cheveigné, 1997, 1993) developed a harmonic cancellation model tuned to the periodicity of the distracter. They showed that the advantage of a pitch difference depends primarily on the harmonicity of the distracter. In order to evaluate the feasibility of the cancellation theory for connected syllables, the current study employed the paradigm described by Vestergaard *et al.* (2009) for voiced and whispered syllables. Whispered speech signals do not provide any acoustic cues to the harmonicity that characterizes speech in voiced phonation (Abercrombie, 1967).

Whispered phonation is produced by allowing turbulent air to flow through partially open glottal folds. Turbulence reduces the gradient of airflow velocities and results in a noise-like excitation of the vocal tract resonances (the formants). In naturally produced speech, whispered syllables are elongated (Schwartz, 1967) and air consumption is dramatically increased (Schwartz, 1968a). Whispered speech is typically 15–20 dB softer than voiced speech (Traunmüller and Eriksson, 2000) and has a spectral tilt of approximately +6 dB/octave (Schwartz, 1970). These features lead to a reduction in the perceptibility of whispered speech while it remains relatively robust. Indeed, whispered speech can convey much of the information that voiced speech can convey. Tartter (1991) found that the intelligibility of whispered vowels was 82%; only 10% lower than for voiced vowels. The recognition of whispered consonants was 64%; much above chance for the 18 consonants in their experiment (Tartter, 1989). Listeners are also able to identify speaker sex given isolated whispered vowels (Schwartz and Rine, 1968) or even when presented with isolated voiceless fricatives (Schwartz, 1968b). Lass *et al.* (1976) reported that the recognition of speaker sex dropped from 96% correct for voiced speech to 75% correct for whispered speech. Moreover, Tartter and Braun (1994) showed that listeners can accurately distinguish "frowned," neutral, and "happy" speech in both voiced and whispered phonations. Thus, while the purpose of whispering is to reduce audibility, whispered speech remains highly functional in other respect. When instructed to adjust the frequency of a pure tone so that it matched the perceived pitch of a whispered vowel [sic], listeners matched the frequency of the pure tone with the frequency of the second formant of the vowel (Thomas, 1969). Thus, in the absence of temporarily defined pitch, listeners revert to an extracted spectral pitch measure as described by Schneider *et al.* (2005). Together, these studies suggest that we can investigate the functional role of harmonicity on the segregation of concurrent syllables by removing the temporal regularity of voiced speech samples and applying a spectral lift, thus simulating whispered speech.

In the current investigation, audibility was varied by testing at different SNRs. Because whispered speech has relatively more high-frequency energy than voiced speech, it has the potential to be a more efficient masker and a more robust target than voiced speech when presented at the same

[a)]Portions of this work were presented at the 153rd Meeting of the Acoustical Society of America, Salt Lake City, UT, 2007.

RMS level. To control such effects, syllable recognition was analyzed as a function of audibility as defined by the speech intelligibility index (ANSI, 1997).

The purpose of the current study was to determine the importance of voicing in the recognition of concurrent speech. It was hypothesized that performance in a syllable recognition task would improve whenever the auditory system can make use of voicing, either to detect a target or reject a competing distracter. Moreover, if the mechanism of cancellation were more effective than the corresponding enhancement mechanism, listeners would be more successful at using voicing to reject a distracter than to detect a target.

II. METHOD

The participants were required to identify syllables spoken by a target voice in the presence of a distracting voice. The target and distracter voices were voiced and whispered in all combinations, and the SNR was varied over a wide range.

The stimuli were taken from the syllable corpus previously described by Ives *et al.* (2005). It contains 180 syllables, divided into 90 consonant-vowel (CV) and vowel-consonant (VC) pairs. Six consonants from each of the three categories (plosives, sonorants,[1] and fricatives) were paired with one of five vowels. The syllables were analyzed and re-synthesized with the STRAIGHT vocoder (Kawahara and Irino, 2004) to simulate voices with different combinations of GPR and VTL. To simulate whispered speech, the STRAIGHT spectrograms were excited with broadband noise and high-pass filtered at 6 dB/octave. This procedure removes pitch from the voiced part of the syllables (in consonants as well as in vowels) and produces an effective simulation of whispered speech. The target voice simulated a tall male speaker (VTL: 182 mm; GPR: 157 Hz when voiced) and the distracting voice sounded like a female of normal height (VTL: 139 mm; GPR: 203 Hz when voiced). The difference in VTL is several just noticeable differences for the discrimination of resonance scale for syllables (Ives *et al.*, 2005), so even when pitch was removed from both the target and distracter voices, it was still easy to hear the difference between the two voices (Vestergaard *et al.*, 2005; Vestergaard, 2007). The target and distracter syllable pairs were presented in triplets to promote the perception of connected speech, and the syllables within a pair were matched according to their phonetic specification to reduce temporal glimpsing (Vestergaard *et al.*, 2009). The listeners responded by clicking on a syllable matrix on a computer screen, and they were given training on the use of the interface before commencing the main experiment. They were seated in an IAC (Winchester, United Kingdom) double-walled, sound-attenuated booth, and the stimuli were presented bilaterally via AKG K240DF headphones at 60-dB sound pressure level.

There were eight listeners (19–21 years old; three male) who all provided informed consent, and the experimental protocol was approved by the Cambridge Psychology Research Ethics Committee (CPREC). Audiograms were recorded at octave frequencies between 500 and 4000 Hz bilaterally, to ensure that the listeners had normal hearing.

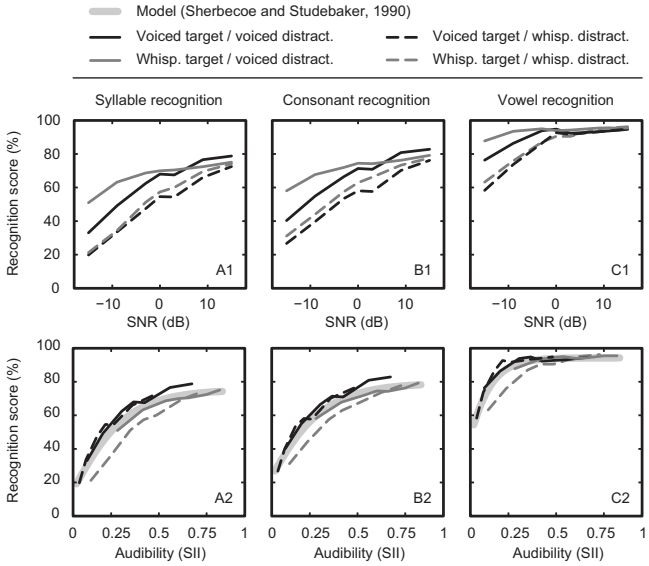


FIG. 1. Recognition scores as a function of (1) SNR (top panels) and (2) SII (bottom panels). The left panels (A#) show syllable recognition, the middle panels (B#) show consonant recognition, and the right panels (C#) show vowel recognition. The black lines show performance for voiced target syllables and the dark gray lines show performance for whispered-target syllables. The solid lines are for voiced distracters and the dashed lines are for whispered distracters. In the bottom panels, the thick light gray curve shows predicted recognition according to the transformation by Sherbecoe and Studebaker (1990). See text for details.

Recognition performance was measured for the target voice in a $2\times2\times7$ factorial design (i.e., 28 conditions). The target voice was either voiced or whispered, and the distracter was either voiced or whispered. Each pair of target and distracter voices was measured at seven SNRs [−15, −9, −3, 0, 3, 9, 15 dB]. The trials were blocked in runs of 40 trials, within which the voice combination and the SNR were constant. Between runs the condition was randomly chosen from the full set of the 28 conditions. Each condition was repeated three times. To increase the sensitivity of the experiment to the variation in voicing, the task was made slightly more difficult by playing the target syllable in either interval 2 or 3. A visual indicator marked the interval to which the listener should respond (for details of the rationale for this paradigm, see Vestergaard *et al.*, 2009).

III. RESULTS

Three different scores were computed: percent correct syllable recognition (the primary task), percent correct consonant recognition, and percent correct vowel recognition. The average values for the three scores are shown as a function of SNR for the four vocal conditions in the top row of Fig. 1. The results show the expected effect of SNR and some notable effects of voicing. To control for the predictable effect of audibility caused by the difference in spectrum between voiced and whispered speech, the following analysis was run: For each trial, an audibility index [the speech intelligibility index (SII) (ANSI, 1997)] was calculated by deriving the spectrum levels for the target and distracter syllables. The distracters' spectrum levels were used as masker

when estimating the audibility of the targets. An important function for English nonsense syllables was used according to the ANSI standard. A transfer function (Sherbecoe and Studebaker, 1990) was then fitted to the data. This analysis allows for a data-driven audibility-controlled prediction of recognition performance. The results of this transform are shown at the bottom row of Fig. 1. They show the effects of voicing once audibility has been taken into account.

The scores from the experiment and the scores from the prediction described above were converted to rationalized arcsine units (RAUs) (Studebaker, 1985; Thornton and Raffin, 1978). The effects of voicing on performance were analyzed by assessing the departure of the observed scores from the corresponding predictions. A three-way repeated-measures analysis of variance [2(target)×2(distracter)×7(SNR)] was run on the prediction mismatch units (observed RAU scores−predicted RAU scores) for the three different scores. Greenhouse–Geisser correction was used to compensate for lack of sphericity, and paired comparisons with Sidak correction were used to analyze effects within condition. For the syllable scores (A panels in Fig. 1), recognition of voiced target syllables was above the predicted value, and recognition of whispered-target syllables was below the predicted value ($F_{1,7}=39.9$, $p=0.001$, $\eta_p^2=0.84$). Prediction accuracy increased with increasing SNR ($F_{6,42}=8.4$, $p=0.002$, $\varepsilon=0.40$, $\eta_p^2=0.55$), and this trend was more pronounced for whispered targets and less pronounced for voiced targets ($F_{6,42}=3.59$, $p=0.018$, $\varepsilon=0.68$, $\eta_p^2=0.33$). Recognition performance was below the predicted value for whispered distracters and above the predicted value for voiced distracters ($F_{1,7}=59.6$, $p<0.001$, $\eta_p^2=0.89$), and this effect was entirely driven by the whispered-target condition ($F_{1,7}=11.8$, $p=0.011$, $\eta_p^2=0.63$). The results of this analysis are illustrated in Fig. 2, which also shows the direction of each effect.

There are two pronounced kinks in the results for the voiced targets at 0–3 dB in Fig. 1, panel A1 (and B1). However, paired comparisons of the prediction mismatch data showed that none of the corresponding troughs in Fig. 2, panel A1 (and B1), was significantly different from the neighboring points, so the kinks would appear to be just statistical fluctuations.

The pattern of consonant scores and vowel scores is similar to the pattern of syllable scores, except the interaction between target and distracter is only significant for vowel recognition ($F_{1,7}=20.8$, $p=0.003$, $\eta_p^2=0.75$; panel C2 in Fig. 2) and not for consonant recognition (panel B2 in Fig. 2). Furthermore, for the vowel data, the effect of SNR interacted not only with the voicing of the target ($F_{6,42}=5.83$, $p=0.012$, $\varepsilon=0.36$, $\eta_p^2=0.46$) but also with the distracter ($F_{6,42}=4.72$, $p=0.011$, $\varepsilon=0.51$, $\eta_p^2=0.40$). Moreover, target and distracter voicing and SNR showed a three-way interaction ($F_{6,42}=3.30$, $p=0.030$, $\varepsilon=0.6$, $\eta_p^2=0.32$) indicating that the effect of SNR on the prediction accuracy of vowel recognition was entirely driven by the whispered-target/whispered-distracter condition (see panel C1 in Fig. 2).

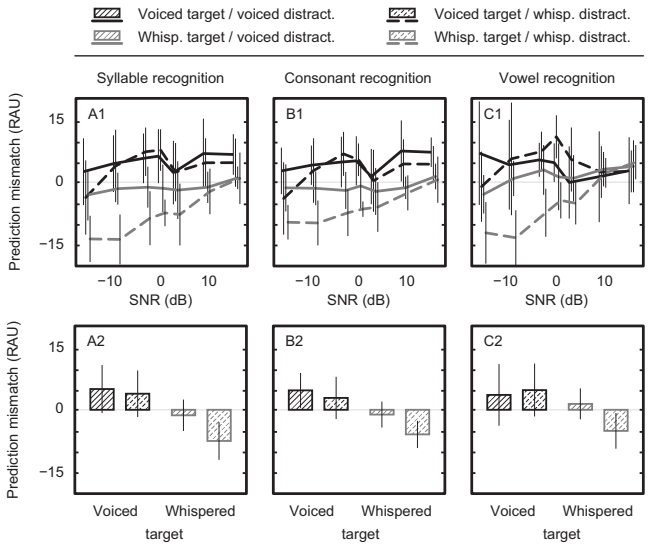


FIG. 2. The prediction mismatch for (A#) syllable, (B#) consonant, and (C#) vowel recognition scores in RAU. The top panels show (1) the interactions of vocal characteristics with SNR, and the bottom panels show (2) the interaction between target voicing and distracter voicing. The black lines and hatching are for the voiced target syllables and gray is for the whispered-target syllables. Solid lines and hatching are for voiced-distracter syllables and dashed lines and hatching are for whispered-distracter syllables.

IV. DISCUSSION

Overall, voiced syllables were better recognized than whispered syllables, and whispered distracters led to lower recognition performance than voiced distracters. It would appear that the listeners could use voicing to reject a distracter as well as to detect a target. However, the effect of voicing for distracters was only present when the target was whispered. Similarly, recognition of voiced syllables was well predicted by the audibility model, whereas for whispered-target syllables, recognition was lower than predicted, and more so at low SNRs. The fact that whispered syllables were less intelligible than voiced syllables corroborates previous studies on the perception of whispered speech (Tartter, 1991, 1989).

The harmonic cancellation model suggests that the benefit of voicing to recognition should be greater for the distracter than the target. Thus, it was expected that there would be an asymmetry in which listeners suffer more from removal of voicing in the distracter than in the target. The interaction of target and distracter in the vowel recognition data of Fig. 2, panel C2, supports this prediction: The difference between the bars with solid-line hatching shows the drop in recognition performance associated with the removal of voicing in the target syllable, and the difference between the gray bars shows the drop in recognition performance associated with the removal of voicing in the distracter syllable. Since the difference between the bars with solid-line hatching is smaller than the difference between the gray bars, it could be argued that the results are compatible with the cancellation model. However, it is also the case that the effect of removing voicing in the distracter is only pronounced when the target itself is whispered. For voiced target syllables, there is no significant effect of removing voicing in

the distracter for any of the scores, possibly because there was already a cue to the target voice provided by its resonance scale. Vestergaard *et al.* (2009) previously showed that when there is a sizable difference in VTL between the competing voices, then the benefit of additional cues is diminished. Overall, while the data do not provide strong evidence for the cancellation hypothesis, they do not rule it out.

V. CONCLUSION

The main conclusion is that listeners use voicing whenever it is present, either to detect the target speech or to reject the distracter. This study also illustrates the importance of controlling for the effects of audibility in experiments with voiced and whispered speech. Direct interpretation of the recognition scores in the top panels of Fig. 1 would lead to an overestimation of the robustness of whispered speech. When the effects of audibility are included, as in the bottom panels of Fig. 1, the effect of voicing is observed to be considerably smaller than when performance is presented as a function of SNR. To wit, three of the four vocal conditions contained voicing in either the target, the distracter, or both, and they show comparable results once audibility has been taken into account. By contrast, in the condition in which both target and distracter were whispered, performance drops off progressively with audibility, especially below an audibility index of 0.5. In other words, audibility predicts the identification of the target when one of the concurrent syllables is voiced, but it leads to an overestimation of the recognition of whispered syllables when the distracter is a whispered syllable.

ACKNOWLEDGMENTS

The research was supported by the United Kingdom Medical Research Council under Grant Nos. G0500221 and G9900369. We would like to thank James Tanner and Sami Abu-Wardeh for help with collecting the data, and Nick Fyson for assistance in producing the programs that ran the experiments.

[1]The category *sonorant* here refers to a selection of consonants from the manner classes: nasal, trill, and approximant (sometimes called semivowels) that are common in the English language ([m], [n], [r], [j], [l], [w]).

Abercrombie, D. (**1967**). *Elements of General Phonetics* (Edinburgh University Press, Edinburgh).

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

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**, 2913–2922.

de Cheveigné, A. (**1993**). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.

de Cheveigné, A. (**1997**). "Concurrent vowel identification. III. A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.

de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997b**). "Concurrent vowel identification. I. Effects of relative amplitude and f_0 difference," J. Acoust. Soc. Am. **101**, 2839–2847.

de Cheveigné, A., McAdams, S., and Marin, C. M. H. (**1997a**). "Concurrent vowel identification. II. Effects of phase, harmonicity and task," J. Acoust. Soc. Am. **101**, 2848–2856.

Ives, D. T., Smith, D. R., and Patterson, R. D. (**2005**). "Discrimination of speaker size from syllable phrases," J. Acoust. Soc. Am. **118**, 3816–3822.

Kawahara, H., and Irino, T. (**2004**). "Underlying principles of a high-quality speech manipulation system straight and its application to speech segregation," in *Speech Separation by Humans and Machines*, edited by P. L. Divenyi (Kluwer Academic, Boston, MA).

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (**1976**). "Speaker sex identification from voiced, whispered, and filtered isolated vowels," J. Acoust. Soc. Am. **59**, 675–678.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleeck, S., Stippich, C., and Rupp, A. (**2005**). "Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference," Nat. Neurosci. **8**, 1241–1247.

Schwartz, M. F. (**1967**). "Syllable duration in oral and whispered reading," J. Acoust. Soc. Am. **41**, 1367–1369.

Schwartz, M. F. (**1968a**). "Air consumption, per syllable, in oral and whispered speech," J. Acoust. Soc. Am. **43**, 1448–1449.

Schwartz, M. F. (**1968b**). "Identification of speaker sex from isolated, voiceless fricatives," J. Acoust. Soc. Am. **43**, 1178–1179.

Schwartz, M. F. (**1970**). "Power spectral density measurements of oral and whispered speech," J. Speech Hear. Res. **13**, 445–446.

Schwartz, M. F., and Rine, H. E. (**1968**). "Identification of speaker sex from isolated, whispered vowels," J. Acoust. Soc. Am. **44**, 1736–1737.

Sherbecoe, R. L., and Studebaker, G. A. (**1990**). "Regression equations for the transfer functions of ANSI s3.5–1969," J. Acoust. Soc. Am. **88**, 2482–2483.

Studebaker, G. (**1985**). "A "rationalized" arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Tartter, V. C. (**1989**). "What's in a whisper?" J. Acoust. Soc. Am. **86**, 1678–1683.

Tartter, V. C. (**1991**). "Identifiability of vowels and speakers from whispered syllables," Percept. Psychophys. **49**, 365–372.

Tartter, V. C., and Braun, D. (**1994**). "Hearing smiles and frowns in normal and whisper registers," J. Acoust. Soc. Am. **96**, 2101–2107.

Thomas, I. B. (**1969**). "Perceived pitch of whispered vowels," J. Acoust. Soc. Am. **46**, 468–470.

Thornton, A. R., and Raffin, M. J. M. (**1978**). "Speech discrimination scores modeled as a binomial variable," J. Speech Hear. Res. **21**, 507–518.

Traunmüller, H., and Eriksson, A. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**, 3438–3451.

Vestergaard, M. D. (**2007**). "The effect of voicing, pitch and vocal tract length on the recognition of concurrent speech," J. Acoust. Soc. Am. **121**, 3200.

Vestergaard, M. D., Fyson, N. R. C., and Patterson, R. D. (**2009**). "The interaction of vocal characteristics and audibility in the recognition of concurrent syllables," J. Acoust. Soc. Am. **125**, 1114–1124.

Vestergaard, M. D., Ives, D. T., and Patterson, R. D. (**2005**). "Identification of voiced syllables masked by whispered syllables and vice versa," in Proceedings of the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness, Cardiff, Wales, p. 58.

On the use of leaky modes in open waveguides for the sound propagation modeling in street canyons

Adrien Pelat,[a] Simon Félix, and Vincent Pagneux
LAUM, CNRS, Université du Maine, avenue Olivier Messiaen, 72085 Le Mans, France

(Received 30 January 2009; revised 11 September 2009; accepted 20 September 2009)

An urban, U-shaped, street canyon being considered as an open waveguide in which the sound may propagate, one is interested in a multimodal approach to describe the sound propagation within. The key point in such a multimodal formalism is the choice of the basis of local transversal modes on which the acoustic field is decomposed. For a classical waveguide, with a simple and bounded cross-section, a complete orthogonal basis can be analytically obtained. The case of an open waveguide is more difficult, since no such a basis can be exhibited. However, an open resonator, as displays, for example, the U-shaped cross-section of a street, presents resonant modes with complex eigenfrequencies, owing to radiative losses. This work first presents how to numerically obtain these modes. Results of the transverse problem are also compared with solutions obtained by the finite element method with perfectly mathed layers. Then, examples are treated to show how these leaky modes can be used as a basis for the modal decomposition of the sound field in a street canyon.
 [DOI: 10.1121/1.3259845]

PACS number(s): 43.20.Mv, 43.28.Js, 43.50.Vt [LLT] Pages: 2864–2872

I. INTRODUCTION

The aim of the present work is to solve the wave equation in a long open rectangular enclosure (Fig. 1). This open waveguide is an idealized domain of propagation modeling a street canyon. This idealized domain being a very simple representation of a real street canyon, competitive effects between guided waves (between street facades), and leaky waves (through the opening on the sky) occurring along the propagation can be more easily apprehended. In this paper, a modal based method is proposed to describe acoustic propagation in rectangular straight open waveguides with a particular attention paid to wave radiation in the free space above. Then, in other future studies in the field of urban acoustics, one could extend the principles of this method to treat more realistic street geometries.

The investigation of sound propagation in urban environments and streets has been the subject of extensive researches in the past 4 decades, as the response to a growing social demand.

After experimental observations in the 1960s,[1–3] the earliest theoretical works on this topic were conducted in the 1970s. Davies,[4] Lee and Davies,[5] Stenackers *et al.*,[6] Lyon[7] used image sources to study multiple sound reflections in a street considered as a channel between two infinite walls. Later, image source method has been improved considering scattering at facade irregularities,[8] diffusely reflecting facades,[9] or coherent image sources.[10] Other energetic approaches were also used in urban acoustics. Kang[11] developed a radiosity based model, Bradley[12] used ray tracing method, and Picaut[13] proposed a method based on a diffusion equation governing sound particles propagation. These energetic approaches give statistical description of sound fields in urban environments and are able to model more or less accurately numerous phenomena occurring in streets. Since these approaches assume high frequency hypothesis, they cannot describe sound fields when the wavelength is in the range of street width. Furthermore, computation costs strongly increase for complex geometries or for three-dimensional (3D) problems.

As pointed out by Lu[10] and Walerian,[14] interference effects are significant for relatively narrow street canyons. Hence, wave methods present real interests for sound propagation modeling. Bullen and Fricke,[15,16] 30 years ago, studied the wave propagation in two-dimensional (2D) street using a modal approach, notably to model junctions of streets. To solve the wave equation in a 2D street canyon, finite element method (FEM) or boundary element method[17] could as well give solutions to the wave equation for realistic geometries, but the high computation costs restrict their use at low frequencies or for 2D problems. The equivalent sources approach[18–20] is field-based rather than ray-based, and apprehends the resonant behavior of a city canyon. In this approach, a set of equivalent sources is used to couple the free half space above the canyon to the cavity inside the canyon. The finite difference time domain (FDTD) method describes the sound field in 2D even 3D problems[21] and can model *a priori* a very large number of phenomena. The parabolic equation coupled with FDTD method can also be useful to take into account meteorological effects.[22] So, even if these wave methods can model sound propagation in realistic situations, the necessary numerical resolution of the wave equation does not provide explicitly links between solution behaviors and domain geometry of the studied problem.

The aim of the present paper is to establish a multimodal description of the wave propagation in a 3D street canyon, regarded as a straight open waveguide.

In the classical case of closed waveguides having simple and bounded cross-section, a complete orthogonal modal ba-

[a]Author to whom correspondence should be addressed. Electronic mail: adrien.pelat.etu@univ-lemans.fr

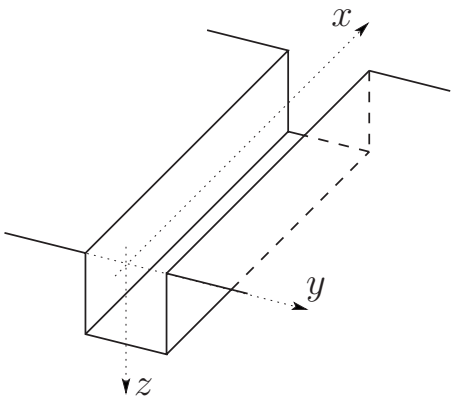


FIG. 1. A straight open waveguide as model of a street canyon.

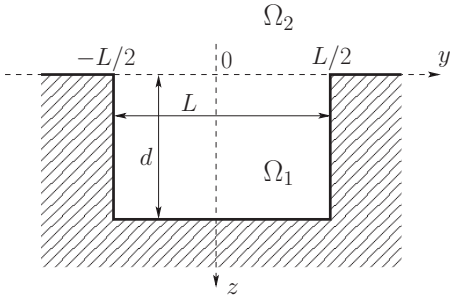


FIG. 2. The cross-section of a street canyon seen as a 2D open rectangular cavity.

sis $\{\phi_i\}$ can be analytically obtained. The case of open waveguides, as the one shown in Fig. 1, with a partially bounded cross-section, is more difficult, since no such a basis can be exhibited. However, such a cross-section being regarded as an open resonator also displays eigenmodes with complex eigenfrequencies, owing to the radiation losses.[5,23–25]

In this paper, we propose to describe how the resonant modes of the open cross-section of an open waveguide (Fig. 1) can be used to give a multimodal formulation of the sound propagation in long open enclosures. A general method to compute the resonant frequencies and mode shapes in the transversal open cross-section of the duct is described. Results of the transverse problem are also compared with FEM computations using perfectly matched layer (PML). Then, the multimodal propagation in a straight open waveguide is formulated and numerical examples are given and discussed.

II. EIGENMODES OF THE TRANSVERSE PROBLEM

A. Theory and formulation

The wave equation to solve is

$$\left(\nabla^2 - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right)p = 0, \tag{1}$$

with ∇^2 the Laplacian operator, c_0 the wave speed, and p the pressure. In a modal approach within a uniform waveguide, elementary solutions for the pressure are written as

$$p_i(x,y,z) = e^{j(\pm k_i x - \omega t)}\phi_i(y,z), \tag{2}$$

where $k_i^2 = k^2 - \alpha_i^2$, $\Re\{k_i\} > 0$, $\Im\{k_i\} > 0$, $k = \omega/c_0$, and (α_i, ϕ_i) are the eigensolutions of the transverse eigenproblem

$$\nabla_\perp^2 \phi = -\alpha^2 \phi, \tag{3}$$

with proper boundary conditions, in the cross-section of the waveguide. Then, a solution in the waveguide can be built as a sum on these elementary solutions [the time dependence $\exp(-j\omega t)$ is omitted]:

$$p(x,y,z,\omega) = \sum_{i\geq 1} (a_i e^{jk_i x} + b_i e^{-jk_i x})\phi_i(y,z), \tag{4}$$

with $i \geq 1$ an integer number and the coefficients a_i and b_i are found as functions of the boundary conditions defined at the waveguide extremities. The transverse modes of the open waveguide shown in Fig. 1 are written as the solutions (α, ϕ) of the eigenproblem (3) with boundary conditions

$$\partial_z \phi = 0 \quad \text{if } z = 0 \quad \text{and} \quad |y| > L/2, \tag{5a}$$

$$\partial_z \phi = 0 \quad \text{if } z = d \quad \text{and} \quad |y| < L/2, \tag{5b}$$

$$\partial_y \phi = 0 \quad \text{if } y = \pm L/2 \quad \text{and} \quad 0 < z < d \tag{5c}$$

in the cross-section of the waveguide, regarded as a 2D rectangular cavity Ω_1 of width L and depth d, open on the semi-infinite space Ω_2 (Fig. 2).

As we present below, the modes $\phi_i(y,z)$ can be found by solving the continuity equation at the interface between Ω_1 and Ω_2. A similar problem, with elastic waves, has been treated by Maradudin and Ryan.[23] In their work, authors have calculated the discrete frequencies of the elastic vibration modes of a 2D rectangular ridge fabricated from one material that is bonded on the planar surface of a substrate of a second material. In the following, part of the equations are derived from this work to be adapted in the case of a fluid resonant open cavity.

In the cavity $(z \geq 0)$, the general solution satisfying boundary conditions can be written as a discrete sum of functions,

$$\phi(y, z \geq 0) = \sum_{n\geq 0} A_n \cos(\beta_n(z-d))\psi_n(y) \tag{6}$$

with d the depth of the cavity, $\beta_n^2 = \alpha^2 - (n\pi/L)^2$ with $n \geq 0$ an integer number and

$$\psi_n(y) = \sqrt{2 - \delta_{n0}}\cos\left(\frac{n\pi}{L}\left(y - \frac{L}{2}\right)\right), \tag{7}$$

where δ_{mn} is the Kronecker symbol. Note that a finite impedance at the walls (absorbing material) could be considered by, e.g., modifying expression (7), although it could result in a less straightforward formulation. An alternative, then, would be to solve the transverse problem using a FEM. This would be also particularly adapted in the case of more complex cavity shapes, modeling facades irregularities.

Above the cavity $(z \leq 0)$, the general solution is written as the spatial Fourier transform

$$\phi(y,z\le 0)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}Be^{j(\alpha_y y-\alpha_z z)}d\alpha_y, \tag{8}$$

where $\alpha_z^2=\alpha^2-\alpha_y^2$ with $\Re\{\alpha_z\}\ge 0$ and $\Im\{\alpha_z\}\ge 0$.

Continuity equations in the interface plane $z=0$, $|y|<L/2$ are

$$\phi(y,z=0^+)=\phi(y,z=0^-) \tag{9}$$

and

$$\partial_z\phi(y,z=0^+)=\partial_z\phi(y,z=0^-). \tag{10}$$

A first relation between the $\{A_n\}$ and B is found by substituting Eqs. (6) and (8) for ϕ in combined Eqs. (5a) and (10):

$$\sum_{n\ge 0}A_n\beta_n\sin(\beta_n d)\psi_n(y)=-j\frac{1}{2\pi}\int_{-\infty}^{+\infty}B\alpha_z e^{j\alpha_y y}d\alpha_y, \tag{11}$$

to find

$$B=j\sum_{n\ge 0}A_n\beta_n\sin(\beta_n d)\frac{S_n(\alpha_y)}{\alpha_z}, \tag{12}$$

where

$$S_n(\alpha_y)=\int_{-L/2}^{L/2}\psi_n(y)e^{-j\alpha_y y}dy. \tag{13}$$

A second relation is found by using the continuity equation (9):

$$\sum_{n\ge 0}A_n\cos(\beta_n d)\psi_n(y)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}Be^{j\alpha_y y}d\alpha_y; \tag{14}$$

whence it follows that

$$A_m\cos(\beta_m d)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}BS_m^*(\alpha_y)d\alpha_y, \tag{15}$$

where $m\ge 0$ is an integer. Then, Eqs. (12) and (15) lead to the set of linear, homogeneous equations (16) for the $\{A_n\}$: $\forall m\in \mathrm{IN}$,

$$A_m\cos(\beta_m d)=j\sum_{n\ge 0}\Pi_{mn}(\alpha)\beta_n A_n\sin(\beta_n d), \tag{16}$$

where

$$\Pi_{mn}=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{S_m^*S_n}{\alpha_z}d\alpha_y. \tag{17}$$

It is easily shown that, for real values of α_y, Π_{mn} vanishes unless m and n have the same parity.[23] Then, Eq. (16) breaks up into two sets of linear equations, one governing the symmetrical eigenmodes (even functions of y with even values of m and n), and the other governing antisymmetrical eigenmodes (odd functions of y with odd values of m and n). The following equation gives a general expression of Π_{mn} for both even or odd indices m and n:

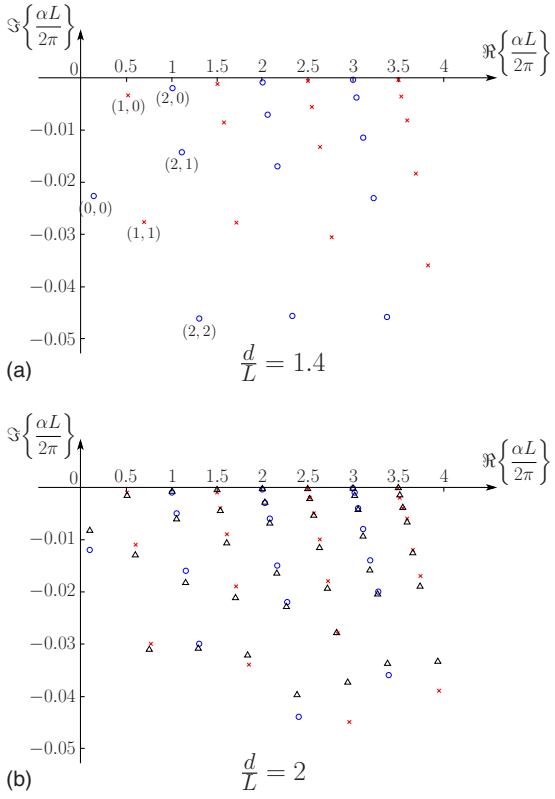


FIG. 3. (Color online) Spectrums of complex eigenvalues of the transverse problem for two different values of the aspect ratio: (a) $d/L=1.4$; (b) $d/L=2$. ○: symmetric modes; ×: antisymmetric modes; △: eigenvalues computed with FEM.

$$\pi_{mn}(\alpha)=(-1)^{(3m+n/2)}\frac{\sqrt{2-\delta_{m0}}\sqrt{2-\delta_{n0}}}{n}\times\int_0^{+\infty}\frac{\alpha_y^2}{\alpha_z\left(\alpha_y+\frac{m\pi}{L}\right)\left(\alpha_y+\frac{n\pi}{L}\right)}\times\mathrm{sinc}\left(\frac{\alpha_y-\frac{m\pi}{L}}{2}\right)\mathrm{sinc}\left(\frac{\alpha_y-\frac{n\pi}{L}}{2}\right)d\alpha_y. \tag{18}$$

Finally, Eq. (16) can be written in the matrix form

$$D\vec{\mathcal{A}}=\vec{0}, \tag{19}$$

where the components of vector $\vec{\mathcal{A}}$ are $\mathcal{A}_n=A_n\sin(\beta_n d)$ and terms of matrix D are

$$D_{mn}(\alpha)=\cot(\beta_m d)\delta_{mn}-j\Pi_{mn}\beta_n. \tag{20}$$

Then, the eigenvalues of the transverse eigenproblem are the values α_i of α for which $\det(D)=0$, and the eigenfunctions ϕ_i are given by the corresponding set of coefficients $A_n^{(i)}$, satisfying Eq. (19) with $D=D(\alpha_i)$.

B. Numerical resolution

After truncation of Eq. (19) at a finite-size matrix problem, zeros of the determinant of D are numerically located in the complex α-plane to compute eigenvalues α_i of the transverse problem (Fig. 3).

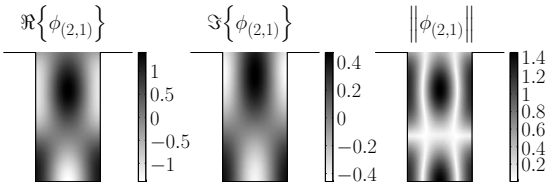


FIG. 4. Symmetric mode $\phi_{(2,1)}$ (real part, imaginary part, and modulus). The indices (2, 1) are chosen following an analogy with the classical eigenmodes in a closed cavity with a Dirichlet condition $\phi=0$ on the upper boundary: these indices denote the number of vertical and horizontal nodal lines.

Owing to the radiation losses in the infinite space Ω_2 (above the waveguide), eigenvalues are complex, lying in the lower half-plane.[5,23–25] The spectrum displays families of eigenvalues corresponding to either symmetric (blue circles ○) or antisymmetric modes (red crosses ×). An analogy with the classical, real, modes of the simple problem with a Dirichlet condition ($\phi=0$) at $z=0$ instead of the exact radiating condition written in Sec. II A, allows us to label the complex modes ϕ_i—at least the modes which eigenvalue is located close enough to the real axis in the complex α-plane—with a couple of integers denoting the number of vertical and horizontal nodal lines (Fig. 4). Following this terminology, the families displayed in Fig. 3 are the $\phi_{(p,q)}$ with p constant.

Figure 3 shows the spectrum of eigenvalues for two different values of the aspect ratio of the cavity: (a) $d/L=1.4$ and (b) $d/L=2$. The pattern in both plots is similar, exhibiting the families of modes $\phi_{(p,q)}$ with p constant. However, in the "deeper" cavity, with the aspect ratio $d/L=2$, the confinement of the modes is more important; thus, eigenvalues α_i have a smaller imaginary part than in the cavity with aspect ratio 1.4.

For comparison, a finite element method is used to solve the transverse eigenproblem (3). The semi-infinite space above the cavity is bounded with PMLs, as used by Hein *et al.*[26] and Koch[5] in a similar problem (Fig. 5).

The results shown in this paper [Figs. 3(b) and 6] have been obtained with parameters $d'=L/2$, $l'=L$, $d_{\mathrm{PML}}=L/2$, and $\tau_y=\tau_z=1+j$ (see Appendix). A Dirichlet condition $\phi=0$ is imposed on the outer boundaries of the PML. Moreover, as the geometry of the cross-section is symmetric about the z axis, one-half only of the domain is meshed, with the appropriate symmetry or antisymmetry condition imposed at $y=0$. Computations have been performed using the partial differential equation toolbox from MATLAB®.

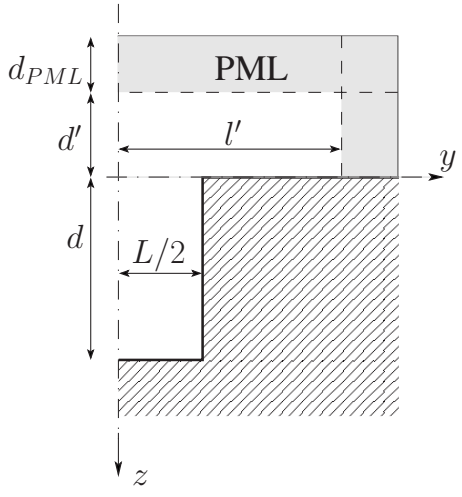


FIG. 5. Geometry of the domain with PML in the FEM computations.

The results of the two compared methods—the resolution of Eq. (19) and the FEM—are in good agreement [Fig. 3(b)]. The discrepancy between the results increases for larger values of the imaginary part $\Im\{\alpha_i\}$. However, eigenvalues that are less well-estimated are associated with modes that will be strongly attenuated when propagated in the waveguide. They are then of secondary importance when considering the transport of energy on a sufficiently long distance. Moreover, it will be shown in the following that the contribution of these modes in the determination of the sets of coefficients $\{a_i\}$ and $\{b_i\}$ is almost negligible.

A comparison between the eigenfunctions ϕ_i deduced from the method detailed in Sec. II A and from the FEM computation shows also a good agreement (Fig. 6). Both methods give very similar results, even when the error on the imaginary part of α_i becomes more significant (Fig. 6, bottom). The figure also clearly shows that low order modes weakly "radiate" in the infinite space Ω_2: the effect of the opening of the waveguide appears as a small perturbation on the classical, nonradiating, solutions that would be obtained by applying a homogeneous Dirichlet condition at the top of the waveguide ($z=0$). For higher order modes, however, eigenfunctions ϕ_i differ more and more from the real, "Dirichlet," solutions. Patterns of nodal lines are more complex. Consequently, the indexation with indices (p,q) become less relevant, and the confinement, which was strong for low order modes, becomes weaker, with the energy increasing near the interface $z=0$.

Now that the transverse eigenmodes are determined, they can be used to write a multimodal formulation of the sound propagation within the open waveguide that displays a street canyon.

III. PROPAGATION ALONG THE STREET

As explained in Sec. I, the transverse modes ϕ_i are used, for given source and radiation conditions at the ends of the waveguide, to built a solution of the wave equation, as written in Eq. (4).

Because the transverse eigenvalues α_i are complex with $\Im\{\alpha_i\}<0$, the propagation constants k_i are also complex, with $\Im\{k_i\}>0$, even for real source frequency ω. Then, all the modes $\phi_i \exp(\pm jk_i x)$ in the waveguide decrease exponentially while propagating, reflecting the radiation losses during the propagation along the open waveguide. This corresponds to leaky modes.

Then, the two sets of coefficients $\{a_i\}$ and $\{b_i\}$ must be found, as functions of the end conditions in the waveguide. At the input end of the waveguide, a source condition is defined as a given acoustic pressure distribution in a plane $x=$const, with frequency ω. For example, at $x=0$: $p(0,y,z,\omega)=p_0(y,z)\exp(-j\omega t)$. At the output end of the waveguide is given a radiation condition.

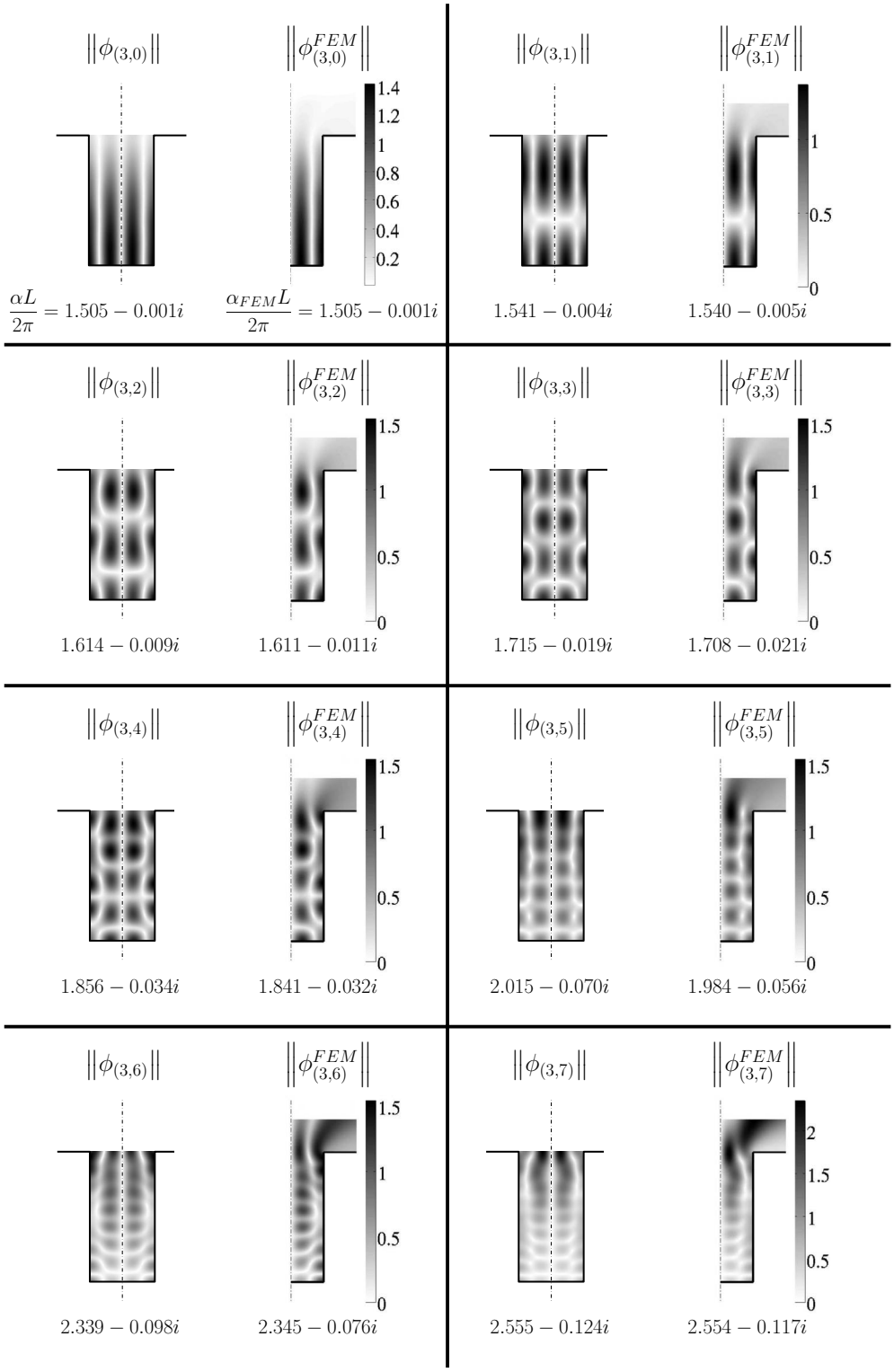


FIG. 6. Antisymmetric eigenfunctions $\phi_{(3,q)}$, $q>0$, and comparison with FEM computations. The PML domain is not shown in the FEM results.

A. Input condition

Let us call $P_i(x)=a_i \exp(jk_i x)+b_i \exp(-jk_i x)$ the coefficients in the development in series (4). Since the modes ϕ_i are not orthogonal, the initial field $p_0(y,z,\omega)$ cannot be projected on the $\{\phi_i\}$ as it is classically made to find the $\{P_i(0)\}$. Thus, after truncation at a finite number $\mathcal{N}$ of terms in the development (4), a least squares method is thus used to find these coefficients:[27]

$$\vec{P}(0) = (\Lambda^{(\phi)})^{-1} \vec{p}_0^{(\phi)}, \tag{21}$$

where the ith component of $\vec{P}(0)$ is $P_i(0)$ and

$$\Lambda_{ij}^{(\phi)} = \langle \phi_i | \phi_j \rangle, \tag{22a}$$

$$p_{0i}^{(\phi)} = \langle \phi_i | p_0 \rangle, \tag{22b}$$

with the product

$$\langle f|g\rangle = \int\int \bar{f} g d\Omega. \tag{23}$$

Practically, we consider source conditions $p_0(y,z,\omega)$ with a compact support, included in Ω_1, i.e., inside the street canyon. Thus, for convenience when determining $\vec{P}(0)$, we consider, as the eigenfunctions ϕ_i, their restriction to the domain Ω_1, that is,

$$\forall (y,z) \in \left[-\frac{L}{2},\frac{L}{2}\right] \times [0,d],$$

$$\phi_i(y,z) = \sum_{n\geq 0} A_n^{(i)} \cos(\beta_n^{(i)}(d-z))\psi_n(y), \tag{24}$$

and the product $\langle \mathrm{l} \rangle$ above is

$$\langle f|g\rangle = \int_0^d \int_{-L/2}^{L/2} \bar{f} g dy dz. \tag{25}$$

Note that in the case of real orthogonal modes, as in classical closed waveguides, the least squares method gives the usual projection coefficients $P_m = \langle \phi_m | p \rangle$.

B. Output condition

Let $Q_i(x) = jk_i(a_i \exp(jk_i x) - b_i \exp(-jk_i x))$ be the coefficients in the development of the x-component of $\vec{\nabla} p$. One assumes that at the output end of the waveguide, say, at $x = x_{\text{end}}$, the condition is given as an admittance matrix Y_{end} fulfilling $\vec{Q}(x_{\text{end}}) = Y_{\text{end}} \vec{P}(x_{\text{end}})$.

Again, as for the formulation at the input end, and due to the non-orthogonality of the eigenmodes, the matrix Y_{end}, for some complex end conditions, may not be straightforwardly calculated. However, usual end conditions—rigid end, non-radiating open end, anechoic termination—can be easily formulated with this type of admittance matrix, generalization for all modes of the usual admittance for the plane wave.[28]

C. Solutions for $\{a_i\}$ and $\{b_i\}$

Now that an input condition $\vec{P}(O)$ and an output condition Y_{end} are known, the vectors $\vec{a}$ and $\vec{b}$ of the $\{a_i\}$ and $\{b_i\}$ in Eq. (4) can be calculated:[29,28]

$$\vec{a} = (1-\delta)^{-1} \vec{P}(0), \tag{26}$$

$$\vec{b} = -\delta(1-\delta)^{-1} \vec{P}(0), \tag{27}$$

where $\delta = D_1(Y_{\text{end}} + Y_c)^{-1}(Y_{\text{end}} - Y_c)D_1$, D_1 is diagonal with terms $D_{1i} = \exp(jk_i x_{\text{end}})$, Y_c is diagonal with terms $Y_{ci} = jk_i$. Thus, with these solutions for $\vec{a}$ and $\vec{b}$, the pressure field in the open waveguide can be calculated. However, terms $\exp(-jk_i x)$ can be the source of numerical problems of convergence. Then, defining $\tilde{\vec{b}} = D_1^{-1}\vec{b}$, the pressure field is written as

$$p^{(N)}(x,y,z,\omega) = \sum_{i\geq 1} (a_i e^{jk_i x} + \tilde{b}_i e^{jk_i(x_{\text{end}}-x)}) \phi_i(y,z). \tag{28}$$

This new formulation depends only on D_1, not on D_1^{-1}, and on exponentials with positive arguments x or $L-x$.[30]

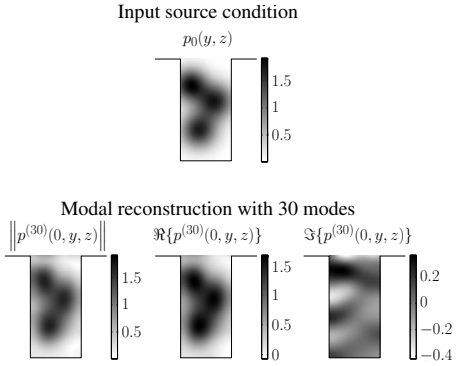


FIG. 7. Top: Initial condition at abscissa $x=0$. The parameters of the three Gaussian functions in Eq. (29) are $(y_1,z_1,\sigma_1)=(0.18L,0.42d,0.24)$, $(y_2,z_2,\sigma_2)=(-0.30L,0.26d,0.22)$, and $(y_3,z_3,\sigma_3)=(-0.14L,0.70d,0.24)$. Bottom: Modal reconstruction of the initial condition $p_0(y,z)$, using $N=\mathcal{N}=30$ modes.

IV. RESULTS

In the following, for simplicity, we will consider the wave field downstream from a source in an infinite waveguide. Then, there are no back propagated waves: $b_i=0$.

A. Input condition

The initial condition at $x=0$ is the pressure distribution shown in Fig. 7 (left part) and given by

$$p_0(y,z) = \sum_{k=1}^{3} \frac{1}{\sigma_k\sqrt{2\pi}} e^{-(y-y_k)^2-(z-z_k)^2/2\sigma_k^2}, \tag{29}$$

where $\sigma_k \in \mathbb{R}^+$ and $(y_k,z_k) \in [-L/2,L/2]\times[0,d]$. The associated dimensionless frequency is $kL/2\pi = 1.2$. This input condition is chosen as a nontrivial solution for the modal formulation. Moreover, it can be seen as a simple and general way to describe several, spatially distributed, sources.

Then, the $\{a_i\}$ are found by substituting Eq. (29) for $p_0(y,z)$ in Eqs. (21) and (22). The modal reconstruction is shown in Fig. 7. Using a basis of $N=\mathcal{N}=30$ modes [we recall that N is the number of terms in the series (28) and $\mathcal{N}$ is the size of the linear problem (21) in the least square estimation of the $\{a_i\}$, so that $N \leq \mathcal{N}$], the input pressure condition is well reproduced with a residual error of 3.6%. This error is due to the high order depth modes of the first families that have deliberately not been considered in the modal basis because of their weak of relevance in the propagation. Furthermore, it will be shown in the following that omitting these modes does not affect significantly the estimation of the $\{a_i\}$ for the modes taken into account.

To evaluate the convergence of the method when increasing the number N of modes taken into account in Eq. (28), from a $\mathcal{N}=30$ modes basis, an error ϵ is defined as

$$\epsilon = \sqrt{\frac{\int_0^d \int_{-L/2}^{L/2} \|p^{(N)} - p_0\|^2 dy dz}{\int_0^d \int_{-L/2}^{L/2} \|p_0\|^2 dy dz}}, \tag{30}$$

where $p^{(N)}$ is the modal solution obtained with N modes [Eq. (28)], and p_0 the reference field. The modes are sorted by

TABLE I. Classification of the modes by increasing value of $\Im\{k_i\}$ at frequency $kL/2\pi=1.2$.

Classification	Couple (m,n)
1st	(1,0)
2nd	(0,0)
3rd	(2,0)
4th	(1,1)
5th	(2,1)
6th	(0,1)
7th	(1,2)
8th	(2,2)
9th	(0,2)
10th	(1,3)

increasing value of the imaginary part of their propagation constant k_i, that is, from the least damped to the most damped leaky mode propagating along the canyon (Table I). This type of classification depends directly on each eigenvalue and the source frequency.

The convergence of the reconstruction of the initial condition $p_0(y,z)$ is shown in Fig. 8 (circles ○). Naturally in such a modal approach, depending on the source distribution, each mode introduced in the computation contributes differently to the reconstruction of $p_0(y,z)$.

B. Propagation within the waveguide

Assuming that the convergence is reached for $N=30$, the field $p^{(30)}(x,y,z,\omega)$ is now taken as the reference field to compute an estimation error at abscissa $x=10L$ (triangles △ in Fig. 8). The variability, depending on the initial condition $p_0(y,z)$, of the contributions of the modes to the convergence is still visible, but, moreover, it clearly appears that only a few modes—the less damped modes—still contribute to the transport of energy at that distance from the "source." Practically, sorting the modes as done in Table I is thus a good choice to increase the convergence, as soon as one is interested by the wave field in the street canyon at a sufficient distance from the source.

Figure 9 shows the field in the cross-section of the waveguide at abscissa $x=10L$, L the width of the waveguide. The left plot is obtained using Eq. (28) and $N=30$—the total number of modes used to perform the least squares estimation, while the right plot is obtained using only the first six

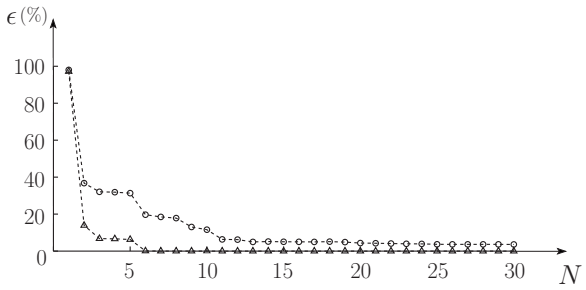


FIG. 8. Evolution of the error indicator ϵ with the number of modes N taken into account in the computation of the solution. ○: the error is computed at $x=0$, with ϵ given in Eq. (30). △: the error is computed at $x=10L$, with $p^{(30)}(10L,y,z,\omega)$ as the reference field.

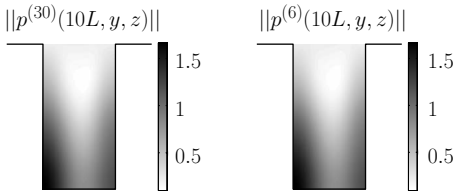


FIG. 9. Modal field at $x=10L$ using $N=30$ modes (left) and $N=6$ modes (right).

modes, with the ordering defined above. Both results are very close: the relative error between them, defined as in Eq. (30), is less than 0.35%.

Since only a few modes are necessary to describe the field at certain distance from the source, it would be advantageous to use a reduced modal basis in the computation. Since modes are not orthogonal, the value of each modal coefficient a_i depends on the size of the "basis" used in the least squares estimation. One shows, however, that this dependency is rapidly weak, in particular, for the first modes, that is, the less damped (Fig. 10). The notation $a_i^{(\mathcal{N})}$ is used to denote the number of modes $\mathcal{N}$ taken into account for the least squares estimation of a_i.

To evaluate the relevance of using a reduced modal basis, two modal solutions of the wave equation in the infinite street canyon with the initial condition (29) at $x=0$ are compared: the solution with $N=\mathcal{N}=10$, and the solution with N

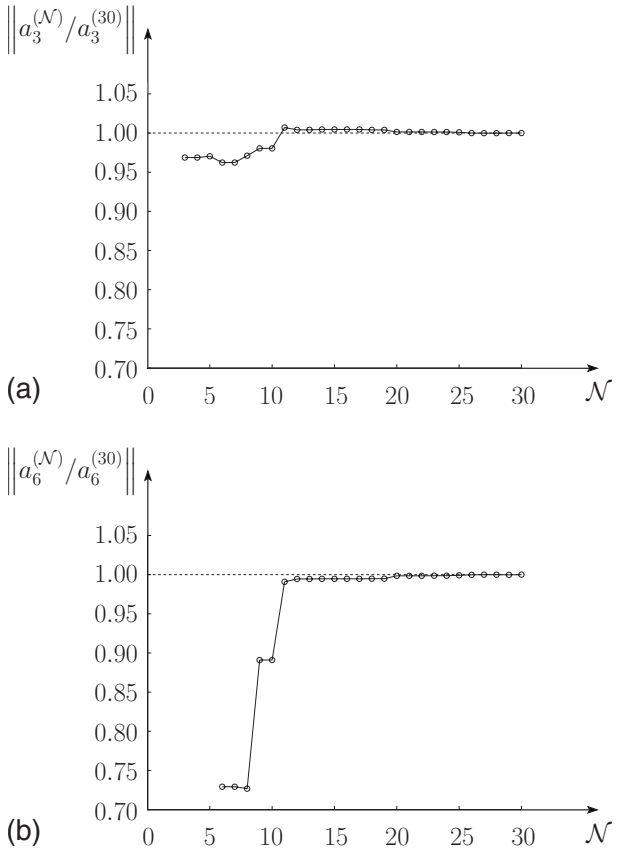


FIG. 10. Evolution of the modal coefficients a_3 (a) and a_6 (b) with the number of modes $\mathcal{N}$ used for the least squares estimation.

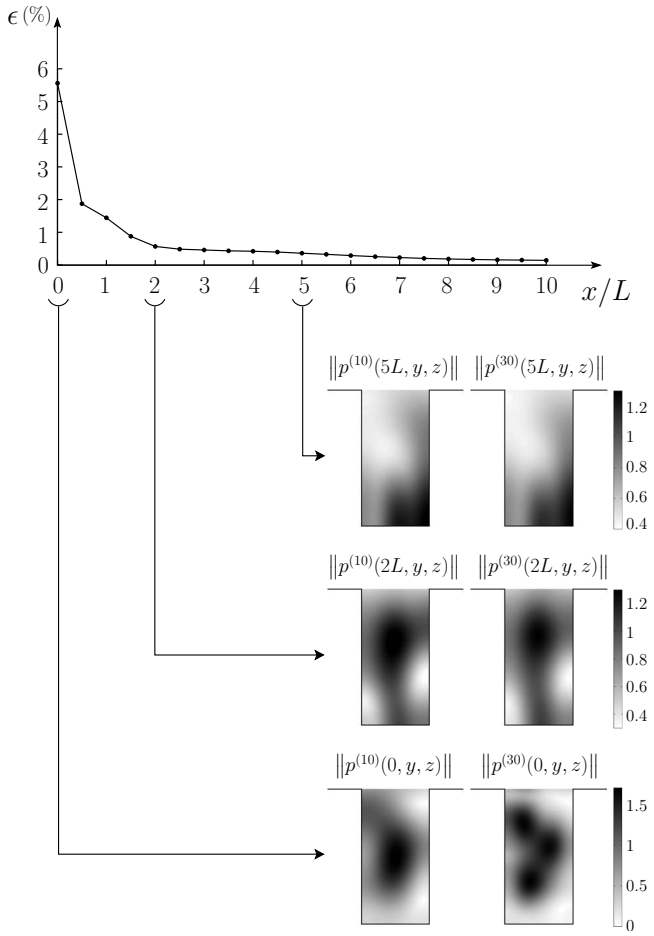


FIG. 11. Relative error [as defined by Eq. (30)] between the modal solution (28) with $N=\mathcal{N}=10$ and the same modal solution with $N=\mathcal{N}=30$, as function of the distance from the source plane.

$=\mathcal{N}=30$. The relative error between these two solutions, as defined in Eq. (30), is plotted in Fig. 11 as function of the distance from the source plane.

Naturally, the source condition $p_0(y,z)$ is badly reconstructed with a limited number of modes, and the relative error near the plane $x=0$ is thus significant. But it decreases rapidly, to be under 1% after less than two widths. It follows that the acoustic field, rapidly, is "carried" by a small number of modes, the less damped modes, that are therefore well confined and guided in the open geometry of the street canyon. This points out a double interest of the modal formulation using leaky modes in a street canyon: first, as a physically relevant approach, describing the competitive effects of confinement and radiation in such an open geometry, second, as an efficient numerical method, since a few modes only is used to accurately model the wave propagation in the waveguide.

V. CONCLUSIONS

The problem of sound propagation in an idealized urban street canyon is solved using a multimodal formulation that gives the acoustic field as a sum on the leaky modes of the open waveguide that displays the canyon. The leaky modes that naturally reflects the competitive effects of confinement and radiation of the wave in such partially bounded geometries can be numerically determined. As these are complex modes that decay exponentially while propagating, the number of modes that effectively carry the wave field (emitted by some source in the waveguide) decreases rapidly, so that only a few modes, at a sufficiently large distance from the source, is necessary to accurately model the wave propagation. This gives this approach a real interest for numerical computations, in addition to its interest as a physically meaningful description of the street as a partially confining and guiding medium for the acoustic waves. As a first step and to clearly point out the principles and interests of our approach, only a uniform open waveguide was considered in this paper. Following, cross-section discontinuities, as model of the junctions between buildings, can be considered by using mode matching techniques.

APPENDIX: FORMULATION IN THE PERFECTLY MATCHED LAYERS

PML are used as a to avoid non-physical reflections at the boundaries of a necessarily finite domain in a numerical computation. The method works as follows: the solution $\phi(y,z)$ of the eigenproblem (3) above the cavity is analytically continued in the PML with respect to variables (y,z) to complex variables $(\hat{y},\hat{z})$. The extended solution $\hat{\phi}$ satisfies

$$\left(\frac{\partial^2}{\partial \hat{y}^2}+\frac{\partial^2}{\partial \hat{z}^2}\right)\hat{\phi}=0. \tag{A1}$$

Complex variables $(\hat{y},\hat{z})$ are now written as

$$\hat{y}(y)=\int_0^y \tau_y(y')dy', \quad \hat{z}(z)=\int_0^z \tau_z(z')dz' \tag{A2}$$

with $\Re\{\tau_{y,z}\}>0$, $\Im\{\tau_{y,z}\}>0$, and $\tau_y(y\le l')=1$, $\tau_z(z\le d')=1$ (Fig. 5). The results in this paper have been obtained with $\tau_y(y>l')=\tau_z(z>d')=1+j$.

[1] F. M. Wiener, C. I. Malme, and C. M. Gogos, "Sound propagation in urban areas," J. Acoust. Soc. Am. **37**, 738–747 (1965).
[2] W. R. Schlatter, "Sound power measurements in a semi-confined space," MS thesis, MIT, Cambridge, MA (1971).
[3] P. R. Donovan, "Sound propagation in urban spaces," Ph.D. thesis, MIT, Cambridge, MA (1976).
[4] H. G. Davies, "Noise propagation in corridors," J. Acoust. Soc. Am. **53**, 1253–1262 (1973).
[5] W. Koch, "Acoustic resonances in rectangular open cavities," AIAA J. **43**, 2342–2349 (2005).
[6] P. Stenackers, H. Myncke, and A. Cops, "Reverberation in town streets," Acustica **40**, 115–119 (1978).
[7] R. H. Lyon, "Role of multiple reflections and reverberation in urban noise propagation," J. Acoust. Soc. Am. **55**, 493–503 (1974).
[8] D. J. Oldham and M. M. Radwan, "Sound propagation in streets," Build. Acoust. **1**, 65–87 (1994).
[9] J. Kang, "Sound propagation in street canyons: Comparison between diffusely and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[10] K. K. Iu and K. M. Li, "The propagation of sound in narrow street canyons," J. Acoust. Soc. Am. **112**, 537–550 (2002).
[11] J. Kang, "Numerical modelling of the sound fields in urban streets with diffusely reflecting boundaries," J. Sound Vib. **258**, 793–813 (2002).
[12] J. S. Bradley, "A study of traffic noise around buildings," Acustica **38**, 247–252 (1977).
[13] J. Picaut, L. Simon, and J. Hardy, "Field modelling in a street with a diffusion equation," J. Acoust. Soc. Am. **106**, 2638–2645 (1999).

[14]E. Walerian, R. Janczur, and M. Czechowicz, "Sound level forecasting for city centers. Part I: Sound level due to a road within an urban canyon," Appl. Acoust. **62**, 359–380 (2001).
[15]R. Bullen and F. Fricke, "Sound propagation in a street," J. Sound Vib. **46**, 33–42 (1976).
[16]R. Bullen and F. Fricke, "Sound propagation at a street intersection in an urban environment," J. Sound Vib. **54**, 123–129 (1977).
[17]J. Defrance and Y. Gabillet, "A new analytical method for the calculation of outdoor noise propagation," Appl. Acoust. **57**, 109–127 (1999).
[18]J. Forssén, "Calculation of noise barrier performance in a turbulent atmosphere by using substitute sources above the barrier," Acta. Acust. Acust. **86**, 269–275 (2000).
[19]M. Ögren and W. Kropp, "Road traffic noise propagation between two dimensional city canyons using an equivalent sources approach," Acta. Acust. Acust. **90**, 293–300 (2004).
[20]M. Hornikx and J. Forssén, "Noise abatement schemes for shielded canyons," Appl. Acoust. **70**, 267–283 (2009).
[21]D. Heimann, "Three-dimensional linearised Euler mode simulations of sound propagation in idealised situations with wind effects," Appl. Acoust. **68**, 217–237 (2007).
[22]T. Van Renterghem, E. Salomons, and D. Botteldooren, "Efficient FDTD-PE model for sound propagation in situations with complex obstacles and wind profiles," Acta. Acust. Acust. **61**, 671–679 (2005).
[23]A. A. Maradudin and P. Ryan, "Shear horizontal acoustic surface resonances," Phys. Rev. B **38**, 3068–3074 (1988).
[24]S. Marburg, "Normal modes in external acoustics. Part I: Investigation of the one-dimensional duct problem," Acta. Acust. Acust. **91**, 1063–1078 (2005).
[25]C. K. W. Tam, "The acoustic modes of a two-dimensional rectangular cavity," J. Sound Vib. **49**, 353–364 (1976).
[26]S. Hein, T. Hohage, and W. Koch, "On resonances in open systems," J. Fluid Mech. **506**, 255–284 (2004).
[27]S. Félix, A. Pelat, and V. Pagneux, "Waveguide modes of a long open enclosure," in Proceedings of International Symposium on Room Acoustics (2007).
[28]V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. Part I. Theory and validation," J. Acoust. Soc. Am. **100**, 2034–2048 (1996).
[29]S. Félix and V. Pagneux, "Sound attenuation in lined bends," J. Acoust. Soc. Am. **116**, 1921–1931 (2004).
[30]K. P. Lee and H. G. Davies, "Monogram for estimating noise propagation in urban areas," J. Acoust. Soc. Am. **57**, 1477–1480 (1975).

Three dimensional finite element modeling of smart foam

Pierre Leroy, Noureddine Atalla, and Alain Berry
Groupe d'Acoustique de l'Universite de Sherbrooke, Sherbrooke, Quebec J1K 2R1, Canada

Philippe Herzog
Laboratoire de Mecanique et d'Acoustique, 13402 Marseille cedex 20, France

(Received 29 August 2008; revised 17 August 2009; accepted 31 August 2009)

The "smart foam" concept and design originate from the combination of the passive dissipation capability of foam in the medium and high frequency ranges and the active absorption ability of piezoelectric actuator (generally polyvinylidene fluoride) in the low frequency range. This results into a passive/active absorption control device that can efficiently operate over a broad range of frequencies. In this paper, a full three dimensional finite element model of smart foam is presented including its experimental validation. The modeling tool uses quadratic poroelastic elements, as well as elastic, fluid, and piezoelectric elements. The weak integral formulation of the different media involved is presented with the associated coupling conditions. A simplified orthotropic model of poroelastic media is presented. To validate the developed model, a prototype of a smart foam has been realized and its passive absorption and radiation measured in an impedance tube and compared to predictions. The experimental validation demonstrates the validity of the model. This modeling tool constitutes a general platform to simulate and optimize various configurations of smart foams. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238241]

PACS number(s): 43.20.Tb, 43.20.Jr, 43.38.Fx, 43.50.Ki [KVH] Pages: 2873–2885

I. INTRODUCTION

Improving low frequency absorption is an important topic in acoustics. Passive materials generally provide adequate absorption at medium and high frequencies whereas active control is efficient to cancel low frequency sound waves. Many devices combining sound absorbing passive materials and active absorption properties have been studied. One can distinguish between two main approaches. The first approach consists in associating a passive porous layer with an active surface separated from the rear face of the porous layer by an air gap. There are here two different control strategies. The first strategy imposes a zero pressure on the back surface of the porous layer.[1–7] The second strategy consists in controlling the surface impedance of the active surface in order to cancel the reflected sound wave.[8–12] These two strategies have proved to be efficient for a broad frequency range but result in a weight and space penalty that may limit their application in industrial sectors such as aerospace. Indeed, the actuators used for these strategies are often composed of a piezoceramic bonded on a metallic plate or flat speaker that has a non-negligible weight and the air gap between the actuator and the absorbing material has to be almost 1 cm to enable the control. Moreover, the efficiency is very low for frequencies below 500 Hz when a piezoceramic is used as actuator. To overcome these limitations, a second approach, called "adaptative foam" or "smart foam," consists in a control actuator [generally a piezoelectric polyvinylidene fluoride (PVDF) membrane] directly embedded in a foam layer[13–18] (Fig. 1). Smart foams can be used to reduce structural radiation or increase sound absorption. The general context of this work is the application of smart foams to sound absorption.

Only few models have been developed to describe the complex acoustical and structural behaviors of smart foams. Mathur *et al.*[19] proposed a two dimensional (2D) finite element model able to couple poroelastic, acoustic, elastic, and piezoelectric media. In their model, the piezoelectric media were modeled with a linear plate element with only one direction of strain. They did not experimentally validate their model. Akl *et al.*[20] developed a 2D model for poroelastic media in order to more easily couple elastic and piezoelectric plates to the foam. They developed a new type of smart foam by replacing the PVDF film with piezoelectric fiber modeled in one dimension. All the above models use a displacement based (u,U) formulation for the porous media. The displacement based formulation (u,U) is computational and space memory consuming because it describes the behavior of the solid and the fluid phase using three degrees of freedom (DOFs) of displacement (translation) for each phase. The previous models use only linear elements in the finite element method implementation, which involves the need of very high mesh discretization in order to achieve the convergence of the model. Moreover, the piezoelectric modeling is limited to one direction of strain, which is erroneous for a PVDF film. In addition, no experimental validation has been directly conducted for the smart foam model in a radiation configuration.

In this paper, a full three dimensional (3D) finite element modeling tool for smart foam is presented. The model predicts acoustical and structural behaviors of coupled media such as piezoelectric, elastic, poroelastic, and fluid. The modeling of the poroelastic media uses the enhanced mixed pressure-displacement (u,p) formulation proposed by Atalla *et al.*,[21,22] where the solid phase is described by three DOFs of displacement (translation) and the fluid phase is described by a DOF of pressure. This formulation is less time-space

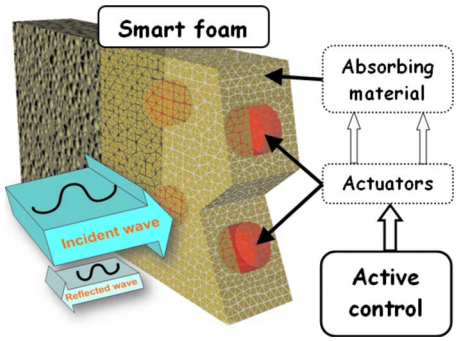


FIG. 1. (Color online) Principe of smart foam for absorption.

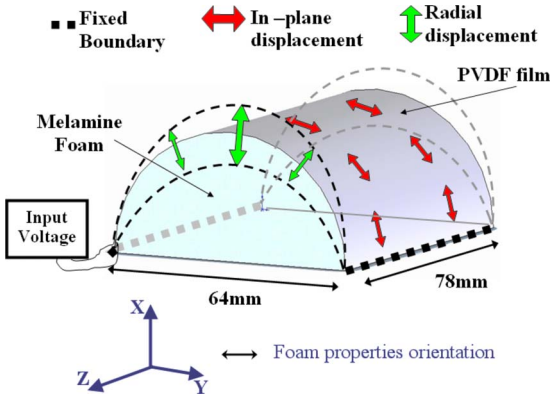


FIG. 2. (Color online) Schematics of the smart foam.

consuming than the (u,U) formulation because it enables the gain of two DOFs by node, and couples easily with elastic and acoustic media. The modeling of the piezoelectric media has been developed in three dimensions using the equations proposed by Tiersten.[23] 2D piezoelectric and elastic plate elements have been also developed. The numerical implementation uses quadratic elements with 20 nodes for volume elements and 8 nodes for surface elements. Various types of loads (point forces, surface forces, and electric potentials) are included in the model.

The presented formulation makes it possible to calculate in local and global scale classical physical indicators, such as spatially averaged quadratic sound pressures and velocities, dissipated power and power exchange in the foam (input and output powers), absorption coefficient, surface impedance, and optimal actuator input in active control. This will help making judicious choices for the optimization of smart foams. However, since this paper is mainly concerned with the presentation and validation of the developed formulation, only local indicators such as acoustic pressure and structural displacement and global indicator such as absorption are presented and compared with experimental measurements. Moreover, the validation focuses on the acoustical and structural behaviors of the coupled smart foam structure without control.

The governing equations of individual domains, as well as coupling conditions, are first presented. The experimental validation is then presented using a prototype realized with a PVDF film bonded onto a melamine foam having a half cylinder shape. The smart foam is backed by a small cavity and radiates in an impedance tube. Passive absorption, radiated sound pressure, and displacement of the foam and the PVDF are presented. Finally, a short numerical example of two smart foam configurations in absorption control situation is presented and analyzed in order to illustrate their control mechanisms.

II. GOVERNING EQUATIONS OF A SMART FOAM

The modeling of adaptive foams is particularly difficult because it involves coupling phenomena relating materials having very different behaviors. The use of numerical methods, such as the finite element method, is then incontournable for the modeling of complex geometrical structures using hybrid active/passive materials.

The finite element method is a computational technique for obtaining approximate solutions to the partial differential equations. The finite element method utilizes a variational problem that involves an integral of the differential equation over the problem domain. This domain is divided into a number of subdomains called finite elements, and the solution of the partial differential equation is approximated by a simpler polynomial function on each element. The variational integral is evaluated as a sum of contributions from each finite element. The main method to obtain the variational problem is the Galerkin method. In order to construct a symmetric variational form the second derivative term is integrated by part and the integral system reduces in a lower degree system. The resulting equation is called the weak integral equation and constitutes the base of the finite element method implementation. In this paper, the weak integral equation has been directly obtained using the principle of virtual work: For any virtual displacement, the sum of the virtual work of the external forces, strain forces, and inertial force must be zero.

Although the equations and the model presented in this paper are general, a simple configuration of smart foam, making it possible to highlight all the involved phenomena, is used here (Fig. 2). The configuration is made up of a half cylinder of melamine foam covered with a PVDF film. The curved shape of the PVDF ensures coupling of in-plane displacement of the film with its efficiently radiating radial displacement.

The general situation of a smart foam involves elastic, poroelastic, piezoelectric, and acoustic domains coupled by connecting surfaces (Fig. 3).

The weak integral formulation for these domains and their relatives' coupling equations are presented in this section. The weak integrals of the piezoelectric and elastic domains are given in two dimensions and three dimensions.

A. Poroelastic domain

A poroelastic medium is an elastic solid (often called matrix) permeated by an interconnected network of pores filled with a viscous fluid. Usually, both the solid matrix and the pore network are assumed to be continuous, so as to form two interpenetrating continua such as in a sponge. Within the

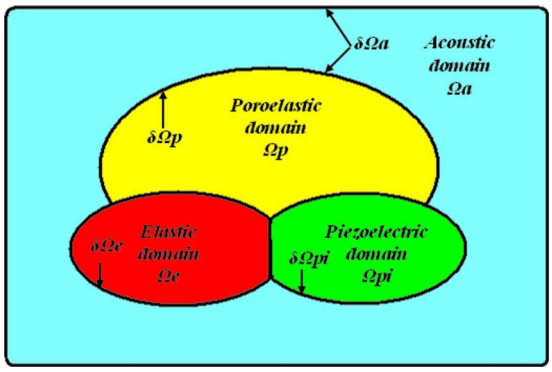


FIG. 3. (Color online) Coupling domains.

framework of the Biot theory, three waves simultaneously propagate in a poroelastic medium: a shear or transverse wave, and two types of longitudinal or compressional waves, which Biot[24,25] called type I and type II waves. The transverse and type I (or fast) longitudinal waves are similar to the transverse and longitudinal waves in an elastic solid, respectively. The slow compressional wave (also known as Biot's slow wave) is specific to poroelastic materials.

Within the Johnson–Allard–Champoux formulation,[26] a poroelastic medium is characterized by its porosity (ϕ), its flow resistivity (σ), its tortuosity (α_∞), and two characteristic lengths [viscous (Λ) and thermal (Λ')], as well as the properties of its constituents (solid matrix and saturating fluid). Assuming isotropy, the solid matrix is characterized by the mass density of the solid matrix (ρ), Young's modulus (E), the Poisson coefficient (ν), and the structural loss factor (η). The fluid is characterized by its density (ρ_f) and speed of sound. These 11 parameters constitute the Biot parameters for an isotropic poroelastic medium. A precise mathematical description of these parameters is given by Allard.[27]

The enhanced weak integral formulation for the mixed (u,p) poroelastic equations proposed by Atalla *et al.*[22] is used here. This formulation allows for a direct coupling between elastic, poroelastic, and piezoelectric materials; i.e., the sum of the surface integral of the poroelastic media and the coupling media is equal to zero. For the coupling with acoustical media, the sum of the surface integral of the acoustical and poroelastic media reduces to the classical fluid-structure coupling surface integral. The weak variational formulation applied to a poroelastic domain Ω_p in the frequency domain results in the following equation:

$$\underbrace{\int_{\Omega_p} \{\varepsilon_{(\delta u)}\}^T [\tilde{H}]\{\varepsilon\} d\Omega}_{\text{work of internal strain forces on the solid part}} - \underbrace{\int_{\Omega_p} \omega^2 \tilde{\rho}\{\delta u\}^T \{u\} d\Omega}_{\text{work of inertial forces on the solid part}}$$
$$+ \underbrace{\int_{\Omega_p} \frac{\phi^2}{\omega^2 \tilde{\rho}_{22}} \{\nabla \delta p\}^T . \{\nabla p\} d\Omega}_{\text{work of inertial forces on the fluid part}} - \underbrace{\int_{\Omega_p} \frac{\phi^2}{\tilde{R}} \delta p . p d\Omega}_{\text{work of internal compressional forces on the fluid part}}$$
$$- \underbrace{\int_{\Omega_p} \frac{\phi}{\tilde{\alpha}} \delta(\{\nabla p\}^T . \{u\})\, d\Omega}_{\text{inertial coupling work between solid and fluid part}} - \underbrace{\int_{\Omega_p} \phi\left(1 + \frac{\tilde{Q}}{\tilde{R}}\right) \delta(p \nabla . u)\, d\Omega}_{\text{elastic coupling work between the solid and fluid part}}$$
$$- \underbrace{\int_{\delta\Omega_p} \phi(U_n - u_n)\delta p\, d\Gamma}_{\text{work of external forces on the fluid part}} - \underbrace{\int_{\delta\Omega_p} \{\delta u\}^T \{\sigma^t\}_n d\Gamma}_{\text{work of external forces on the elastic part}} = 0 \quad \forall (\delta \underline{u}, \delta p) \tag{1}$$

The unknowns for the poroelastic domain are the solid phase displacement field u and the fluid phase pressure field p. δu and δp are two admissible variations of the solid phase displacement field and fluid phase pressure field, respectively. The poroelastic domain and its boundary are denoted by Ω_p and $\delta\Omega_p$. ω is the angular frequency. $\underline{n}$ is the unit normal vector outwards of the surface $\delta\Omega_p$. The subscript n denotes the normal component of a vector. U_n and u_n are the normal macroscopic displacements of the fluid phase and the solid phase on $\delta\Omega_p$, respectively. $\{\varepsilon\}$ is the strain vector of the solid phase. $\{\sigma^t\}$ is the total stress vector of the material and $\{\sigma^t\}_n = \underline{\underline{\sigma^t}} \cdot \underline{n}$ represents the external surface forces per unit area along the normal direction on $\delta\Omega_p$. The tilde symbol is used to indicate that a quantity is complex. $[\tilde{H}]$ is the stiffness matrix of the solid phase. $\tilde{Q}$ is an elastic coupling coefficient between the solid and fluid phases. $\tilde{R}$ is the bulk modulus of the air occupying a fraction ϕ of the volume of the material. $\tilde{\rho}$ and $\widetilde{\rho_{22}}$ are the effective densities of the solid phase and the fluid phase, respectively. $\tilde{\alpha}$ is the dynamic tortuosity. The expressions of $\tilde{Q}$, $\tilde{R}$, $\tilde{\rho}$, $\widetilde{\rho_{22}}$, and $\tilde{\alpha}$ are given in Refs. 21 and 22.

The last two terms in the left hand side of Eq. (1) express the external loads on the poroelastic domain. In the

case of an imposed pressure field p^i applied on $\delta\Omega_p$, the continuity of normal stress leads to

$$\{\sigma^t\}_n = -p^i \underline{n}, \quad p = p^i, \tag{2}$$

and the surface integrals of Eq. (1) simplify to

$$-\int_{\delta\Omega_p} \phi(U_n - u_n)\,\delta p\, d\Gamma - \int_{\delta\Omega_p} \{\delta u\}^T \{\sigma^t\}_n d\Gamma = \int_{\delta\Omega_p} p^i\,\delta u_n d\Gamma. \tag{3}$$

The condition $p=p^i$ has to be explicitly applied in addition to integral (3). In a finite element implementation, this is done through assembling. The case of an imposed point force can be easily deduced from this general case.

In the case of an imposed displacement field $\{u\}^i$ applied on $\delta\Omega_p$, the continuity of the normal displacement between the solid phase and fluid phase and the continuity between the imposed displacement vector and the solid phase displacement vector impose

$$\phi(U_n - u_n) = 0, \quad \{u\} = \{u\}^i, \tag{4}$$

and the surface integrals of Eq. (1) simplify to

$$-\int_{\delta\Omega_p} \phi(U_n - u_n)\,\delta p\, d\Gamma - \int_{\delta\Omega_p} \{\delta u\}^T \{\sigma^t\}_n d\Gamma = 0. \tag{5}$$

This condition is natural and only the kinematic condition $\{u\}=\{u\}^i$ has to be explicitly applied.

Depending on the used model, matrix $[\tilde{H}]$ can be written in the isotropic or orthotropic form.

Mechanical properties of melamine foams are known to be direction dependent.[28] Because of the manufacturing process, several foams are not isotropic; there is a difference between the Biot properties in the rise direction and the other two directions. Anisotropy of the acoustical properties (flow resistivity and tortuosity) is small and negligible.[29] Orthotropy is assumed here and is limited to the mechanical properties. The proposed orthotropy model has been found to better capture the mechanical behavior of the foam. The associated flexibility matrix is presented as follows for the sake of clarity:

$$[\tilde{H}]^{-1} = 1/(1+i\eta) \times \begin{bmatrix} 1/E_x & -\nu_{xy}/E_x & -\nu_{xz}/E_x & 0 & 0 & 0 \\ -\nu_{yx}/E_y & 1/E_y & -\nu_{yz}/E_y & 0 & 0 & 0 \\ -\nu_{zx}/E_z & -\nu_{zy}/E_z & 1/E_z & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/G_{yz} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/G_{xz} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/G_{xy} \end{bmatrix}, \tag{6}$$

where E_i is Young's modulus in the ith direction, G_{ij} is the shear modulus in the ij plane, and ν_{ij} is the Poisson ratio expressing the strain in the jth direction due to a stress in the ith direction. The factor $(1+i\eta)$ is introduced to take into account structural damping, where η is the structural loss factor in the frequency domain. The ν_{ij} satisfy the following equation:

$$\frac{\nu_{ij}}{E_i} = \frac{\nu_{ji}}{E_j}. \tag{7}$$

It is assumed here that the shear moduli G_{xy}, G_{xz}, and G_{yz} take a form similar to the isotropic case

$$G_{ij} = \frac{E_i}{2(1+\nu_{ij})}. \tag{8}$$

B. Acoustic domain

The weak integral of an acoustic fluid domain is[30]

$$\underbrace{\int_{\Omega_a} \frac{1}{\omega^2\rho_0}\{\nabla\delta p\}^T.\{\nabla p\}\, d\Omega}_{\text{work of inertial forces}} - \underbrace{\int_{\Omega_a} \frac{1}{\rho_0 c_0^2}\delta p p\, d\Omega}_{\text{work of compressional forces}} - \underbrace{\int_{\delta\Omega_a} \frac{1}{\omega^2\rho_0}\delta p \frac{\partial p}{\partial n} d\Gamma}_{\text{work of external forces}} = 0 \quad \forall(\delta p) \tag{9}$$

where p is the unknown acoustic pressure and δp is the admissible variation of the pressure field p, ρ_0 is the density of the fluid medium, and c_0 is the speed of sound in the fluid. Ω_a and $\delta\Omega_a$ refer to the acoustic domain and its boundary. n is the outward unit normal vector of the surface $\delta\Omega_a$.

The last term in the left hand side of Eq. (9) represents the action of an imposed pressure field p^i applied on $\delta\Omega_a$. In this case, the continuity of pressure imposes $p=p^i$, such that

$$\int_{\delta\Omega_a} \frac{1}{\omega^2\rho_0}\delta p \frac{\partial p}{\partial n} d\Gamma = 0. \tag{10}$$

Only the condition $p=p^i$ has to be explicitly applied.

In the case of an imposed displacement field $\{u\}^i$ applied on $\delta\Omega_a$, the continuity of displacement imposes

$$\frac{1}{\omega^2\rho_0}\frac{\partial p}{\partial n} = u_n^i, \tag{11}$$

and the surface integral of Eq. (9) simplifies to

$$\int_{\delta\Omega_a} \frac{1}{\omega^2\rho_0}\delta p\frac{\partial p}{\partial n}d\Gamma = \int_{\delta\Omega_a} \delta p u_n d\Gamma. \tag{12}$$

C. Piezoelectric domain

1. Weak integral for the piezoelectric domain in three dimensions

The weak integral for a piezoelectric solid can be deduced from the Hamilton principle applied to linear piezoelectric media,[23,31,32] and is presented here in the frequency domain as

$$\underbrace{\int_{\Omega_{pi}} \{\delta\varepsilon\}^T[H]\{\varepsilon\}\, d\Omega}_{\text{work of internal strain forces}} - \underbrace{\int_{\Omega} \rho_{pi}\omega^2\{\delta u\}^T\{u\}d\Omega}_{\text{work of inertial forces}}$$
$$- \underbrace{\int_{\Omega_{pi}} [\{\delta\varepsilon\}^T[e]^t\{E\} + \{\delta E\}^T[e]\{\varepsilon\} + \{\delta E\}^T[\epsilon_d]\{E\}]\, d\Omega}_{\text{work of internal piezoeletric and dielectric forces}}$$
$$- \underbrace{\int_{\delta\Omega_{pi}} \{\delta u\}^T\{\sigma\}_n\, d\Gamma}_{\text{work of external elastic forces}} + \underbrace{\int_{S_\Phi} \delta\Phi\{D\}^T\{n\}d\Gamma}_{\text{work of external electric forces}} = 0 \quad \forall(\delta\underline{u}, \delta\Phi) \tag{13}$$

where Ω_{pi} refers to the piezoelectric domain and $\delta\Omega_{pi}$ to its boundary. S_Φ is the charge imposition surface. Φ is the electric potential and $\{E\}$ is the electric field vector given by $\{E\}=\{\nabla\Phi\}$. $\{D\}$ is the electric charge density displacement vector. $[\epsilon_d]$ is the dielectric matrix and $[e]$ is the piezoelectric matrix. $\{\sigma\}_n=\underline{\underline{\sigma}}\cdot\underline{n}$ represents the external surface forces per unit area along the normal direction. $\{\varepsilon\}$ is the strain vector. $[H]$ is the stiffness matrix. ρ_{pi} is the density of the piezoelectric media.

2. Weak integral for the piezoelectric domain in two dimensions

The PVDFs used for smart foam applications are generally thin (of the order of tens of microns). To be efficient in terms of radiation, the PVDF surface has to be quite large and curved. It is therefore not realistic to model the PVDF with solid finite elements. Such an approach would result in (i) a very large elemental aspect ratio that results in a degradation of the numerical results and (ii) a very large number of DOFs. In this paper, the thin plate hypothesis is used to model the thin piezoelectric materials present in smart foam. The curvature of the PVDF structure is approximated using flat elements.

Within the framework of Mindlin's plate theory, and assuming small thickness in the z direction, the displacement field $\{u\}$ can be written[33] as

$$\begin{Bmatrix} u(x,y,z) \\ v(x,y,z) \\ w(x,y,z) \end{Bmatrix} = \begin{Bmatrix} u(x,y) \\ v(x,y) \\ w(x,y) \end{Bmatrix} + z\begin{Bmatrix} \beta_x \\ \beta_y \\ 0 \end{Bmatrix}. \tag{14}$$

The thin plate theory utilizes five kinematic variables: the in-plane displacements u and v in the reference plane $z=0$, the transverse displacement w, and the rotations β_x and β_y.

The strain field is divided into in-plane, bending, and shear deformations. It can be written as

$$\{\varepsilon\} = \{\varepsilon_m\} + z\{\chi\}, \tag{15}$$

where

$$\langle\varepsilon\rangle = \langle\varepsilon_x, \varepsilon_y, 2\varepsilon_{xy}\rangle, \tag{16}$$

$$\langle\varepsilon_m\rangle = \langle u_{,x}, v_{,y}, u_{,y} + v_{,x}\rangle, \tag{17}$$

$$\langle\chi\rangle = \langle\beta_{x,x}, \beta_{y,y}, \beta_{x,y} + \beta_{y,x}\rangle, \tag{18}$$

where $\langle\varepsilon_m\rangle$ is the in-plane strain field of in-plane type, $z\langle\chi\rangle$ is the bending strain field, and $\langle\chi\rangle$ is the curvature vector. The shear strain is

$$\langle\gamma\rangle = \langle\beta_x + w_{,x}, \beta_y + w_{,y}\rangle. \tag{19}$$

In order to use this plate element in 3D problems, a drilling DOF of rotation is added.[33]

As for the electric conditions for thin plate, it is assumed that the electric field E is applied across the plate thickness. It will be written E_z as follows:

$$E_z = \frac{\Phi}{h}, \tag{20}$$

where h is the thickness and Φ is the electrical potential.

The electrical variable of the problem is the potential Φ. The electrical equations also use the surface electric charges $\varrho=\{D\}\{n\}$. In the plate theory, the electrical displacement becomes a scalar, noted D_z. The dielectric permittivity matrix $[\epsilon_d]$ reduces to the component of the permittivity along the z axis ϵ_{33}. The matrix of the piezoelectric coupling coefficients $[e]$ becomes a column vector

$$\{e_c\}=\begin{Bmatrix} e_{31} \\ e_{32} \\ 0 \end{Bmatrix}. \tag{21}$$

The weak integral for the piezoelectricmedia in two dimensions becomes

$$\begin{aligned} &\underbrace{-\int_S \omega^2\{\delta u\}^T[\rho I]\{u\}dS}_{\text{work of the inertial forces}} \\ &+\int_S[\ \underbrace{\{\delta\varepsilon_m\}^T[H_m]\{\varepsilon_m\}}_{\text{work of the in-plane elastic forces}}+\underbrace{\{\delta\chi\}^T[H_f]\{\chi\}}_{\text{work of the bending elastic forces}}+\underbrace{\{\delta\gamma\}^T[H_c]\{\gamma\}}_{\text{work of the shearing elastic forces}}\ ]\,dS \\ &\underbrace{-\int_S[\{\delta\varepsilon_m\}^T\{e_c\}E_z+\delta E_z\{e_c\}^T\{\varepsilon_m\}+\delta E_z\epsilon_{33}E_z]\,h\,dS}_{\text{work of the piezoelectric and dielectric forces}} \\ &-\underbrace{\int_S\{\delta u\}^T\{\sigma\}_n dS}_{\text{work of the elastic external forces}}+\underbrace{\int_{S_\Phi}\delta\Phi D_z dS}_{\text{work of the electrical external forces}}=0\ \ \forall(\delta\underline{u},\delta\Phi) \end{aligned} \tag{22}$$

where S is the piezoelectric media surface and S_Φ is the surface over which the electric charge is imposed.

The matrices $[H_m]$, $[H_f]$, and $[H_c]$ represent the in-plane stiffness matrix, the bending stiffness matrix, and the shearing stiffness matrix, respectively.

$$[H_m]=h\frac{E}{1-\nu^2}\begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix}, \tag{23}$$

$$[H_f]=\frac{h^3}{12}\frac{E}{1-\nu^2}\begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix}, \tag{24}$$

$$[H_c]=\frac{5}{6}h\frac{E}{2(1+\nu)}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{25}$$

The coefficient 5/6 in Eq. (25) is the shear correction coefficient. Finally, $[\rho I]$ is a diagonal matrix defined by

$$[\rho I]=\begin{bmatrix} \rho_{\text{pi}}h & 0 & 0 & 0 & 0 \\ 0 & \rho_{\text{pi}}h & 0 & 0 & 0 \\ 0 & 0 & \rho_{\text{pi}}h & 0 & 0 \\ 0 & 0 & 0 & \dfrac{\rho_{\text{pi}}h^3}{12} & 0 \\ 0 & 0 & 0 & 0 & \dfrac{\rho_{\text{pi}}h^3}{12} \end{bmatrix}. \tag{26}$$

D. Elastic domain

In the context of smart foams, elastic domains can be used to represent solid inclusions, impervious layers, and membranes in the poroelastic domain, or bonding (glue) layers between the foam and PVDF. The weak integrals for elastic domains in three dimensions and two dimensions are classical. They can be viewed here as particular cases of the piezoelectric domain and can be obtained by fixing the dielectric and piezoelectric constants and the electrical potential to 0.

The weak integral for a 3D elastic media is[33]

$$\int_{\Omega_e}\{\varepsilon_{(\delta u)}\}^T[H]\{\varepsilon\}d\omega-\int_{\Omega_e}\rho_e\omega^2\{\delta u\}^T\{u\}d\omega -\int_{\delta\Omega_e}\{\delta u\}^T\{\sigma\}_n d\Gamma=0,\quad\forall(\delta\underline{u}), \tag{27}$$

where Ω_e and $\delta\Omega_e$ refer to the elastic domain and its boundary. ρ_e is the density of the elastic media. $[H]$ is the stiffness

matrix and $\{\varepsilon\}$ is the strain vector. $\{\sigma\}_n = \underline{\underline{\sigma}} \cdot \underline{n}$ represents the external surface forces along the normal direction.

By analogy with the 2D piezoelectric case, Eq. (22), the weak integral for a 2D elastic domain becomes

$$\int_S [\{\delta\varepsilon_m\}^T[H_m]\{\varepsilon_m\} + \{\delta\chi\}^T[H_f]\{\chi\} + \{\delta\gamma\}^T[H_c]\{\gamma\}]dS - \int_S \omega^2\{\delta u\}^T[\rho I]\{u\}dS - \int_S \{\delta u\}^T\{\sigma\}_n dS = 0, \quad \forall(\delta\underline{u}), \tag{28}$$

where S is the elastic media surface.

The terms $[H_m]$, $[H_f]$, $[H_c]$, $\{\varepsilon_m\}$, $\{\chi\}$, $\{\gamma\}$, and $\{\sigma\}_n$ are similar to those given for the piezoelectric media.

III. COUPLING CONDITIONS

In addition to the above governing equations for the individual domains of a smart foam, various coupling conditions between these domains have to be imposed. They are of three types: (i) poroelastic-acoustic, (ii) poroelastic-structural, and (iii) structural-acoustic. Here, structural denotes both elastic and piezoelectric domains. The details of the coupling conditions are given by Debergue *et al.*[34] Here, in accordance with the presented experimental validation case and for the sake of conciseness, only poroelastic-piezoelectric (poroelastic-structural) coupling conditions are presented.

When a poroelastic material is bonded to an elastic or piezoelectric domain, the three following conditions have to be respected. The first equation ensures continuity of the total normal stress. The second expresses the fact that the relative mass flux between the solid and fluid phases across the interface is null. The third equation expresses the continuity of the solid displacement vector.

$$\{\sigma_p^t\}_n = \{\sigma_e\}_n,$$
$$\phi(U_n - u_n) = 0,$$
$$\{u_p\} = \{u_e\}, \tag{29}$$

where $\{\sigma_e\}_n$ represents the stress normal tensor of the elastic (piezoelectric) domain. Introducing these equations in the weak integral formulations of poroelastic [Eq. (1)] and piezoelectric [Eqs. (13) and (22)] or elastic [Eqs. (27) and (28)] domains, the following condition is obtained:

$$-\int_{\delta\Omega_{ps}} \phi(U_n - u_n)\delta p dS - \int_{\delta\Omega_{ps}} \{\delta u\}^T\{\sigma_p^t\}_n dS + \int_{\delta\Omega_{ps}} \{\delta u\}^T\{\sigma_e^t\}_n dS = 0, \tag{30}$$

where $\delta\Omega_{ps}$ is the contact surface between the poroelastic and structural domains. Thus, the coupling between the elastic and the poroelastic domains is natural. In a finite element implementation, no additional surface integral term has to be added to the volume integral terms on the interface. Only the kinematic condition $\{u_p\}=\{u_e\}$ has to be explicitly imposed on the contact surface. This condition will be automatically accounted for in the assembling process.

IV. FINITE ELEMENT IMPLEMENTATION

Using the presented formulations, poroelastic, elastic, fluid, and piezoelectric quadratic elements have been developed. The poroelastic element is a 20 node element (Hexa20) based on an orthotropic definition of the solid phase. The element has 4 DOFs per node: 3 translations and 1 pressure resulting in an 80 DOF poroelastic element. The fluid element is a classical Hexa20 with one pressure DOF per node. Two isotropic elastic elements were developed. The first is a solid quadratic element based on the Hexa20 shape functions. It has three DOFs per node (three translations). The second is an eight node plate element (Quad8) with six DOFs per node (three translations and three rotations). Finally, Hexa20 volume and Quad8 surface piezoelectric elements have been implemented. The implementation assumes isotropic behavior for the piezoelectric medium. The volume element has four DOFs per node (three translations and one electrical potential) while the piezoelectric surface element has seven DOFs per node (three translations, $2 \Rightarrow 3$ rotations, and one electrical potential).

For the sake of conciseness and to highlight some subtleties in the numerical implementation, the description is limited here to the global coupled system of equations. All the mathematical developments are given in Leroy's Ph.D. thesis.[35]

The global coupled system has the following form:

$$\begin{bmatrix} [\mathrm{Ac}] & [C_{fs}] & [C_{fs}] & [C_{fs}] & 0 \\ [C_{fs}]^T & [\mathrm{Po}] & 0 & 0 & [L] \\ [C_{fs}]^T & 0 & [\mathrm{Ela}] & 0 & [C] \\ [C_{fs}]^T & 0 & 0 & [\mathrm{Pi}] & [M] \\ 0 & [L & C & M] & 0 \end{bmatrix} \cdot \begin{Bmatrix} \{U_{\mathrm{Ac}}\} \\ \{U_{\mathrm{Po}}\} \\ \{U_{\mathrm{Ela}}\} \\ \{U_{\mathrm{Pi}}\} \\ \{\lambda\} \end{Bmatrix} = \begin{Bmatrix} \{F_{\mathrm{Ac}}\} \\ \{F_{\mathrm{Po}}\} \\ \{F_{\mathrm{Ela}}\} \\ \{F_{\mathrm{Pi}}\} \\ \{h\} \end{Bmatrix}, \tag{31}$$

where [Ac], [Po], [Ela], and [Pi] are the system matrix of acoustic, poroelastic, elastic, and piezoelectric domains, respectively. $[C_{\mathrm{fs}}]$ are the coupling fluid-structure matrices (acoustic-poroelastic, acoustic-piezoelectric, and acoustic-elastic). The Lagrange multiplier method is used to account for pressure discontinuity at a foam-plate (elastic and piezoelectric)-foam interface. The constraint is imposed as $[\mathrm{LCM}]\{u\}=\{h\}$ where [LCM] is the Lagrange constraint matrix and $\{h\}$ is the expression of the constraint related to the Lagrange matrix and the DOFs $\{u\}$. The Lagrange multiplier vector is noted $\{\lambda\}$. $\{U_{\mathrm{Ac}}\}$, $\{U_{\mathrm{Po}}\}$, $\{U_{\mathrm{Ela}}\}$, and $\{U_{\mathrm{Pi}}\}$ represent the vectors of the unknown DOFs of acoustic, poroelastic, elastic, and piezoelectric domains, respectively. $\{F_{\mathrm{Ac}}\}$, $\{F_{\mathrm{Po}}\}$, $\{F_{\mathrm{Ela}}\}$, and $\{F_{\mathrm{Pi}}\}$ denote the load vectors applied to acoustic, poroelastic, elastic, and piezoelectric domains, respectively.

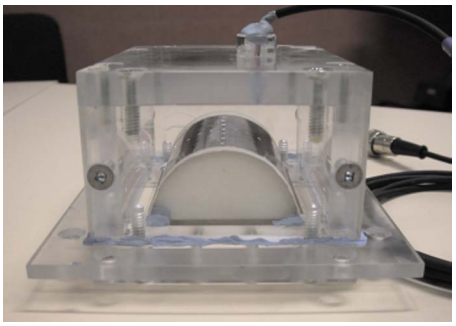


FIG. 4. (Color online) Active cell with smart foam.

The boundary conditions are of three types: (i) fixed displacement, (ii) fixed pressure, and (iii) fixed electrical potential. These variables can be fixed either to zero or to an imposed value. In the latter case the boundary condition can be interpreted as excitation. For example, imposing the electrical potential on the PVDF surface will induce a strain of the PVDF. The imposition of the displacement on a surface can also be imposed in order to create a source. The imposition of the pressure on a poroelastic surface is more delicate and requires the addition of an external load vector on the solid part of poroelastic (3). All imposed DOFs are implemented using the partitioning method.[33]

V. EXPERIMENTAL VALIDATION

A. Methodology

Again for the sake of conciseness, the numerical and experimental validations of the individual elements, using existing codes and published data, are not presented. Instead, the paper will concentrate on validating the presented model versus experimental measurement in the simple smart foam configuration presented in Fig. 2. The PVDF is bonded on the foam and mounted in a small cavity (interior dimensions: height, 55 mm; width, 85 mm; and depth, 110 mm) equipped with electric connections in order to feed the PVDF. Plexiglass flanges are used to ensure tightness between the foam and the back cavity. The smart foam and the cavity form a so-called active cell (Fig. 4). The back cavity setup has been chosen to facilitate the pressure measurement and the comparison with the numerical model.

The passive absorption and the radiation of the cell are measured in an impedance tube (Fig. 5) with a cut off frequency of 2200 Hz. The frequency range of interest is 100–1500 Hz. The tube can be used in a radiation measurement configuration by closing the end of the tube with a plexiglass plate. 100 V are then applied to the PVDF. As the plexiglass plate is transparent, it enables the measurement of the normal velocity of the foam surface using a Doppler laser vibrometer. However, since the melamine foam is highly porous, the laser beam penetrates in the foam. To prevent this phenomenon, a light reflector (a drop of liquid paper) is put on the foam surface at the measurement position. The measurement of velocity of the PVDF surface center is done the same way. The acoustic pressures at the end of the tube opposite to the smart foam and in the back cavity of the cell are also measured.

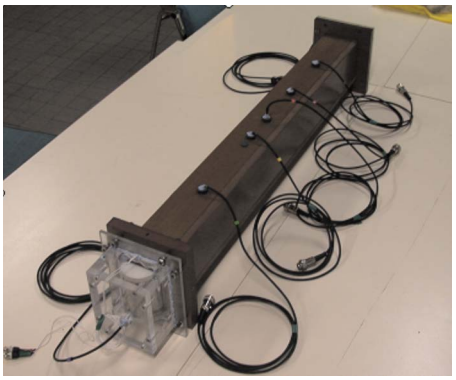


FIG. 5. (Color online) Measurement tube in the radiation configuration.

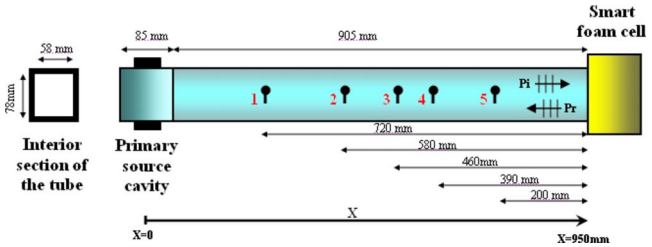


FIG. 6. (Color online) Closed waveguide for active absorption experiments, with microphone positions on the tube.

The tube can be also used in a sound absorption measurement configuration by adding a loudspeaker at the end opposite to the active cell (Fig. 6). The absorption measurement of the smart foam is done using the Chung and Blaser method[36] with several microphone pairs in the tube (Fig. 6). Four microphone pairs (1-2, 2-3, 3-4, and 4-5) are used and the absorption coefficient values are averaged depending on the validity frequency range of the pairs. This measurement gives information on the passive behavior of the cell.

Since the melamine is highly porous, its surface was conditioned with a heat-reactivatable membrane before the PVDF was bonded onto it using a double-sided tape. This method allows for a durable and controlled bonding. However, the bonding layer significantly adds mass and stiffness to the PVDF film. This has been accounted for in the finite element model as an elastic layer.

B. Numerical model

The mesh is based on a maximum frequency of 1500 Hz. At this frequency, the wavelength of the acoustic wave in the fluid is 23 cm. The size of the fluid element in the tube is equal to 3 cm, which corresponds to more than seven quadratic elements per wavelength. Closer to the foam surface (10 cm from the foam surface) and in the back cavity, the size of the fluid element is reduced to 0.7 cm in order to take account of the local effect of evanescent waves. Using the smallest Young's modulus of the three orthotropic directions of the foam, the wavelengths of the slow and fast compressional waves at 1500 Hz are found equal to 7.5 and 14.5 cm, respectively. The wavelength of the shear wave propagating in the solid phase is equal to 3 cm. Based on six elements per smallest wavelength, the average size of the poroelastic element has been chosen equal to 0.5 cm. The calculation of the

TABLE I. Structural and acoustical parameters of the melamine foam.

Parameters	Name	Value	Units
ϕ	Porosity	0,96	SU
σ	Resistivity	15 300	N s m^{-4}
α_∞	Tortuosity	1,02	SU
Λ	Viscous characteristic length	105×10^{-6}	m
Λ'	Thermic characteristic length	205×10^{-6}	m
ρ	Mass density	9	kg m^{-3}
Ex	Young's modulus in x direction	400 000	N m^{-2}
Ey	Young's modulus in y direction	180 000	N m^{-2}
Ez	Young's modulus in z direction	55 000	N m^{-2}
ν_{xy}	Poisson ratio	0,4	SU
ν_{xz}	Poisson ratio	0,4	SU
ν_{yz}	Poisson ratio	0,4	SU
η	Structural damping	0,1	SU

wavelength in the piezoelectric shell coupled with the bonding layer and the foam is more delicate since the stiffness added by the bonding layer and the foam considerably changes the behavior of the piezoelectric shell. The compressional wavelength in the piezoelectric shell alone is greater than 110 cm, and the shear wavelength is around 0.8 cm. These values would impose a very small element size (around 0.2 cm) for the piezoelectric plate elements, thus resulting in a very large number of DOF. In provision of the stiffness added by the bonding layer and the foam and for the sake of a compatible mesh between the shell and the foam, the piezoelectric shell element size has been chosen equal to 0.5 cm. A convergence analysis has been conduced to verify that the chosen mesh size was acceptable.

Using these elements, the final model is composed of 18 121 nodes, 480 porous elements, 128 piezoelectric elements, 128 bonding elements, and 3072 fluid elements with a total of 28 390 DOFs.

C. Smart foam parameters

The physical properties of the melamine foam have been measured[29] and are given in Table I. The measurement revealed that the melamine foam was anisotropic. It was, however, assumed that the foam anisotropy affects only the elastic parameters (Young's modulus and Poisson ratio). The acoustic parameters (flow resistivity, porosity, tortuosity, and viscous and thermic characteristic lengths) are assumed the same for all three directions.[28,29] Young's moduli for the three directions have been obtained by measuring the absorption coefficients of a cubic foam sample oriented along the orthotropic directions. The three Young's moduli have then been adjusted in order to fit the resonance peak of the absorption coefficient given by a commercial code with the measured resonance peak of the absorption. Special care was given to the mounting of the foam in the tube since it is known to affect the absorption resonance (a controlled air gap was left around the foam lateral surface).

The used PVDF films have 28 μm thickness and Cu–Ni electrodes. They are made by Measurement Specialties, Inc. Their properties were taken from published data.[37] They are given in Table II.

The foam-PVDF bonding layer is modeled as an isotropic plate. The properties of this layer (Young's modulus,

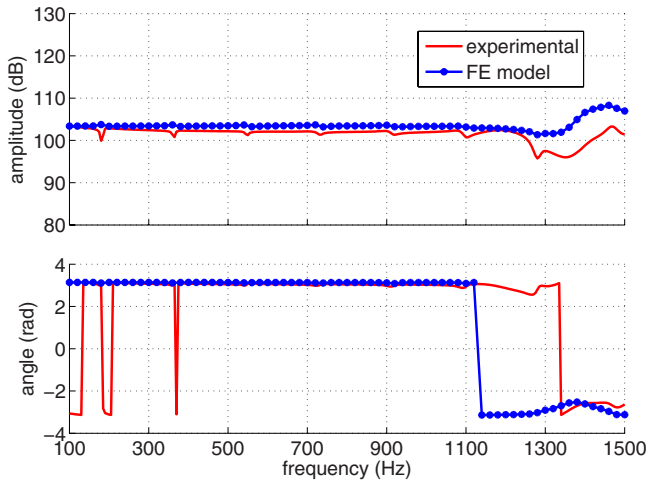


FIG. 7. (Color online) Acoustic pressure in the back cavity of the cell, for 100 V applied to the smart foam.

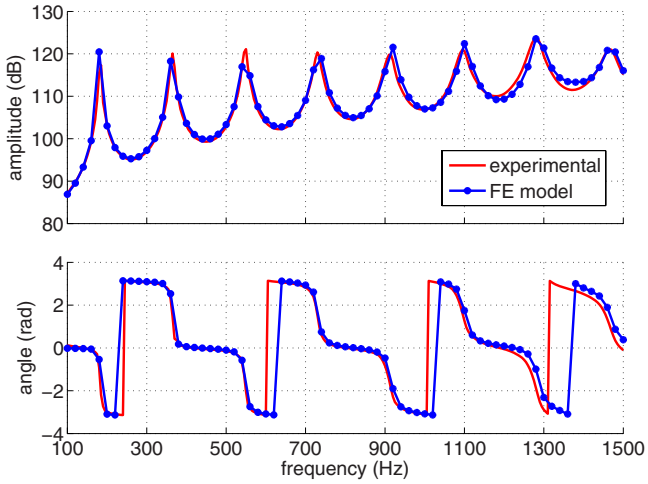


FIG. 8. (Color online) Acoustic pressure at the rigid end of the tube, for 100 V applied to the smart foam.

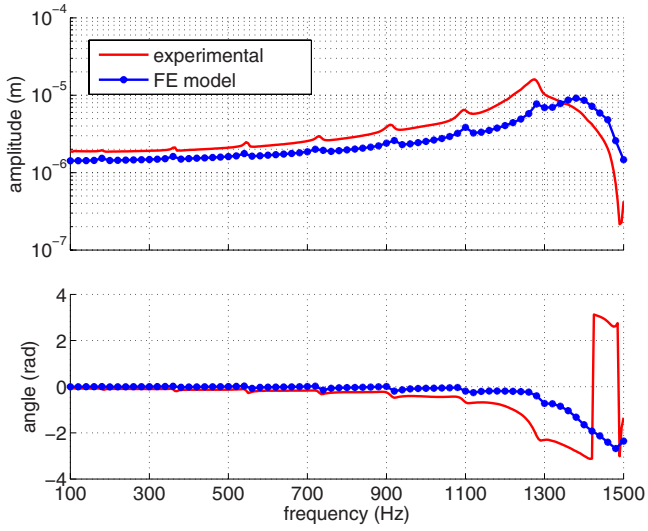


FIG. 9. (Color online) Normal displacement of foam surface center, for 100 V applied to the smart foam.

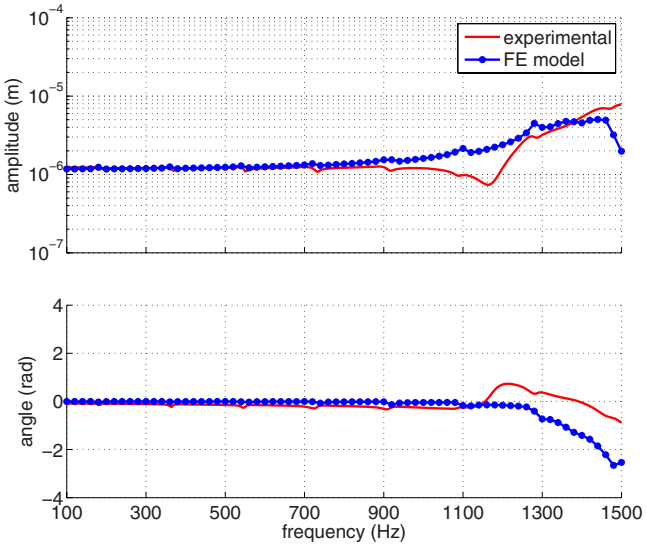


FIG. 10. (Color online) Normal displacement of PVDF surface center, for 100 V applied to the smart foam.

Poisson ratio, and structural loss factor) were identified using a parameter study comparing numerical results and experimental measurements of the radiated pressure from a smart foam placed in a small rectangular cavity (interior dimensions: height, 20 mm; width, 64 mm; and depth, 78 mm) (Table III). These parameters have been found to have a negligible influence on the passive absorption of the smart foam.

D. Results and discussion

The results of prediction versus measurements are given in Figs. 7–10 for the radiation configuration with the smart foam excited with a 100 V input. Figures 7 and 8 show the numerical and experimental results for the acoustic pressure in the back cavity and at the rigid end of the tube, respectively. Figures 9 and 10 show the numerical and experimental comparisons for the normal displacement of the center of the foam surface and the center of the PVDF surface, respectively. Figure 11 shows the experimental and calculated absorption coefficients for the passive absorption configuration.

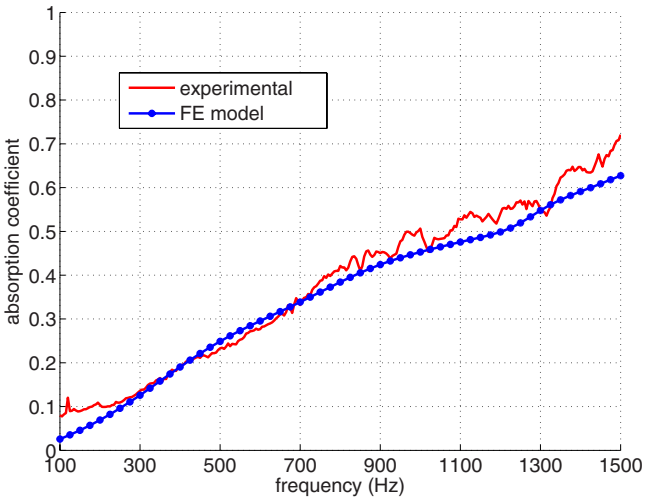


FIG. 11. (Color online) Passive absorption.

TABLE II. Elastic and electric parameters of the PVDF (Ref. 37).

Parameters	Name	Value	Units
E	Young's modulus	$5,4\times10^{9}$	N m^{-2}
ρ	Mass density	1780	kg m^{-3}
η	Structural damping	0,05	SU
ν	Poisson ratio	0,18	SU
t	Thickness	28.10^{-6}	m
e_{31}	Piezoelectric coupling coefficient in x direction	0,03 599	C m^{-2}
e_{32}	Piezoelectric coupling coefficient in y direction	0,13 087	C m^{-2}
ε_{33}	Dielectric coefficient in z direction	1×10^{-10}	F m^{-1}

Concerning the radiated pressure generated by the cell, the comparison of experimental and numerical data is excellent (Figs. 7 and 8). The radiated pressure was the result of the displacement of the entire surface; thus, the pressure averages the error that can be locally done on displacement. The pressure in the back cavity is constant up to 1250 Hz (Fig. 7) where the second resonance mode of the foam-PVDF system begins to affect the displacement field. The influence of the tube resonances is also seen in Fig. 7. At the other end of tube (Fig. 8), the radiated pressure is controlled by the modal behavior of the tube. All the peaks correspond to tube resonances. Note that the pressure level is around 100 dB, which is quite important. In the back cavity it is around 103 dB.

It is observed that the model follows very well the trends in the measured displacement data (Figs. 9 and 10). However, there are few discrepancies. For the foam displacement, the model seems to slightly overestimate the amplitude. The peak of the experimental data is well represented by the numerical model but it is located at a slightly higher frequency (Fig. 9). For the PVDF displacement, the FE model predicts very well the amplitude but does not fully capture the shape of the peak (Fig. 10). However, this may be due to the uncertainty on the measurement position.

In the passive absorption configuration, the measured absorption coefficient of the smart foam is compared to predictions in Fig. 11. Overall the comparison is very good. The main differences are observed at frequencies lower than 200 Hz and higher than 1000 Hz. At low frequency, possible leaks in the tube may have disturbed the measurements. Moreover, at this low frequency, the relative influence of the other dissipation mechanisms in the tube (speaker suspension, join, wall, etc.) is important. At frequencies higher than 1000 Hz, the uncertainties on the microphone position become important. Moreover, the modal behavior of the foam-PVDF system is significant and affects the absorption. And since the structural behavior above 1000 Hz is not perfectly predicted by the model, the predicted absorption coefficient is slightly different from measurement.

Despite the excellent performance of the presented model, it is important to note three sources of uncertainties that affect the predictions. The first is the uncertainty about the parameters of the used materials. The most important seems to be on the bonding parameters. These parameters

TABLE III. Elastic parameters of the bonding.

Parameters	Name	Value	Units
E	Young modulus	1×10^{9}	N m^{-2}
ρ	Mass density	1300	kg m^{-3}
η	Structural damping	0,1	SU
ν	Poisson ratio	0,4	SU
t	Thickness	2×10^{-4}	m

have been approximated but an accurate characterization is still needed. The PVDF properties, taken from literature, are also a source of uncertainty. Another source is related to the observed anisotropy of the foam and the assumptions made to limit this anisotropy to its mechanical coefficients. The orthotropic model of the foam has been found essential when the geometry of the structure induces an important modal coupling. The second category of uncertainty is about the fabrication of the smart foam. The bonding is very difficult to realize. As mentioned previously, the foam surface was conditioned with a coating to allow for the bonding of the PVDF using a double-sided tape. This bonding method is, however, not perfect since microbubbles form between the PVDF and foam making the bonding imperfect. The third category of uncertainty is on the boundary conditions. At the base of the PVDF, the PVDF is considered as perfectly embedded in the jaws (clamped boundary condition). Considering the very small thickness of the PVDF (around 28 μm) and the fact that the PVDF stretches, the perfect embedment is very hard to obtain. The most important boundary condition uncertainty is on the joining between the PVDF and the flanges. The PVDF is bonded to the flanges with powerful adhesive, and the tightness is assured by adding tape on the joining. It is very hard to characterize this boundary condition. In the model, the translations of the PVDF were blocked, which is only an approximation of the real boundary condition.

VI. NUMERICAL EXAMPLE

To illustrate the possible use of the model, two different smart foams are compared in an active control situation. The first smart foam corresponds to the one used in the experimental validation and will be named the "impervious smart foam." The second smart foam has the same shape but the PVDF is permeable (perforated). It will be named the "permeable smart foam." The smart foams are placed in a small cavity fixed at the end of the impedance tube previously described in the absorption configuration. The primary source is modeled with a piston with a imposed displacement of 1 μm. The optimal filter, giving a perfect absorption, is calculated using the source superposition method. The amplitude of the filter represents the ratio of the control voltage to the incident pressure amplitude, and the phase represents the angle between the control voltage and the displacement of the air particles near the foam surface (Fig. 12). The displacement of the PVDF is in phase with the applied voltage. Figure 13 shows the power dissipated by viscous effects in the foam and Fig. 14, the quadratic pressure in the tube, the foam, and in the back cavity. These two last figures are presented for a frequency of 25 Hz in order to represent the behavior of the foam in the low frequency range with a small influence of the phase change due to the propagation of the wave within the foam. The dissipated power is calculated using the formulations presented by Sgard *et al.*[38] They are based on the breakdown of the power dissipated within a foam in terms of power dissipated by structural, viscous, thermal, and radiation effects. The expressions of these contributions are well known for the (u,U) formulations.[39] For the (u,p) formulations, a simple derivation of the expression of these powers was given by Sgard *et al.*[38] Recently, a formal derivation was given by Dazel *et al.*[40]

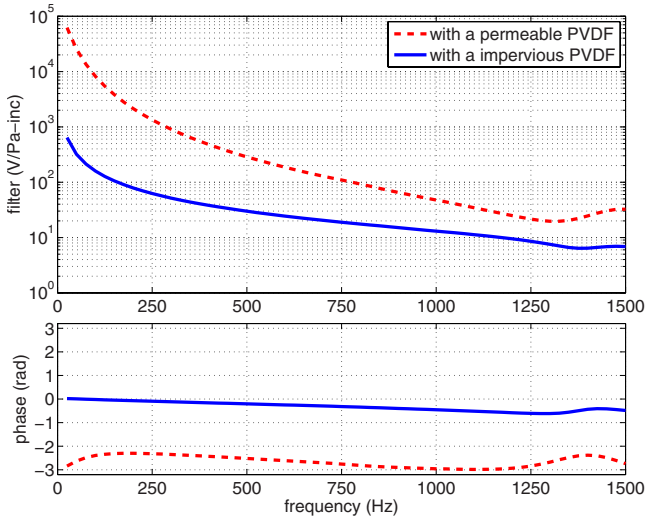


FIG. 12. (Color online) Optimal filter.

Figure 12 illustrates the action mode of the two smart foams. First, the permeable smart foam needs a much higher control voltage than the impervious smart foam to absorb the same incident pressure. At 250 Hz the control voltage of the permeable smart foam is 25 times more important. Second,

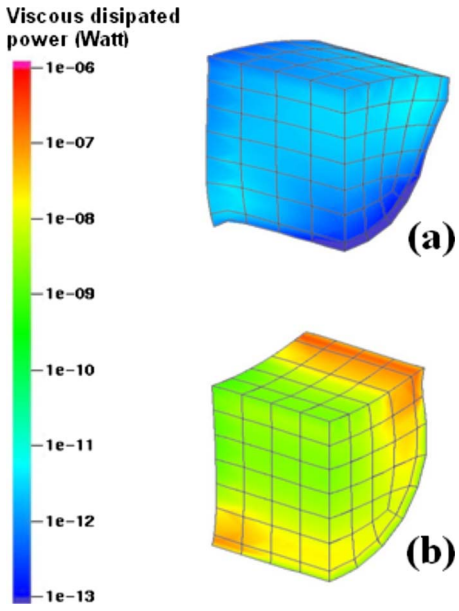


FIG. 13. (Color online) Viscous dissipated power at 25 Hz in the smart foams for an optimal control: (a) smart foam with an impervious PVDF and (b) smart foam with a permeable PVDF (1/4 of the geometry is represented).

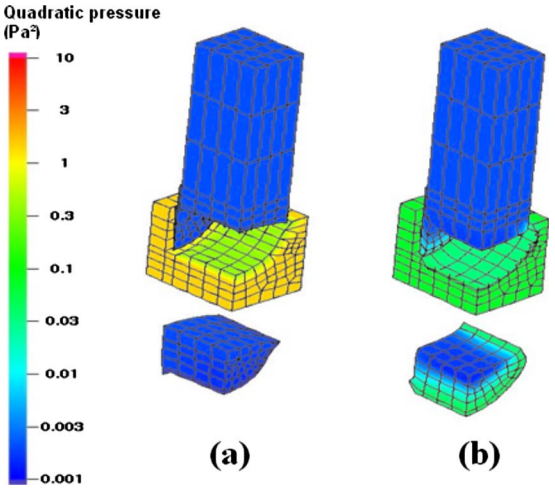


FIG. 14. (Color online) Space averaged quadratic pressure at 25 Hz in the system (tube, smart foam, and back cavity) for an optimal control: (a) smart foam with an impervious PVDF and (b) smart foam with a permeable PVDF (1/4 of the geometry is represented and the smart foam is separated from the cavity and the tube for more visibility).

the phase of the optimal filter of the impervious smart foam indicates that the PVDF displacement (voltage) is in phase with the displacement of the air particles in front of the foam. In this configuration, the smart foam adapts its surface impedance to the characteristic impedance of air. At the opposite, the phase of the optimal filter of the permeable smart foam indicates that the displacement of the PVDF and the foam is approximately out of phase with the displacement of the air particles in front of the foam surface. The smart foam maximizes the viscous dissipation in the foam. This observation is confirmed in Fig. 13 showing the power dissipated within the foam by viscous effects for the two configurations at low frequency. This figure shows that the viscous dissipated power is much more important in the case of the permeable smart foam and very small in the case of the impervious PVDF. The remaining question is the whereabouts of the incident acoustic energy in the case of the impervious smart foam, since it is not dissipated within the foam.

The response is partially given in Fig. 14 showing the space averaged quadratic pressure in the tube, the foam, and the back cavity. It appears that the quadratic pressure in the back cavity is much more important in the case of the impervious smart foam than in the case of the permeable smart foam. In the impervious smart foam case, the pressure absorbed in front of the smart foam is transmitted in the back cavity, while it is dissipated for a large part by viscous effect in the permeable smart foam case.

VII. CONCLUSION

A full 3D smart foam model was developed and experimentally validated. Despite few uncertainties on the fabrication of the smart foam and the physical properties of its materials, the model gives results in very good agreement with the measurement of absorption and radiated pressures, and approaches very well the vibratory behavior of the smart foam. The comparison is especially perfect in the low frequency range where the smart foam is needed for active control. The proposed model appears as an efficient tool to understand the control mechanisms of smart foams and to design optimized hybrid control systems. This model is general and can be used for any coupled structure combining fluid, elastic, poroelastic, and piezoelectric media. It can be used, for instance, to develop any type of hybrid passive-active material such as absorption control, structure radiation control, and transmission control devices.

ACKNOWLEDGMENTS

The research is partially funded by AUTO21 (Canada) and by Project No. ANR-06-BLAN-0081 (ANR, France). The authors wish to thank the members of the "Centre de Calcul Scientifique de l'Université de Sherbrooke" and especially HuiZhong Lu for his help with the numerical implementation of the code.

[1]H. F. Olson and E. G. May, "Electronic sound absorber," J. Acoust. Soc. Am. **25**, 1130–1136 (1953).
[2]D. Guicking and E. Lorentz, "An active absorber with porous plate," ASME J. Vib., Acoust., Stress, Reliab. Des. **106**, 389–392 (1984).
[3]D. Guicking, K. Karcher, and M. Rollwage, "Coherent active methods for application in rooms acoustics," J. Acoust. Soc. Am. **78**, 1426–1434 (1985).
[4]M. Furstoss, D. Thenail, and M. A. Galland, "Surface impedance control for sound absorption: Direct and hybrid passive/active strategies," J. Sound Vib. **203**, 219–236 (1997).
[5]K. Henrioulle, D. Desplenere, P. Hemsoote, and P. Sas, "Design of a quarter wavelength resonance absorber using a flat loudspeaker," ACTIVE 99 (1999).
[6]O. Hilbrunner, M. A. Galland, and N. Sellen, "Optimisation of a hybrid acoustic liner for noise reduction of engine aircraft nacelles," ACTIVE 2002, ISRV, UK (2002).
[7]M. A. Galland, N. Sellen, and B. Mazeaud, "Hybrid passive/active absorbers for flow ducts," Appl. Acoust. **66**, 691–708 (2005).
[8]S. Beyene and A. Burdisso, "A new passive/active noise absorption system," J. Acoust. Soc. Am. **101**, 1512–1515 (1997).
[9]J. P. Smith, B. D. Johnson, and R. A. Burdisso, "A broadband passive-active sound absorption system," J. Acoust. Soc. Am. **106**, 2646–2652 (1999).
[10]P. Cobo, A. Fernandez, and O. Doutres, "Low-frequency absorption using a two-layer system with active control of input impedance," J. Acoust. Soc. Am. **114**, 3211–3216 (2003).
[11]P. Cobo, J. Pfretzschner, M. Cuesta, and D. K. Anthony, "Hybrid passive-active absorption using microperforated panels," J. Acoust. Soc. Am. **116**, 2118–2125 (2004).
[12]J. Yuan, "Causal impedance matching for broadband hybrid noise absorption," J. Acoust. Soc. Am. **113**, 3226–3234 (2003).
[13]C. R. Fuller, M. J. Bronzel, C. A. Gentry, and D. E. Whittington, "Control of sound radiation/reflection with adaptive foams," National Conference on Noise Control Engineering (1994), pp. 429–436.
[14]C. R. Fuller, C. A. Gentry, and C. Guigou, "Foam-PVDF Smart Skin for Active Control of Sound," Proc. SPIE **2721**, pp. 26–37 (1996).
[15]C. Guigou and C. R. Fuller, "Foam-PVDF Smart Skin for Aircraft Interior Sound Control," Proc. SPIE **3044**, 68–78 (1997).
[16]C. Guigou and C. R. Fuller, "Control of aircraft interior broadband noise with FOAM-PVDF smart skin," J. Sound Vib. **220**, 541–557 (1999).
[17]C. Guigou and C. R. Fuller, "Adaptive feedforward and feedback methods for active/passive sound radiation control using smart foam," J. Acoust. Soc. Am. **104**, 226–234 (1998).
[18]Y. J. Kang, "Smart foam for active noise control in a duct and device equipped with the same," U.S. Patent No. 6,560,342 (May 6, 2003).
[19]G. P. Mathur, C. L. Chin, M. A. Simpson, and J. T. Lee, "Structural Acoustic Prediction and Interior Noise Control Technology," Report No. NASA/CR-2001-211247, Hampton, VA (2001).
[20]W. N. Akl, A. M. Baz, Y. Whang, J. Zhu, and M. Zhang, "Smart foam for active vibration noise control," NOISE-CON 2004 (2004), pp. 164–173.
[21]N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–

1452 (1998).
[22]N. Atalla, M. A. Hamdi, and R. Panneton, "Enhanced weak integral formulation for the mixed (u,p) poroelastic equations," J. Acoust. Soc. Am. **109**, 3065–3068 (2001).
[23]H. Tiersten, *Hamiltons Principle for Linear Piezoelectric Media* (IEEE, New York, 1967), pp. 1523–1524.
[24]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).
[25]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).
[26]J. F. Allard, P. Herzog, D. Lafarge, and M. Tamura, "Recent topics concerning the acoustics of fibrous and porous materials," Appl. Acoust. **39**, 3–21 (1993).
[27]J. F. Allard, *Propagation of Sound in Porous Media, Modelling Sound Absorbing Materials* (Elsevier, London, 1993).
[28]N. Geebelen, L. Boeckx, G. Vermeir, W. Lauriks, J. F. Allard, and O. Dazel, "Measurement of the rigidity coefficients of a melamine foam," Acta Acust. **93**, 738–788 (2007).
[29]M. Melon, M. Mariez, C. Ayrault, and S. Sahraoui, "Acoustical and mechanical characterization of anisotropic open-cell foams," J. Acoust. Soc. Am. **104**, 2622–2627 (1998).
[30]M. Petyt, J. Lea, and G. H. Koopman, "A finite element method for determining the acoustic modes of irregular shaped cavities," J. Sound Vib. **45**, 495–502 (1976).
[31]D. Marinkovic, H. Koppe, and U. Gabbert, "Finite element development for generally shaped piezoelectric active laminates part I—Linear approach," Facta Universitatis; Series: Mechanical Engineering **2**, 11–24 (2004).
[32]V. Piefort, "Finite element modelling of piezoelectric active structure," Ph.D. thesis, ULB, Liege, Belgium (2001).
[33]J. L. Batoz and G. Dhatt, *Modélisation des Structures par Élémens Finis (Structure Modeling Using the Finite Element Method)* (Les Presses de l'Universite Laval, Quebec, Canada, 1990).
[34]P. Debergue, R. Panneton, and N. Atalla, "Boundary conditions for the weak formulation of the mixed (u,p) poroelasticity problem," J. Acoust. Soc. Am. **106**, 2383–2390 (1999).
[35]P. Leroy, "Les mousses adaptatives pour l'amélioration de l'absorption acoustique: Modélisation, mise en œuvre, mécanismes de contrôle ("Smart foams" for enhancing acoustic absorption: Modeling, experimental implementation, action mode)," Ph.D. thesis, Universite de Sherbrooke, Sherbrooke, Canada (2008).
[36]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).
[37]K. C. Bailo, D. E. Brei, and K. Grosh, "Investigation of curved polymeric piezoelectric active diaphragms," J. Vibr. Acoust. **125**, 145–154 (2003).
[38]F. Sgard, N. Atalla, and J. Nicolas, "A numerical model for the low-frequency diffuse field sound transmission loss of double-wall sound barriers with elastic porous lining," J. Acoust. Soc. Am. **108**, 2865–2872 (2000).
[39]N. Dauchez, S. Sahraoui, and N. Atalla, "Dissipation mechanisms in a porous layer bonded onto a plate," J. Sound Vib. **265**, 437–449 (2003).
[40]O. Dazel, F. Sgard, F.-X. Beckot, and N. Atalla, "Expressions of dissipated powers and stored energies in poroelastic media modeled by u,U and u,P formulations," J. Acoust. Soc. Am. **123**, 2054–2063 (2008).

Short-range shock formation and coalescence in numerical simulation of broadband noise propagation

Micah R. Shepherd[a)]
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Kent L. Gee
Department of Physics and Astronomy, Brigham Young University, N-243 ESC, Provo, Utah 84602

Mark S. Wochner
Applied Research Laboratories, The University of Texas at Austin, P.O. Box 8029, Austin, Texas 78713

(Received 19 January 2009; revised 9 September 2009; accepted 10 September 2009)

The number of jet and rocket noise studies has increased in recent years as researchers have sought to better understand aeroacoustic source and radiation characteristics. Although jet and rocket noise is finite-amplitude in nature, little is known about the existence of shock formation and coalescence close to the source. A numerical experiment is performed to propagate finite-amplitude noise and determine the extent of the nonlinearity over short distances with spherical spreading. The noise is filtered to have a haystack shape in the frequency domain, as is typical of such sources. The effect of the nonlinearity is compared in both the temporal and frequency domains as a function of distance. Additionally, the number of zero-crossings and overall sound pressure level is compared at several distances. The results indicate that the center frequency plays a particularly important role in the amount of coalescence and spectral redistribution that occurs. The general applicability of these results to actual near-field finite-amplitude jet and rocket noise experiments is also presented. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3243466]

PACS number(s): 43.25.Cb [RR] Pages: 2886–2893

I. INTRODUCTION

The propagation of intense random acoustic noise has important applications in the present day. Military jet aircraft and launch vehicle technologies are rapidly advancing to meet new challenges, often causing such aircraft to produce higher thrust and consequently high noise levels which propagate nonlinearly. Characterization of these high-amplitude aeroacoustic sources is a necessary part of predicting structural vibration of the vehicle as well as the impact of noise on the surrounding community and environment.

Early experimental studies of noise produced by jets and rockets[1–6] began in the 1950s. The effort to characterize and ultimately predict far-field noise propagation has continued to the present with a wave of recent research.[7–13] These studies, along with other analytical studies (see, e.g., Tam *et al.*[14]), have established the spectral characteristics common to jet and rocket noise, namely, an increase according to f^2 up to some center frequency followed by a decrease according to f^{-2}. This spectral shape is commonly referred to as a haystack spectrum and is a result of the large scale turbulent structure of the jet.[15] Note that waveforms with haystack spectra and waveforms with flat, white-noise spectra with the same rms levels will have appreciably different peak pressures. The haystack-based waveform has greater peak pressures because the relative levels around the center frequency must be greater to produce the same rms level.

Nonlinear effects in the propagation of jet noise propagation were first studied by Morfey and Howell[16] and more recently by Gee *et al.*[7] McInerny and Ölçmen[17] also showed evidence of nonlinear propagation in rocket noise. Published overall sound pressure levels (OASPLs) of military jet aircraft noise (F-22A Raptor) are 143 dB at 23 m,[7] while published Titan IV rocket noise levels are 140 dB at 820 m.[17] By removing spreading back to 1 m, these values are approximately 170 dB for the jet aircraft and 198 dB for the rocket. All levels here and in the rest of the article are referenced to 20 μPa. Although spherical spreading will not hold at 1 m due to the spatially extended nature of the turbulence, these numbers are nevertheless illustrative of the extremely high amplitudes near the jet or rocket noise source. Acoustical measurements of static, horizontally fired four-segment reusable solid rocket motors (like those used on Space Shuttles) reveal levels that are likely slightly lower than the Titan IV.[10] However, note that the proposed ARES I and ARES V vehicles will use 5-segment and 5.5-segment versions of the current four-segment motors, which may result in a slight increase in level.

The first laboratory study of intense broadband noise propagation was performed by Pestorius and Blackstock.[18] Band-limited broadband noise was propagated down a 29.3 m (96 ft) tube and compared to numerically propagated noise. The algorithm and experiment showed good agreement of shock formation and propagation at sound pressure levels up to 160 dB. Other laboratory experiments of finite-amplitude noise propagation are summarized by Gurbatov and Rudenko.[19]

[a)]Author to whom correspondence should be addressed. Electronic mail: mrs30@psu.edu

Many numerical studies of nonlinear propagation have also been performed,[20–22] though only a few have studied the nonlinear propagation of noise. As mentioned, Pestoruis and Blackstock developed a numerical solution of the generalized Mendouse–Burgers equation and studied one-dimensional propagation of band-limited finite-amplitude noise. However, this study did not include geometrical spreading or discuss the effects of the initial spectral shape of the waveform. To more accurately model the actual propagation of jet noise through the atmosphere, the Pestorius and Blackstock algorithm was further developed to include spherical spreading by Blackstock.[23] Most recently, Gee *et al.*[7] made further developments to the algorithm and performed in-depth studies of far-field jet noise propagation according to engine status and angle. The input waveforms used in the latter study were measured data recorded 23 m (75.4 ft) from an F-22A Raptor.

When including spherical spreading in a calculation, there will be substantially less waveform steepening and shock formation compared to the one-dimensional propagation scenario studied by Pestorius and Blackstock.[18] Because of this, they postulated that significant low frequency buildup due to the loss of zero-crossings resulting from shock coalescence will not occur in practical applications. When spherical spreading is included in the finite-amplitude propagation of noise, they argued that spreading losses will prevent low frequency buildup from being significant.

The shock formation distance using weak shock theory[24] for monofrequency waves seems to concur with the postulation from Pestorius and Blackstock since geometrical spreading requires much longer distances for significant nonlinear distortion and shock formation to occur than is necessary for an equivalent non-spreading case. This idea is illustrated well when the shock formation distance ($\bar{r}$) for monofrequency spherical waves is expressed in terms of shock formation distance ($\bar{x}$) for non-spreading monofrequency waves.

$$\bar{r} = r_0 e^{(1/\beta\epsilon k r_0)} = r_0 e^{(\bar{x}/r_0)}. \tag{1}$$

Here, r_0 is some distance where the waveform is known, β is the coefficient of nonlinearity, ϵ is the acoustic Mach number, and k is the acoustic wavenumber. The definition of the narrowband noise shock formation distance is similar to Eq. (1), except that the rms pressure is used as opposed to the maximum pressure.[19]

Care must be taken when using the shock formation distance of narrowband noise to gain insight into the behavior of nonlinear propagation of broadband noise since the sound pressure level is based on the rms pressure and does not tell the entire story. For a Gaussian process, the instantaneous pressure will be greater than the rms pressure 31.7% of the time. Furthermore, the actual pressure will be 6 dB higher than rms (i.e., greater than 2 std from mean) 4.6% of the time and 9.5 dB higher than rms 0.3% of the time. Although this may seem inconsequential, for a 1 s recording of Gaussian noise sampled at 20 kHz there will likely be over 800 samples 6 dB higher than the rms pressure and 20 samples 9.5 dB higher than rms pressure. Therefore a broadband noise signal of finite amplitude will likely have waveform steepening and shock formation faster than estimates based on narrowband noise shock formation distances. The more important indicators of shock formation are the peak pressures (i.e., the extreme-valued pressures).

Supersonic jet and rocket noise, however, is not Gaussian distributed but has been shown to have an asymmetric distribution. McInerny[25] found rocket noise to exhibit nonzero skewness, while Petitjean *et al.*[26] and Gee *et al.*[27] found similar trends from model-scale and full-scale jet noise experiments, respectively. This indicates that there will likely be more extreme-valued pressures in jet and rocket noise than in a Gaussian process.

Once the extreme-valued pressures have propagated a distance sufficient to form shocks, the decay of the shock will no longer follow geometrical spreading, and the speed of the shock U_{sh} is dictated by weak shock theory[24] to be

$$U_{\mathrm{sh}} = c_0 + \beta \frac{p_a + p_b}{2\rho_0 c_0}, \tag{2}$$

where c_0 is the small signal sound speed, p_a is the pressure in front of the shock, p_b is the pressure behind the shock, and ρ_0 is the ambient fluid density. The shocks from these extreme-valued pressures may travel faster than other parts of the waveform and eventually overtake them, causing coalescence. Therefore, the extreme-valued pressures, which can easily be 8–10 dB greater than the rms pressure, will play an important role in the amount of shock coalescence.

Only a handful of studies have addressed shock coalescence either analytically or numerically. The Pestorius and Blackstock[18] algorithm predicted a downward shift in center frequency due to the loss of zero-crossing resulting from shock coalescence. Lighthill[28] performed an analytical study for conical shock waves, describing coalescence in terms of shock "bunchings," "unions," and "proness to form further unions." Khokhlova *et al.*[29] proposed a statistical characterization of shock coalescence using an analogy to kinetic theory of inelastic particles. However, it has not been experimentally determined if significant shock coalescence could occur near a military jet aircraft or launch vehicle.

Since our understanding of jet and rocket noise relies on measurements near the source, the question must be raised: Are nonlinear effects important in the short-range propagation of jet and rocket noise and if so, how will this influence typical characterization methods? Furthermore, how significant of a role does the center frequency play in the propagation? In this article, a numerical experiment is presented in which haystack noise waveforms at amplitudes similar to those expected near military aircraft and launch vehicles are propagated over short distances. This will shed insight into whether significant shock formation and coalescence can realistically occur close to a finite-amplitude jet or rocket noise source despite the spherical spreading. Once this is established, a simple linear reconstruction is performed to reveal the spectral changes in the waveform as a function of short-range propagation distance and center frequency.

II. PROPAGATION METHODS

The propagation algorithm used for this study was recently developed by Wochner[30] and builds off prior work by Sparrow and Raspet.[20] To outline the physics modeled in the propagation algorithm, the equation set and numerical methods used in this research will be briefly reviewed. A detailed derivation is found in Ref. 31.

The equation set is an extended Navier–Stokes equation set comprised of conservation equations. They are defined for spherical waves in one dimension as

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u}{\partial r} + 2\frac{\rho u}{r} = 0, \tag{3}$$

$$\frac{\partial \rho u}{\partial t} + \frac{\partial \rho u^2}{\partial r} = -\frac{\partial p}{\partial r} + \mu_B \frac{\partial^2 u}{\partial r^2} + \mu \frac{\partial \phi_{rr}}{\partial r} - \frac{2\rho u^2}{r}, \tag{4}$$

$$\frac{\partial \rho s_{\mathrm{fr}}}{\partial t} + \frac{\partial \rho u s_{\mathrm{fr}}}{\partial r} = \sigma_s - \sum_{\nu} \frac{\rho}{T_\nu} c_{u\nu} \frac{DT_\nu}{Dt} + \frac{\partial}{\partial r}\left(\frac{\kappa}{T}\frac{\partial T}{\partial r}\right) - \frac{2\rho u s}{r}, \tag{5}$$

and

$$\frac{\partial \rho T_\nu}{\partial t} + \frac{\partial \rho u T_\nu}{\partial r} = \frac{\rho}{\tau_\nu}(T - T_\nu) - \frac{2\rho u T_\nu}{r}. \tag{6}$$

In the above equations, ρ is the total fluid density, u is the particle velocity, μ and μ_B are the shear and bulk viscosities, ϕ_{rr} is the rate of shear tensor, s_{fr} is the total frozen entropy, T is the total absolute temperature, κ is the coefficient of thermal conduction, $c_{v\nu}$ is the specific heat constant of the ν-type molecule, σ_s is the entropy source term, and T_ν and τ_ν are the apparent vibration temperature and relaxation time of the ν-type molecule.

To solve the equation set, the temporal and spatial variables are grouped together to form a matrix equation of this form

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial \mathbf{F}}{\partial r} = \mathbf{H}, \tag{7}$$

where $\mathbf{w}$ is a matrix of the time-dependent variables, $\mathbf{F}$ is a matrix of the r-dependent variables, and $\mathbf{H}$ is a matrix of the remaining source-like terms. The acoustic pressure p is then found using the van der Waals form of the equation of state,

$$p = c^2\left[(\rho - \rho_0) + \frac{\gamma - 1}{2\rho_0}(\rho - \rho_0)^2 + \frac{\rho\beta T}{c_p}(s_{\mathrm{fr}} - s_{\mathrm{fr}_0})\right], \tag{8}$$

which relates the acoustic pressure, density, and entropy of a perturbation, where the subscript "0" represents the ambient values. As noted by Wochner,[30] this Eqs. (3)–(6) are an extension of the Yano and Inoue[21] equation set in that thermoviscous and molecular relaxation losses are added to the Euler equations. Important to note is that no weak nonlinearity or far-field assumptions are made in the derivation of this equation set.

Equation (7) is solved using a weighted essentially nonoscillatory (WENO) scheme[32] in space and third-order Runge–Kutta scheme in time. The WENO scheme is able to stably propagate discontinuities and therefore will not go unstable when shocks with near infinite slopes form. More details regarding the implementation of the WENO scheme can be found in Ref. 30. Previous uses of the WENO scheme in acoustics include computational aeroacoustics and Mach stem formation.

An input waveform was read into the algorithm with 200 points per wavelength at 20 kHz. Since the number of points per wavelength directly relates to the computation time, these parameters were chosen to allow for sufficient frequency resolution (related to points per wavelength) and propagation distance (related to the number of time steps by the Courant number) to allow for a reasonable amount of computation time. The WENO scheme's ability to propagate discontinuities also inherently includes an initial smoothing effect of the input waveform. The numerical smoothing for the input waveforms for the parameters previously stated alters the decay of the input spectrum at frequencies above approximately 2 kHz. Since the f^{-2} decay is necessary to have an initial haystack spectral shaped spectrum, the frequencies above 2 kHz are not shown. It must be noted that the WENO method does not have trouble creating frequencies above 2 kHz that occur with nonlinear distortion, but that the initial smoothing is the WENO scheme's attempt to ensure adequate smoothness of the broadband input waveform.

Three input waveforms were created using an array of 2^{16} Gaussian-distributed random numbers and filtered to have a haystack spectrum with center frequencies of 100, 300, and 500 Hz. The 100 Hz center frequency was chosen to be in the upper range of launch vehicles and the lower range of military jet aircraft while the 300 and 500 Hz center frequencies were chosen in the middle-to-high range of military jet aircraft. Center frequencies below 100 Hz become increasingly difficult to resolve temporally in terms of spectral density estimates for the computational and time resources available. The time series was scaled to have OASPL of approximately 165 dB, which is equivalent to a rms pressure just above 3.5 kPa.

The crest factor has been used to characterize the extreme-valued samples in rocket noise data to provide an indication of potential nonlinearity.[25] The crest factor, defined as

$$\mathrm{Cr} = 20\log\left(\frac{p_{\max}}{p_{\mathrm{rms}}}\right), \tag{9}$$

is shown for the input waveforms in Table I, revealing that the peak pressures are above 11.5 kPa (175 dB) for all three cases. As previously stated, the extreme pressures are more important than rms values in estimating nonlinearity. The crest factor for the waveforms used in the study approach the

TABLE I. Crest factor of input waveforms.

Waveform	Center frequency (Hz)	Crest factor (dB)
1	100	9.84
2	300	11.42
3	500	11.80

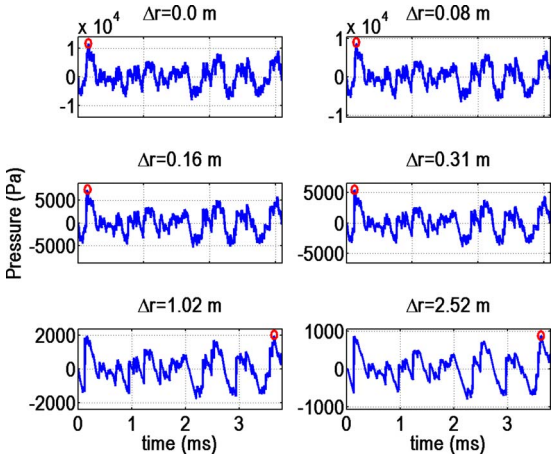


FIG. 1. (Color online) The first 30 000 points in the $f_c=100$ Hz waveform at six propagation distances. Nonlinear propagation has caused a shock to form after 0.31 m and shock coalescence to occur after 2.52 m. The maximum value (circled) changes peaks after 1.02 m.

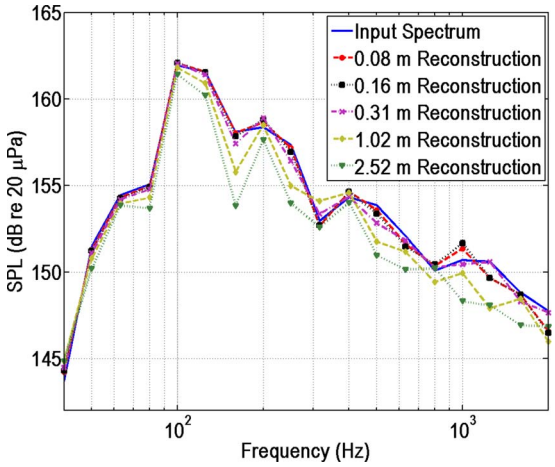


FIG. 2. (Color online) The reconstructed one-third octave spectra of the $f_c=100$ Hz waveform initially at 165 dB OASPL after propagating 0.08, 0.16, 0.31, 1.02, and 2.52 m along with the input spectrum at the source. Since the spectral shape is generally maintained despite the waveform steepening, the reconstructed spectra match the input spectra well.

valid limit for weak shock theory[33] but are handled correctly since the propagation algorithm is not based on weak shock theory.

The propagation was started 0.32 m from the spreading origin with absorbing boundary conditions to prevent numerical reflections. The waveforms were propagated 2.52 m, and the pressure waveform was sampled every 0.08 m. Because the propagation was one-dimensional, the linear reconstruction was performed by removing the $1/r$ magnitude decay due to the spherical spreading. This simple reconstruction scheme serves as a method to study the physics of the propagation, while estimating its effects on spectral characteristics over distance and the three center frequencies. For the remainder of this paper, the propagation distance (Δr) will be referenced from the propagation starting point, 0.32 m from the origin.

III. RESULTS

Figure 1 shows the first 30 000 points of the $f_c=100$ Hz waveform after propagating 0.08, 0.16, 0.31, 1.02, and 2.52 m. The maximum value in the waveform of 11.6 kPa is contained in this sample. Also displayed in Fig. 1 is the input time waveform to show the waveform's initial shape. Nonlinear propagation effects are clearly visible as portions of the waveform have become steepened, and the first shock forms from the maximum value after propagating 0.31 m. After 2.52 m, there is one main shock followed by a random set of steepened/shocked noise-fronts of smaller amplitude. Shock coalescence has clearly occurred after propagating 2.52 m as many of the higher frequency undulations/shocks have disappeared during the propagation. The energy from these frequencies has collected at the larger shock locations as a result of the nonlinear phase coupling. Also of note is that the peak value (circled in the figure) shifts from being at the beginning of the snapshot to a peak toward the end of the snapshot after 1.02 m.

The autospectrum of the $f_c=100$ Hz waveform was calculated, and the magnitude reconstruction back to the source location was applied. The original and reconstructed autospectra are compared on a one-third octave scale in Fig. 2. The reconstruction from 0.08, 0.15, and 0.31 m matches the source spectra very well, while the reconstruction from 1.02 and 2.52 m deviates slightly from the original spectrum. The deviations come from changes in the time waveform due to waveform steepening, but do not significantly affect the slope of the decay. This is because the energy transfer due to waveform steepening causes a high frequency roll-off proportional to f^{-2}, which the initial waveform already had. Additionally, the amplitude at the center frequency did not experience a shift in amplitude or frequency.

Figure 3 shows 20 000 points of the $f_c=300$ Hz waveform after propagating 0.08, 0.16, 0.31, 1.02, and 2.52 m along with the input time waveform. This set contains the waveform's maximum value of 11.7 kPa. The first shock forms after propagating 0.16 m. Shock coalescence has clearly occurred after 1.02 m of propagation, and the shocks appear to be interacting on a slightly larger scale than was seen in Fig. 1 so that when the waveform has reached

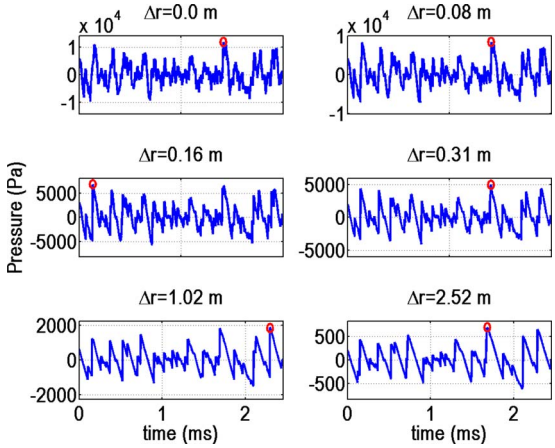


FIG. 3. (Color online) The first 20 000 points in the $f_c=300$ Hz waveform at six propagation distances. Nonlinear propagation has caused a shock to form after 0.16 m and shock coalescence to occur after 1.02 m. The maximum value (circled) changes peaks several times during the propagation.

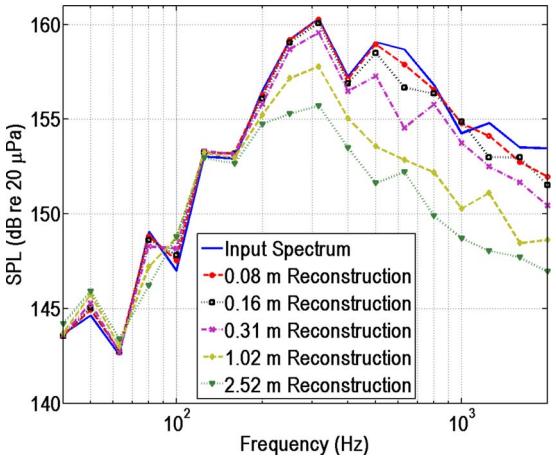


FIG. 4. (Color online) The reconstructed one-third octave spectra of f_c =300 Hz waveform initially at 165 dB OASPL after propagating 0.08, 0.16, 0.31, 1.02, and 2.52 m along with the input spectrum at the source. The amplitude of the center frequency drops for 1.02 and 2.52 m propagations due to energy transfer to high frequencies.

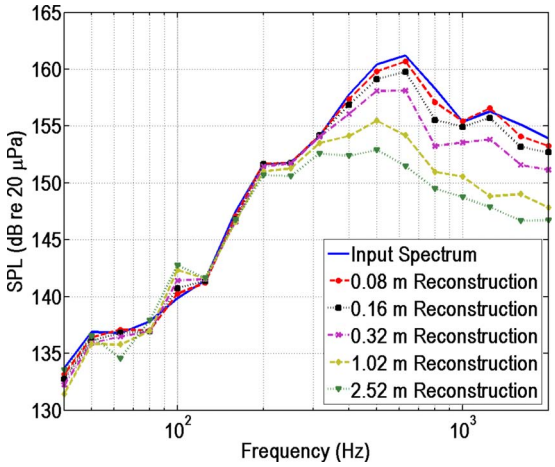


FIG. 6. (Color online) The reconstructed one-third octave spectra of f_c =500 Hz waveform initially at 165 dB OASPL after propagating 0.08, 0.16, 0.31, 1.02, and 2.52 m along with the input spectrum at the source. The amplitude of the center frequency drops in amplitude and frequency for 1.02 and 2.52 m propagations due to energy transfer high frequencies and loss in zero-crossings.

2.52 m, it consists of random arrays of shocks varying in amplitude. The waveform maximum value shifts peaks several times during the propagation.

The autospectrum of the f_c=300 Hz waveform was calculated, and the magnitude reconstruction back to the source location was applied, with the input and reconstructed autospectra compared on a one-third octave scale in Fig. 4. The reconstruction from 1.02 and 2.52 m deviates significantly from the input at and above the center frequency (300 Hz). The energy transfer from the center frequency to high frequencies was significant enough to cause the center frequency to drop almost 5 dB after 2.52 m. However, the center frequency does not shift downward.

The f_c=500 Hz waveform was also propagated out to the distance of 2.52 m at an initial OASPL of 165 dB, with the first 10 000 points shown in Fig. 5. A shock forms at the maximum value, initially at 12.3 kPa, after 0.08 m. Visual comparison reveals that the nonlinear distortion is significantly greater in Fig. 5 than in Fig. 1. This is to be expected

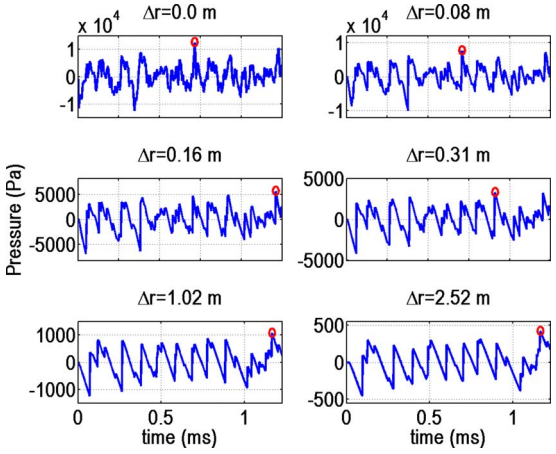


FIG. 5. (Color online) The first 10 000 points of the f_c=500 Hz waveform shown at six propagation distances. Nonlinear effects have caused shocks to form after 0.08 m and shock coalescence to occur by 1.02 m. The maximum value (circled) changes peaks several times during the propagation.

given the general trend that shock formation distance decreases as frequency increases. Shock coalescence is again evident after 1.02 m. After 2.52 m, the waveform is composed of nearly equispaced shocks of comparable amplitude and begins to partially resemble a sawtooth wave. The waveform maximum value has again shifted peaks several times during the propagation.

Figure 6 shows the reconstructed one-third octave spectra for the f_c=500 Hz waveform from the same distances shown previously. The center frequency amplitude begins to drop in amplitude after 0.08 m and begins to shift downward after 1.02 and 2.52 m. The downward shift in frequency is typically attributed to a decrease in zero-crossings as a result of the shock coalescence. The crest factor again shifts position during the propagation.

The overall sound pressure level is compared as a function of distance for all waveforms in Table II. After 2.52 m of linear propagation with spherical spreading, the OASPL for would be expected to drop 19 dB to the value of 146 dB. The f_c=100 Hz waveform has decayed 3 dB more than expected with spherical spreading alone, while the f_c=300 Hz

TABLE II. OASPL of the waveforms after propagating several distances compared to the levels resulting from just spherical spreading at those distances (rounded to the nearest decibel). Note that Δr is referenced to the propagation starting point and Δr+0.32 represents the distance from the origin.

Δr (m)	Spreading alone	100 Hz	300 Hz	500 Hz
0.00	165	165	165	165
0.08	163	163	163	163
0.16	162	162	161	160
0.31	159	159	158	157
1.02	153	152	150	148
2.52	146	145	142	139

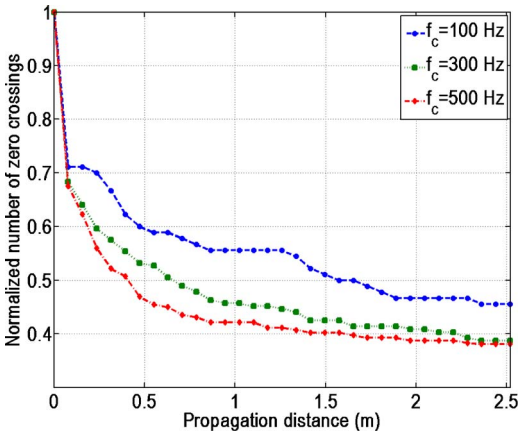


FIG. 7. (Color online) The number of zero-crossings normalized by the initial number of zero-crossing for all three waveforms. A decrease in the number of zero-crossings is shown for each waveform even though only the f_c=500 Hz waveform experiences a downward shift in center frequency (see Fig. 6).

and f_c=500 Hz waveforms have decayed 4 and 7 dB more, respectively. This discrepancy is a direct result of nonlinear shock attenuation.[24]

By visual inspection, all three waveforms involved at least some shock coalescence. For more thorough investigations of the existence of shock coalescence, technical metrics are typically used such as a decrease in the number of zero-crossings[34] or an increase in the characteristic time scale (which is inversely proportional to the center frequency).[19] However, for haystack-shaped noise, the center frequency dependence and random nature of the zero crossings make a direct comparison between the three case studies here impossible. Therefore, the number of zero-crossings normalized by their initial value is shown in Fig. 7 as a function of distance. All three waveforms share a decreasing trend in the number of zero-crossings as a function of propagation distance, with the f_c=100 Hz waveform having the lowest percentage of zero-crossings lost and the f_c=500 Hz having the highest. Since all three waveforms have a similar order of magnitude for the percent change in zero-crossings, but only the f_c=500 Hz waveform experienced a downward shift in frequency, this suggests that the percent change in zero-crossings alone does not give conclusive evidence of a downward frequency shift. Conversely, the absence of a downward shift in center frequency does not conclusively say that shock coalescence has not occurred.

In summary, for the same spectral shapes and OASPL but different center frequencies, the amount of nonlinear distortion is very different. This is expected since nonlinear effects occur over shorter distances for higher frequencies. The relatively good reconstruction for the f_c=100 Hz waveform indicates that although the time waveform changes, the spectral shape can generally be maintained despite the effects of nonlinearity. This occurs since the asymptotic dependence of low frequencies is proportional to f^2, and the high frequency roll-off resembles that of sawtooth wave spectrum, which decay according to the harmonic number.[19] This implies that haystack spectra may retain their spectral shape during magnitude reconstruction since the change in phase due to nonlinear coupling will not be detected. However, the results from the f_c=300 and 500 Hz waveforms indicate that when shock coalescence is more significant, the center frequency will shift downward and cause the high and low frequencies to be offset from the original spectrum, even though the high frequency slopes are similar.

For significant nonlinear propagation of waveforms with differing spectral shapes, the initial spectral shape would naturally not be retained. As shocks form, the nonlinearity will in essence filter the high and low frequency regions of the spectrum and cause significant changes in the spectral shape.[35]

IV. DISCUSSION

The results of this study have shown that shock coalescence can occur despite spherical spreading over short propagation distances near the spreading origin. Given these results, several observations will be discussed in a broader context.

First, the results presented above illustrate how important nonlinear behavior of haystack spectra will not be captured when studying spectrally white waveforms. The most obvious difference between the two is that the peak pressure levels of haystack spectra will be higher than a flat spectrum given the OASPL, therefore causing more nonlinear effects over shorter propagation distances. An effect entirely missed when studying spectrally white noise is how the center frequency influences the propagation, particularly how quickly the waveform will become random sawtooth waves. Comparison of Figs. 1 and 5 for propagation out to 2.52 m illustrates the significance of center frequency with respect to the waveform, transforming into a series of random sawtooth waves. Since the center frequency is partially dictated by the jet nozzle size and corresponding turbulence scales,[15] rocket noise typically has a lower center frequency in the jet noise. However, rocket noise levels are typically much higher than jet noise levels.

Next, several remarks must be made on the use of crest factors in nonlinear propagation of random noise of any spectral shape. As previously noted, McInerny[25] used the crest factor in analyses of rocket noise data to provide an indication of potential nonlinearity. However, Figs. 1, 3, and 5 show that the maximum value changes peaks during propagation. This illustrates how the maximum value is directly influenced by the behavior of the waveform samples directly preceding it. For example, a large negative pressure immediately prior to a large positive pressure will cause significant steepening to occur in both the positive and negative pressure regions, resulting in the formation of one large shock. The relative symmetry of the shock may cause the speed to be near the small amplitude sound speed [see Eq. (2)] while nonlinear attenuation will cause the amplitude to decay and lessen its long-term likelihood of overtaking other portions of the noise. Alternatively, if the large negative pressure was instead a large positive pressure, the shock speed, coalescence, and nonlinear attenuation would be completely different. Therefore, the crest factor is not necessarily a complete

metric for characterizing nonlinear propagation of noise. More detail on the suppression of large outliers due to shock attenuation is found in Ref. 19.

Finally, it has been noted how Pestorius and Blackstock[18] reasoned that spherical spreading will weaken nonlinear effects in the finite-amplitude propagation of noise so that energy transfer will be limited to high frequencies for most practical problems. Although this paper has clearly shown that substantial shock coalescence can occur over short propagation distances despite spherical spreading, a discussion of the assumptions made in this study is merited to determine whether this numerical experiment represents a practical problem.

Some reasoning suggests that the results from this experiment might represent an upper bound on the amount of nonlinearity that might occur near a military jet aircraft or launch vehicle. Center frequencies can be much lower than 500 or even 100 Hz, especially for noise caused by large rockets. This would result in larger propagation distances required for shock formation. Also, the extended nature of the source may cause superposing crossing waves to produce high levels without interacting long enough to produce cumulative nonlinear distortion. Additionally, partial source coherence and diffraction could prevent the waveform from developing shocks strong enough to rapidly coalesce near the source.

Despite these arguments, there are several factors which suggest that the results presented here may represent realistic amounts of nonlinearity near the finite-amplitude sources of interest. First, the amplitudes used in this study are comparable to the levels expected within several meters of military jet aircraft and launch vehicles. Second, the extended nature of the source will result in less than spherical decay near the plume, resulting in nonlinearity occurring relatively more quickly. Also, jet aircraft and rocket noise can exhibit significant positive skewness which, as mentioned previously, are due to large outliers. The higher number of large outliers will cause shock formation and coalescence more quickly than the Gaussian distribution presented here.

As a final point, there is relatively high coherence in the peak radiation direction of jet noise,[15,36] suggesting that the radiating waves from the large structures are not traveling in random directions. Because of this, waves propagating along this angle may have sufficient time for nonlinear interactions. Evidence supporting this idea has recently been observed in near-field, model-scale supersonic jet noise data,[37] where nonlinear effects were detected along the peak radiation direction. All of these arguments suggest that nonlinear interactions, including significant shock coalescence, can be expected within several meters of the plume region of military jet aircraft and rockets, specifically along the peak radiation angle.

V. CONCLUSION

This numerical experiment shows that nonlinear propagation effects can be significant enough to induce significant shock formation and coalescence in less than 3 m of propagation at amplitudes and center frequencies similar to those of military jet aircraft and rocket noise. Additionally, the center frequency of haystack spectra plays an important role in the amount of shock coalescence and how soon the waveform will transform into a series of random sawtooth waves. As previously discussed, arguments exist both for and against the likelihood of significant shock coalescence occurring in actual finite-amplitude jet or rocket noise. However, the results of this paper reveal the importance of seriously considering the finite-amplitude nature of jet or rocket noise sources in the near-field of the source. Furthermore, since nonlinear effects vary with amplitude, distance, and frequency, every source characterization technique performed on high-amplitude noise sources must carefully consider the relative importance of waveform steepening and shock formation before determining the accuracy of the results.

Since the exact role of shock coalescence in jet and rocket noise propagation is experimentally unverified, the results of this study suggest the need for further investigations that specifically target the existence and role of shock coalescence in the near-field of actual jet and rocket noise sources. Studies of the amount of shock coalescence required to appreciably alter a spectrum may help determine threshold sound pressure levels or distances for significant shock coalescence level. Statistics could be used to infer how characteristics of the noise change during propagation. Additionally, knowledge of the specific effects of shock coalescence and the scalability of shock formation and coalescence may help refine efforts in jet noise reduction and aid other advancements in vibroacoustic modeling of military aircraft and launch vehicles.

ACKNOWLEDGMENTS

The authors wish to gratefully acknowledge the financial support from NASA in collaboration with STI Technologies. Additionally, the authors thank Daniel Jensen and the Fulton Supercomputing Laboratory at Brigham Young University for computer support.

[1]J. N. Cole, H. E. von Gierke, D. T. Kyrazis, and A. J. Humphrey, "Noise radiation from fourteen types of rockets in the 1,000 to 130,000 pounds thrust range," WADC Technical Report No. 57-354 AD 130794, Wright Air Development Center, Wright-Patterson AFB, OH, 1957.
[2]W. H. Mayes, W. E. Lanford, and H. H. Hubard, "Near-field and far-field noise surveys of solid-fuel rocket engines of a range of nozzle exit pressures," NASA Technical Report No. TN D-21, NASA Langley RC, Langley Field, VA, 1959.
[3]R. C. Potter and M. J. Crocker, "Acoustic prediction methods for rocket engines, including the effects of clustered engines and deflected exhaust flow," NASA Report No. CR-566, NASA Marshall SFC, Huntsville, AL, 1966.
[4]R. C. Potter, "Noise field for shadowgraph model rocket experiments," NASA Report No. CR-76267, NASA, Washington, DC, 1966.
[5]R. C. Potter, "Investigation to locate the acoustic sources in a high speed jet exhaust stream," NASA Report No. CR-101105, NASA, Washington, DC, 1968.
[6]K. Eldred, "Acoustic loads generated by the propulsion system," NASA Technical Report No. SP-8072, Wyle Laboratories, NASA Langley RC, Hampton, VA, 1971.
[7]K. L. Gee, V. W. Sparrow, M. M. James, J. M. Downing, C. M. Hobbs, T. B. Gabrielson, and A. A. Atchley, "The role of nonlinear effects in the propagation of noise from high-power jet aircraft," J. Acoust. Soc. Am. **123**, 4082–4093 (2008).
[8]K. L. Gee, J. H. Giraud, J. D. Blotter, and S. D. Sommerfeldt, "Energy-

based acoustical measurements of rocket noise," AIAA Paper No. AIAA 2009-3165.
[9]R. H. Schlinker, S. A. Liljenberg, D. R. Polak, K. A. Post, C. T. Chipman, and A. M. Stern, "Supersonic jet noise source characteristics propagation: Engine and model scale," AIAA Paper No. AIAA 2007-3623.
[10]R. J. Kenny, C. Hobbs, K. Plotkins, and D. Pilkey, "Measurement and characterization of space shuttle solid rocket motor plume acoustics," AIAA Paper No. AIAA 2009-3161.
[11]S. Saxena, P. J. Morris, and K. Viswanathan, "Algorithm for the nonlinear propagation of broadband jet noise," AIAA J. **47**, 186–194 (2009).
[12]M. Kandula, "Near-field acoustics of clustered rocket engines," J. Sound Vib. **309**, 852–857 (2008).
[13]D. Casalino, M. Barbarino, M. Genito, and V. Ferrara, "Hybrid empirical/computational aeroacoustic methodology for rocket noise modeling," AIAA J. **47**, 1445–1460 (2009).
[14]C. K. W. Tam, N. N. Pastouchenko, and R. H. Schlinker, "Noise source distribution in supersonic jets," J. Sound Vib. **291**, 192–201 (2006).
[15]C. K. W. Tam, K. Viswanathan, K. K. Ahuja, and J. Panda, "The sources of jet noise: Experimental evidence," AIAA Paper No. AIAA 2007-3641.
[16]C. L. Morfey and G. P. Howell, "Nonlinear propagation of aircraft noise in the atmosphere," AIAA J. **19**, 986–992 (1981).
[17]S. A. McInerny and S. M. Ölçmen, "High-intensity rocket noise: Nonlinear propagation, atmospheric absorption, and characterization," J. Acoust. Soc. Am. **117**, 578–591 (2005).
[18]F. M. Pestorius and D. T. Blackstock, "Experimental and theoretical study of propagation of finite-amplitude noise in a pipe," in *Finite-Amplitude Wave Effects in Fluids*, edited by L. Bjorno (IPC Science and Technology Press, Ltd., Guildford, UK, 1973).
[19]S. N. Gurbatov and O. V. Rudenko, "Statistical phenomena," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic Press, San Diego, CA, 1998).
[20]V. W. Sparrow and R. Raspet, "A numerical method for general finite amplitude wave propagation in two dimensions and its application to spark pulses," J. Acoust. Soc. Am. **90**, 2683–2691 (1991).
[21]T. Yano and Y. Inoue, "Numerical study of strongly nonlinear acoustic waves, shock waves and streaming caused by a harmonically pulsating sphere," Phys. Fluids **6**, 2831–2844 (1994).
[22]Y. Jing and R. O. Cleveland, "Modeling the propagation of nonlinear three-dimensional acoustic beams in homogeneous media," J. Acoust. Soc. Am. **122**, 1352–1364 (2007).
[23]D. T. Blackstock, "Nonlinear propagation in jet noise," in *Proceedings of the 3rd Interagency Symposium on University Research in Transportation Noise* (University of Utah, Salt Lake City, UT, 1975), pp. 389–397.
[24]D. T. Blackstock, M. F. Hamilton, and A. D. Pierce, "Progressive waves in lossless and lossy fluids," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic Press, San Diego, CA, 1998).
[25]S. A. McInerny, "Launch vehicle acoustics Part 2: Statistics of the time domain data," J. Aircr. **33**, 518–523 (1996).
[26]B. P. Petitjean, K. Viswanathan, and D. K. McLaughlin, "Acoustic pressure waveforms measured in high speed jet noise experiencing nonlinear propagation," AIAA Paper No. AIAA 2005-209.
[27]K. L. Gee, V. W. Sparrow, A. A. Atchley, and T. B. Gabrielson, "On the perception of crackle in high-amplitude jet noise," AIAA J. **45**, 593–598 (2007).
[28]J. Lighthill, "The inaugural theodorsen lecture: Some aspects of the aeroacoustics of high-speed jets," Theor. Comput. Fluid Dyn. **6**, 261–280 (1994).
[29]V. A. Khokhlova, O. V. Rudenko, and O. A. Sapozhnikov, "Sawtooth waves: One dimensional statistical ensembles and thermal self-focusing of the beam," in *Frontiers of Nonlinear Acoustics: Proceedings of the 12th ISNA*, edited by M. F. Hamilton and D. T. Blackstock (Elsevier Science Publications, Ltd., London, UK, 1990).
[30]M. S. Wochner, "Numerical simulation of multi-dimensional acousitc propagation in air including the effects of molecular relaxation," Ph.D. thesis, The Pennsylvania State University, University Park, PA (2006).
[31]M. S. Wochner, A. A. Atchley, and V. W. Sparrow, "Numerical simulation of finite amplitude wave propagation in air using a realistic atmospheric absorption model," J. Acoust. Soc. Am. **118**, 2891–2898 (2005).
[32]C. W. Shu, "Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws," in *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, edited by A. Quarteroni (Springer, Berlin, Germany, 1998).
[33]F. M. Pestorius and S. B. Williams, "Upper limit on the use of weak-shock theory," J. Acoust. Soc. Am. **55**, 1334–1335 (1974).
[34]B. P. Petitjean and D. K. McLaughlin, "Experiments on the nonlinear propagation of noise from supersonic jets," AIAA Paper No. AIAA 2003-3127.
[35]M. R. Shepherd, "The effects of nonlinear propagation on near-field acoustical holography," MS thesis, Brigham Young University, Provo, UT (2007).
[36]K. Viswanathan, "Investigation of the sources of jet noise," AIAA Paper No. AIAA 2007-3601.
[37]K. L. Gee, M. R. Shepherd, L. E. Falco, A. A. Atchley, L. S. Ukeiley, B. J. Jansen, and J. M. Seiner, "Identification of nonlinear and near-field effects in jet noise using nonlinearity indicators," AIAA Paper No. AIAA 2007-3653.

Geostatistical modeling of sound propagation: Principles and a field application experiment

Olivier Baume[a)]
Department of Environmental Sciences, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands

Benoit Gauvreau and Michel Bérengier
Section Acoustique Routière et Urbaine, Laboratoire Central des Ponts et Chaussées, Route de Bouaye, B.P. 4129, 44341 Bouguenais Cedex, France

Fabrice Junker
Département Analyses Mécaniques et Acoustique, EDF Recherche & Développement, 1 avenue du Général de Gaulle, 92141 Clamart Cedex, France

Hans Wackernagel and Jean-Paul Chilès
Centre de Géosciences/Géostatistique, Ecole des Mines de Paris, 35 rue Saint-Honoré, 77300 Fontainebleau, France

(Received 13 August 2008; revised 9 September 2009; accepted 14 September 2009)

The assessment of noise sources for environmental purposes requires reliable methods for mapping. Numerical models are well adapted for sophisticated simulations and sensitivity analyses; however, real-time mapping of large frequency bands must be based on fast and acceptable computations and honor *in situ* measurements. In this paper, a real-time mapping procedure of noise exposure is proposed. The procedure is based on geostatistical modeling of spatial variations and applied to a case study taken from an experimental campaign, where a point source was placed on a flat meadow. An analytical approximation of the acoustic field was first computed with the Embleton model. The difference between this approximation and the actual measurements ($L_{\mathrm{eq15\ min}}$ 1/3-octave bands samples from 19 microphones spread over the meadow) showed spatial structure, which has been modeled with a variogram. Finally, the geostatistical technique of kriging with external drift provided an optimal interpolation of the acoustic field data while encapsulating the first approximation from the Embleton model. Systematic geostatistical inference and real-time mapping with the proposed procedure can be envisaged in simple cases.
 [DOI: 10.1121/1.3243301]

PACS number(s): 43.28.Tc, 43.28.Lv, 43.28.En [KVH] Pages: 2894–2904

I. INTRODUCTION

Real-time mapping of sound exposure in semi-open areas has never been attempted although it is of high interest for practical purposes. The assessment of sound exposure in semi-open areas is a difficult task. In semi-open areas, micrometeorology and ground properties have a major influence on sound propagation. Additionally micrometeorology and ground properties change continuously in time and space. Therefore, with the use of experimental data and a limited number of measurement locations in space, propagation conditions can only be known approximately. Prediction models need to take advantage of deterministic models with a growing complexity of the influence of physical parameters: ground effect,[1–3] mean refraction indices, and knowledge on the turbulence structure.[4,5] The notion of randomness already exists in numerical propagation models: for example, the scattering effect of turbulence is considered through the inclusion of physical randomness.[6,7]

In a prediction context, where generally time is hardly involved, highly complex models require a large amount of input data.[8,9] For the purpose of real-time mapping, the use of such models is unrealistic. Rather, a trade-off between the loss of complexity in the physical model and a possible estimation of the embedded error must be found to obtain, in the long run, an appropriate assessment method.

Séchet showed in a previous study[10] that partial information on mean refraction conditions and on the turbulence structure may lead to a loose statistical relationship between micrometeorological measurements and sound level records. He also concluded that, for the purpose of assessment, sound fields in semi-open areas should be modeled as random fields. The spatial structure of the sound field is statistically much easier to evaluate, though, when setting the integration time step of acoustical time series at 15 min.[10,11] At this time scale, mean vertical sound celerity gradients are representative of mean propagation conditions.

The objective of our research was to develop a well-adapted method to assess and map the acoustical impact of sound sources in semi-open areas. We intended to work first on a simplified case study of noise pollution where the

a)Author to whom correspondence should be addressed. Electronic mail: obaume@hotmail.com

source characteristics are known and constant in time, a case where the ground is flat and has slowly varying acoustical properties, and still a case where micrometeorological conditions are measured as well. We wanted to demonstrate that it is practically feasible to model the sound pressure as a random function and, finally, we aimed at evaluating potential applications.

In semi-open areas, the complex influence of micrometeorological conditions and impedance conditions requires to combine both physical modeling and statistical knowledge.[8,9] Physical models provide the main spatial characteristics of the sound field, and statistical methods introduce an uncertainty estimation of the final assessment. In our context, computation time was also considered to be an important factor for practical applications. Hence we considered simplified physical models to be of interest. The prediction provided by a simplified physical model is an approximation of reality. The actual difference between a deterministic approximation of reality and measurements is then viewed as the outcome of a random process, and the spatial structure of such a random process may be characterized by means of statistical inference.

Geostatistics emerged as a collection of statistical techniques for spatial data in a mining industry context.[12] Matheron[13] developed the theoretical basis for its use (see Ref. 14 for a more complete historical review). Applications of geostatistics in environmental issues are numerous. For instance, geostatistical modeling proved efficient in assessing air pollution.[15] In particular, the interest of geostatistics is when the mean underlying physical process is known and can be modeled. This is the case in sound propagation. One issue remains in the choice of the acoustical model to be included. In a method that combines a physical model with statistical tools, it is crucial to find a good trade-off between simplicity, computation time, and compatibility with geostatistical assumptions.

In this article, we aim at introducing geostatistical principles and consider opportunities to infer a geostatistical model for automatic mapping in a simple context. The choice is oriented toward a physical model that does not include micrometeorology in the present case study: the Embleton model.[3] Then, we show how geostatistics can handle differences between a simple physical model such as the Embleton model and measurements which vary under the influence of wind and temperature.

The remainder of the article is structured as follows. In Sec. II, we introduce the principles of geostatistics with an emphasis on its main analysis tool called the variogram and its use in statistical interpolation. In Sec. III, we present the case study, a simple example of sound propagation from a point source in a flat meadow field. In Sec. IV, we describe the different computational steps of the geostatistical study. The computational steps include the choice of an adequate physical model as a first guess of the sound field, the analysis of the fluctuations between this guess and the measurements ($L_{\mathrm{eq15\ min}}$), and finally an interpolation of the data taking account of the guess field. The illustration concerns mid-frequency propagation (1/3-octave band centered on 1 kHz) where the sound field is more erratic than at lower frequency and has an important impact on the environment. Results are interpreted and discussed in Sec. V.

II. INTRODUCTION TO GEOSTATISTICS

The main fact that leads to the application of geostatistics for mapping is the observation of spatial auto-correlation in the studied phenomenon: the closer two measurement locations, the more likely their acoustic levels will be correlated. Geostatistics developed a whole theoretical framework for optimizing mapping in such a situation. We present in this section the basis on which sound propagation can be modeled with geostatistical tools.

A. The variogram

The spatial and temporal uncertainties in micrometeorological conditions, ground acoustical properties, and sound propagation are a major challenge in assessing the acoustic exposure in a spatial domain. Geostatistical methods are designed for solving estimation problems based on the spatial data and numerical model output. Geostatistical methods rely on a probabilistic framework that is appropriate for describing phenomena in space and time.

The starting point is a regionalized variable $p(\mathbf{x})$ describing the value of the acoustic pressure field at location $\mathbf{x}$ in a spatial domain $\mathcal{D}$. In the present paper, $p(\mathbf{x})$ will be expressed as the sound pressure relative to a reference pressure measured or calculated 10 m from the sound source. This relative sound pressure is expressed in decibels (dB) for each 1/3-octave band. In addition, we define the spatial domain $\mathcal{D}$ of geostatistical study as a two-dimensional (2D) horizontal domain so that $\mathbf{x}=\{x_1,x_2\}\in\Re^2$.

Reality $p(\mathbf{x})$, from which data are sampled, is considered as a realization of a random function, by which we mean a collection of random variables $\{P(\mathbf{x}),\mathbf{x}\in\mathcal{D}\}$. To ease notation, the curly brackets are dropped and the random function is simply written $P(\mathbf{x})$.

A convenient assumption is to consider that the variance of the difference $P(\mathbf{x})-P(\mathbf{x}')$ for any two locations $\mathbf{x}$ and $\mathbf{x}'$ is translation invariant, $\mathrm{var}(P(\mathbf{x})-P(\mathbf{x}'))=2\gamma(\mathbf{h})$, i.e., the variance only depends on the vector $\mathbf{h}=\mathbf{x}-\mathbf{x}'$ and not on the position of the vector in $\mathcal{D}$. If we further assume that the expectation of the increments is zero, $E[P(\mathbf{x})-P(\mathbf{x}')]=0$, we obtain the expression

$$\gamma(\mathbf{h})=\tfrac{1}{2}E[(P(\mathbf{x})-P(\mathbf{x}'))^2], \tag{1}$$

which is called the variogram. The above assumption is called *intrinsic stationarity*.

B. Kriging

The problem at hand is to attempt the reconstruction of a phenomenon over a continuous spatial domain based on the measured values, only available at a limited number of sampling locations. The procedure known as *kriging* [after D. G. Krige who proposed it in 1951 (Ref. 12)] consists in searching for the linear estimator that bears the smallest estimation variance. This leads to a system of linear equations (kriging

equations) whose solution provides the weights corresponding to a minimum estimation variance (kriging variance[16,17]). There are close links between splines and kriging despite differences in their practical implementation.[16]

The starting point for estimating P at an unsampled location $\mathbf{x}_0$ is a linear estimator P^* assigning weights w_α to data at locations $\mathbf{x}_\alpha$, $\alpha=1,\ldots,n$,

$$P^*(\mathbf{x}_0)=\sum_{\alpha=1}^{n} w_\alpha P(\mathbf{x}_\alpha). \tag{2}$$

Second, two assumptions will lead to express the statistical moments of the estimator through a linear estimation of weights w_α. First assumption, the estimator is unbiased at all locations of the study domain $\mathcal{D}$, hence writing the particular case at measurement locations

$$P^*(\mathbf{x}_\alpha)=P(\mathbf{x}_\alpha)=\sum_{\alpha=1}^{n} w_\alpha P(\mathbf{x}_\alpha), \quad \alpha=1,\ldots,n. \tag{3}$$

The latter relationship is independently built on the value of P. Thus, at any location $\mathbf{x}_0$ of domain $\mathcal{D}$, the weights w_α are solution of

$$1=\sum_{\alpha=1}^{n} w_\alpha. \tag{4}$$

Second assumption, the intrinsic stationarity [stationarity of the second moment of $P(\mathbf{x})$] allows to use the definition of the variogram $\gamma(\mathbf{x})$ given in Eq. (1).

If the statistical mean of $P(\mathbf{x})$ (in the sense of a Gaussian distribution) is further assumed to be constant over $\mathcal{D}$, the most commonly used estimator is called *ordinary kriging*. In the case of ordinary kriging, weights w_α are solution of the linear system

$$\begin{cases} \displaystyle\sum_{\beta=1}^{n} w_\beta \gamma(\mathbf{x}_\beta-\mathbf{x}_\alpha)+\mu=\gamma(\mathbf{x}_0-\mathbf{x}_\alpha), & \alpha=1,\ldots,n, \\ \displaystyle\sum_{\alpha=1}^{n} w_\alpha=1. \end{cases} \tag{5}$$

In its left-hand side, the system incorporates the variogram values $\gamma(\mathbf{x}_\beta-\mathbf{x}_\alpha)$ between all sample point pairs, while in its right-hand side, we find the variogram values $\gamma(\mathbf{x}_0-\mathbf{x}_\alpha)$ between the estimation location and each sampling location. The second equation that ensures unbiasedness of the estimator explains the presence of an additional variable (the Lagrange parameter μ), which brings an extra degree of freedom in the linear system.

The expression of the minimum estimation variance in case of ordinary kriging is

$$\sigma^2_{\mathrm{OK}}(\mathbf{x}_0)=\sum_{\alpha=1}^{n} w_\alpha \gamma(\mathbf{x}_0-\mathbf{x}_\alpha)+\mu. \tag{6}$$

C. Kriging with external drift

The assumption of intrinsic stationarity may not be suitable if there is a systematic trend in the spatial domain. Such a systematic change can often be described by a numerical model implementing physical equations. The output $\rho(\mathbf{x})$ from such a mechanistic model provides a tentative description of reality $P(\mathbf{x})$ that can be used as another piece of information in addition to measurement values $P(\mathbf{x}_\alpha)$. In such a situation, we will assume that the random function is decomposed into two components

$$P(\mathbf{x})=m(\mathbf{x})+\epsilon(\mathbf{x}) \tag{7}$$

where $\epsilon(\mathbf{x})$ is a zero-mean intrinsic random function called the residuals and $m(\mathbf{x})$ is a deterministic drift. The objective is to use the physical model output $\rho(\mathbf{x})$ in order to obtain a physically sound estimate of the mean $m(\mathbf{x})$, while $\epsilon(\mathbf{x})$ bears the second moment characteristics of $P(\mathbf{x})$—in the sense of Gaussian distribution the variance is a sufficient descriptor of a statistical distribution.

To better assess the trend (on the basis of measurement values), $m(\mathbf{x})$ is assumed to be linearly related to the model output and to spatial coordinates $\mathbf{x}=(x_1,x_2)$ in a 2D horizontal domain:

$$E[P(\mathbf{x})]=m(\mathbf{x})=a+b\rho(\mathbf{x})+c_1x_1+c_2x_2. \tag{8}$$

If we constrain the estimator (2) to interpolate the drift component without error whatever the coefficients a and b, we obtain the equations of kriging with external drift:

$$\begin{cases} \displaystyle\sum_{\beta=1}^{n} w_\beta \gamma(\mathbf{x}_\alpha-\mathbf{x}_\beta)+\mu_1+\mu_2\rho(\mathbf{x}_\alpha)+\mu_3x_{1\beta}+\mu_4x_{2\beta} \\ \qquad =\gamma(\mathbf{x}_\alpha-\mathbf{x}_0), \quad \alpha=1,\ldots,n, \\ \displaystyle\sum_{\beta=1}^{n} w_\beta=1, \\ \displaystyle\sum_{\beta=1}^{n} w_\beta\rho(\mathbf{x}_\beta)=\rho(\mathbf{x}_0), \\ \displaystyle\sum_{\beta=1}^{n} w_\beta x_{1\beta}=x_{10}, \\ \displaystyle\sum_{\beta=1}^{n} w_\beta x_{2\beta}=x_{20}. \end{cases} \tag{9}$$

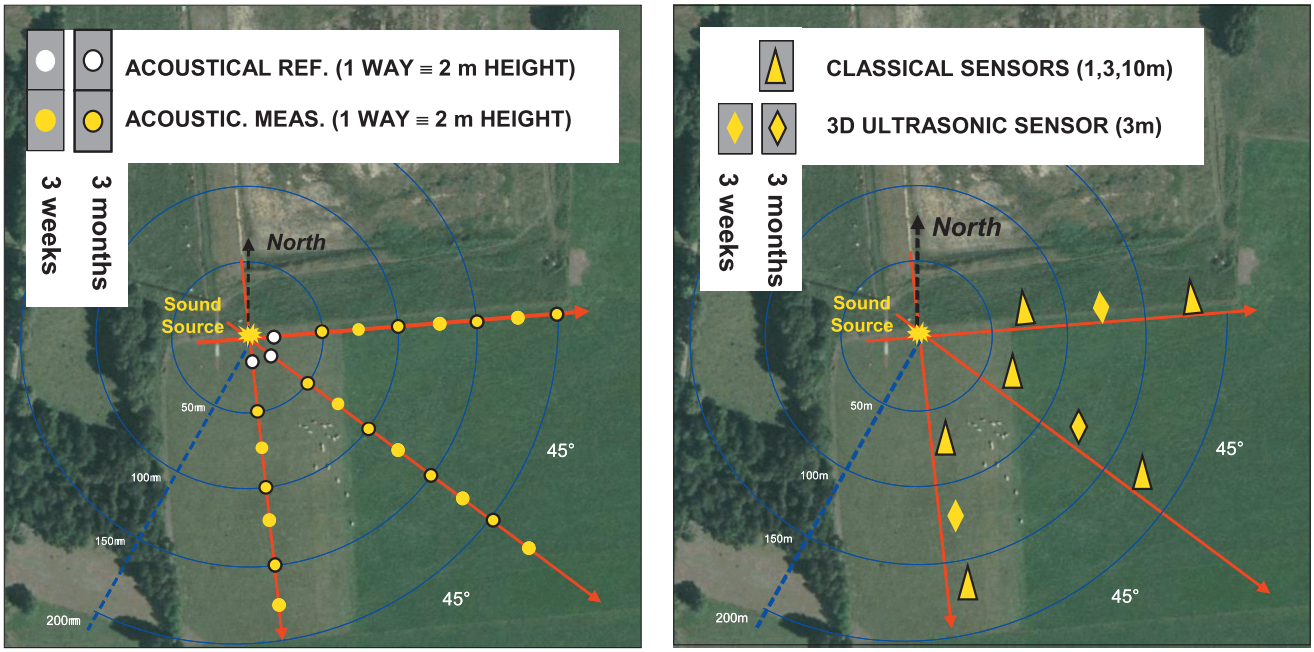
It should be noted that the coefficients a, b, c_1, and c_2 of the external drift do not appear in the system as they are estimated implicitly. An explicit estimation of these coefficients is possible, using a different set of equations[17,18] (see Sec. IV B).

D. Variogram model

An empirical variogram $\hat{\gamma}(h)$ can be computed from the data using the formula

$$\hat{\gamma}(h)=\frac{1}{2N(h)}\sum_{|\mathbf{x}_\alpha-\mathbf{x}_\beta|\approx h}(P(\mathbf{x}_\alpha)-P(\mathbf{x}_\beta))^2, \tag{10}$$

where $N(h)$ is the number of pairs of measurements.



(a) (b)

FIG. 1. Experimental setup of Lannemezan campaign in 2005: the sound source and the microphones are placed 2 m above the flat ground: (a) microphones and (b) micrometerological towers.

In the case of ordinary kriging, a theoretical variogram is fitted to the empirical variogram of $P(\mathbf{x})$, while in the case of kriging with external drift, the theoretical variogram is adjusted to the empirical variogram of the residuals $\epsilon(\mathbf{x})$. A variogram is a conditionally negative definite function and several classes of variogram models are available, allowing to comply with different behaviors of the empirical variogram and different physical contexts.

III. THE CASE STUDY

Sound propagation is a non-stationary process in space and time. Geometrical spreading of wave fronts yields a decrease in sound energy with the distance from the source. The first case study of geostatistical inference needs to fulfill two main criteria:

- to be a well-known example that is easily approximated with a simple analytical model and
- to have a high sensitivity with respect to micrometeorology and ground effects.

Point sources are easier to model than extended sources (e.g. lineic sound sources) and their advantage is that they bear the highest variability of the sound field in space, especially due to the wind direction changes. Thus installing a point source above a flat meadow field seemed a reasonable and well-adapted case study for a first geostatistical work on outdoor sound propagation.

A. The experimental setup and validation of the measurements

We tuned a Bruel & Kjaer source, type 4292 to deliver a 118 dB(A) pink noise during the experiment, its emission power being controlled in a continuous manner with reference measurements 10 m away [see Fig. 1(a)]. The source was placed 2 m above a flat meadow located on a plateau in the area of the Pyrénées mountains (Lannemezan, France). The experiment lasted 3 months on a continuous basis (from early June to late August, 2005). A spatial mean of acoustical properties of the ground was estimated from 13 locations. Measurements of impedance were carried out almost every day during the first 3 weeks (in June) and once a week in the last part of the campaign. One additional impedance monitoring point was placed close to the source (one measurement each 4 h during 3 months).

Sound pressure levels were recorded with 19 microphones during the period of June. Measurements were spatially sampled following three propagation axes which draw a 90° angle around the source [Fig. 1(a)]. We chose this design in order to record acoustical levels under various propagation conditions. The microphones were placed 2 m above the ground. $L_{\mathrm{eq1\ s}}$ samples have been stored for 1/3-octave bands centered on frequencies from 50 Hz to 4 kHz. Propagation conditions were recorded with micrometeorological towers equipped with classical sensors (1 Hz sampling frequency devices as vane anemometers, weather vanes, and ventilated temperature probes) and ultrasonic sensors (20 Hz sampling frequency). Sensors were located as in Fig. 1(b). Micrometeorological data were not included in the geostatistical model but we relied on them for physical interpretation of the results.

A strict validation process was designed and carried out in order to leave out any suspicious sample from the final acoustical and micrometeorological databases. Rainy or too windy conditions (wind speed >5 m/s) were excluded. We selected turbulence data in the validity domain of the Monin–Obukhov similarity theory[19] ($-2<z/L_{\mathrm{MO}}<1$, where z is the height and L_{MO} the Monin–Obukhov length). More-

over Monin–Obukhov theory has been used to derive vertical sound speed profiles from vertical wind and temperature profiles using universal functions with Businger–Dyer empirical parameters.[20] Those experimental (turbulent) data have been cross-correlated with meteo-towers (classical) profiles, in order to verify (or not) the range dependence of the sound speed profiles along each propagation axis during each 15 min sample.

We also filtered out extraneous acoustical events (trains passing by in the valley, road traffic close by, insects noise on the meadow during summertime). The automatic procedure for validation included the use of a characteristic calibration spectrum (envelop spectrum based on the overall temporal mean) at each measurement location. More details on the data validation procedure for this experimental campaign can be found in Ref. 21.

B. The acoustic samples and ground measurements

Although 1-s samples were available from the database ($L_{\mathrm{eq1\ s}}$), the most easily interpretable time series has a 15-min integration interval as mentioned in the Introduction. Thus an integration of acoustical samples was carried out ($L_{\mathrm{eq15\ min}}$). After the validation process, the database was lacking spatial information for a number of time steps. Yet the geostatistical study based on the modeling of spatial autocorrelation should be done with a maximum number of spatial samples.

After the validation process described in Sec. III A, we selected a whole day of acoustical data: the 22 June 2005. Micrometeorological conditions changed substantially from clear night conditions to sunny morning and sunny afternoon conditions of early summer (temperature vertical gradient from positive to negative values), and back to clear night conditions during the evening. We based the case study on two typical 1/3-octave bands for road traffic or industrial noise outdoor propagation: a characteristic band for low frequencies (100 Hz) and a characteristic band for mid-frequencies (1 kHz). Higher 1/3-octave bands could have been selected and included in the sequel. However, considering our sound source, signal to noise ratio is poor for high frequencies (1/3-octave bands superior to 2 kHz) and for long distances (receivers farther than 100 m). After the validation procedure, the experimental database is reduced for those 1/3-octave bands and receiver locations. Thus only a few 15-min samples were validated at high frequencies for this kriging procedure.

We also selected the most representative ground properties values (for a more detailed study on ground impedance measurements for this campaign see Ref. 22). These correspond to the impedance measurements carried out the 20 June 2005. No major precipitation event happened to modify ground properties of the meadow between the 20th and the 22nd in June. Finally, in order to build the final acoustical database for the study, measured acoustic levels (up to 200 m from the source) were expressed in sound pressure levels (SPLs) relative to the reference microphone of the corresponding propagation axis. This allowed us to minimize the effect of spatial directivity and temporal fluctuations (amplitude and spectrum) of the sound source.

Sections IV and V present the application of kriging modeling to this case study. A detailed description of the method is given in Sec. IV for the 1 kHz 1/3-octave band sample ($L_{\mathrm{eq15\ min}}$), recorded between 23:45 and 24:00, on 22 June 2005. The 1 kHz database is more critical than the 100 Hz example because the spatial structure is more irregular (interference patterns due the ground reflection) and more fluctuating as well. Same steps in the modeling process are followed at 100 Hz. In Sec. V, results and discussion also include the low frequency case.

IV. MODELING THE ACOUSTIC FIELD WITH KRIGING

The non-stationarity context of the case study (in the spatial sense) suggests the use of kriging with external drift. The statistical distribution of the measurements for one time sample (19 $L_{\mathrm{eq15\ min}}$ for the SPL) cannot show the Gaussian distribution characteristics of a stationary process because SPL measurements have various positions in terms of source-receiver distance. The interpolation procedure follows five steps.

- Step 1: A description of the deterministic part of the acoustic field is generated using the Embleton propagation model and the Delany and Bazley ground model. A mean value of ground properties over the study domain is used.
- Step 2: The Embleton model is fitted to the measurements using a least squares procedure. This step leads to a preliminary estimation of the drift (the deterministic part of the geostatistical model).
- Step 3: The difference (called residuals) between measurements and the fitted Embleton model (both expressed in SPL relative to the reference microphone located 10 m from the source and 2 m above the ground) is computed. The variogram of the residuals is computed as well.
- Step 4: A variogram model is fitted to the sample variogram of the residuals.
- Step 5: Kriging is used for mapping.

The first four steps can be seen as the inference steps; the model is being built up and its parameters are estimated. The last step is the interpolation step using the inferred model.

A. Step 1: Computation of the first-order physical model

We computed a first-order approximation of the acoustic attenuation (SPL relative to a reference microphone) with the Embleton model.[3] The acoustic attenuation is computed relatively to the reference microphone 10 m away from the source located on the same propagation direction. The general form of the acoustic pressure above a porous half-space in a 2D horizontal domain at location $\mathbf{x}=(x_1,x_2)$ is given by

$$p_E(\mathbf{x}) = \frac{A_d}{r_d(\mathbf{x})}\exp(ik_d r_d(\mathbf{x})) + Q\frac{A_r}{r_r(\mathbf{x})}\exp(ik_r r_r(\mathbf{x})). \quad (11)$$

r_d is the direct travel distance from the source [located at $\mathbf{x}=(0,0)$] to the receiver location $\mathbf{x}=(x_1,x_2)$. r_r is the travel distance of the reflected wave. A_d and A_r are the amplitudes of the direct and reflected waves, respectively, and k_d and k_r are the associated wave numbers. To compute the drift, we assume no turbulence on the wave path. Hence we have $A_d=A_r$ and $k_d=k_r=k_0$, where k_0 is the wave number in the air. Direct and reflected sound ray distances are, respectively,

$$r_d=\sqrt{\mathbf{x}^2+|h_S-h_R|^2},$$

$$r_r=\sqrt{\mathbf{x}^2+(h_S+h_R)^2}, \tag{12}$$

where h_S and h_R are the source and receiver heights. In the specific case of a study in a 2D horizontal plane, source and receiver heights are equal, hence simplifying the previous to

$$r_d=|\mathbf{x}|,$$

$$r_r=\sqrt{\mathbf{x}^2+4h_S^2}. \tag{13}$$

Q is the complex form of the spherical reflection coefficient of the porous ground.[23] In the considered case (propagation distance greater than the wave length), the expression of Q can be approximated by

$$Q=R_p+(1-R_p)F(w), \tag{14}$$

where the plane wave reflection coefficient R_p is

$$R_p=\frac{\sin\phi_i-\mathcal{Z}_0/\mathcal{Z}_G}{\sin\phi_i+\mathcal{Z}_0/\mathcal{Z}_G} \tag{15}$$

and

$$F(w)=2i\sqrt{w}e^{-w}\int_{-i\sqrt{w}}^{\infty}e^{-u^2}du. \tag{16}$$

ϕ_i is the incidence angle with the ground, $\mathcal{Z}_0$ the air impedance, and $\mathcal{Z}_G$ the ground impedance. The numerical distance w reads

$$w=\frac{2ik_0r_r}{(1-R_p)^2\cos^2\phi_i}(\mathcal{Z}_0/\mathcal{Z}_G)^2\left(1-\frac{k_0^2}{k^2}\cos^2\phi_i\right), \tag{17}$$

where k the characteristic wave number of the ground.

Through the numerical distance w and the reflection coefficient R_p, the expression of the field is dependent on the frequency and the ratio $\mathcal{Z}_0/\mathcal{Z}_G$. The computation for 1/3-octave bands is made using several calculations (from 1 to 7 monochromatic calculations for each 1/3-octave band[24]) to avoid local interference patterns. From the impedance measurements, we took into account a spatial average, whose value slowly fluctuates with time.[22] The computation consisted in giving a map of the relative SPL levels 2 m above the ground, following the height of the source and microphones.

Impedance of the ground is an input variable of the Embleton model. In practice, the impedance is assessed through a ground model that defines a relationship between the impedance and other parameters that are fitted with the actual measurements. In our case study, we computed the impedance with the Delany and Bazley phenomenological model.[1] We applied a two-parameter fitting procedure. The first parameter is the phenomenological parameter of the Delany and Bazley model: the specific airflow resistivity σ_G. The second parameter is the layer thickness e, assuming a perfectly reflective surface above e.[25]

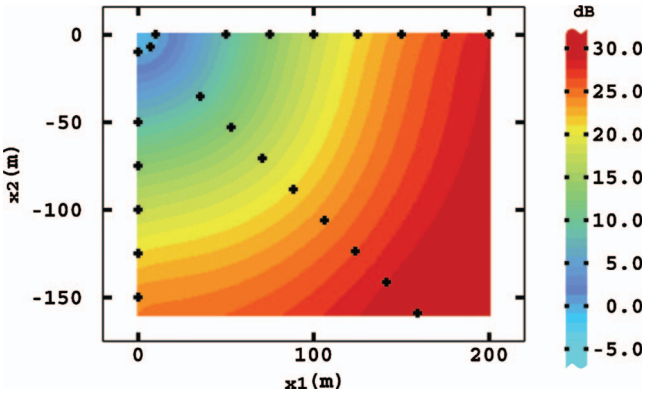


FIG. 2. Computation of SPL rel. to ref. microphone ($L_{eq15\ min}$) using the Embleton model at 1 kHz with mean ground properties: $\sigma_G=151$ kN s m^{-4} and $e=0.02$ m. Prediction height is 2 m above ground. Reference level is taken 10 m from the source and 2 m above ground.

The fitting procedure focused on the frequency range from 100 Hz to 1.5 kHz, leading to the following spatial average values for 20 June 2005:

$$\sigma_G=151\ \text{kN sm}^{-4},$$

$$e=0.02\ \text{m}.$$

Figure 2 displays a map of the SPL values using the Embleton model for the 1 kHz 1/3-octave band. It shows an isotropic propagation around the omni-directional source located in the upper left corner. Interference patterns are barely noticeable for this frequency, even close to the reference measurements mainly because of a coarse grid definition (2×2 m^2).

B. Step 2: Estimation of the trend

The Embleton model does not include micrometeorological conditions. The temperature field can lower or raise SPL values, in an isotropic manner, around the source, whereas the wind field has a vectorial effect. A way to make the Embleton predictions p_E statistically sensitive to micrometeorological conditions is to apply an additive factor a and a multiplicative factor b. In addition, in order to mimic more specifically the influence of wind conditions, a linear combination of the spatial coordinates $c_1\ x_1+c_2\ x_2$ can be added as well. The coefficients a, b, c_1, and c_2 are obtained through a regression procedure. Hence, we arrive to an external drift model of the form

$$P(\mathbf{x})=a+bp_E(\mathbf{x})+c_1x_1+c_2x_2+\varepsilon(\mathbf{x}). \tag{18}$$

P represents the statistical distribution of the acoustic field at any location $\mathbf{x}$, and $\varepsilon(\mathbf{x})$ represents a zero-mean random function associated to P. In a kriging procedure $\varepsilon(\mathbf{x})$ is assumed to be intrinsically stationary.

The regression procedure (to estimate the coefficients a, b, c_1 and c_2) uses two types of data: the computed Embleton model (from step 1) and measured acoustical data. The Em-

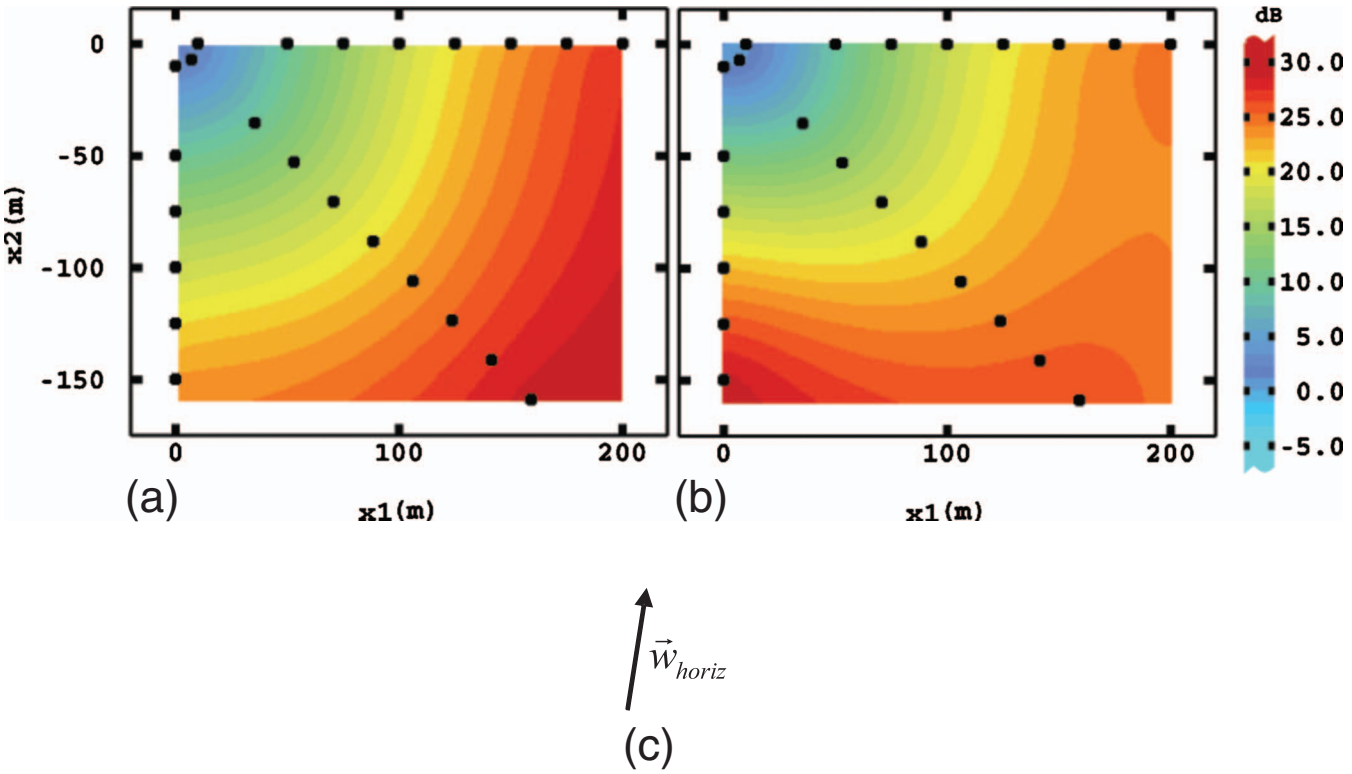


FIG. 3. Lannemezan campaign 22 June 2005 between 23:45 LST and 24:00 LST: fit of Embleton model at 1 kHz (SPL rel. to ref. microphone, $L_{\text{eq15 min}}$): (a) without linear drift and (b) including a linear drift. (c) Direction of mean horizontal wind speed. Prediction height is 2 m above ground.

bleton model is computed 2 m above the ground in terms of SPL relative to a reference level located 10 m from the source ($L_{\text{eq15 min}}$ for each 1/3-octave band, in dB). In the case of measurements, the reference level is taken from the microphone located 10 m from the source and 2 m above the ground on each propagation line (see the experiment protocol in Fig. 1).

Ordinary least squares is a classical method to estimate regression coefficients. At measurement locations, the mathematical expectation of P equals the measurement value. Hence, writing the mathematical expectation of P using Eq. (8) and applying the relationship at n measurement locations lead to

$$\begin{pmatrix} p(\mathbf{x}_1) \\ \vdots \\ p(\mathbf{x}_n) \end{pmatrix} = \begin{pmatrix} p(x_{11},x_{12}) \\ \vdots \\ p(x_{n1},x_{n2}) \end{pmatrix} = \begin{pmatrix} 1 & p_E(x_{11},x_{12}) & x_{11} & x_{12} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & p_E(x_{n1},x_{n2}) & x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} a \\ b \\ c_1 \\ c_2 \end{pmatrix}, \tag{19}$$

whose matrix form is

$$p = \mathbf{X}d. \tag{20}$$

Applying the least squares formalism leads to the estimate $\hat{d}$ of the regression coefficients:

$$\hat{d} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'p. \tag{21}$$

In practice, the linear trend $c_1x_1+c_2x_2$ adds some flexibility in case of an anisotropic behavior of the acoustic field. We also checked the introduction of a quadratic component, but this proved to be irrelevant. In the sequel, we will only compare modeling with and without a linear component in the drift. These estimated drift maps are presented in Fig. 3. Figure 3(c) indicates the direction of the mean horizontal wind speed between 23:45 LST and 24:00 LST.

C. Step 3: Analysis of the variogram of the residuals

The empirical variogram of the residuals ε measures the part of the variability of the studied phenomenon that is not explained by the drift. A high explanatory power of the drift is associated with a low level (or sill) of the variogram of the residuals. Another important characteristic of the variogram is its correlation length (known as the range): namely, the distance at which the variogram reaches its sill. When computing the empirical variogram, one has to carefully choose the distance classes h_d. We tried various distance classes and finally retained classes of 15 m. Figure 4 presents the sample variogram of the residuals when the drift model does not include a linear trend (left plot) or includes a linear trend (right plot). The present case study at 1 kHz shows a clear difference (decreasing of the sill by a factor of 4) when including the linear drift into the drift model. The range then drops substantially from 120 to 90 m.

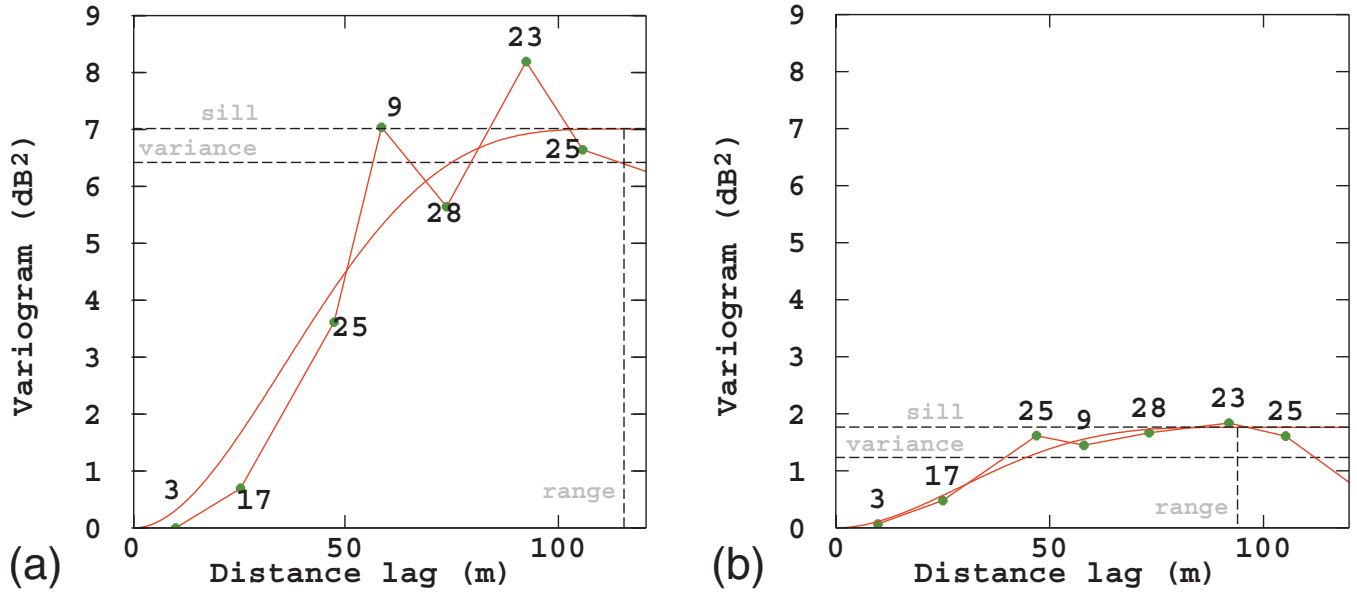


FIG. 4. Lannemezan campaign 22 June 2005 between 23:45 LST and 24:00 LST: empirical variogram of the residuals and associated model at 1 kHz (a) without linear drift and (b) including a linear drift. The horizontal dashed line shows the variance of the residuals. The vertical dashed line indicates the range of the variogram. The zero value at the origin shows that the process is spatially continuous, its parabolic behavior at short distances that it is even differentiable. Distance lag on the x coordinate axes means distance between measurements.

D. Step 4: Variogram fitting

Figure 4 presents the variograms of the residuals (computed from relative sound pressure levels) for both drift models for the 1 kHz data. The usefulness of a linear component in the drift is obvious, because it decreases the variance of the residuals (represented by a horizontal dashed line) as well as the variogram by a factor of 5. Both variograms present no discontinuity at the origin, which means that data do not seem to be affected by measurement errors. Both variograms show a parabolic behavior at short distances, which is the expression of a spatial continuity and differentiability of the acoustic field, which is consistent with physical intuition. They reach a sill at a distance of about 100 m (shorter when a linear component is included in the drift). That distance, called the range, measures the correlation length of the random part of the phenomenon. Kriging makes use of a theoretical variogram model that we have to fit to the sample variogram of the residuals. A great variety of theoretical models exist in the literature, but since we are interested in models with a parabolic behavior at the origin, we only checked the Gaussian model and the cubic model. The latter appeared as the most appropriate. Its expression is

$$\gamma(h) = \begin{cases} v\left(7\frac{h^2}{r^2} - 35/4\frac{h^3}{r^3} + 7/2\frac{h^5}{r^5} - 3/4\frac{h^7}{r^7}\right) & \text{for } h < r \\ v & \text{for } h \geq r, \end{cases} \tag{22}$$

where h is the spatial separation—also called lag—between two locations in meters, r is the range of the variogram (in meters), and v the sill of the variogram (in dB^2). The range and sill of all sample variograms were fitted manually by eye-fit. For eye-fitting, the three most important features are the value at lag 0 (the nugget effect), the range, and the total sill of the sample variogram. Depending on the geostatistical software, the fitting procedure may be more or less flexible depending on the interactive interface of the software. For an automatic process of kriging, automatic procedures also exist,[16] using maximum likelihood methods for instance.

E. Step 5: Prediction

Prediction was achieved by the application of kriging formulas (prediction and prediction error) at any location on the prediction grid. The prediction grid had a 2-m sized discrete mesh (7500 interpolation points). The kriging weights w_α were computed from the system of equations (9). Figure 5 depicts kriging results with both external drift models. Top graphs are the interpolation results. Bottom graphs are the kriging standard-deviation maps. Prediction locations closer to the measurement locations obtain a lower standard deviation value (becoming zero at the measurement locations as exactness is by design a property of the kriging predictor—see Sec. II).

V. RESULTS AND DISCUSSION

The above procedure was repeated for each 15-min validated sample ($L_{\text{eq15 min}}$) during the whole day of 22 June 2005 for 100 Hz and 1 kHz 1/3-octave bands. At both frequency bands, the method worked well. All variograms showed a parabolic behavior at the origin, indicating differentiability of the sound field even at mid-frequency (see Fig. 4). They were easily fitted with a cubic variogram model (with varying range and sill). At 100 Hz, the sill of the variogram is ten times lower than at 1 kHz. Such a difference indicates a better approximation of the field at low frequency when the external drift includes the Embleton model. At low frequency, spatial interferences between direct field and reflected field have larger structures in a semi-open area.

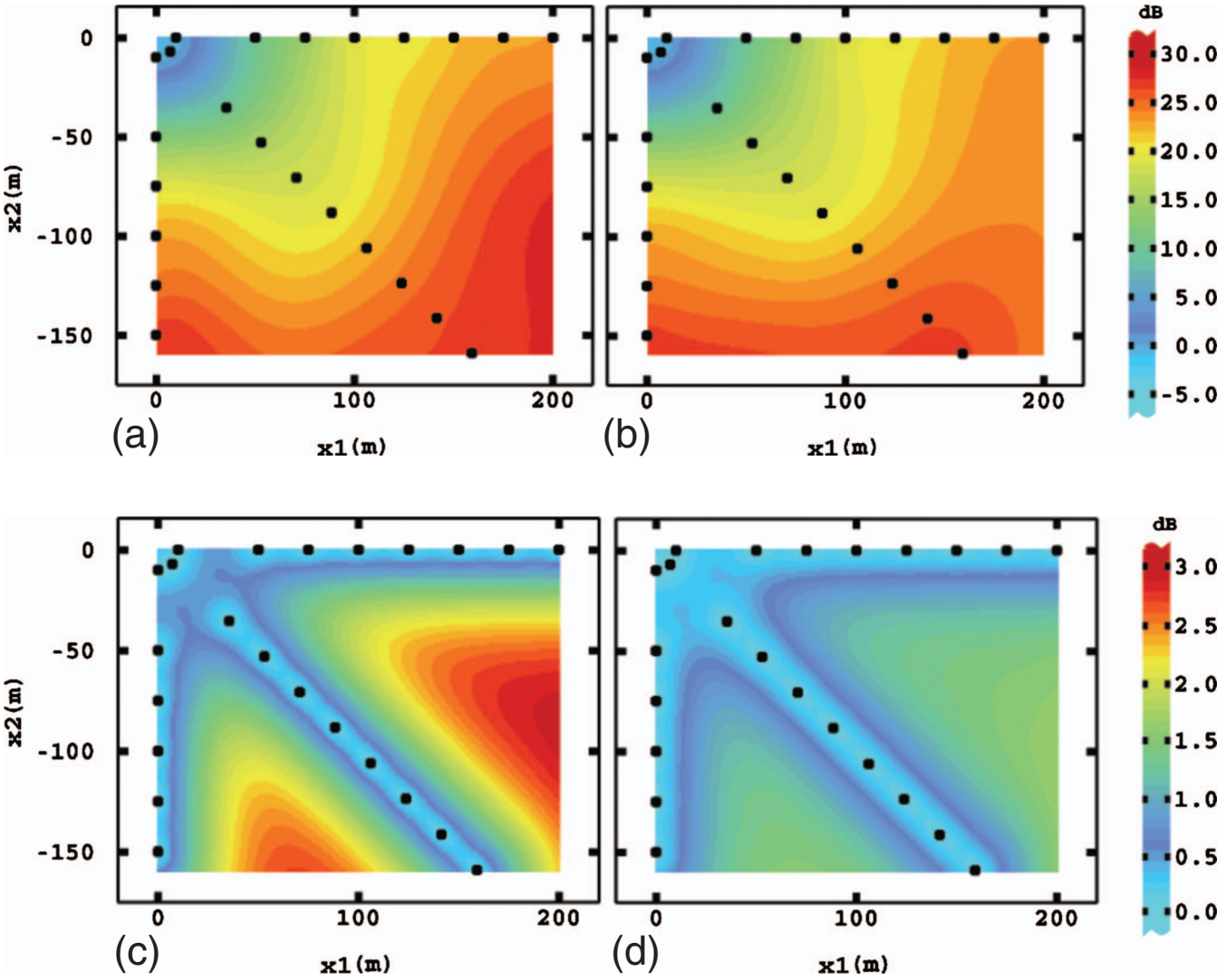


FIG. 5. Lannemezan campaign 22 June 2005 between 23:45 LST and 24:00 LST: final interpolation maps of sound attenuation (SPL rel. to ref. microphone, $L_{\text{eq15 min}}$) at 1 kHz (a) without linear drift and (b) including a linear drift; corresponding standard-deviation maps (c) without linear drift and (d) including a linear drift. Prediction height is 2 m above ground.

Hence we may rely the physical interpretation of the variogram on the physical structure of the acoustic field.

The Embleton model is not designed to follow the temporal variation in the acoustic field but provides a rough sketch of the structure of the acoustic field in our simple case study. It produces a symmetric output model (Fig. 2), whereas reality shows some anisotropy. This anisotropy was modeled through the introduction of a linear drift component in the drift model. The inclusion of the linear drift yields shorter sill and shorter range of the variogram of residuals, and lower kriging variances. A linear trend (at the scale of the data domain) is sufficient to adequately reflect the anisotropy of the acoustic field which can be observed in this case study with integrated $L_{\text{eq15 min}}$ samples. Neglecting turbulence and intermittence during the 15-min periods (first-order approximation), atmospheric effects are mainly related to vertical mean wind and vertical temperature gradients. The map of the external drift with a linear component [Fig. 3(b)] reflects this anisotropy and is close to the corresponding final kriging map [Fig. 5(b)].

In the case study (point source, flat and almost homogeneous ground), the anisotropy of the acoustic field is mostly due to wind conditions. Note that the drift is not static. Its components are estimated at each mapping time step (15 min), which allows to orientate the drift following the orientation of the wind.

Interpolation values have to be interpreted together with the corresponding standard deviation. If we assume a Gaussian distribution for the kriging error, kriging prediction values are not the reality but the most likely values of the sound field. The kriging standard deviation measures the accuracy of the kriging estimate. Thus kriging results have to be interpreted with care. When the interpolation location is far from any measurement, the uncertainty attached to the interpolated value becomes high. As an illustration, one can read the results at location $\{x_1=200, x_2=-100\}$ (following the coordinate system of the maps), where standard deviation is the highest. The interpolated value with the 95% confidence interval (twice the standard deviation) is

30.0 ± 6.0 dB without a linear drift,

24.0 ± 3.2 dB including a linear drift.

Without a linear drift, the interpolation result is too uncertain. In contrast when the linear drift is included, the confidence interval decreases by 2 and is more acceptable in that extrapolation situation.

A. Benefits from geostatistics

Through the construction of a hybrid model (a physical model included in a statistical model), geostatistics leads to interpretative maps of acoustical levels (sound attenuation from reference measurements in our case study). If a Gaussian distribution of the kriging error can be assumed, the outcome of kriging is the local conditional distribution of the acoustical level, which gives access to probability analysis (threshold level detection, for instance, confidence intervals).

Results indicate that a cubic variogram model enables a good fit of the sample variogram of the residuals for the present case study. The inclusion of the linear drift leads to low uncertainty even in extrapolation zones and improves the description of the acoustic field. This leads to an automatic mapping procedure. Input data for the whole procedure are the mean specific air flow resistivity and acoustically equivalent thickness of the ground over the whole domain of study, as well as measurements. The external drift estimation and the kriging prediction on the grid are fast to compute. As a consequence, geostatistical methods can provide real-time mapping procedures for a large band spectrum whose accuracy is given by the kriging standard-deviation map.

We applied the kriging procedure with the major assumption that the non-stationarity of the acoustical propagation can be contained in the drift only, and therefore, we further assumed that the rest (i.e., the residuals) was stationary. Stationarity of the residuals sounds peculiar and may contradict the physical intuition. However, the external drift estimation model actually does not fit better the measured values close to the source than further away, which makes our assumption reasonable in our application.

Kriging is not designed for extrapolation. Even though it is tempting to extrapolate outside the data domain, extrapolation of kriging results is not recommended. A pragmatic reason is that the statistical model, which relies on measured values, does not have information about the acoustic field apart from the data. Especially when the drift does not contain enough physical knowledge, the risk of false prediction may increase. In order to optimize the model inference, a relevant analysis of the kriging errors would require more measurement locations. Such analysis is planned in a forthcoming study.

B. Perspective of applicability

We have provided a simple physical interpretation of the interpolated values. It is much less obvious for the uncertainties. The main reason is the stationarity of residuals. We can question the fact that, from a physical perspective, mapping uncertainty may be of the same level close and further away from the source. Further validation work on that point is needed. For instance, a comparison between random perturbations of numerical prediction with a parabolic equation model and kriging uncertainty would help to envisage the issue.

More complex situations are envisaged as new case studies to test the generalization of the method. We think of varying properties of the ground (for instance, with sharp impedance jumps between the source and the measurement locations). Other existing analytical models could be introduced in the drift as an alternative to the Embleton model. Different topography and source-receiver geometry should also be studied.

We handled data sampled along three transects 45° apart, around the point source. A new modeling attempt should be carried out with sensors all around the source in order to deal with upwind and downwind conditions at the same time.

VI. CONCLUSION

Modeling sound propagation with geostatistical tools has been shown to be feasible in a simple experimental setup (omni-directional source, flat, open and almost homogeneous meadow). Thanks to the integration of an analytical model, kriging yields maps that have a sound physical interpretation. The degree of confidence of the interpolation is indicated by the kriging standard deviation; therefore, statistics in the long run become possible. However, for a clear physical interpretation of kriging uncertainty, some validation work should be pursued to compare the mapping uncertainty associated with kriging maps with that of numerical simulations from a reference model.

The method that we developed so far does not rely on meteorological measurements (the analytical model does not take micrometeorology into account). This could be seen as a conceptual drawback, but it has the practical advantage to be operational when a study must be carried out in the field in a limited time. The Embleton model is adequate for the very simple setting of the Lannemezan 2005 case study. More elaborate models might be needed as approximations of the sound field in other situations. A geostatistical application in a more complex case study of environmental acoustics is in progress. For assessing pollution from an industrial plant or from car traffic, a trade-off must be found between complexity of physical modeling and well-adapted statistical methods.

ACKNOWLEDGMENTS

The authors acknowledge The French Ministry of Ecology and Sustainable Development for its financial support to carry out the "Lannemezan 2005" experimental campaign. The authors are grateful to two anonymous reviewers and especially to Editor Kirill Horoshenkov for fruitful comments that substantially improved the content of this paper.

[1]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbant materials," Appl. Acoust. **3**, 105–116 (1970).
[2]K. Attenborough, S. I. Hayek, and J. M. Lawther, "Propagation of sound over a porous half space," J. Acoust. Soc. Am. **68**, 1493–1501 (1980).

[3]T. F. W. Embleton, "Effective flow resistivity of ground surfaces determined by acoustical measurements," J. Acoust. Soc. Am. **74**, 1239–1244 (1983).
[4]V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN Spon, London, UK, 1999).
[5]E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer, Dordrecht, The Netherlands, 2001).
[6]D. K. Wilson, J. G. Brasseur, and K. E. Gilbert, "Acoustic scattering and the spectrum of atmospheric turbulence," J. Acoust. Soc. Am. **105**, 30–34 (2007).
[7]D. Juvé, P. Blanc-Benon, and P. Chevret, "Numerical simulation of sound propagation through a turbulent atmosphere," in 5th International Symposium on Long Range Sound Propagation, Milton Keynes, UK, pp. 282–286 (1992).
[8]C. L. Pettit and D. K. Wilson, "Proper orthogonal decomposition and cluster weighted modeling for sensitivity analysis of sound propagation in the atmospheric surface layer," J. Acoust. Soc. Am. **122**, 1374–1390 (2007).
[9]D. K. Wilson, E. L. Andreas, J. W. Weatherly, C. L. Pettit, E. G. Patton, and P. P. Sullivan, "Characterization of uncertainty in outdoor sound propagation predictions," J. Acoust. Soc. Am. **121**, EL177–EL183 (2007).
[10]E. Séchet, "Modeling imprecise knowledge about the influence of meteorological conditions on sound propagation from experimental data," Ph.D. thesis, Université de Paris IX, France (1996).
[11]O. Baume, B. Gauvreau, F. Junker, H. Wackernagel, M. Bérengier, and J.-P. Chilès, "Statistical exploration of small scale variation in acoustic time series taking into account micrometeorological conditions," in Forum Acusticum, Budapest, Hungary (2005), pp. 117–122.
[12]D. G. Krige, "A statistical approach to some mine valuation and allied problems on the Witwatersrsand," MS thesis, University of Witwatersrand, South Africa (1951).
[13]G. Matheron, "Principles of geostatistics," Econ. Geol. **58**, 1246–1266 (1963).
[14]N. A. C. Cressie, "The origins of kriging," Math. Geol. **3**, 239–252 (1990).
[15]H. Wackernagel, C. Lajaunie, N. Blond, C. Roth, and R. Vautard, "Geostatistical risk mapping with chemical transport model output and ozone station data," Ecol. Modell. **179**, 177–185 (2004).
[16]J.-P. Chilès and P. Delfiner, *Geostatistics, Modeling Spatial Uncertainty*, Wiley Series on Probability and Statistics (Wiley, New-York, 1999).
[17]H. Wackernagel, *Multivariate Geostatistics*, 3rd Ed. (Springer-Verlag, Berlin, 2003).
[18]G. Hudson and H. Wackernagel, "Mapping temperature using kriging with external drift: Theory and an example from Scotland," Int. J. Climatol. **14**, 77–91 (1994).
[19]R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer Academic, Dordrecht, The Netherlands, 1988).
[20]T. Foken, *Micrometeorology* (Springer-Verlag, Berlin, 2006), p. 306.
[21]F. Junker, B. Gauvreau, C. Cremezi-Charlet, and P. Blanc-Benon, "Classification of relative influence of physical parameters for long range sound propagation," in Forum Acusticum, Budapest, Hungary (2005).
[22]O. Baume, "Geostatistical approach to assess the influence of physical parameters on long range sound propagation," Ph.D. thesis, Université du Maine, France (2006).
[23]J. E. Piercy, T. F. W. Embleton, and L. C. Sutherland, "Review of noise propagation in the atmosphere," J. Acoust. Soc. Am. **61**, 1403–1418 (1977).
[24]F. Junker, C. Cremezi-Charlet, C. Gérault, B. Gauvreau, P. Blanc-Benon, B. Cotté, and D. Écotière, "Classification of relative influence of physical parameters for long range acoustic propagation experimental and numerical results," in Euronoise 2006, Tampere, Finland (2006).
[25]K. Attenborough, K. M. Li, and K. V. Horoshenkov, *Predicting Outdoor Sound* (Taylor & Francis, London, UK, 2007).

Measurements and inversion of acoustic scattering from suspensions having broad size distributions

Benjamin D. Moate[a] and Peter D. Thorne
Proudman Oceanographic Laboratory, Joseph Proudman Building, 6, Brownlow Street, Liverpool L3 5DA, United Kingdom

(Received 21 October 2008; revised 10 August 2009; accepted 8 September 2009)

Measurements are presented from a multi-frequency acoustic backscatter study of aqueous suspensions of irregularly shaped quartz sediments having broad particle size distributions. Using the backscattered sound from a homogenous suspension, measurements of the ensemble backscatter form function and ensemble normalized total scattering cross section were obtained. Three different size distribution types are examined; namely Gaussian, log-normal, and bi-modal distributions, each covering a range of particle sizes similar to those observed in sandy marine environments near the seabed. The measurements of ensemble scattering are compared with theoretical predictions, derived by integrating the intrinsic scattering properties of the sediments over the probability density functions of the size distributions used in the present study. The results show that the ensemble scattering parameters are significant functions of both the width and type of size distribution in suspension. The impact of errors in size distribution width on inversion predictions of both mean size and suspended concentration is also examined. The validity of the theoretical predictions is discussed, along with the implication of the inversion results for using acoustic backscatter data to measure suspended concentration and particle size in sandy marine environments.
 [DOI: 10.1121/1.3242374]

PACS number(s): 43.30.Ft, 43.20.Fn, 43.30.Gv, 43.30.Pc [KGF] Pages: 2905–2917

I. INTRODUCTION

Suspended marine sands significantly scatter underwater sound at megahertz frequencies, with the suspended concentration and size controlling the backscattered intensity (Hay, 1991; He and Hay, 1993; Thorne and Buckingham, 2004). Utilizing this premise, monostatic acoustic backscatter systems (ABS) have been developed in recent years, designed to collect profiles of suspended sediments in the bottom 1–2 m above the bed (Hess and Bedford, 1985; Hay, 1991; Thorne and Hardcastle, 1997). Acoustics offer the advantages of non-intrusive measurements, with centimetric resolution, at turbulent and intra-wave time scales (Thorne and Hanes, 2002).

Early ABS measurements at sea relied on empirical calibrations using locally collected sediments, though this approach could not account for changes in suspended size distribution (Young *et al.*, 1982; Vincent *et al.*, 1991). More recently, analytical inversions of multi-frequency ABS data have facilitated non-empirical estimates of suspended concentration and size (Crawford and Hay, 1993; Thorne and Hardcastle, 1997; Thosteson and Hanes, 1998). Such inversions require knowledge of the acoustic scattering properties of the particles, typically characterized by two dimensionless parameters: the backscatter form function, f, and normalized total scattering cross section, χ. Physically, f describes the backscattering characteristics of a particle relative to its geometrical size, while χ quantifies a particle's total scattering over all angles, relative to its geometric cross section, and is proportional to particle scattering attenuation losses. Both parameters are further detailed by Thorne and Meral (2008).

For regularly shaped particles such as spheres, algebraic expressions describing χ and f can be derived from theory (Sheng and Hay, 1988; Thorne *et al.*, 1993). Theoretical predictions show that χ and f vary significantly with scatterer size and sound wavelength, often being expressed in terms of the dimensionless parameter $x=ka$, where $k=2\pi/\lambda$, with λ the acoustic wavelength in water, and a the particle radius. Measurements of f for single spheres have shown close agreement with theoretical predictions, over a range of x, and for a variety of materials (Thorne *et al.*, 1992). For irregularly shaped particles however, such as marine sands, no analytical theoretical solution exists to describe χ and f. Consequently, to facilitate inversion of marine ABS data, χ and f have been determined experimentally for suspensions of sandy sediments (Flammer, 1962; Hay, 1991; He and Hay, 1993; Schaafsma and Hay, 1997; Thorne and Buckingham, 2004). To assess the variation of χ and f with x, these experiments used sediments sieved into $\frac{1}{4}\,\Phi$ size fractions, where $\Phi=-\log_2(d)$, with d the particle diameter in millimeters. These $\frac{1}{4}\,\Phi$ size fractions produce nominally a single size in suspension, with the standard deviation of the distribution and range of sizes present being approximately 5% and $\pm 10\%$, respectively, relative to the mean size.

Marine suspensions contain broader size ranges than those present in $\frac{1}{4}\,\Phi$ size fractions, however (Bale and Morris, 1987; Krank and Milligan, 1991; Stavn and Keen, 2004). Size distributions of near-bed suspended sediments are often controlled by those present in the local bed material (Sengupta, 1979; Krank and Milligan, 1991), which are consistently broad in the marine environment (Sval'nov and Alek-

[a]Author to whom correspondence should be addressed. Electronic mail: bdm@pol.ac.uk

seeva, 2006). For size distributions, the ensemble χ, defined here as χ_0, and ensemble f, defined here as f_0, are theoretically determined by integrating values of χ and f (being the single size or $\frac{1}{4}\,\Phi$ size fraction values) over the probability density function (PDF) of the suspension (Sheng and Hay, 1988; Thorne and Campbell, 1992; Thorne and Meral, 2008). Derivation of the integration expressions for χ_0 and f_0 includes an assumption that the particles are spherically shaped; however, there has been no validation of these expressions by measurement of χ_0 and f_0 for broad size distributions of either spherical or irregularly shaped particles. Modeling studies show that χ_0 and f_0 vary significantly between narrow and broad size distributions, suggesting correct representation of χ_0 and f_0 is important for analytical inversions (Thorne, 2006; Thorne and Meral, 2008).

As ABS use continues to increase, an evaluation of theoretical estimates of χ_0 and f_0 relative to measured values is timely, as is an assessment of the impact of errors in the size distribution width on inversion predictions of suspended concentration and mean size. These are the aims of the present study. Here, we present measurements of χ_0 and f_0 obtained from aqueous suspensions of irregularly shaped sandy sediments, having broad size distributions comparable to those observed in marine environments. Three distribution types are examined: log-normal, bi-modal, and Gaussian. In addition, we assess whether size distribution type *per se*, rather than width, has any effect on the ensemble scattering. To this end, the ensemble scattering parameters are modeled and compared for log-normal and bi-modal distributions, having identical values of mean size and relative standard deviation. Finally, we examine the impact of size distribution errors on acoustic inversions, for modeled sandy suspensions, using the general intrinsic scattering properties for quartz sand as reported previously in the literature (see Thorne and Meral, 2008, for a recent review and synthesis). These general results are compared to predictions obtained by inverting our measured acoustic scattering data, obtained from the broad size distributions presented herein, and specific to the quartz sand used here.

II. THEORY

At concentrations low enough for multiple scattering to be ignored, the ensemble mean-square backscattered pressure $\overline{P_b^2}$ for a suspension of spheres is (Sheng and Hay, 1988)

$$\overline{P_b^2} = \int_0^{\pi}\int_0^{2\pi}\int_{r_1}^{r_2} N\left(\int_0^{\infty} P_b(a)P_b^*(a)n(a)da\right) \times r^2 \sin\vartheta dr d\phi d\vartheta, \tag{1}$$

where P_b is the backscattered pressure from a single sphere, N is the number of particles in an elemental volume, $n(a)$ is the particle size PDF, θ and φ describe the angular location of the volume relative to an axis normal to the transducer face in spherical polar coordinates, $r_1 = r-(rc\tau/4)$ and $r_2 = r+(rc\tau/4)$, τ is the pulse duration, c is the speed of sound in water, r is the range from the transducer face, and * denotes the complex conjugate. The mass M of spheres of density ρ is

$$M = \frac{4N\rho\pi}{3}\int_0^{\infty} a^3 n(a)da; \tag{2}$$

hence, substituting the usual expression for P_b (Thorne and Campbell, 1992) and N from Eq. (2) into Eq. (1), and assuming $c\tau \ll r$, the far field root-mean-square backscattered pressure, P_{rms}, is

$$P_{\mathrm{rms}} = \frac{K_P}{r}\left(\frac{M}{\rho}\right)^{1/2}\left(\frac{\int_0^{\infty} a^2 f(x)^2 n(a)da}{\int_0^{\infty} a^3 n(a)da}\right)^{1/2} e^{-2\alpha r}, \tag{3}$$

with

$$K_P = P_0 r_0\left(\frac{3c\tau}{16}\right)^{1/2}\left[\int_0^{\pi}\left(\frac{2J_1(ka_t\sin\vartheta)}{ka_t\sin\vartheta}\right)^4 \sin\vartheta d\vartheta\right]^{1/2}, \tag{4}$$

where P_0 is the pressure at a reference range r_0, J_1 is the first order Bessel function and describes the directivity function for a piston transducer, with a_t the transducer radiating aperture, θ is the angle the sphere subtends to the acoustic axis, f is the backscattering form function, α is the total attenuation over the range r, with k and x as previously defined. The total attenuation includes contributions from water absorption, α_W, and particle scattering losses, α_S, which add linearly, e.g., $\alpha = \alpha_W + \alpha_S$ (Sheng and Hay, 1988; Thorne and Campbell, 1992). α_S is proportional to the total scattering cross section, σ_t, of the suspended particles, with $\alpha_S = N\sigma_t/2$ (Sheng and Hay, 1988; Thorne and Campbell, 1992). Conventionally, σ_t is divided by twice the geometrical area of the particles to obtain the normalized total scattering cross section χ. Substituting for N from Eq. (2) thus yields

$$\alpha_S = \frac{3M\int_0^{\infty}\chi(x)a^2 n(a)da}{4\rho\int_0^{\infty} a^3 n(a)da}. \tag{5}$$

For the special case of a single size of sphere in suspension, Eqs. (3) and (5) reduce to

$$P_{\mathrm{rms}} = \frac{K_P f(x)}{r}\left(\frac{M}{a\rho}\right)^{1/2} e^{-2\alpha r} \tag{6}$$

and

$$\alpha_S = \frac{3M\chi(x)}{4a\rho}. \tag{7}$$

Hence, comparison of Eq. (5) with Eq. (7), and Eq. (3) with Eq. (6) shows that for a suspension of spherical particles with some arbitrary size distribution, the ensemble normalized total scattering cross section, χ_0, and ensemble form function, f_0, can be defined as

$$\chi_0(x_0) = \frac{\int_0^\infty an(a)da \int_0^\infty a^2\chi(x)n(a)da}{\int_0^\infty a^3 n(a)da} \tag{8}$$

and

$$f_0(x_0) = \left[\frac{\int_0^\infty an(a)da \int_0^\infty a^2 f(x)^2 n(a)da}{\int_0^\infty a^3 n(a)da} \right]^{1/2}, \tag{9}$$

where $x_0 = ka_0$, with a_0 being the mean size of the suspended size distribution. The mean size and standard deviation, σ, of the suspended size distribution are defined from the PDF as

$$a_0 = \int_0^\infty an(a)da, \tag{10a}$$

$$\sigma = \left(\int_0^\infty (a - a_0)^2 n(a)da \right)^{1/2}, \tag{10b}$$

$$\sigma_0 = \sigma / a_0, \tag{10c}$$

with σ_0 the relative standard deviation. Thus, to determine the ensemble scattering properties at a given sound frequency (and hence x_0), for size distributions consisting of irregularly shaped sands, χ_0 and f_0 are typically evaluated using empirical heuristic expressions to describe the intrinsic scattering properties of the sediments, χ and f, on the right hand side of Eqs. (8) and (9), respectively (Thorne and Meral, 2008).

III. METHODS

A. Description of the particles and size distributions examined

Irregularly shaped quartz sediments were sourced from a commercial supplier; to ensure the desired size, compositional and shape characteristics were obtained. These sediments consisted of quarried quartz sand that had been washed and graded, with the primary composition being 99% silica. Figure 1 shows a scanning electron micrograph (SEM) typical of these irregularly shaped particles. The supplied quartz sediments were sieved into narrow $\frac{1}{4}$ Φ size fractions, providing 20 sub-populations with nominal radii across the range 11.25–327.5 μm in the standard $\frac{1}{4}$ Φ intervals. Acoustic measurements were initially obtained on suspensions of the individual $\frac{1}{4}$ Φ size fractions, to determine the intrinsic scattering properties, χ and f, specific to the sediments used here. The variation of χ and f with frequency and size, x, was characterized by fitting a heuristic expression to the resulting measurements, to enable the evaluation of Eqs. (8) and (9), respectively, at all values of x.

All broad size distributions examined were created by mixing appropriate proportions of the $\frac{1}{4}$ Φ size fractions. Gaussian, log-normal, and bi-modal size distributions are common in the marine environment (Horn and Walton, 2007; Jonasz, 1983; Soulsby, 1997), and the ensemble scattering parameters for these three distribution types were measured. PDFs were modeled for Gaussian, $n_g(a)$, log-normal, $n_l(a)$, and bi-modal, $n_b(a)$, distributions using

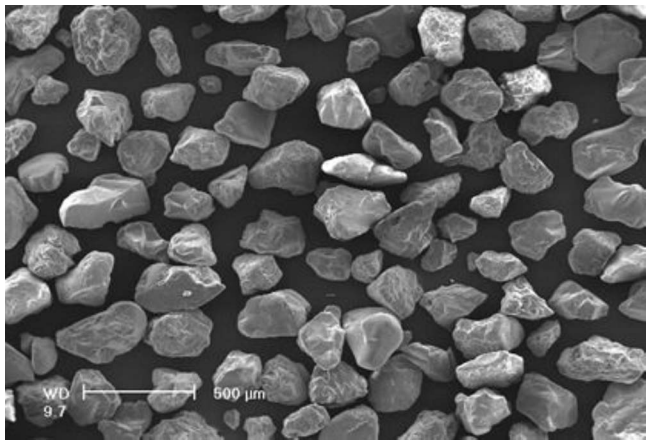


FIG. 1. SEM of the irregularly shaped quartz sediments used in the experiments.

$$n_g(a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(a-a_0)^2/2\sigma^2}, \tag{11}$$

$$n_l(a) = \frac{1}{a\sqrt{2\pi\xi}} e^{-(\log_e(a-m_0)^2)/2\xi^2}, \tag{12a}$$

with

$$\xi = \sqrt{\log_e(\sigma_0^2 + 1)}, \quad m_0 = \log_e(a_0^2/\sqrt{a_0^2 + \sigma^2}) \tag{12b}$$

and

$$n_b(a) = \frac{1}{2\sigma_1\sqrt{2\pi}} e^{-((a-x_1)^2)/2\sigma_1^2} + \frac{1}{2\sigma_2\sqrt{2\pi}} e^{-((a-x_2)^2)/2\sigma_2^2}, \tag{13}$$

with

$$x_1 = 0.375a_0, \quad x_2 = 1.1a_0, \quad \sigma_1 = 2.7\sigma_0 x_1, \quad \sigma_2 = 2.4\sigma_0 x_2,$$

where a_0, σ, and σ_0 are as defined in Eq. (10). To generate a broad size distribution of a given type, Eqs. (11), (12a), (12b), and (13) were used as an initial guide by modeling the PDF at 0.25 μm intervals from 0 to 1000 μm. Modeled PDF values at each $\frac{1}{4}$ Φ size were then used to determine the relative proportions of $\frac{1}{4}$ Φ size fractions to mix, in terms of the number of particles. However, as the metric width of $\frac{1}{4}$ Φ size fractions varies from one size fraction to the next, and as the smallest $\frac{1}{4}$ Φ size fraction in the present study had a nominal radius of 11.25 μm (spanning 10–12.5 μm), the PDF of the size distribution resulting from mixing the available $\frac{1}{4}$ Φ size fractions would not necessarily equal the original modeled PDF. Therefore, the actual PDF was calculated for each broad size distribution at each of the $\frac{1}{4}$ Φ sizes present within the distribution, using

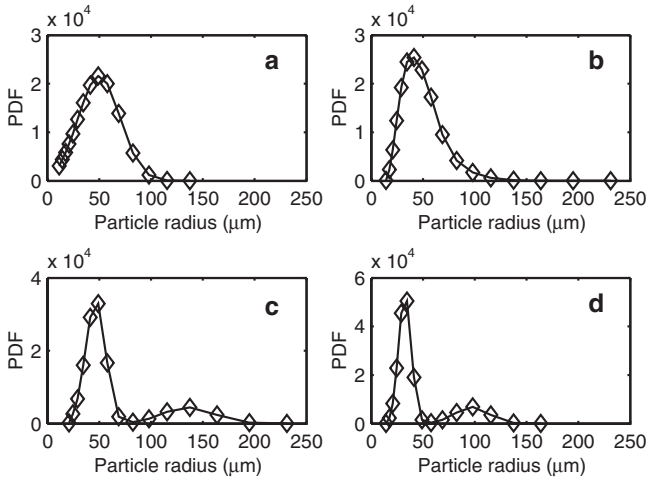


FIG. 2. PDF for each size distribution examined. Distribution number and type were as follows: (a) i, Gaussian; (b) ii, log-normal; (c) iii, bi-modal; and (d) iv, bi-modal. Adjoining lines are for visualization purposes only.

$$n(a_i) = \frac{N_i}{N_T \Delta a_i}, \tag{14}$$

where $n(a_i)$ was the PDF value for size fraction i, a_i was the nominal radius of size fraction i, N_i was the number of particles to be contributed by size fraction i, Δa_i was the metric interval between a_i and a_{i-1}, and N_T was the total number of particles in the resulting broad size distribution. The resulting PDFs of each broad size distribution used in the experiments are presented in Fig. 2. All the PDFs presented in Fig. 2 satisfied the criterion that

$$\int_{i=1}^{i=20} n(a_i) da_i = 1, \tag{15}$$

where da_i is in meters. To physically mix the distributions, the number of particles for each $\frac{1}{4}$ Φ size fraction present in a given distribution was subsequently converted to a mass, M_i, using

$$M_i = \frac{4\pi}{3} a_i^3 N_i \rho, \tag{16}$$

where ρ was the density of the particles being mixed (2650 kg m^{-3}). The required mass of each $\frac{1}{4}$ Φ size fraction was then weighed, and the broad size distribution created by mixing together the individual contributions from each $\frac{1}{4}$ Φ size fraction present. Values of a_0 and σ_0 were calculated for each broad size distribution following Eq. (10), using the appropriate PDF (see Fig. 2). Additionally, broad size distributions were further characterized in terms of the corresponding mass distribution statistics. The mass distributions were described using the median diameter (d_{50}), the geometric standard deviation (Soulsby, 1997), the inclusive graphic skewness, and the graphic kurtosis (Krumbein, 1939; Folk and Ward, 1957). The last two parameters are often used by sedimentologists when describing grain size distributions by mass. The inclusive graphic skewness is positive or negative when more fine or coarse materials are present, respectively, than would be present in a Gaussian mass distribution. Following Krumbein (1939), relatively flat mass distributions have high graphic kurtosis, while those of a peaked nature have lower values. A summary of the statistics for each size distribution shown in Fig. 2 is provided in Table I. The SEM image of the particles presented in Fig. 1 were obtained from size distribution iv [see Table I and Fig. 2(d)].

B. Experimental arrangement

Scattering measurements were obtained in a sediment tower specifically designed for ABS studies, shown in Fig. 3. The sediment tower has been described previously (see Thorne and Buckingham, 2004), and is briefly summarized here. The tower was filled with mains supply water (~133 l), passed through a 20 μm aperture filter, and allowed to warm to room temperature and degas for 3–4 days before measurements commenced. A known mass of scatterers were added to the tower and maintained in suspension by continuously circulating water from the bottom back to the top via the pump assembly (see Fig. 3). The pumps operated at half maximum capacity to prevent pump cavitation and minimize the possibility of introducing air bubbles into the system. The pumped suspension was re-introduced below the upper water surface via a mixing chamber designed to homogenize the suspended particles without the entrainment of air. Providing an additional means of homogenization, a rotating mixing assemblage was mounted in the lower half of the tower, rotating at a constant rate of ~1 rev/s. This arrangement generated a homogenous suspension within the immediate 1 m or so below the transducers, the uniformity of which has been examined elsewhere (Thorne and Buckingham, 2004; Betteridge *et al.*, 2008).

Scattering measurements were collected using a four frequency AQUATEC® AQUAscat ABS, operating at 0.5, 1, 2, and 4 MHz. The AQUAscat transducers were mounted in the top of the tower, facing directly down to insonify the central

TABLE I. Summary of grain size distribution statistics for each distribution shown in Fig. 2. Distribution type, mean size (a_0), and relative standard deviation (σ_0) describe the PDF of the particle size distribution. Median size (d_{50}), geometric standard deviation (σ_g), inclusive graphic skewness (Sk_i), and graphic kurtosis (Kq_g) describe the statistics in terms of the corresponding mass distributions.

Distribution	Type	a_0 (μm)	σ_0	d_{50} (μm)	σ_g	Sk_i	Kq_g
i	Gaussian	50	0.40	136	1.3	0.13	1.0
ii	Log-normal	50	0.40	144	1.5	−0.01	1.0
iii	Bi-modal	69	0.61	288	1.3	0.41	1.5
iv	Bi-modal	49	0.61	202	1.2	0.43	1.6

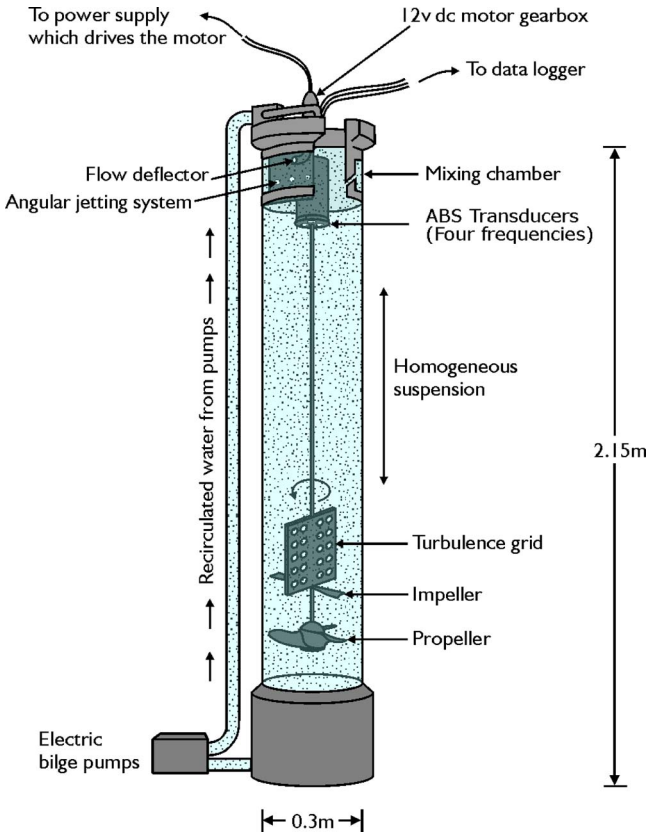


FIG. 3. (Color online) The sediment tower, showing the ABS transducers, mixing assemblage, and pumps.

axis (see Fig. 3). The AQUAscat measured the envelope of the backscattered signal at intervals of 0.01 m over a range of 1.28 m. A pulse repetition frequency of 4 Hz was used (to allow each transmission to fully dissipate before the next transmission) with a system generated ensemble average produced over 32 transmissions, generating 1 recorded profile every 8 s. Prior to the introduction of any scatterers, backscattered profiles were recorded for 1 h to measure the background signal due to any residual scatterers not removed by the 20 μm aperture filter. After adding a known mass of scatterers to the tower, the system was left to homogenize for 1 h before backscattered profiles were collected. Pump samples were collected at four depths beneath the transducers (0.25, 0.45, 0.65, and 0.85 m), to measure the mass concentration in suspension. Pumped sediments were returned to the tower, which was left to re-equilibrate overnight before a second batch of particles was added to obtain measurements at a second concentration the following day, after which the tower was again pump sampled as before. Sediment tower water temperature was monitored throughout the experiments using a standard mercury filled thermometer, placed behind the transducers. Temperature was measured to enable accurate calculation of the speed of sound and, hence, wavenumber k, and x. After each experiment, all particles with radii >10 μm were removed using a 20 μm aperture net. The tower was then emptied and rinsed, with the above process repeated for the next experiment.

TABLE II. Operating frequencies, radiating aperture, a_t, and mean system calibration constants, K_t, for the four frequency AQUAscat ABS. S.D. denotes 1 standard deviation about the mean.

Operating frequency (MHz)	a_t (mm)	K_t (V m$^{2/3}$) $\pm$ S.D.
0.500	10.0	0.0126 $\pm$ 0.0014
1.015	9.0	0.0264 $\pm$ 0.0031
2.015	4.8	0.0164 $\pm$ 0.0020
4.020	4.9	0.0163 $\pm$ 0.0017

C. Obtaining χ_0 and f_0 from the ABS measurements

For a homogenous suspension, as was the case here, then from Eqs. (3), (8), and (9), the root-mean-square voltage recorded by an ABS, V_{rms}, received from multiple spherical scatters with an arbitrary size distribution at range r is

$$V_{\mathrm{rms}}(r) = \frac{K_t K_S M^{1/2}}{r\psi(r)} e^{-2r(\alpha_W+\alpha_S)}, \tag{17}$$

with

$$K_S = \left(\frac{f_0^2}{a_0\rho}\right)^{1/2} \tag{18}$$

and

$$\alpha_S = \frac{3M\chi_0}{4a_0\rho}, \tag{19}$$

where $K_t = gK_PR$, with K_P as defined in Eq. (4), R is the receive sensitivity, g is the electronic gain of the system, ψ accounts for the transducer near field correction (Downing *et al.*, 1995), and all other terms are as defined earlier. Equation (17) requires knowledge of the system constants K_t. Hence, the ABS was calibrated following the procedure of Betteridge *et al.* (2008), with K_t determined before and after the experiments, to check system stability. Mean system constants and other transducer characteristics are provided in Table II, for each operating frequency.

Taking the log transformation of Eq. (17) gives (Thorne and Buckingham, 2004)

$$\log_e(V_{\mathrm{rms}} r\psi) = \log_e(K_t K_S M^{1/2}) - 2r(\alpha_W + \alpha_S). \tag{20}$$

Thus, a linear regression of $\log_e(V_{\mathrm{rms}} r\psi)$ on range yields a slope of $-2(\alpha_W+\alpha_S)$, which, provided M, ρ, and a_0 are known, can be solved for χ_0 (or equivalently, χ for suspensions containing nominally a single size), with α_W taken from the literature (Kaye and Laby, 1986). Here, the slope, α, was calculated using ABS data obtained at ranges between 0.2 and 0.8 m. The lower limit was set to fully avoid transmitter-receiver cross-talk interference following transmission, with the upper limit set to reduce signal to noise problems at longer ranges. In this way, accurate estimates of χ_0 could be obtained providing $\alpha_S \geq \alpha_W$. Where $\alpha_S < \alpha_W$, small errors in α diminish the accuracy of the estimated χ_0. This limitation, combined with the maximum sediment concentration that would not damage the pumps being ~ 2 gl^{-1}, resulted in measurements of χ and χ_0 being usually obtainable at 4 MHz only. To utilize ABS data at lower operating

frequencies, f_0 was calculated using V_{rms} and modeled values of χ_0 to compute α_S, rearranging Eq. (17) as

$$f_0 = \frac{V_{\mathrm{rms}} r \psi \sqrt{a_0 \rho}}{K_t M^{1/2}} e^{2r(\alpha_W + \alpha_S)}. \tag{21}$$

The modeled values of χ_0 were calculated by integrating the heuristic estimates of χ over the appropriate PDF. This approach is valid only if the modeled values of χ_0 are suitably accurate. Numerical simulations showed a 10% error in χ_0, resulted in maximum errors of 0.1%, 0.6%, 2.8%, and 8.8% in the resulting values of f_0 for sound frequencies of 0.5, 1, 2, and 4 MHz, respectively, at the maximum concentrations used of ~ 1 gl^{-1} (data not shown). Similarly, Eq. (21) was used to obtain f for suspensions containing nominally a single size, using the heuristic estimates of χ to compute α_S.

D. Acoustic inversion

Acoustic inversions were conducted using a recently described technique (Thorne *et al.*, 2007). Following this technique, the ratio of the mean of the mass concentration predicted at each operating frequency, $\langle M \rangle$, and the standard deviation about this mean, σ_M, was calculated across a range of mean particle sizes, 5–400 μm, in 0.25 μm intervals. The inversion predicted mean size was obtained from the minimum in the ratio $\sigma_M / \langle M \rangle$. The inversion technique calculates mass concentration iteratively, until convergence at each particle size examined is achieved. The scheme steps through the ABS profile, completing the estimates of mean size and suspended concentration at range r_n before moving on to range r_{n+1}. For measured profiles of V_{rms}, the inversion technique was calculated initially to obtain profiles of predicted mean size, a_I. The profile mean a_I was then used to calculate the inversion to produce smoothed profiles of predicted suspended concentration, M_I. In the present study, the inversion was conducted for ABS operating frequencies of 0.5, 1, and 2 MHz. The 4 MHz data were not used because the high attenuation at this operating frequency produced inversion instabilities. Due to the problem of cross talk already mentioned above, inversion results from the first ten ABS bins were also not used.

IV. RESULTS

A. Intrinsic scattering properties of the irregularly shaped quartz sediments

Initially, acoustic backscatter data were collected from suspensions consisting of the $\frac{1}{4}\,\Phi$ size fractions. The nominal radii of the $\frac{1}{4}\,\Phi$ size fractions for which measurements were obtained were 49, 82.5, 98, 115.5, 137.5, 163.75, and 195 μm, thus spanning the very fine sand to medium sand size classes on the Wentworth grain size scale. From a regression of $\log_e(V_{\mathrm{rms}} \psi r)$ on r, and using Eqs. (19) and (20), measured values of χ were derived from the 4 MHz acoustic backscatter data for all $\frac{1}{4}\,\Phi$ size fractions, and also at 2 MHz

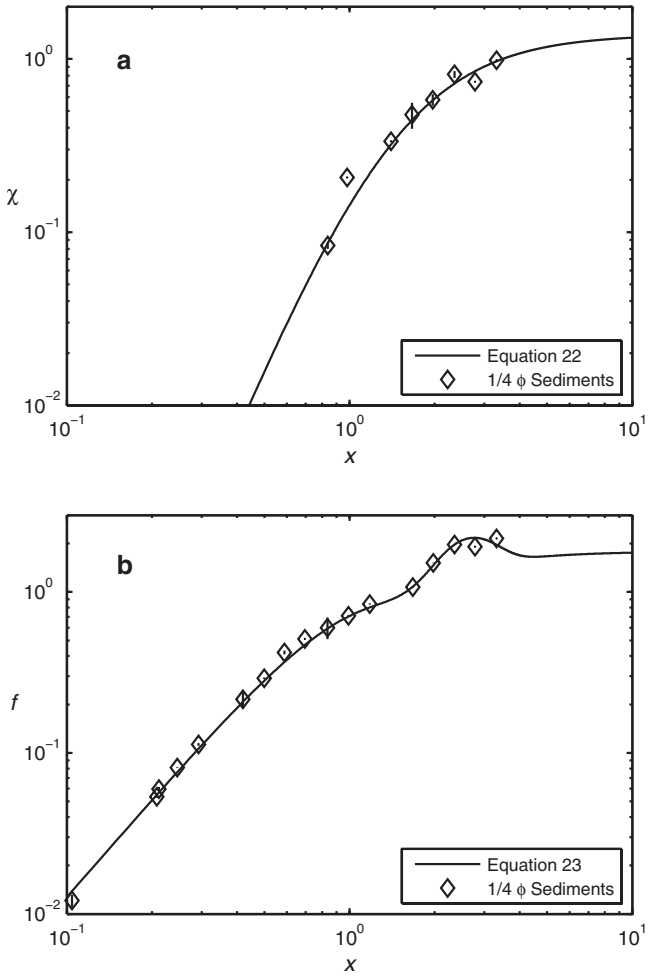


FIG. 4. Measurements of (a) χ and (b) f, obtained from $\frac{1}{4}\,\Phi$ suspensions of irregularly shaped quartz sediments. The heuristic fits to the measurements are also shown (—).

for the 115.5 μm size fraction. The measurements are presented in Fig. 4(a), with each value of χ shown representing the mean of at least two runs in the sediment tower, and error bars denoting ± 1 standard deviation about this mean. Measurements of χ at the other ABS operating frequencies were not derived due to $\alpha_S \leq \alpha_W$ at the suspended concentrations used. The heuristic expression fitted to the measured χ was similar to those reported by others (Sheng and Hay, 1988; Thorne and Meral, 2008), being algebraically

$$\chi(x) = \frac{0.29x^4}{0.92 + 0.9x^2 + 0.21x^4}. \tag{22}$$

This expression, shown as the solid line in Fig. 4(a), reproduced the observed variation of χ with x, increasing from the Rayleigh regime ($x \ll 1$) to a steady plateau in the geometric scattering regime ($x \gg 1$). For the Rayleigh regime, Eq. (22) reduces to $\chi(x) = 0.32x^4$, while for the geometric regime, $\chi(x)$ tended to a constant value of 1.4.

Measurements of f, derived from the ABS data obtained from the same suspensions, are shown in Fig. 4(b). Figure 4(b) shows measurements of f obtained from all four ABS

operating frequencies. This was achieved by modeling χ using Eq. (22), at all operating frequencies, and deriving f from the measured V_{rms} using Eq. (21). The 0.5 MHz transducer was unavailable during the 82.5, 163.75, and 195 μm size fraction measurements. The heuristic expression fitted to the measured f values was

$$f(x) = \frac{v_1 v_2 v_3 x^2}{1 + 0.56x^2} \tag{23}$$

with

$$v_1 = 1 - (0.21e^{((x-1.6)/0.55)^2})$$

$$v_2 = 1 + (0.3e^{((x-0.3)/0.8)^2})$$

$$v_3 = 1 + (0.53e^{((x-2.55)/0.95)^2})$$

which again reproduced the observed features of the data, namely, the increase in f from small x, the point of inflexion in the $x=1-2$ region, and the plateau at large x. For the Rayleigh regime, Eq. (23) reduces to $f(x)=1.26x^2$, while for the geometric regime, $f(x)$ tended to a constant value of 1.8.

B. Scattering from suspensions of irregularly shaped quartz sediments with broad particle size distributions

With the intrinsic scattering properties of the irregularly shaped quartz sediments determined [see Eqs. (22) and (23)], acoustic backscatter data were collected from four separate suspensions of these sands, with the suspensions having size distributions i, ii, iii, and iv, as defined in Fig. 2 and Table I. Figure 5 presents the resulting measured values of χ_0. Figure 5(a) presents χ_0 obtained from Gaussian distribution i and log-normal distribution ii, while Fig. 5(b) presents χ_0 obtained from both bi-modal distributions iii and iv. Measurements of χ_0 were obtained at 4 MHz for all distributions, and additionally at 2 MHz for bi-modal distribution iii. While the number of measurements was limited, Fig. 5 shows close agreement between the measurements and theoretical predictions [derived from Eqs. (8) and (22)]. Error bars denote ± 1 standard deviation about the mean of two runs, and were typically smaller than the graphic symbol. While the mean size of bi-modal distribution iii was different from bi-modal distribution iv, the theoretical prediction curves for these two size distributions were identical [the solid line in Fig. 5(b)], due to the relative standard deviation and distribution type of their respective PDFs being the same (see Table I). The theoretical predictions shown in Fig. 5 clearly illustrate the effect the presence of a size distribution has on the ensemble scattering. Relative to the intrinsic scattering properties of the irregularly shaped quartz sediments (the dotted line in Fig. 5), Fig. 5 shows that χ_0 was elevated at $x_0<1.6$ while generally being reduced at $x_0>1.6$ for all four size distributions. Indeed, Fig. 5(b) shows that at $x_0 \approx 0.25$, the theoretical predictions of χ_0 for the bi-modal distributions are enhanced by an order of magnitude relative to the intrinsic χ represented by Eq. (22). Figure 5 also illustrates that as the size distribution width increases (from $\sigma_0=0.4$ for the log-normal and Gaussian, to $\sigma_0=0.61$ for both bi-modal distributions), the degree to which the ensemble scattering is changed relative to the intrinsic scattering also increases.

Measurements of f_0 derived from the same suspensions are presented in Fig. 6. Figure 6(a) presents f_0 obtained from Gaussian distribution i and log-normal distribution ii, while Fig. 6(b) presents f_0 obtained from both bi-modal distributions iii and iv. Figures 6(a) and 6(b) show measurements of f_0 obtained from all four ABS operating frequencies, obtained by modeling χ_0 using Eq. (22) in Eq. (8) employing the appropriate PDFs (see Fig. 2), and deriving f_0 from the measured V_{rms} using Eq. (21). Figure 6 shows close agreement between the measured values of f_0 derived in this way, and the theoretical predictions [derived from Eqs. (9) and (23)] for all size distributions and across an order of magnitude range in x_0. The effect of the presence of a size distribution on the backscattering is similar to that described above for χ_0, with backscattering generally enhanced at $x_0<1.8$ and reduced at $x_0>1.8$, with the degree of change increasing as distribution width (σ_0) increases.

It is worth noting that despite both the Gaussian and log-normal distributions (i and ii) having the same mean size and relative standard deviation (see Table I), Figs. 5(a) and 6(a) show that at values of x_0 below ~ 1, the enhancement of

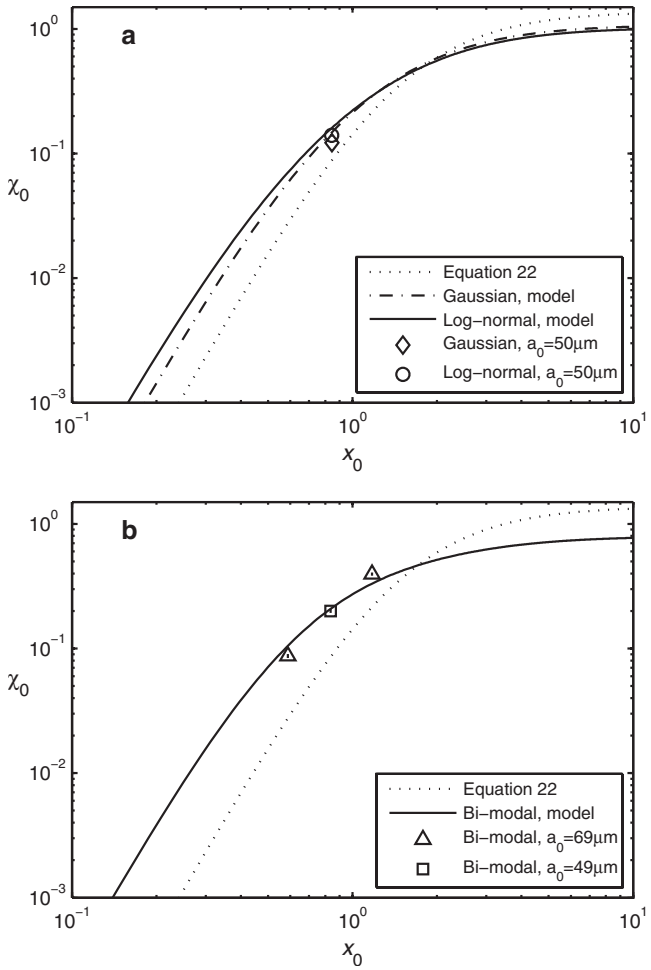


FIG. 5. Measurements of χ_0 obtained from distributions (a) i, Gaussian ($\diamond$) and ii, log-normal ($\circ$), and (b) iii, bi-modal ($\triangle$) and iv, bi-modal ($\square$). Theoretical predictions are shown for distributions (a) i (-.-) and ii (—), and (b) iii and iv (—). Equation (22) is also shown (…).

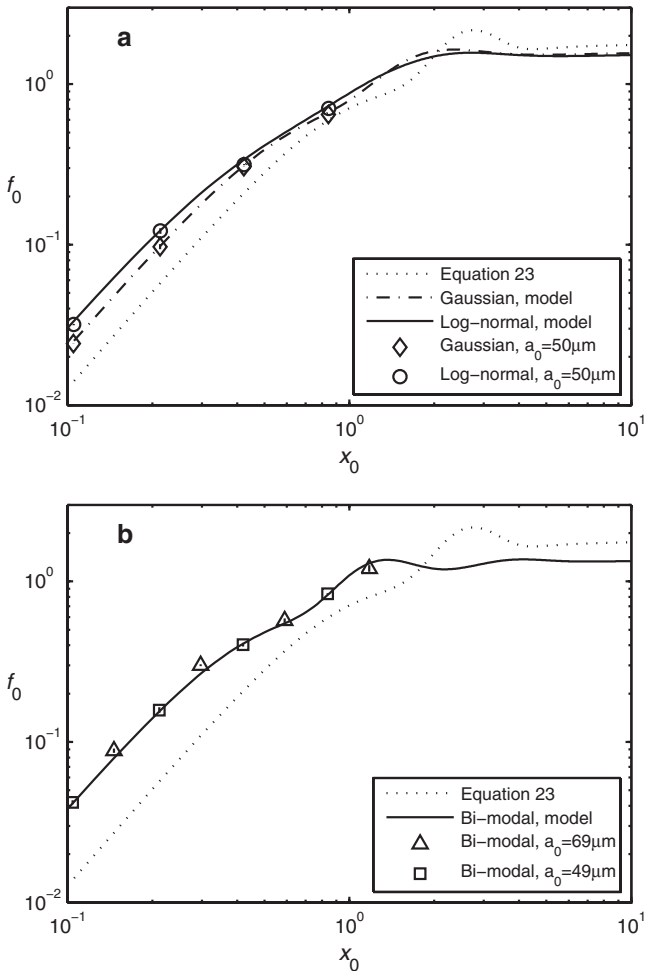


FIG. 6. Measurements of f_0 obtained from distributions (a) i, Gaussian (◇) and ii, log-normal (○), and (b) iii, bi-modal (△) and iv, bi-modal (□). Theoretical predictions are shown for distributions (a) i (-.-) and ii (—), and (b) iii and iv (—). Equation (23) is also shown (…).

χ_0 and f_0 for the log-normal distribution relative to the intrinsic scattering properties was greater than that observed for the Gaussian distribution. It is encouraging that the measurements for distributions i and ii generally resolved this small difference, though it is acknowledged that not all measured values were in perfect agreement with their respective theoretical predictions.

C. Modeling evaluation of the effect of size distribution type on χ_0 and f_0

Figures 5(a) and 6(a) showed that for size distributions with identical a_0 and σ_0, the resulting ensemble scattering parameters were to some degree dependent on the size distribution type, with both χ_0 and f_0 elevated for the log-normal distribution relative to the Gaussian distribution. To further establish the degree to which distribution type can influence ensemble scattering, we compared estimates of χ_0 and f_0 obtained from modeled log-normal and bi-modal distributions having identical a_0 and σ_0. Modeled PDFs were calculated for log-normal and bi-modal distribution types using Eqs. (12) and (13), respectively. For the bi-modal distributions,

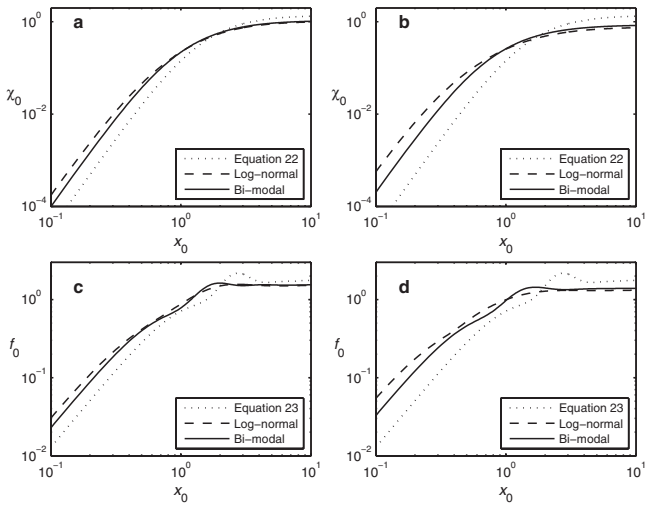


FIG. 7. Modeled predictions of ensemble scattering for log-normal (---) and bi-modal (—) size distributions having $a_0=100$ μm for (a) χ_0 with $\sigma_0=0.4$, (b) χ_0 with $\sigma_0=0.6$, (c) f_0 with $\sigma_0=0.4$, and (d) f_0 with $\sigma_0=0.6$. The intrinsic scattering properties are also shown (…).

$$x_1 = (1 - [0.15 + 0.8\sigma_0])a_0,$$

$$x_2 = (1 + [0.15 + 0.8\sigma_0])a_0, \quad \sigma_1 = 0.5\sigma, \quad \sigma_2 = \sigma, \quad (24)$$

where from Eq. (10), $\sigma=\sigma_0 a_0$. While the form of x_1 and x_2 may seem somewhat complex, this was necessary to ensure that at small σ_0, the two peaks in the bi-modal PDFs were sufficiently close together, while at large σ_0, the two peaks did not merge.

Two log-normal distributions were modeled using Eq. (12), with $a_0=100$ μm and σ_0 set to 0.4 and 0.6. Similarly, two bi-modal distributions were modeled using Eq. (13) for the same set of a_0 and σ_0 values. Using Eqs. (8) and (9), respectively, χ_0 and f_0 were calculated for each modeled size distribution. The resulting theoretical predictions for each combination of a_0 and σ_0 are presented for χ_0 in Figs. 7(a) and 7(b), and f_0 in Figs. 7(c) and 7(d). Figure 7 shows that the effect of size distribution type on the ensemble scattering parameters is greatest at low x_0, and increases as the width (σ_0) of the distributions increases. Figure 7 also shows that in the Rayleigh scattering regime ($x_0 \ll 1$), both χ_0 and f_0 are greater for a log-normal distribution compared to the bi-modal distribution, while in the geometric scattering regime ($x_0 \gg 1$), χ_0 and f_0 are slightly greater for the bi-modal distribution. For the broadest size distributions modeled ($\sigma_0=0.6$), Fig. 7 shows that significant differences exist between the log-normal and bi-modal distributions at $x_0<0.5$ for χ_0 [Fig. 7(b)], and at $x_0<0.3$ for f_0 [Fig. 7(d)].

D. Quantifying the effects of size distribution errors on acoustic inversions

The impact of errors in size distribution width on ABS inversion estimates of mean size and suspended concentration was examined. Using the scattering measurements collected from log-normal size distribution ii (see Fig. 2 and Table I), inversions were computed with the relative standard deviation of the size distributions used in the inversions, σ_I, ranging from 0.1 to 0.8. Figure 8 presents profiles of inver-

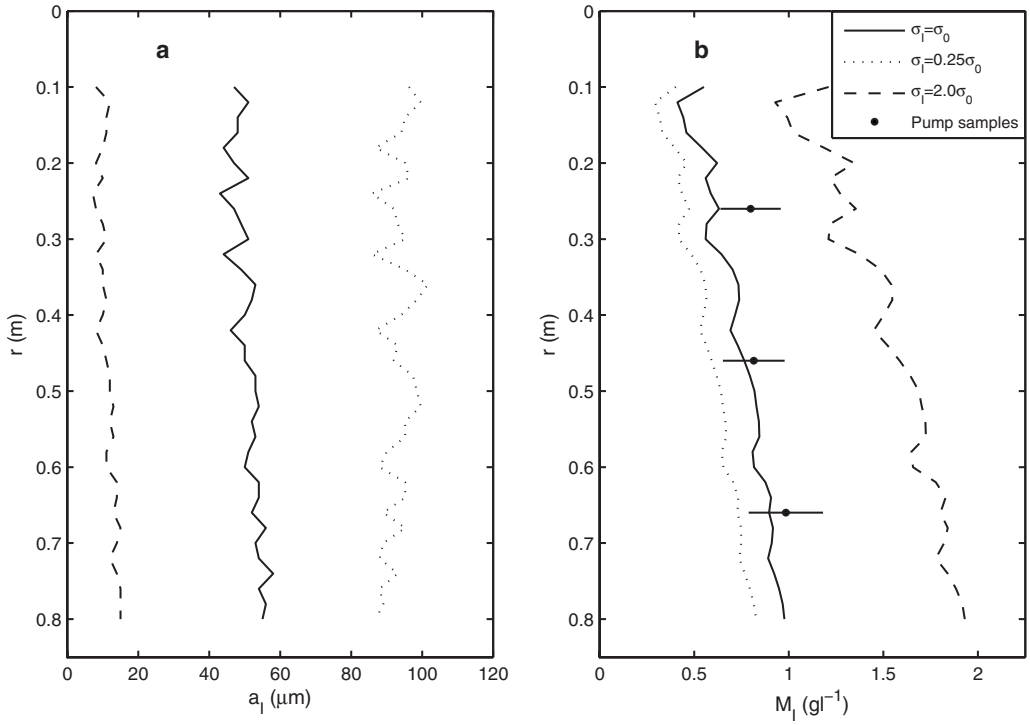


FIG. 8. Inversion predicted profiles of (a) mean size, a_I, and (b) suspended concentration, M_I, derived from the laboratory data obtained from size distribution ii. Inversion results are shown for $\sigma_I=\sigma_0$ (—), $\sigma_I=0.25\sigma_0$ (…), and $\sigma_I=2\sigma_0$ (---), along with the pump sampled concentrations in (b).

sion predicted mean size, a_I, and suspended concentration, M_I, obtained from size distribution ii by calculating the inversion with $\sigma_I=\sigma_0$, with $\sigma_I=0.25\sigma_0$, and with $\sigma_I=2\sigma_0$, where $\sigma_0=0.4$ for size distribution ii (see Table I). Figure 8(b) includes the pump sampled suspended concentrations, with the error bars indicating the typical accuracy achieved using this technique, determined from previous tests in the sediment tower (±20%). With the size distribution properly accounted for, $\sigma_I=\sigma_0$, Fig. 8(a) shows that the inversion predicted mean size was close to the known mean size of distribution ii (50 μm, see Table I). Similarly, Fig. 8(b) shows that with $\sigma_I=\sigma_0$, the inversion predicted suspended concentrations were in close agreement to the concentrations obtained from the pump samples. In contrast, Fig. 8 shows that with the size distribution width underestimated, $\sigma_I=0.25\sigma_0$, predicted mean sizes were overestimated, while predicted suspended concentrations were underestimated. This pattern was reversed if the size distribution width was overestimated, $\sigma_I=2\sigma_0$, and it is worth highlighting that both predicted mean sizes and predicted suspended concentrations were most sensitive to overestimated distribution widths.

Further to the above inversions of the measured scattering data obtained from size distribution ii, additional inversion results obtained from modeled scattering data were also examined. Modeled scattering data were included for two reasons: to enable a more detailed assessment than was practical using laboratory measurements alone, and to enable the generation of inversion results representative of quartz based sands in general, by using the average of the intrinsic scattering properties observed for quartz sand reported elsewhere in the literature (see Thorne and Meral, 2008). Thus, V_{rms} was modeled using Eqs. (17)–(19), and employing the heuristic expressions for χ and f presented in Thorne and Meral, 2008. The modeled V_{rms} was calculated for log-normal size distributions having $\sigma_0=0.4$, being equal to σ_0 for size distribution ii (see Table I), and with a_0 in the range 50–150 μm, and $M=0.5$ gl^{-1}. This choice of a_0 and M is typical of the range of near-bed values reported in previous ABS studies (Crawford and Hay, 1993; Thorne *et al.*, 1993; Vincent, 2007).

The results of inverting the modeled scattering data are presented in Fig. 9 (lines). Figure 9 presents inversion predicted mean sizes, a_I, and suspended concentrations, M_I, normalized by a_0 and M, respectively, obtained by computing the inversions with σ_I ranging from 0.1 to 0.8, and using the appropriate (Thorne and Meral, 2008) scattering properties. Figure 9 includes a_I/a_0 and M_I/M for modeled suspensions having three different mean sizes of $a_0=50$, 100, and 150 μm. Figure 9 shows that while the variation of a_I/a_0 and M_I/M with σ_I/σ_0 was similar for all three modeled suspensions, the degree to which a_I and M_I were altered by errors in σ_I was a function of the mean size in suspension, increasing as the mean size of the suspension decreased. For comparison with these modeling results, Fig. 9 also includes inversion estimates of a_I/a_0 and M_I/M as a function of σ_I/σ_0, obtained by inverting the scattering measurements from size distribution ii. Thus, the symbols in Fig. 9(a) represent the profile mean a_I/a_0 obtained from size distribution ii, with error bars denoting ±1 standard deviation about this mean. In Fig. 9(b), the symbols represent the mean M_I/M obtained from size distribution ii, at the three pump sample depths within the profile, obtained using the profile mean a_I. The error bars in Fig. 9(b) indicate the range of M_I/M, obtained by calculating the inversions with a_I in the range $a_I\pm\sigma(a_I)$. Figure 9 shows that as could reasonably be ex-

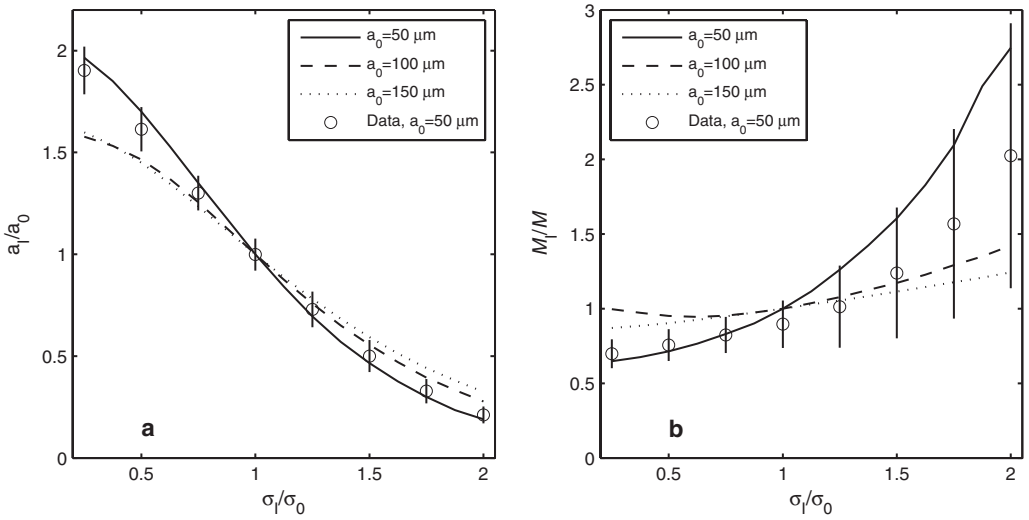


FIG. 9. Normalized inversion predictions of (a) mean particle size, a_I/a_0, and (b) suspended concentration, M_I/M, as a function of the normalized inversion size distribution width, σ_I/σ_0, for a log-normal size distribution. Results are shown for modeled suspensions at three mean sizes (lines) and for the laboratory data obtained from size distribution ii (o).

pected, the inversion predictions from size distribution ii were in closest agreement to the modeled data with $a_0 = 50$ μm, equal to the mean size of distribution ii (see Table I).

A comparable assessment of errors in size distribution width on inversion estimates for the Gaussian and bi-modal distributions presented within this paper was not possible. This was because overestimated distribution widths for both distribution types, for mean sizes in the range studied here, lead to significant portions of the PDF being cut off at 0 μm, resulting in Eq. (15) no longer being satisfied. Additionally for bi-modal distributions, both under and overestimation of distribution width caused the two modes to merge for size distributions iii and iv. However, it was possible to examine the impact of underestimating the size distribution width for distribution i (Gaussian), and it was considered useful to also explore the impact of erroneously including an additional coarser mode in the size distribution used in the inversion. The later error could conceivably occur in the analysis of field data, for example, if only the fine mode of a bi-modal bed distribution were resuspended, with the bed distribution being used to calculate the inversion. Figure 10

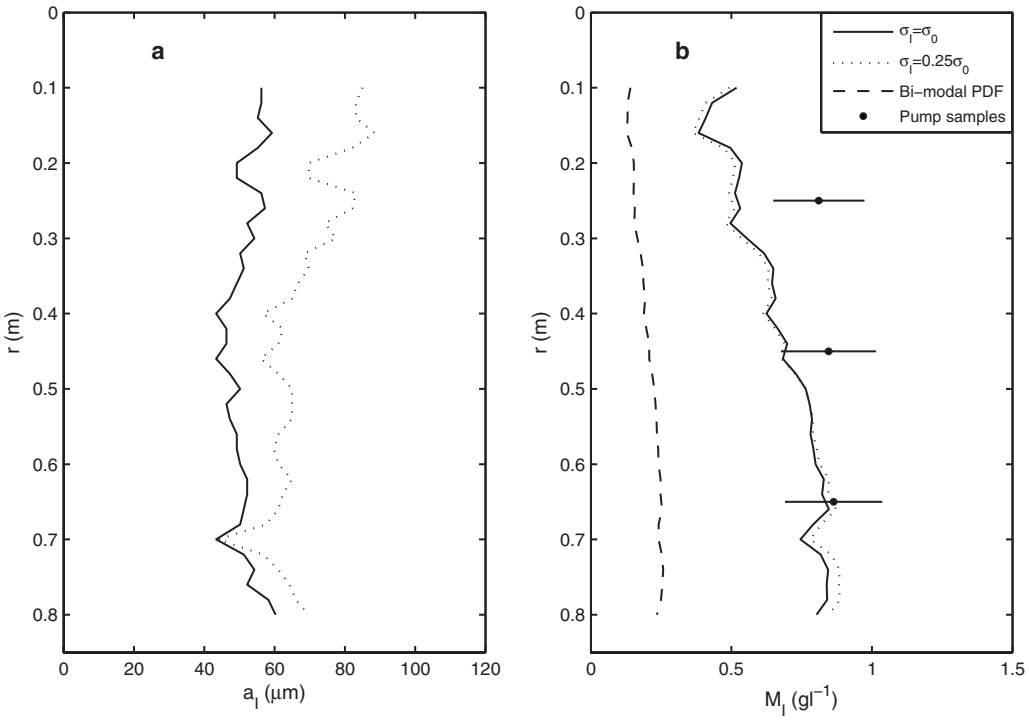


FIG. 10. Inversion predicted profiles of (a) mean size, a_I, and (b) suspended concentration, M_I, derived from the laboratory data obtained from size distribution i. Inversion results are shown for $\sigma_I=\sigma_0$ (—), $\sigma_I=0.25\sigma_0$ (…), and using bi-modal distribution iii in the inversion (---), along with the pump sampled concentrations in (b).

presents profiles of a_I and M_I obtained from distribution i, by calculating the inversion with $\sigma_I=\sigma_0$ and $\sigma_I=0.25\sigma_0$, and by calculating the inversion using bi-modal distribution iii shown in Fig. 2(c). Figure 10 shows that with the size distribution width underestimated, predicted mean sizes were overestimated, as observed for the log-normal distributions above. However, predicted suspended concentrations for $\sigma_I=0.25\sigma_0$ were relatively unchanged compared with the $\sigma_I=\sigma_0$ case [see Fig. 10(b)]. With the scattering measurements obtained from size distribution i, inverted using the bi-modal distribution and mean size of distribution iii (see Fig. 2 and Table I), predicted suspended concentrations were significantly underestimated relative to the pump sampled concentration. These results demonstrate the importance of properly accounting for the suspended size distribution in non-empirical analytical inversion schemes.

V. DISCUSSION AND CONCLUSION

The successful application of acoustics to the measurement of suspended sediments hinges on the analytical inversions used to extract the parameters of interest, namely, the suspended concentration and mean size. Most reported inversion techniques account for the presence of a size distribution by employing the spherical integration approach described by Eqs. (8) and (9) (Hay and Sheng, 1992; Thorne and Hardcastle, 1997; Thosteson and Hanes, 1998). Previously, these expressions have received indirect operational validation only, by comparing ABS predictions to estimates obtained from independent methods, such as optical backscatter. Favorable comparisons were not always obtained, however, particularly in high energy environments where sources of uncertainties are difficult to pinpoint (Hay and Sheng, 1992; Schat, 1997; Thorne *et al.*, 2007). As the ensemble scattering expressions are derived for spheres, it would seem reasonable to assess this spherical integration approach for natural marine sediments, which are typically irregularly shaped and of an angular nature. Perhaps of greater significance, the effects of uncertainties in suspended size distribution on acoustic inversions have not been previously demonstrated. The present study is the first to address these two interrelated issues, providing the first independent evaluation of the spherical integration approach for suspensions of irregularly shaped particles having broad size distributions, and quantifying the impact of errors in size distribution width on inversion predictions of mean size and suspended concentration.

The intrinsic scattering properties of the irregularly shaped quartz sediments used in the present study could not be described theoretically. Consequently, these were characterized following the approach adopted by others (Crawford and Hay, 1993; Thorne and Buckingham, 2004), using suspensions of $\frac{1}{4}\ \Phi$ size fractions to obtain discrete measurements (Fig. 4). Our resulting measured values of χ and f fall well within the range reported by others for quartz based sandy sediments, and show a similar variation with x (Thorne and Meral, 2008). The heuristic expressions fitted to the χ and f observed in the present study differ only slightly in form to those presented in Thorne and Meral, 2008, with χ being a maximum of $\sim$1.2 times greater, and f being $\sim$1.6 times greater here than the average values reported by Thorne and Meral (2008), in the geometric scattering regime. No significant difference existed at values of $x<0.7$.

Figure 5 showed close agreement between the modeled and measured values of χ_0, though some scatter about the model predictions was apparent. The degree of scatter observed in Fig. 5, however (root-mean-square deviation of 0.015), was smaller than the scatter of measured χ around the fitted heuristic expression shown in Fig. 4(a) (root-mean-square deviation of 0.031). Therefore, at least some of the scatter in Fig. 5 was likely to be due to uncertainties in the theoretical predictions of χ_0. This overall close agreement between modeled and measured values of χ_0 (Fig. 4) justified the approach of modeling χ_0 to obtain f_0 from the ABS measurements, and suggested that this did not introduce any significant source of error, or biasing. Generally, the measured f_0 showed even closer agreement to the modeled values, though again, a small degree of scatter was observed (Fig. 6). Sources of this scatter likely include both the uncertainties in χ_0, as well small uncertainties in the measured ABS system constants, K_t (presented in Table II). As the measured values of f_0 were inversely proportional to K_t [see Eq. (21)], any error in K_t would propagate during the calculation of f_0, with the error being potentially different for each operating frequency. In contrast, the measured values of χ_0 were completely independent of K_t, as χ_0 was obtained from the slope of Eq. (20) only. Following the usual rules of uncertainty propagation, i.e., that fractional uncertainties add in quadrature (Taylor, 1997), we estimate that if the K_t values presented in Table II were offset by only 1 standard deviation, the corresponding uncertainty in f_0 would be $\sim$10% for each of the four operating frequencies. Errors of this magnitude would account for most of the departures from the theoretical predictions observed in Fig. 6. The results presented in Figs. 5 and 6 suggest that the spherical integration approach is valid for use with irregularly shaped particles, and is applicable to the inversion of marine ABS data.

The measurements and theoretical predictions show that the effect of a size distribution is to typically enhance scattering in the Rayleigh regime, while reducing scattering in the geometrical regime. It is worth highlighting the physical reason as to why this behavior occurs, which is due to the nature of the variation of the intrinsic scattering properties, χ and f, with x. At low values of x, χ and f increase rapidly with increasing x ($\chi\propto x^4, f\propto x^2$), while above $x\sim 2$, χ and f reach a plateau and thereafter remain relatively constant with increasing x (see Fig. 4 and Sec. IV A). Hence, for x_0 $(=ka_0)<2$, the scattering properties of a single size particle are strongly dependent on x_0, and, consequently, for a given broad size distribution with the same mean size, scattering from the larger particles more than offsets the reduced scattering from the smaller particles, resulting in increased scattering relative to the single size case. In contrast, for $x_0>2$, the single size case has scattering properties defined by the constant region, and consequently for a broad size distribution with the same mean size, the scattering by the larger particles does not compensate for the reduction in scattering that occurs due to the smaller particles, with the net effect

being a decrease in the scattering relative to the single size case. The dependency on size distribution type observed in Figs. 5(a), 6(a), and 7 can be understood by considering this physical behavior for, e.g., log-normal and Gaussian distributions with the same mean size and relative standard deviation. As the size range of the log-normal distribution is biased toward larger particles compared with the Gaussian distribution [see Figs. 2(a) and 2(b)], the ensemble scattering at $x_0 < 2$ is consequently greater for the log-normal distribution than for the Gaussian, as was observed in both the measurements and theoretical predictions [see Figs. 5(a) and 6(a)].

The above mechanism is also responsible for the impact of errors in size distribution width being a function of the mean size of the suspension (see Fig. 9). At the ABS operating frequencies used here for the inversions, 0.5, 1, and 2 MHz, scattering from suspensions with mean sizes in the very fine sand size range is firmly in the Rayleigh regime, with $x_0 < 0.5$. Hence, at these low x_0 values, errors in size distribution width result in significant modifications to the ensemble scattering (see Figs. 5 and 6), leading to substantial errors in the inversion predictions of mean size and suspended concentration. In contrast, scattering from suspensions with mean sizes in the medium sand size range has x_0 values in the region of unity and extending into the geometric scattering regime. Hence, as Figs. 5(a) and 6(a) show, errors in size distribution width have a reduced impact on the ensemble scattering for log-normal size distributions at these higher x_0 values, compared with $x_0 < 0.5$, and consequently the resulting errors in the inversion predictions of mean size and suspended concentration are therefore reduced.

The agreement between the modeled and measured inversion results shown in Fig. 9 is perhaps to be expected, given the small differences between the Thorne and Meral (2008) intrinsic scattering expressions, used in the model, and those specific to the sand used here, Eqs. (22) and (23), as quantified above. Both the measured and model inversion results show that for underestimated size distribution widths, predicted mean sizes were overestimated, while for overestimated size distributions the predicted mean sizes were underestimated. This occurs because an underestimated size distribution width results in reduced ensemble scattering, and hence to account for the observed scattering, an increase in inverted mean size results. Similarly, an overestimated size distribution width results in enhanced ensemble scattering, and hence to account for the observed scattering a decrease in inverted mean size results. The differences in the effect of underestimated size distribution width on predicted mean size observed for distributions ii and i [Figs. 8(a) and 10(a), respectively], can therefore be related to the differences observed in the ensemble scattering properties for these two distributions [see Figs. 5(a) and 6(a)]. For Gaussian distribution i, χ_0 and f_0 are lower than for log-normal distribution ii, despite the two size distributions having the same relative standard deviation and mean size. Consequently, for $\sigma_I = 0.25\sigma_0$, the predicted mean size from distribution i is not overestimated as much as a_I obtained from distribution ii to account for the observed scattering. This in turn results in a lower error in the predicted suspended concentrations for distribution i.

Figure 9 shows what at first glance may appear to be an apparent contradiction in the variation of uncertainties with σ_I/σ_0 for the measured inversion predictions. In Fig. 9(a), the uncertainties in a_I/a_0 are greatest for underestimated size distribution widths, $\sigma_I/\sigma_0 < 1$, while in Fig. 9(b) the uncertainties in M_I/M are greatest for overestimated size distribution widths, $\sigma_I/\sigma_0 > 1$. This behavior is also due to the nature of the variation of the scattering properties with x. For underestimated size distribution widths, the predicted mean sizes are overestimated, moving x_0 toward the geometric scattering regime. Here, errors in size distribution width have a reduced impact on the ensemble scattering, and small changes in x_0 have little effect on χ_0 and f_0. Hence, uncertainties in a_I do not result in large uncertainties in M_I. In contrast, for overestimated size distribution widths, the predicted mean sizes are underestimated, moving x_0 further into the Rayleigh scattering regime. Here, errors in size distribution width have a substantial impact on the ensemble scattering properties, and small changes in x_0 result in large changes in χ_0 and f_0 (see Figs. 5 and 6). Hence, even small uncertainties in a_I result in large uncertainties in M_I, as observed in Fig. 9.

These results have significant implications for the inversion of ABS data collected at sea. Numerous observations in the marine environment have shown that suspended size distribution can vary over inter-tidal time scales due to advection, flocculation, and resuspension processes (Whitehouse, 1995; Li *et al.*, 1999; Ellis *et al.*, 2004). Our assessment of the effect of size distribution errors on ABS inversions illustrates that if such temporal changes in size distribution are not properly accounted for, significant offsets in ABS predictions of suspended concentration and mean size can result (see Figs. 8–10), particularly if the mean size in suspension is in the very fine sand size range, or below. One avenue to counter this problem would be to collect simultaneous and independent measurements of suspended particle size distribution, to ensure the correct size distribution can be supplied to the ABS inversion at any given time, as has been attempted recently (Thorne *et al.*, 2007). *In-situ* measurements of suspended particle size distribution can now be routinely obtained at sea using laser diffraction based instruments, such as the Sequoia® LISST (Laser In-Situ Scattering and Transmissiometry), though it is currently unclear whether optical and acoustic techniques observe size distributions in a compatible manner. It is also unclear as to how the scattering properties of flocculated particles differ, if at all, to those of disaggregated particles of equal size. Investigations to address both of these issues are required to expand the quantitative use of ABS in the marine environment.

ACKNOWLEDGMENTS

This work was jointly funded by the UK Natural Environment Research Council, as part of the FLOCSAM and FORMOST projects, and by the European HydralabIII-SANDS program. The authors would like to thank

AQUATEC® SUBSEA Ltd. for the loan of the 0.5 MHz transducer, and also Dr. Richard D. Cooke (Proudman Oceanographic Laboratory) for his assistance.

Bale, A. J., and Morris, A. W. (**1987**). "*In situ* measurements of particle size in estuarine waters," Estuarine Coastal Shelf Sci. **24**, 253–263.

Betteridge, K. F. E., Thorne, P. D., and Cooke, R. D. (**2008**). "Calibrating multi-frequency acoustic backscatter systems for studying near-bed suspended sediment transport processes," Cont. Shelf Res. **28**, 227–235.

Crawford, A. M., and Hay, A. E. (**1993**). "Determining suspended sand size and concentration from multifrequency acoustic backscatter," J. Acoust. Soc. Am. **94**, 3312–3324.

Downing, A., Thorne, P. D., and Vincent, C. E. (**1995**). "Backscattering from a suspension in the near field of a piston transducer," J. Acoust. Soc. Am. **97**, 1614–1620.

Ellis, K. M., Bowers, D. G., and Jones, S. E. (**2004**). "A study of the temporal variability in particle size in a high-energy regime," Estuarine Coastal Shelf Sci. **61**, 311–315.

Flammer, G. H. (**1962**). "Ultrasonic measurements of suspended sediments," Geological Survey Bulletin No. 1141-A, United States Government Printing Office, Washington, DC.

Folk, R. L., and Ward, W. C. (**1957**). "Brazos River bar: A study in the significance of grain size parameters," J. Sediment. Petrol. **21**, 127–130.

Hay, A. E. (**1991**). "Sound scattering from a particle laden, turbulent jet," J. Acoust. Soc. Am. **90**, 2055–2074.

Hay, A. E., and Sheng, J. (**1992**). "Vertical profiles of suspended sand concentration and size from multifrequency acoustic backscatter," J. Geophys. Res. **97**, 15661–15677.

He, C., and Hay, A. E. (**1993**). "Broadband measurements of the acoustic backscatter cross section of sand particles in suspension," J. Acoust. Soc. Am. **94**, 2247–2254.

Hess, F. R., and Bedford, K. W. (**1985**). "Acoustic backscatter system (ABSS): The instrument and some preliminary results," Mar. Geol. **66**, 357–379.

Horn, D. P., and Walton, S. M. (**2007**). "Spatial and temporal variations of sediment size on a mixed sand and gravel beach," Cont. Shelf Res. **202**, 509–528.

Jonasz, M. (**1983**). "Particle-size distributions in the Baltic," Tellus B Chem. Phys. Meterol. **35**, 346–358.

Kaye, G. W. C., and Laby, T. H. (**1986**). *Tables of Physical and Chemical Constants* (Longman, London).

Krank, K., and Milligan, T. G. (**1991**). "Grain size in oceanography," in *Principles, Methods, and Application of Particle Size Analysis*, edited by J. P. M. Syvitski (Cambridge University Press, Cambridge), pp. 332–345.

Krumbein, W. C. (**1939**). "Graphic presentation and statistical analysis of sedimentary data," *Recent Marine Sediments*, edited by P. D. Trask (Thomas Murby Co., London).

Li, B. G., Eisma, D., Xie, Q. Ch., Kalf, J., Li, Y., and Xia, X. (**1999**). "Concentration, clay mineral composition and Coulter counter size distribution of suspended sediment in the turbidity maximum of the Jiaojiang River estuary, Zhejiang, China," J. Sea Res. **42**, 105–116.

Schaafsma, A. S., and Hay, A. E. (**1997**). "Attenuation in suspensions of irregularly shaped sediment particles: A two-parameter equivalent spherical scatterer model," J. Acoust. Soc. Am. **102**, 1485–1502.

Schat, J. (**1997**). "Multifrequency acoustic measurement of concentration and grain size of suspended sand in water," J. Acoust. Soc. Am. **101**, 209–217.

Sengupta, S. (**1979**). "Grain-size distribution of suspended load in relation to bed materials and flow velocity," Sedimentology **26**, 63–82.

Sheng, J., and Hay, A. E. (**1988**). "An examination of the spherical scatterer approximation in aqueous suspensions of sand," J. Acoust. Soc. Am. **83**, 598–610.

Soulsby, R. (**1997**). *Dynamics of Marine Sands* (Thomas Telford, London).

Stavn, R. H., and Keen, T. R. (**2004**). "Suspended minerogenic particle distributions in high-energy coastal environments: Optical implications," J. Geophys. Res., C **109**, C05005.

Sval'nov, V. N., and Alekseeva, T. N. (**2006**). "Average grain-size parameters of unlithified sediments in the world ocean," Oceanology (Engl. Transl.) **46**, 545–556.

Taylor, J. R. (**1997**). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements* (University Science Books, Mill Valley, CA).

Thorne, P. D. (**2006**). "Acoustic backscattering from suspended sediments with narrow, broad and bi-modal size distributions," in Proceedings of the Eighth European Conference on Underwater Acoustics, Carvoeiro, Portugal, 12–15 June, edited by S. M. Jesus and O. C. Rodriguez.

Thorne, P. D., Agrawal, Y. C., and Cacchione, D. A. (**2007**). "A comparison of near-bed acoustic backscatter and laser diffraction measurements of suspended sediments," IEEE J. Ocean. Eng. **32**, 225–235.

Thorne, P. D., and Buckingham, M. J. (**2004**). "Measurements of scattering by suspensions of irregularly shaped sand particles and comparison with a single parameter modified sphere model," J. Acoust. Soc. Am. **116**, 2876–2889.

Thorne, P. D., and Campbell, S. C. (**1992**). "Backscattering by a suspension of spheres," J. Acoust. Soc. Am. **92**, 978–986.

Thorne, P. D., and Hanes, D. M. (**2002**). "A review of acoustic measurement of small-scale sediment processes," Cont. Shelf Res. **22**, 603–632.

Thorne, P. D., and Hardcastle, P. J. (**1997**). "Acoustic measurements of suspended sediments in turbulent currents and comparison with in-situ samples," J. Acoust. Soc. Am. **101**, 2603–2614.

Thorne, P. D., Hayhurst, L., and Humphrey, V. F. (**1992**). "Scattering by non-metallic spheres," Ultrasonics **30**, 15–20.

Thorne, P. D., Manley, C., and Brimelow, J. (**1993**). "Measurements of the form function and total scattering cross section for a suspension of spheres," J. Acoust. Soc. Am. **93**, 243–248.

Thorne, P. D., and Meral, R. (**2008**). "Formulations for the scattering properties of suspended sandy sediments for use in the application of acoustics to sediment transport processes," Cont. Shelf Res. **28**, 309–317.

Thosteson, E. D., and Hanes, D. M. (**1998**). "A simplified method for determining sediment size and concentration from multiple frequency acoustic backscatter measurements," J. Acoust. Soc. Am. **104**, 820–830.

Vincent, C. E. (**2007**). "Measuring suspended sand concentration using acoustic backscatter: A critical look at the errors and uncertainties," in *Coastal and Shelf Sediment Transport*, edited by P. S. Balson and M. B. Collins (Geological Society of London, London), Vol. **274**, pp. 7–15.

Vincent, C. E., Hanes, D. M., and Bowen, A. J. (**1991**). "Acoustic measurements of suspended sand on the shoreface and the control of concentration by bed roughness," Mar. Geol. **96**, 1–18.

Whitehouse, R. (**1995**). "Observations of the boundary layer characteristics and the suspension of sand at a tidal site," Cont. Shelf Res. **15**, 1549–1567.

Young, R. A., Merrill, J. T., Clarke, T. L., and Proni, J. R. (**1982**). "Acoustic profiling of suspended sediments in the marine bottom boundary layer," Geophys. Res. Lett. **9**, 175–178.

Automatic recognition of fin and blue whale calls for real-time monitoring in the St. Lawrence

Xavier Mouy[a)]
Marine Sciences Institute, University of Québec at Rimouski, 310 Allée des Ursulines, Rimouski, Québec G5L-3A1, Canada

Mohammed Bahoura
Department of Mathematics, Informatics and Engineering, University of Québec at Rimouski, 300 Allée des Ursulines, Rimouski, Québec G5L-3A1, Canada

Yvan Simard
Marine Sciences Institute, University of Québec at Rimouski, 310 Allée des Ursulines, Rimouski, Québec G5L-3A1, Canada and Maurice Lamontagne Institute, Fisheries and Oceans Canada, 850 Route de la Mer, Mont-Joli, Québec G5H-3Z4, Canada

(Received 20 April 2009; revised 30 September 2009; accepted 6 October 2009)

Monitoring blue and fin whales summering in the St. Lawrence Estuary with passive acoustics requires call recognition algorithms that can cope with the heavy shipping noise of the St. Lawrence Seaway and with multipath propagation characteristics that generate overlapping copies of the calls. In this paper, the performance of three time-frequency methods aiming at such automatic detection and classification is tested on more than 2000 calls and compared at several levels of signal-to-noise ratio using typical recordings collected in this area. For all methods, image processing techniques are used to reduce the noise in the spectrogram. The first approach consists in matching the spectrogram with binary time-frequency templates of the calls (coincidence of spectrograms). The second approach is based on the extraction of the frequency contours of the calls and their classification using dynamic time warping (DTW) and the vector quantization (VQ) algorithms. The coincidence of spectrograms was the fastest method and performed better for blue whale A and B calls. VQ detected more 20 Hz fin whale calls but with a higher false alarm rate. DTW and VQ outperformed for the more variable blue whale D calls.
 [DOI: 10.1121/1.3257588]

PACS number(s): 43.30.Sf, 43.60.Bf [WWA] Pages: 2918–2928

I. Introduction

The Atlantic blue whale (*Balaenoptera musculus*) and fin whale (*Blaenoptera physalus*) are listed as endangered and of special concern, respectively, according to the Canadian species at risk act (Canada, 2008). The Northwest Atlantic population sizes are estimated in the low hundreds for the blue whale and between 3500 and 6300 for the fin whale (Perry *et al.*, 1999; Sears and Calambokidis, 2002). Among the threats for the survival of these species are collisions with ships (Laist *et al.*, 2001; Jensen and Silber, 2004) and the increasing low-frequency anthropogenic noise in the oceans (Andrew *et al.*, 2002; McDonald *et al.*, 2006), which is possibly affecting mysticetes, known to produce and probably perceive low-frequency sounds (Richardson *et al.*, 1995; Croll *et al.*, 2001; NRC, 2003; Simard *et al.*, 2008; Au and Hastings, 2008).

The St. Lawrence Estuary is an intensively used habitat of the North Atlantic fin and blue whale populations. This traditional feeding ground (Simard and Lavoie, 1999) is an important whale watching area (Hoyt, 2001). It is also part of an important North American seaway connecting the Great Lakes and the Atlantic. The impact that the anthropogenic activities might have on the survival of fin and blue whales summering in this area is not well known. Basic information on the long-term spatial and temporal distribution of the whales in the region is lacking but is a matter of concern. The development of efficient methods to continuously monitor the whales over the whole basin would be a contribution to fill this gap.

The emergence of passive acoustics monitoring (PAM) technology offers the possibility of addressing this challenging time-space distribution of whales over their extensive life domains in the oceans (Mellinger *et al.*, 2007). The manual detection of the vocalizations on recordings performed aurally and by manual inspection of spectrograms is a long and laborious task leading to an inconstant bias in the analysis dependent on the experience and the degree of fatigue of the operator. Therefore, the development of efficient and robust automatic detection and classification methods is required. Their performance is dependent on the complexity and diversity of the sounds to recognize and the acoustic properties of the environment such as the characteristics of the background noise and sound propagation in the area. Matched filters appear to be efficient for detecting stereotypical signals in low Gaussian noise conditions (Stafford *et al.*, 1998; Mellinger and Clark, 2000). The correlation of spectrograms

[a)]Author to whom correspondence should be addressed. Electronic mail: xavier.mouy@gmail.com

with a time-frequency kernel representing the call has been used on right whale (Mellinger, 2004; Munger *et al.*, 2005), and North Pacific blue whale calls (Mellinger and Clark, 2000; Wiggins *et al.*, 2005). These methods are well adapted for calls presenting low variability in duration and frequency. Chirplet transform shows an interesting potential for classifying blue whale calls (Bahoura and Simard, 2008). Artificial neural networks (ANNs) and hidden Markov modeling (HMM) show good recognition results on elephants (Clemins *et al.*, 2005), birds (Kogan and Margoliash, 1998), and whale calls (Potter *et al.*, 1994; Mellinger and Clark, 2000; Mellinger, 2004). These two methods have been tested using various attributes for characterizing the calls: Fourier coefficients (Mellinger, 2004), Mel-frequency cepstral coefficients (Clemins *et al.*, 2005), and linear predictive coefficients (Clemins and Johnson, 2006). ANN and HMM can accommodate variability within the calls. The time-frequency contour approach extracts the call shape attributes from the spectrogram (e.g., maximum frequency, minimum frequency, duration, etc.) to feed the classification algorithms. The contours can be extracted using an edge detector (Gillespie, 2004) or through instantaneous frequency tracking algorithms (Sturtivant and Datta, 1995a, 1995b; Brown *et al.*, 2006; Halkias and Ellis, 2006, 2008). Automatic contour extraction by tracking algorithms can be computationally expensive (Brown and Zhang, 1991). Classification methods used so far for this approach include discriminant analysis (Bazúa-Durán and Au, 2004; Gillespie, 2004) and dynamic time warping (DTW) (Buck and Tyack, 1993; Brown *et al.*, 2006). Urazghildiiev and Clark (2006, 2007) proposed another approach involving a generalized likelihood ratio test (GLRT) detector and a finite impulse response (FIR) filter operating on spectrograms, which outperformed the human operator for right whale calls.

The blue whale vocal repertoire in the St. Lawrence is similar to the one found in the North Atlantic. It is composed of three basic calls, the A, B, and D calls. (Edds, 1982; Mellinger and Clark, 2003; Berchok *et al.*, 2006). The A call is an 8-s tonal sound at 18 Hz. The B call is an 11-s down sweep from 18 to 15 Hz [Fig. 1(a)]. Although these two calls usually follow each other in AB phrases repeated every 74 s, it is possible to observe repetitions of A calls or B calls only (Mellinger and Clark, 2003; Berchok *et al.* 2006). The two calls are then separated by an ~5-s gap, but this short silence is sometimes absent. Berchok *et al.* (2006) called these calls the hybrid calls. These latter calls are hereafter referred to as *AB calls*. The D call is less stereotyped than the two previous ones and consists of a down sweep in the ~120–40 Hz frequency range. It sometimes starts with an upsweep forming an arch in the spectrogram. The duration, and the starting and ending frequencies are variable [Fig. 1(c)]. The D calls are not repeated in regular sequences as the A and B calls; however, some time patterns can be observed. As noticed by Berchok *et al.* (2006), the blue whale calls in the St. Lawrence are more variable than the ones reported in the North Atlantic. A lot of them are not continuous but segmented into several fragments. Berchok *et al.* (2006) believed that this segmentation is not due to propagation effects but rather a property of the whale call. The other rare blue

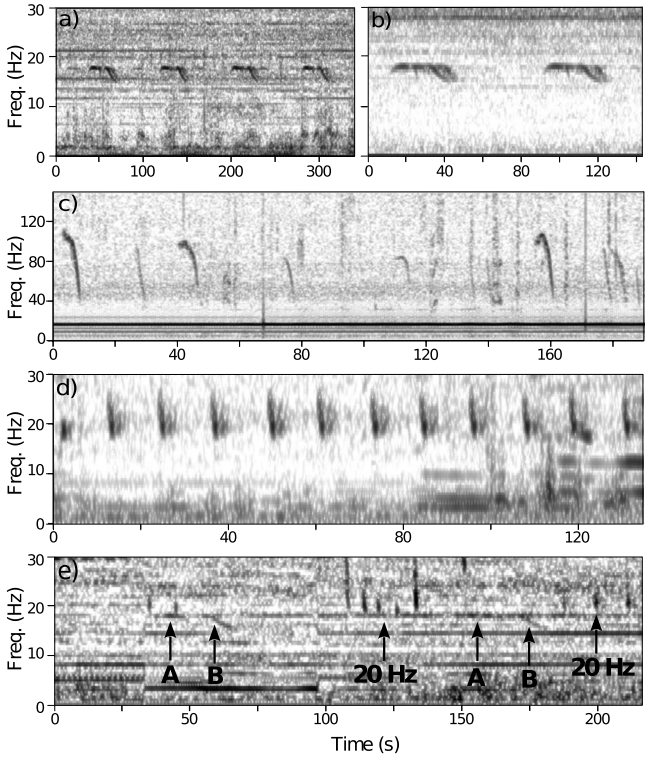


FIG. 1. Blue and fin whales calls. Spectrograms representing (a) 4 blue whale AB phrases, (b) 2 blue whale AB phrases with strong multipath echoes, (c) 11 very variable blue whale D calls, and (d) 12 fin whale 20 Hz calls with multipath echoes. (e) Example of a noisy recording containing blue and fin whale calls.

whale calls such as the 9 Hz sounds reported by Mellinger and Clark (2003), as well as the *blurps* and *grunts* reported by Berchok *et al.* (2006), were not considered here because of their low interest for monitoring applications.

The 20 Hz pulses are the most frequent fin whale calls recorded in the St. Lawrence (Samaran, 2004). They are similar to the ones reported in the other areas of the world (Schevill *et al.*, 1964; Thompson and Friedl, 1982; Edds, 1988; Thompson *et al.*, 1992), except for the occasional lagging lower frequency backbeat in the call bouts (Samaran, 2004; Simard and Roy, 2008). This 20 Hz call consists of a 1-s downsweep from 23 to 18 Hz [Fig. 1(d)]. The 140 Hz call reported by Edds (1988) and Samaran (2004) and the backbeat are less frequent and are therefore not considered in this paper.

Although several methods were already tested to automatically detect and classify fin and blue whale calls, none were thoroughly tested with the specific calls of the St. Lawrence Estuary and its particular acoustic conditions where intense low-frequency shipping noise in the vocalization band causes most recorded calls to have low signal-to-noise ratios (SNRs) [e.g., Fig. 1(e); Simard *et al.*, 2006a, 2008]. The presence of a sound channel in the upper half of the ~350-m water column, resulting from the summer cold intermediate layer and steep walls at the head of the deep Laurentian Channel, favors multipath propagation and the presence of echoes [e.g., Fig. 1(b); Simard and Roy, 2008]. Consequently, the calls often appear distorted and warped in time, which hinders the detection and classification algorithms.

TABLE I. Time-space coordinates of the acoustic recordings.

Station	Recording period	Latitude (°N)	Longitude (°W)	Hydrophone mean depth (m)	Bottom depth (m)
A	Aug.–Sept. 2003	48.2683	69.4663	125	130
B	Aug.–Sept. 2003	48.2687	69.4636	188	193
C	Sept.–Oct. 2003	48.3544	69.1167	54	88
D	Sept.–Oct. 2003	48.2504	69.2225	53	55
E	Sept.–Oct. 2003	48.1900	69.5925	43	60
F	Aug.–Oct. 2004	48.3783	69.3300	155	239
G	Aug.–Oct. 2004	48.1233	69.5433	126	135

This paper explores the capability of signal processing methods that could be implemented in real-time to detect fin and blue whale calls in an intense shipping noise background, and to cope with the call duration and frequency variability observed in the St. Lawrence. Two approaches based on time-frequency representations are tested. The classical spectrogram correlation method, previously used for blue and fin whale call detection, is adapted to the St. Lawrence acoustic conditions. The second approach, consisting in the extraction and classification of the call's time-frequency contours, tries to improve some weaknesses of this adapted classical approach. This work is part of the real-time monitoring project "Whale On the Web" (Simard *et al.*, 2006b).

II. Materials and Methods

This section describes the acoustic data used for this work and the various steps involved in the two classification approaches that were developed. The steps for the first approach are (a) the calculation of the spectrogram, (b) the noise removal, and (c) the coincidence of spectrograms. The steps for the second approach are (a) the calculation of the spectrogram, (b) the noise removal, (c) the extraction of the call contours, and (d) their classification. Two contour classification methods are investigated, dynamic time warping and vector quantization (VQ).

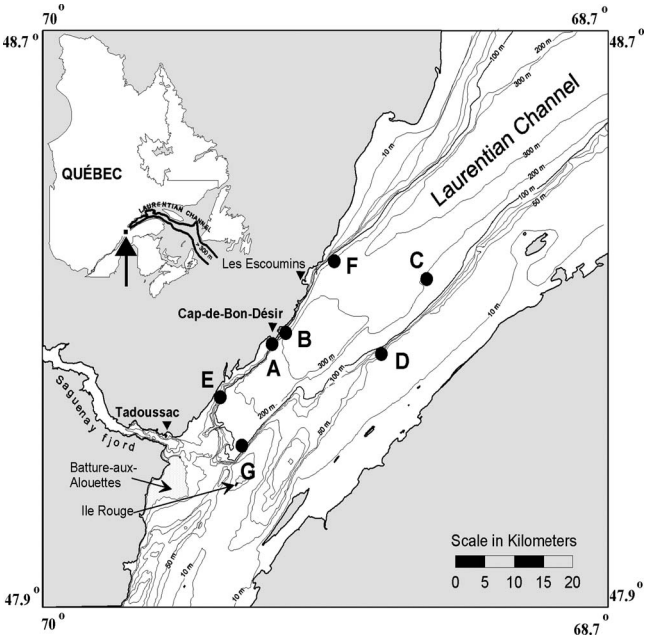


FIG. 2. Map of the study area with bathymetry and locations of the acoustic recording devices.

A. Data collection and data sets

All the data were collected in summer at seven locations in the lower St. Lawrence Estuary (Table I). Locations A and B are part of a hydrophone costal array deployed on the seabed at Cap-de-Bon-Desir in 2003 (Fig. 2; Simard and Roy, 2008). Locations C, D, E, F, and G had aural autonomous hydrophones (Multi-Electronique Inc., Rimouski, QC, Canada) moored around the sound channel depth in 2003 and 2004. The recordings were digitized at 16 bits using a sampling rate of 20 kHz for locations A and B and 2 kHz for the other locations. All recordings were decimated to 2000, 400, or 200 Hz for blue and fin whale call processing. Different sampling rates were used to get optimal time-frequency resolutions for each call type.

Among these recordings, 2472 calls were manually identified, labeled, and time stamped by an experienced observer (XM) through visual inspection of the spectrogram and aural validation when needed. A total of 2063 calls were used to test the performance of the algorithms while the other 409 calls were used for training the algorithms and adjusting their parameters (Table II). These calls were selected to represent the different noise and multipath echo conditions encountered in this part of the St. Lawrence. Calls were carefully chosen to ensure that data for testing and training were recorded at different locations and different times. To test the false alarm rate of the algorithms, 14 h of recordings containing different kinds of typical local noises but no calls were selected. This noise data set was made up of 90 wave files lasting 5–10 min each, collected at different locations of the estuary and containing intense low-frequency shipping noise. Files were usually selected at the closest point of approach of the ship (i.e., center of Lloyd's mirror "V-pattern," Urick, 1983). These recordings corre-

TABLE II. Number of calls of the training and test data sets by call type.

	A call	B call	D call	20 Hz call	Total
Training	117	91	97	104	409
Test	568	490	510	495	2063

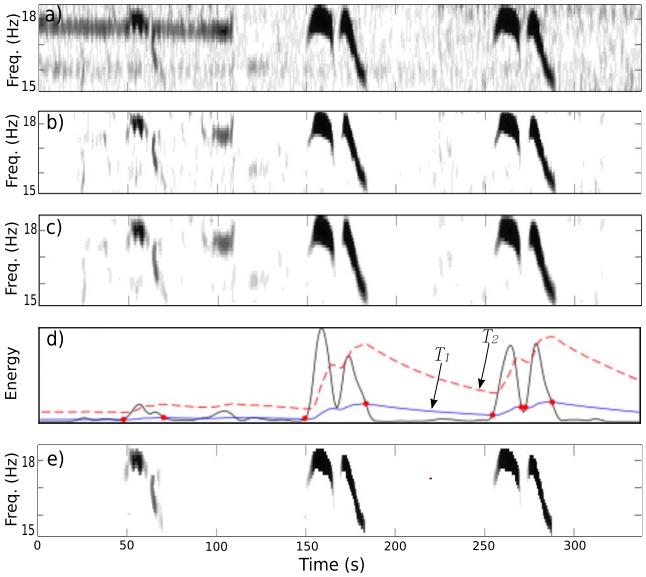


FIG. 3. (Color online) Sequence of steps of the noise reduction process. (a) The original spectrogram representing three blue whale AB phrases, time resolution=0.64 s, frequency resolution=0.13 Hz; (b) the spectrogram after equalization; and (c) Gaussian smoothing. (d) Representation of the energy e (bold line) and the adaptive thresholds T_1 (thin line) and T_2 (dashed line). The dots represent the intersections of e with T_1. (e) Final spectrogram after applying the thresholds.

spond to the most problematic conditions for the detection and classification algorithms.

B. Noise reduction

The time-frequency representation used short-time Fourier transform for both approaches [Fig. 3(a)]. The three noise reduction steps that were applied to both approaches consist of spectrogram equalization, Gaussian kernel smoothing, and adaptive energy thresholding. The spectrogram resolutions differed for each call type and method. First, the equalization consisted of subtracting the time-averaged spectrogram from the original spectrogram. The time-averaged spectrogram was calculated by applying a moving average of the energy values for each frequency band of the spectrogram matrix (Mellinger, 2004). Length of the moving average window, Δt, is reported in Tables III and IV. This reduced the long constant spectral rays produced by the fundamental rotational frequencies of the ship propulsion machinery (Arveson and Vendittis, 2000) and by mooring vibration [Fig. 3(b)]. Second, the spectrogram was smoothed using a Gaussian kernel (Gillespie, 2004). The spectrogram matrix was convolved with a two-dimensional Gaussian mask [Fig. 3(c)]. The size of the kernel, $K_x \times K_y$ (time $\times$ frequency), varied depending on call type and classification method (Tables III and IV). Non-square kernels smooth the spectrogram in a particular direction, improving the continuity of truncated calls along the time or frequency axis. Third, an adaptive energy threshold was applied to discard the parts of the spectrogram with locally non-significant energy. The energy at each time step n of the spectrogram was smoothed using a moving average to generate a smoothed energy curve $e(n)$ [bold line in Fig. 3(d)]. The moving average was performed with a 5-s window for the A and B calls and a 1-s window for the D and 20 Hz calls. Then, two adaptive thresholds T_1 and T_2 are defined, as in Renevey and Drygajlo, 2001.

$$T_1[n] = E[n] + \delta_1 E[n] \tag{1}$$

and

$$T_2[n] = E[n] + \delta_2 E[n], \tag{2}$$

where δ_1 and δ_2 are constants in the range [0,1], and where $E[n]$ is defined by

$$E[n] = \lambda E[n-1] + (1-\lambda)e[n]. \tag{3}$$

In Eq. (3), λ is the adaptive coefficient. The adaptive threshold T_2 allows detection of events while the crossing points of the two series T_1 and $e(n)$ are used to define the start and the end of these events [dots in Fig. 3(d)]. The spectrogram values outside the call windows are blanked. Finally, the pixels exceeding the adaptive threshold T_3 are retained at each time step

$$T_3[n] = \mu[n] + \delta_3 \sigma[n], \tag{4}$$

where μ and σ are the mean and the standard deviation of the spectrogram energy at the step n and δ_3 is a constant in the range [0,1]. The final result of this noise reduction process is shown in Fig. 3(e). Values of parameters δ_1, δ_2, δ_3, and λ were chosen empirically (Tables III and IV).

TABLE III. Values of the parameters used in the first approach (i.e., coincidence of spectrograms).

Parameter	A call	B call	D call	20 Hz call
Sample frequency (Hz)	200	200	200	200
		Spectrogram		
Frame size (points)	512	512	128	64
Overlap (%)	87.5	87.5	75	87.5
FFT size (points)	4608	4608	128	192
Frequency resolution (Hz)	0.04	0.04	1.56	1.04
Time resolution (s)	0.32	0.32	0.16	0.04
		Noise reduction		
Δt (s)	55	15	55	15
K_x (bins)	9	6	6	5
K_y (bins)	3	3	3	3
δ_1	−0.06	−0.06	−0.1	−0.06
δ_2	0.9	0.6	0.6	1.2
δ_3	0.92	0.85	0.92	0.92
λ	0.99	0.986	0.97	0.994
		Coincidence of spectrograms		
f_1 (Hz)	18.1	17.5	80	26
f_2 (Hz)	17.8	15.8	40	20
D_{call} (s)	7	9	3	1
Δf (Hz)	0.3	0.3	4	1
D_{blank} (s)	2	0.7	1	1
T_{CS} (%)	77	75	67	74

C. First approach: Coincidence of spectrograms

The coincidence of spectrograms method consists of finding a specific type of sound represented by a time-

TABLE IV. Values of the parameters used in the second approach (i.e., extraction and classification of call contours).

Parameter	A and B calls	D call	20 Hz call
Sample frequency (Hz)	2000	400	400
	Spectrogram		
Frame size (points)	4096	64	128
Overlap (%)	93	75	87.5
FFT size (points)	12 288	192	384
Frequency resolution (Hz)	0.16	2.08	1.04
Time resolution (s)	0.14	0.04	0.04
Minimum frequency (Hz)	15	30	18
Maximum frequency (Hz)	18.5	115	30
	Noise reduction		
Δt (s)	30	15	35
K_x (bins)	8	11	11
K_y (bins)	4	3	3
δ_1	−0.16	−0.16	−0.16
δ_2	0.7	0.5	0.9
δ_3	0.25	0.6	0.7
λ	0.996	0.992	0.993
	Contour extraction		
T_c	0.2	0.3	0.7
T_{con} (s)	6.5	2.5	0.5
T_{min} (s)	3.5	1	0.5
T_{max} (s)	35	9	3
	Dynamic time warping classification		
T_{DTW}	0.2	5	1.3
	Vector quantization classification		
T_{VQ}	1	18	4

frequency kernel (Mellinger and Clark, 1997, 2000). Various operations could be used to match the kernel with the spectrogram. Here the logical AND operation is used because of the very low computational cost of this operation (Tiemann *et al.*, 2004). The tested spectrogram is first binarized by setting the high-energy parts that exceed a threshold to 1 and the rest to 0. Then, the time-frequency kernel representing a candidate call template is defined by a synthetic binary image generated from the call starting frequency f_1, ending frequency f_2, duration D_{call}, frequency span Δf, and silence duration before and after the call D_{blank} (Table III). Finally the matching of the call kernel with the spectrogram is the arithmetic sum of the bits resulting from a logical AND between the binary kernel and spectrogram at each time step. A perfect match corresponds to a coincidence rate of 100%. The call is detected when the coincidence rate exceeds the empirically chosen threshold value T_{CS} (Table III).

D. Second approach: Extraction and classification of call contours

This second approach proceeds by extracting the call contours followed by their classification. Two classification algorithms are tested: DTW and VQ.

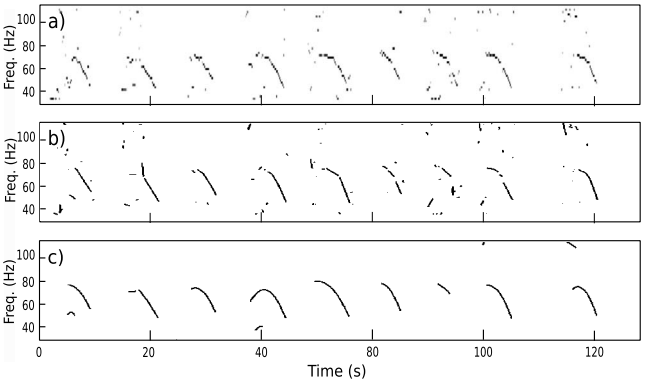


FIG. 4. Example of frequency contour extraction for a series of nine blue whale D calls. (a) Extraction of the local energy maxima, (b) connection of the maxima by time contiguity and frequency proximity, and (c) contours retained after connection of the call segments and duration restriction. The retained contours are approximated by a second-order polynomial.

1. Extraction of the contours

Extraction of time-frequency contours has been used with musical signals (Brown and Zhang, 1991; Brown, 1992), speech (Rabiner *et al.*, 1976; Rabiner, 1977), and odontocete calls (Sturtivant and Datta, 1995a, 1995b; Datta and Sturtivant, 2002; Brown *et al.*, 2006; Halkias and Ellis, 2006). Because real-time implementation is one constraint of the present work, the algorithm has to be optimized for both accurate contour extraction and short computation time. Drawing from Halkias and Ellis, 2006, the algorithm is divided into three steps: identification of the local energy maxima on the spectrogram, the delineation of call segments, and their conditional connection. First, the local energy maxima for each time step of the spectrogram are defined as the spectrum peaks [Fig. 4(a)]. Second, the maxima separated by less than four frequency bins in frequency at adjacent time steps are connected together to form segments [Fig. 4(b)]. To remove the irregularities of the extracted contour segments, their frequency time-series were smoothed with a moving average using a sliding window length of 3 s for the A and B calls, 1 s for the D call, and 0.1 s for the 20 Hz call. Third, the last step consists of connecting some of the isolated segments into more complete call contours. This connection is based on a stochastic model. The probability of connection is evaluated for each close segment pair. Each x_i pair is defined by its starting and ending slopes, α_{i1} and α_{i2}. Two frequency continuity distances β_{i2} and β_{i1} are measured for each possible linear connection; β_{i2} corresponds to the gap in frequency when the end of the first segment is linearly extended to the beginning of the second segment, and β_{i1} corresponds to the gap in frequency measured when the beginning of the second segment is linearly extended to the end of the first segment. The distance β_i is the minimum of these two distances. Two series of observations L_α and L_β representing the parameters α and β calculated on N pairs of connecting segments are extracted from the training database

$$L_\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{i1} & \alpha_{i2} \\ \vdots & \vdots \\ \alpha_{N1} & \alpha_{N2} \end{bmatrix} \quad \text{and} \quad L_\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_i \\ \vdots \\ \beta_N \end{bmatrix}.$$

Two normal distributions $\Theta_\alpha(\mu_\alpha, \Sigma_\alpha)$ and $\Theta_\beta(\mu_\beta, \sigma_\beta)$ are defined to represent the observations L_α and L_β, where Σ_α, μ_α, σ_β, and μ_β are, respectively, the covariance matrix and mean vector of L_α and the standard deviation and mean of L_β. The connection of an unknown call segment pair x is evaluated by the likelihood $P(x)$ defined by

$$P(x) = P(x|\Theta_\alpha)P(x|\Theta_\beta), \tag{5}$$

with $P(x|\Theta_\alpha)$ and $P(x|\Theta_\beta)$ being, respectively, the likelihood that the connection x can be generated by the models Θ_α and Θ_β. For convenience, the maximum of each of the two distributions is individually normalized to 1. Two segments are connected only if the likelihood $P(x)$ exceeds a threshold value T_c empirically adjusted by the operator (Table IV). Among all resulting valid possibilities, the connection proceeds with the segment that maximizes the likelihood. Once a connection is accepted, the two segments are connected by interpolation using an order 2 polynomial [Fig. 4(c)]. Connection is considered only for segments separated by less than the duration T_{con}. Contours shorter than the duration T_{min} or longer than the duration T_{max} are not considered for classification (Table IV).

2. Classification using DTW

The DTW algorithm is used by the speech processing community for isolated word recognition (Rabiner and Juang, 1993). It was applied in bioacoustics for the classification of dolphin sounds (Buck and Tyack, 1993), orca sounds (Brown *et al.*, 2006), and bird sounds (Ito *et al.*, 1996; Anderson *et al.*, 1996). The algorithm consists of classifying an unknown sound T by comparing it to a set R of k reference sounds from a dictionary, obtained from manually identified calls of the training data set. The T and R_k calls are represented by a feature vector composed of the instantaneous frequency f, the speed f', and the acceleration f'' of the call contour at each time step (Fig. 5). The feature vectors were not normalized.

$$T:[T[i];i = 1,2, \ldots ,I],$$

$$R_k:[R_k[j];j = 1,2, \ldots ,J_k],$$

where I and J_k represent the number of time frames of the T and R_k calls, respectively. The comparison of T with each template R_k is performed by calculating the dissimilarity (or distortion) between them based on their feature vectors. During this step, the algorithm has the particularity to take into consideration the possible compressions or extensions in time of the calls. A matrix X_k of dimension $I \times J_k$, is fitted with the Euclidean distances, $d(T[i],R_k[j])$ separating all the vectors $T[i]$ from $R_k[j]$ (Fig. 5). For the DTW algorithm, calculating the similarity between T and R_k consists of finding the shortest monotonic path w, of length C, within the matrix X_k, progressing from the point (1,1) to the point (I,J_k),

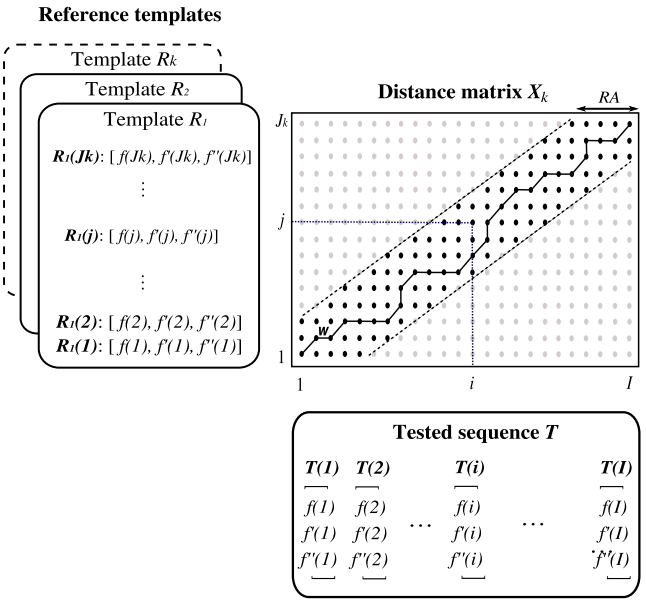


FIG. 5. (Color online) Scheme of the DTW comparison process for an unknown call T and call reference templates R_k. For each template R_k, a matrix X_k representing the Euclidian distances between the feature vectors of T and R_k is calculated. The dissimilarity between the two compared calls is represented by the path w going from the point (1,1) to (I,J_k) of the matrix X_k that minimizes the accumulated distance. The path w is found by calculating recursively the minimum cumulative distance for each point using local constraints. To avoid exaggerated distortions, the path search area is restricted by the two lines parallel to the diagonal of X_k.

$$w:[i(c),j(c)];c = 1,2, \ldots ,C,$$

with $i(1)=1$, $j(1)=1$, $i(C)=I$, and $j(C)=J_k$, such that its associated distance, $D_w(T,R_k)$, defined by

$$D_w(T,R_k) = \frac{\sum_{c=1}^{C} d\{T[i(c)],R_k[j(c)]\}g(c)}{N(g)}, \tag{6}$$

is the minimum. In Eq. (6), $g(c)$ is a weighting function and $N(g)$ is a normalization factor (Rabiner and Juang, 1993). Instead of calculating all the possible paths, the DTW algorithm recursively calculates the minimum cumulative distance for each point using local constraints (Vintsyuk, 1968; Myers *et al.*, 1980; Ney, 1984). The local constraint used in this study limits the connections to the right, the top, and the top right diagonal from the current point. To avoid exaggerated distortions and reduce computing time, a global constraint is applied by a restriction to the area defined by two lines parallel to the diagonal of the matrix X_k. The positions of these two lines are defined by a parameter RA such as $|i - j| \leq RA$ (Sakoe and Chiba, 1978) (Fig. 5). The parameter RA was empirically chosen to be 5 for all the whale call types. The unknown vocalization is classified as the reference template of the dictionary for which the distance $D_w(T,R_k)$ is the minimum.

$$T \triangleq \arg\min_k D_w(T,R_k). \tag{7}$$

The dictionary is composed of the blue whale A, B, D, and AB calls and fin whale 20 Hz call. Noise templates were not

included in the dictionary since noise is highly variable and cannot be extensively represented by a limited number of time-frequency contours. Therefore, extracted contours corresponding to noise events were classified as such when the minimum distance $D_w(T,R_k)$ exceeded the threshold T_{DTW}, defined empirically (Table IV).

The choice of the reference templates is known to be a critical phase of this contour classification method. Subjective choices of the reference templates by an operator usually do not perform well. Here we used a semi-supervised selection of the reference templates (Rabiner and Juang, 1993). For each call type, the operator first randomly assigns a call from the training data set as the reference template. Then, each contour of the training data set is compared to the chosen reference templates using the DTW algorithm. If the contour is properly classified it is then properly represented by the chosen template. If its classification is wrong, the tested contour is added to the call template dictionary. This routine is repeated for all calls of the training data set. This procedure removes the subjectivity and minimizes the redundancy of the dictionary. The final dictionary used in this study was composed of 12 A calls, 17 B calls, 7 AB phrases, 17 D calls, and 2 20 Hz pulses.

3. Classification using VQ

VQ is a clustering/classification algorithm that was used for speaker identification (Pan *et al.*, 1985; Soong *et al.*, 1985), respiratory sound classification (Bahoura and Pelletier, 2003), and hand writing recognition (Camastra and Vinciarelli, 2001). It is comprised of two steps: the training and the classification. During the training step, each contour extracted from the training database is represented by a set of features. Here, these features are its minimal frequency, maximum frequency, duration, and frequency shift. Thereby, all the calls are characterized by a set of four-dimensional vectors that can be represented in a four-dimensional feature space. The objective of the training is to represent the cloud of scattered points for each call class, by a small set, C_k (also called codebook) of local centroids, c_{ki}. To do so, the Linde–Buzo–Gray (LBG) algorithm is used (Linde *et al.*, 1980). In contrast to the well known *k -means* algorithm, the initialization of the LBG algorithm is not performed randomly. During the classification step, the set of features from an unknown call contour T is extracted and represented in the feature space. Then, the distance $D(T,C_k)$ separating T from each codebook C_k is calculated by

$$D(T,C_k)=\frac{1}{M}\sum_{i=1}^{M} d(T,c_{ki}), \tag{8}$$

where M is the number of centroids in the codebook C_k, and $d(T,c_{ki})$ is the Euclidean distance separating T from the centroids c_{ki}. The call T is classified as the class for which the distance $D(T,C_k)$ is minimum.

$$T \triangleq \arg\min_k D(T,C_k). \tag{9}$$

Codebooks were created using all the vocalizations of the training data set. The same 5 call classes used for the DTW method were considered and each codebook was represented by 16 centroids ($M=16$). To avoid classifying noise, the tested time-frequency contour is discarded if its distance $D(T,C_k)$ exceeds the empirically determined threshold T_{VQ} (Table IV).

E. Performance evaluation

The false negatives and false positives were counted for each method by running the algorithms on the test data sets. The false negatives correspond to the calls in the test data set that are not detected or that are considered as noise by the classification process. For a given call class, the false negative rate is defined as the ratio of the number of missed calls to the total number of calls in that class. The false positives correspond to periods of noise classified as calls (false alarms). It is calculated by testing each method on the noise data set. The number of false alarms divided by the duration of the noise data set (14 h) gives an average number of false positives per hour. The performance results are presented for several SNR intervals. The SNR calculation is inspired by Mellinger, 2004 and Mellinger and Clark, 2006. The call power is defined on the spectrogram by the average power of the call in the 15–18 Hz frequency range for the A and B calls, 18–26 Hz for the 20 Hz pulses, and 30–100 Hz for the D calls. The noise power is calculated by the average power before and after the call, in the same frequency band for periods equal to the call durations. This calculation is done on the original spectrogram (i.e., without noise reduction). The performances are reported separately for SNR intervals <0 dB, [0, 5 dB], [5, 10 dB], and >10 dB.

To quantify the relative processing time of each method, a real-time index I_r was defined as

$$I_r=\frac{T_p}{T_r}, \tag{10}$$

where T_p is the duration of the detection and classification process, and T_r is the recording duration. The index I_r is <1 if the processing is faster than the real-time and ≥ 1 otherwise. Processing was done with an Intel Pentium® 4 processor working at 3.2 GHz with 1 Gbyte of random access memory (RAM). The computer was running the Microsoft Windows XP® Professional operating system and the processing was performed in MATLAB© (MathWorks, Inc., Natick, MA). The real-time index was computed for each recording of the test data set and analyzed by call type. The A and B calls were processed simultaneously on the same spectrogram and their computation time could not be separated.

III. Results

A. Performance of classification

The test data set represents an accumulated recording of 17 h of A calls, 17 h of B calls, 7 h of D calls, and 3 h of 20 Hz calls. Their respective SNR distribution is given in the first column of Fig. 6. The A and B calls with a SNR between 0 and 10 dB were the most frequent in the data set. The 20 Hz pulses were equally distributed in the SNR intervals >0 dB. The majority of D calls had a SNR of 0–5 dB.

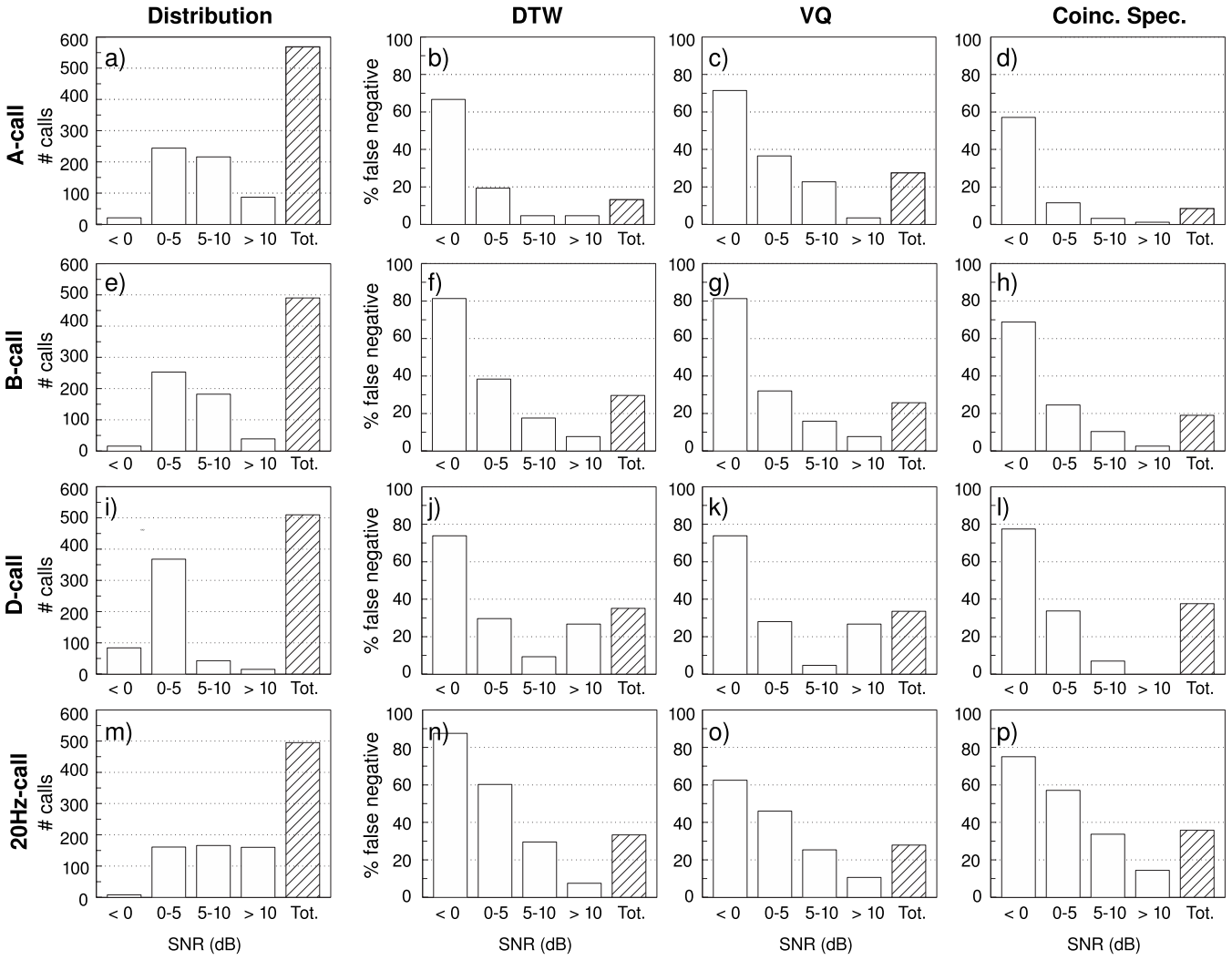


FIG. 6. Signal-to-noise ratio distributions of the (a) A calls, (e) B calls, (i) D calls, and (m) 20 Hz calls used to test the performance of the algorithms. False negative rates for the frequency contour [(b), (f), (j), and (n)] DTW and [(c), (g), (k), and (o)] VQ methods, and [(d), (h), (l), and (p)] the coincidence of spectrograms method, separately for four signal-to-noise ratio intervals and for all calls (hatched bars).

False negative rates for A and B calls decreased when the SNR increased for all methods as expected [Figs. 6(b)–6(d) and 6(f)–6(h)]. The coincidence of spectrograms had the smallest false negative rates with overall values of 8.5% for the A calls and 19% for the B calls. When tested on the noise data set, the three methods had comparable false detection rates for these two call types; coincidence of spectrograms (3 and 2 false alarms per hour for calls A and B, respectively) and DTW (4 and 1 false alarms per hour for calls A and B, respectively) having less false detections than VQ (Table V). For the D calls, VQ had a lower average false negative rate (33.5%) than the other methods [Figs. 6(j)–6(l)]. DTW and VQ showed unexpected increases in false negative rates (26.7%) at SNR > 10 dB. These two methods resulted in 11 and 12 false detections per hour, respectively, compared to 22 for the coincidence of spectrograms method (Table V). For the 20 Hz pulses all three methods presented a gradual decrease in false negative rate with increasing SNR [Figs. 6(n)–6(p)]. Lowest global false negative rates were obtained by VQ and DTW. These methods had, however, a very high number of false alarms per hour (63 with DTW and 74 with VQ, Table V).

TABLE V. Average number of false positives per hour for the three methods tested with the noise data set.

Method	A call	B call	D call	20 Hz call
DTW	4	1	11	63
VQ	4	3	12	74
Coincidence of spectrograms	3	2	22	23

B. Processing time

The coincidence of spectrograms method had the lowest computation time index of the three methods with average values of 0.0083, 0.0039, and 0.013 for the A and B calls, D calls, and 20 Hz calls, respectively [Fig. 7(a)]. The processing time of the VQ method was the same as DTW for the AB and 20 Hz calls with average real-time index value smaller than 0.1. However, VQ real-time index for the D call (average of 0.28) was significantly faster than DTW (average of 0.44) [Figs. 7(b) and 7(c)].

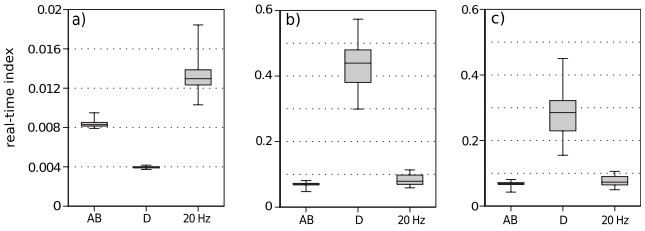


FIG. 7. Box–Whisker plots of the real-time index computation time of the (a) coincidence of spectrograms, (b) DTW, and (c) VQ methods for each call type. Note the different scale in (a).

IV. DISCUSSION

The coincidence of spectrogram method showed the best performance for the A and B calls. For these calls, the different methods differed mainly by their false negative rates. These missed calls were generally resulting from the contour extraction step. Some call contours were truncated into several non-connecting contour segments. The smallest segments were ignored because of the time constraint parameter but the other ones were presented to the classification step and they were classified as noise because of their excessively fragmentary representation of the actual call contour. For the 20 Hz calls, the contour approach, especially VQ, performed better for the false negative criteria, but produced three times more false alarms than the coincidence of spectrogram method. The majority of these false alarms were caused by the low-frequency impulsive noises usually coming from nearby transiting merchant ships that were not filtered out at the noise reduction step. The contour extraction approach best performed for the recognition of D calls. The fixed template used by the coincidence of the spectrograms appeared inappropriate to cope with the variations in duration, frequency sweep rate, and frequency band of the D call. The coincidence of spectrogram method was also more vulnerable to false alarms due to occasional strumming noise in this frequency band from the hydrophone mooring system. As expected, the false negative rates generally increased with decreasing SNRs, except for the D call with VQ and DTW for SNRs >10 dB. This exception appeared to result from the distortions of the calls caused by multipath propagation generating overlapping echoes of the calls that modified the extracted classification features. The low number of D calls in this SNR interval (15 calls) does not really allow a definitive conclusion about the performance of the different algorithms at high SNRs.

Computation costs depend on call frequency range (85 Hz for D calls versus 3.5 Hz for A and B calls) for time-frequency contour extraction, and on spectrogram resolution for the coincidence of spectrogram. DTW requires more computation than VQ and its cost is function of the number of call templates used. Recording sampling frequency differs according to call types and methods (Tables III and IV) and therefore also influences the computation times.

The classical method of coincidence of spectrograms was adapted to the noisy acoustic condition of the St. Lawrence by adding an initial step of noise reduction that was critical for successful classification. The simplicity of the logical AND operation makes this algorithm very fast, which fulfills the real-time requirements of this study. Choosing the optimal call templates empirically can be laborious; thus performing statistical analysis of the acoustic parameters of the calls to produce the kernel is recommended. The additional computation time for the contour approach is variable and depends on the complexity of the recordings' content. However, the calls considered in this paper have a fairly simple structure without harmonics, so the computing time remains relatively small. The training steps to determine the parameters for connecting the contour segment and building the dictionary of call models were easy and robust since they were semi-supervised on a representative call database. This study shows that this method has a promising potential for the recognition of marine mammal calls. Its performance is still slightly lower than the classical approach for the stereotyped calls; however, several modifications could be implemented to improve it. To reduce DTW computation time, some clustering techniques can be used to minimize the number of reference templates to consider for the same call type. Representing all calls with the same time-frequency scheme independently of their duration and frequency band would also improve all methods, which presently required computing spectrograms with different resolutions. It may be worth investigating other time-frequency representation techniques such as the Wigner–Ville transform (Caimi and Hassan, 2000). The number of parameters involved in the methods can be cumbersome to manually tune to optimize the classification results. Statistical optimization methods such as the *design of experiment* (Lochner and Matar, 1990) would be a simpler and more accurate way of finding the optimal set of parameter to maximize the algorithm performance. The algorithm could incorporate an additional step to identify and filter out or exploit the frequent delayed copies of the calls from reverberations. Use of propagation models and matched field techniques (e.g., Thode *et al.*, 2000) is among possible methods. This approach is computer intensive and attached to a particular environment.

What is the effect of the observed detection/classification results for PAM application in the St. Lawrence Estuary given fin and blue whale call occurrence frequency? These animals are known to produce very regular repeated calls (Mellinger and Clark, 2003; Samaran, 2004; Berchok *et al.*, 2006), although nonregular patterns are sometimes observed (Oleson *et al.*, 2007). The A and B calls are usually repeated every ~74 s (Mellinger and Clark, 2003). If a blue whale is calling for 1 h without interruptions, it would produce about 51 A and B calls. Referring to Table V and Fig. 6, the DTW classification method would be expected to detect 48 real A calls with a SNR above 5 dB, and 4 false alarms. The same reasoning is followed to assess the detections expected by the different methods and for the other calls (Table VI). By assuming a swimming speed between 1 and 5 m s^{-1} (Goldbogen *et al.*, 2006), all methods correctly detect enough calls for accurately tracking the animals producing A and B calls with a SNR higher than 5 dB. However, for the 20 Hz calls, only the coincidence of spectrograms appears to be suitable at low SNRs. The false alarms could be rejected using a tracking algorithm taking advantage of detection at

TABLE VI. Number of A, B, and 20 Hz calls expected and detected per recording hour, assuming that the animal calls regularly without interruptions, for the three tested algorithms.

Call type	Method	No. of expected calls	No. of correct detections SNR (dB) <0	0–5	5–10	>10	No. of false alarms
A call	DTW	51	17	41	48	48	4
	VQ	51	14	32	39	49	4
	Coincidence of spectrograms	51	21	45	49	50	3
B call	DTW	51	9	31	42	47	1
	VQ	51	9	34	42	47	3
	Coincidence of spectrograms	51	15	38	45	49	2
20 Hz call	DTW	360	45	143	253	333	63
	VQ	360	135	194	269	322	74
	Coincidence of spectrograms	360	90	154	238	308	23

different locations by several instruments differently affected by noise from transiting ships (Seebaruth, 2006; Simard *et al.*, 2008). It is to be noticed that the values in Table V assume a uniform distribution of the false alarms in time, which is not the case in the real world. Because the D calls are not repeated on predictable patterns, it is not possible to follow the same reasoning for this call type.

For assessing the actual density of calls per unit of time or trying to assess the whale density from the calls, the performance of the call detection algorithm under different noise conditions must be taken into account. The relation of the missed detection rate with the SNR level observed with the different methods could then serve to correct the "apparent call density" measured.

ACKNOWLEDGMENTS

This work has been founded by the FQRNT Québec research grant to M.B. and Y.S., Fisheries and Oceans Canada, and NSERC Discovery and Ship time grants to Y.S. We thank the scientific teams and the crews of the RV Coriolis II, CCGS Isle Rouge, and Cap d'Espoir for their generous contribution to the work at sea. Thanks to Professor Jean Rouat (Sherbrooke University) for his very constructive comments on this project. Many thanks to the two anonymous reviewers for their contribution in improving the clarity and rigorousness of this paper. Thanks to Terry Deveau for refining the English.

Anderson, S. E., Dave, A. S., and Margoliash, D. (**1996**). "Template-based automatic recognition of birdsong syllables from continuous recordings," J. Acoust. Soc. Am. **100**, 1209–1219.

Andrew, R. K., Howe, B. M., Mercer, J. A., and Dzieciuch, M. A. (**2002**). "Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast," ARLO **3**, 65–70.

Arveson, P. T., and Vendittis, D. J. (**2000**). "Radiated noise characteristics of a modern cargo ship," J. Acoust. Soc. Am. **107**, 118–129.

Au, W. L., and Hastings, M. C. (**2008**). *Principles of Marine Bioacoustics* (Springer, New York).

Bahoura, M., and Pelletier, C. (**2003**). "New parameters for respiratory sound classification," Canadian Conference on Electrical and Computer Engineering, 2003. IEEE CCECE 2003, Vol. **3**, pp. 1457–1460.

Bahoura, M., and Simard, Y. (**2008**). "Chirplet transform applied to simulated and real blue whale (Balaenoptera musculus) calls," in *Image and Signal Processing*, Lecture Notes in Computer Science Vol. **5099** (Springer, Berlin), pp. 296–303.

Bazúa-Durán, C., and Au, W. L. (**2004**). "Geographic variations in the whistles of spinner dolphins (*Stenella longirostris*) of the Main Hawaiian Islands," J. Acoust. Soc. Am. **116**, 3757–3769.

Berchok, C. L., Bradley, D. L., and Gabrielson, T. B. (**2006**). "St. Lawrence blue whale vocalizations revisited: Characterization of calls detected from 1998 to 2001," J. Acoust. Soc. Am. **120**, 2340–2354.

Brown, J. C., and Zhang, B. (**1991**). "Musical frequency tracking using the methods of conventional and 'narrowed' autocorrelation," J. Acoust. Soc. Am. **89**, 2346–2354.

Brown, J. C. (**1992**). "Musical fundamental frequency tracking using a pattern recognition method," J. Acoust. Soc. Am. **92**, 1394–1402.

Brown, J. C., Hodgins-Davis, A., and Miller, P. J. O. (**2006**). "Classification of vocalizations of killer whales using dynamic time warping," J. Acoust. Soc. Am. **119**, EL34–EL40.

Buck, J. R., and Tyack, P. L. (**1993**). "A quantitative measure of similarity for Tursiops truncatus signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506.

Canada (**2008**). "Species at risk public registry," http://www.registrelep.gc.ca/ (Last viewed 4/20/2009).

Caimi, F. M., and Hassan, G. A. (**2000**). "Pattern classification approach to underwater acoustic communications based on the Wigner-Ville distribution," Proc. SPIE **4045**, 167–174.

Camastra, F., and Vinciarelli, A. (**2001**). "Cursive character recognition by learning vector quantization," Pattern Recogn. Lett. **22**, 625–629.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2005**). "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," J. Acoust. Soc. Am. **117**, 956–963.

Clemins, P. J., and Johnson, M. T. (**2006**). "Generalized perceptual linear prediction features for animal vocalization analysis," J. Acoust. Soc. Am. **120**, 527–534.

Croll, D. A., Clark, C. W., Calambokidis, J., Ellison, W. T., and Tershy, B. R. (**2001**). "Effect of anthropogenic low-frequency noise on the foraging ecology of Balaenoptera whales," Anim. Conserv. **4**, 13–27.

Datta, S., and Sturtivant, C. R. (**2002**). "Dolphin whistle classification for determining group identities," Signal Process. **82**, 251–258.

Edds, P. (**1982**). "Vocalizations of the blue whale, *Balaenoptera musculus*, in the St. Lawrence River," J. Mammal. **63**, 345–347.

Edds, P. (**1988**). "Characteristics of finback (*Balaenoptera physalus*) vocalizations in the St. Lawrence estuary," Bioacoustics **1**, 131–149.

Gillespie, D. (**2004**). "Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram," Can. Acoust. **32**, 39–47.

Goldbogen, J. A., Calambokidis, J., Shadwick, R. E., Oleson, E. M., McDonald, M. A., and Hildebrand, J. A. (**2006**). "Kinematics of foraging dives and lunge-feeding in fin whales," J. Exp. Biol. **209**, 1231–1244.

Halkias, X. C., and Ellis, D. P. (**2006**). "Call detection and extraction using Bayesian inference," Appl. Acoust. **67**, 1164–1174.

Halkias, X. C., and Ellis, D. P. (**2008**). "A comparison of pitch extraction methodologies for dolphin vocalization," Can. Acoust. **36**, 74–80.

Hoyt, E. (**2001**). *Whale Watching 2001: Worldwide Tourism Numbers, Expenditures and Expanding Socioeconomic Benefits* (International Fund for Animal Welfare, Yarmouth Port, MA).

Ito, K., Mori, K., and Iwasaki, S. (**1996**). "Application of dynamic programming matching to classification of budgerigar contact calls," J. Acoust. Soc. Am. **100**, 3947–3956.

Jensen, A., and Silber, G. K. (**2004**). "Large whale ship strike database," NOAA Technical Memorandum NMFS-OPR, U.S. Department of Commerce, Silver Spring, MA.

Kogan, J. A., and Margoliash, D. (**1998**). "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," J. Acoust. Soc. Am. **103**, 2185–2196.

Laist, D. W., Knowlton, A. R., Mead, J. G., Collet, A. S., and Podesta, M. (**2001**). "Collision between ships and whales," Marine Mammal Sci. **17**, 35–75.

Linde, Y., Buzo, A., and Gray, R. (**1980**). "An algorithm for vector quantizer design," IEEE Trans. Commun. **28**, 84–95.

Lochner, R. H., and Matar, J. E. (**1990**). *Designing for Quality: An Intro-*

duction to the Best of Taguchi and Western Methods of Statistical Experimental Design (Quality Resources, New York).
McDonald, M. A., Hildebrand, J. A., and Wiggins, S. M. (**2006**). "Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California," J. Acoust. Soc. Am. **120**, 711–718.
Mellinger, D. K. (**2004**). "A comparison of methods for detecting right whale calls," Can. Acoust. **32**, 55–65.
Mellinger, D. K., and Clark, C. W. (**1997**). "Methods for automatic detection of mysticete sounds," Mar. Freshwater Behav. Physiol. **29**, 163–181.
Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529.
Mellinger, D. K., and Clark, C. W. (**2003**). "Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic," J. Acoust. Soc. Am. **114**, 1108–1119.
Mellinger, D. K., and Clark, C. W. (**2006**). "MobySound: A reference archive for studying automatic recognition of marine mammal sounds," Appl. Acoust. **67**, 1226–1242.
Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P., and Matsumoto, H. (**2007**). "An overview of fixed passive acoustic observation methods for cetaceans," Oceanogr. **20**, 36–45.
Munger, L., Mellinger, D., Wiggins, S., Moore, S., and Hildebrand, J. (**2005**). "Performance of spectrogram correlation in detecting right whale calls in long-term recordings from the Bering Sea," Can. Acoust. **33**, 25–34.
Myers, C., Rabiner, L., and Rosenberg, A. (**1980**). "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition," IEEE Trans. Acoust., Speech, Signal Process. **28**, 623–635.
Ney, H. (**1984**). "The use of a one-stage dynamic programming algorithm for connected word recognition," IEEE Trans. Acoust., Speech, Signal Process. **32**, 263–271.
NRC (**2003**). *Ocean Noise and Marine Mammals* (The National Academy Press, Washington, DC).
Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of call production by eastern North Pacific blue whales," Mar. Ecol.: Prog. Ser. **330**, 269–284.
Pan, K.-C., Soong, F., and Rabiner, L. (**1985**). "A vector-quantization-based preprocessor for speaker-independent isolated word recognition," IEEE Trans. Acoust., Speech, Signal Process. **33**, 546–560.
Perry, S. L., DeMaster, D. P., and Silber, G. K. (**1999**). "The great whales: History and status of six species listed as endangered under the U.S. Endangered Species Act of 1973," Mar. Fish. Rev. **61**, 1–82.
Potter, J. R., Mellinger, D. K., and Clark, C. W. (**1994**). "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262.
Rabiner, L., Cheng, M., Rosenberg, A., and McGonegal, C. (**1976**). "A comparative performance study of several pitch detection algorithms," IEEE Trans. Acoust., Speech, Signal Process. **24**, 399–418.
Rabiner, L. (**1977**). "On the use of autocorrelation analysis for pitch detection," IEEE Trans. Acoust., Speech, Signal Process. **25**, 24–33.
Rabiner, L., and Juang, B. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).
Renevey, P., and Drygajlo, A. (**2001**). "Entropy based voice activity detection in very noisy conditions," in Proceedings of the 7th European Conference on Speech Communication and Technology EUROSPEECH '2001, Aalborg, Denmark, pp. 1887–1890.
Richardson, W. J., Greene, C. J., Malme, C., and Thomson, D. (**1995**). *Marine Mammals and Noise* (Academic, San Diego, CA).
Samaran, F. (**2004**). "Détectabilité des vocalisations de rorquals communs (*Balaenoptera physalus*) à partir d'une station côtière dans la voie maritime de l'estuaire du Saint-Laurent [Detectability of fin whale calls (*Balaenoptera physalus*) from a coastal array in the St. Lawrence Estuary]," MSc thesis, University of Québec in Rimouski, QC, Canada.
Sakoe, H., and Chiba, S. (**1978**). "Dynamic programming algorithm optimization for spoken word recognition," IEEE Trans. Acoust., Speech, Signal Process. **26**, 43–49.
Schevill, W., Watkins, W., and Backus, R. (**1964**). "The 20-cycle signal and Balaenoptera (fin whales)," in *Marine Bioacoustics*, edited by W. N. Tavolga, Vol. **1** (Pergamon, Oxford), pp. 147–152.
Sears, R., and Calambokidis, J. (**2002**). "Update COSEWIC status report on the blue whale Balaenoptera musculus in Canada," COSEWIC assessment and update status report on the blue whale *Balaenoptera musculus* in Canada, Committee on the Status of Endangered Wildlife in Canada, Ottawa, Canada, http://www.registrelep.gc.ca/virtual_sara/files/cosewic/sr%5Fblue%5Fwhale%5Fe%2Epdf (Last viewed 4/20/2009).
Seebaruth, D. (**2006**). "Algorithmes de localisation de baleines pour le Saint-Laurent (Localization algorithms for the whales of the St. Lawrence)," MSc thesis, Departement de Génie Électrique et de Génie Informatique (GEGI), University of Sherbrooke, QC, Canada.
Simard, Y., Bahoura, M., Park, C. W., Rouat, J., Sirois, M., Mouy, X., Seebaruth, D., Roy, N., and Lepage, R. (**2006b**)."Development and experimentation of a satellite buoy network for real-time acoustic localization of whales in the St. Lawrence," in *OCEANS'06 MTS/IEEE*, Boston, MA (IEEE, Piscataway, NJ), pp. 1–6.
Simard, Y., and Lavoie, D. (**1999**). "The rich krill aggregation of the Saguenay–St. Lawrence Marine Park: Hydroacoustic and geostatistical biomass estimates, structure, variability, and significance for whales," Can. J. Fish. Aquat. Sci. **56**, 1182–1197.
Simard, Y., and Roy, N. (**2008**). "Detection and localization of blue and fin whales from large-aperture autonomous hydrophone arrays: A case study from the St. Lawrence estuary," Can. Acoust. **36**, 104–110.
Simard, Y., Roy, N., and Gervaise, C. (**2006a**). "Shipping noise and whales: World tallest ocean liner vs largest animal on earth," in *OCEANS'06 MTS/IEEE*, Boston, MA (IEEE, Piscataway, NJ), pp. 1–6.
Simard, Y., Roy, N., and Gervaise, C. (**2008**). "Passive acoustic detection and localization of whales: Effects of shipping noise in Saguenay–St. Lawrence Marine Park," J. Acoust. Soc. Am. **123**, 4109–4117.
Soong, F., Rosenberg, A., Rabiner, L., and Juang, B. (**1985**). "A vector quantization approach to speaker recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '85, Vol. **10**, pp. 387–390.
Stafford, K. M., Fox, C. G., and Clark, D. S. (**1998**). "Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean," J. Acoust. Soc. Am. **104**, 3616–3625.
Sturtivant, C. R., and Datta, S. (**1995a**). "Techniques to isolate dolphin whistles and other tonal sounds from background noise," Acoust. Lett. **18**, 189–193.
Sturtivant, C. R., and Datta, S. (**1995b**). "The isolation from background noise and characterization of bottlenose dolphin (Tursiops truncatus) whistles," J. Acoust. Soc. India **23**, 199–205.
Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). "Matched field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 278–286.
Tiemann, C. O., Porter, M. B., and Fraser, L. N. (**2004**). "Localization of marine mammals near Hawaii using an acoustic propagation model," J. Acoust. Soc. Am. **115**, 2834–2843.
Thompson, P. O., and Friedl, W. A. (**1982**). "A long term study of low frequency sound from several species of whales off Oahu, Hawaii," Cetology **45**, 1–19.
Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). "20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.
Urazghildiiev, I. R., and Clark, C. W. (**2006**). "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," J. Acoust. Soc. Am. **120**, 1956–1963.
Urazghildiiev, I. R., and Clark, C. W. (**2007**). "Acoustic detection of North Atlantic right whale contact calls using spectrogram-based statistics," J. Acoust. Soc. Am. **122**, 769–776.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Vintsyuk, T. K. (**1968**). "Speech discrimination by dynamic programming," Kibernetika **4**, 81–88.
Wiggins, S. M., Oleson, E. M., McDonald, M. A., and Hildebrand, J. A. (**2005**). "Blue whale (*Balaenoptera musculus*) diel call patterns offshore of Southern California," Aquat. Mamm. **31**, 161–168.

Eigenmodal resonances of polydisperse bubble systems on a rigid boundary

Suhith Illesinghe and Andrew Ooi[a)]
Department of Mechanical Engineering, The University of Melbourne, Parkville, Victoria 3010, Australia

Richard Manasseh
Fluid Dynamics Group, CSIRO Materials Science and Engineering, Highett, Victoria 3190, Australia

(Received 5 March 2009; revised 15 July 2009; accepted 5 October 2009)

This paper presents theory and experimental data on the resonance frequency of systems consisting of different-sized air bubbles attached to a rigid wall. Effects of the change in resonant frequency with bubble size and distance between the bubbles were studied. It was found that the symmetric mode resonance frequency of the bubble system decreased with increasing $r=R_{02}/R_{01}$, where R_{01} and R_{02} are the equilibrium radii of bubbles in the system. Both the symmetric and antisymmetric modes of oscillation were detected in the experiments, with the resonant frequency of the symmetric mode dominant at small bubble separation and the frequency of the antisymmetric mode dominant when the bubbles were farther apart. A linear coupled-oscillator theoretical model was used to describe the oscillations of the bubble system, in which the method of images was used to approximate the effects of the wall. It was found that there was fair to good agreement between the predictions of the coupled-oscillator model with the experimental data.
 [DOI: 10.1121/1.3257581]

PACS number(s): 43.35.Bf, 43.20.Fn, 43.20.Px, 43.20.Bi [RR] Pages: 2929–2938

I. INTRODUCTION

The expansion and compression of gas bubbles within a liquid medium was first studied theoretically by Lord Rayleigh[1] in 1917. The mathematical model originally considered by Rayleigh has been modified over the years by many researchers.[2–4] One such model is the nonlinear Rayleigh–Plesset equation which like Rayleigh's original derivation assumes that the bubbles remain spherical. It has been well-known that the nonlinear dynamics of the Rayleigh–Plesset equation is only relevant for small (micron sized) bubbles. Large (millimeter sized) bubbles typically undergo very small oscillations, thus experiencing simpler dynamics governed by a linear second order ordinary differential equations, derived by linearizing the Rayleigh–Plesset equation. The natural frequency of this linearized model is commonly used to estimate the resonant frequency of an isolated bubble oscillating in response to acoustic excitation. This frequency, commonly known as the Minnaert's frequency, was also independently derived by Minnaert.[5]

The early studies mentioned above only considered cases where the bubbles were assumed to exist in isolation. More recently, systems consisting of many interacting bubbles have been studied by various investigators.[6–15] A short review of some recent studies can be found in the article by Manasseh and Ooi.[16] As one might expect, bubbles can interact acoustically with each other, causing the natural frequencies of the system to change from that of isolated bubbles, and it was found that the interaction has a greater effect when the bubbles are closer. (These modifications to the system frequencies are linear effects; it had been known for some time that nonlinearities can cause interacting bubbles to move, a phenomenon called the secondary Bjerknes effect, e.g., Ref. 17.) In order to model the linear interaction, Feuillade[8] used the coupled oscillator approach by modeling the system as a coupled set of ordinary differential equations, using the self-consistent formulation introduced by Tolstoy[7] and showed that the predicted data agree well with available experimental data. In the same article, Feuillade conducted mathematical analyses of two and three bubble systems and showed that there are various natural oscillating modes of the bubbles such as symmetric, or "+," (where all bubbles oscillate in phase) and antisymmetric, or "−," (where all bubbles oscillate 180° out of phase) modes. If the bubble system oscillates in the − mode, a phenomena known as super-resonances can occur. In this state, the overall damping of the bubble system is small and the bubbles can undergo very large oscillations (see Ref. 18 for more detailed explanation of this phenomena).

The studies mentioned above investigated the behavior of a bubble or a group of bubbles in an infinite domain. However, in many practical applications, including novel biomedical applications (see Ref. 16 for references), bubbles are likely to be close to or attached to a wall or boundary. The presence of a wall would undoubtedly influence the oscillatory dynamics of the bubbles. Thus, there have been some studies investigating the influence of the wall on bubble oscillations (see review article by Blake and Gibson[19] and references therein) where the bubbles undergo a change in shape under large pressure fluctuations. For small pressure fluctuations, the main effect of the solid boundary is to change the natural frequencies of the collective bubble system. Recent experimental studies by Payne *et al.*[20] have mea-

[a)]Author to whom correspondence should be addressed. Electronic mail: a.ooi@unimelb.edu.au

sured the resonant frequency of identical bubbles attached to a rigid boundary. In the same study, Payne *et al.*[20] proposed to use an image-bubble concept, formulated as a coupled-oscillator problem, to model the behavior of a group of bubbles close to a wall. This approach was originally introduced by Strasberg[21] where he showed that a bubble oscillating close to a rigid boundary can be modeled by introducing a mirror-image bubble oscillating in phase with the original bubble. Thus, the net effect of the mirror-image bubble (and hence the wall) was to reduce the resonance frequency of the bubble. Payne *et al.*[20] showed that the mirror-image bubble approach can also be successfully extended to model the behavior of a group of bubbles close to a wall; predictions showed good agreement with the experimental data.

All the experimental studies conducted thus far have been for systems consisting of bubbles that are of the same size. For this type of system, it has been shown that the coupled-oscillator model predicts resonant frequencies that are in good agreement with experimental results. In the present work, an experimental and theoretical investigation was carried out to investigate the behavior of resonant frequencies for a system consisting of two or three different-sized bubbles on a rigid boundary. The theoretical model considered by Payne *et al.*[20] was extended to explain the experimental data of different-sized bubbles on a rigid boundary.

II. THEORETICAL DEVELOPMENT

It is commonly accepted that oscillations of large (millimeter sized) bubbles is governed by the following set of linear, second order ordinary differential equations

$$\ddot{\delta}_i(t) + \epsilon_i \dot{\delta}_i(t) + \omega_{oi}^2 \delta_i(t) = -\sum_{i \neq j}^{N} \frac{R_{oj}^2}{R_{oi} d_{ji}} \ddot{\delta}_j(t) \tag{1}$$

(see Ref. 22 and references therein) where $\delta_i(t) = R_i(t) - R_{oi}$, $R_i(t)$ is the radius of bubble i as a function of time, R_{oi} is the equilibrium radius of bubble i, and

$$\epsilon_i = \frac{4\mu}{\rho R_{0i}^2} \tag{2}$$

are the damping coefficients. Note that, in general, the damping is quite small and the mathematical expression for ϵ_i should include the radiation, thermal, and viscous damping. These expressions can get quite lengthy (see published works by Devin[23] and Eller[24]). In order to simplify the analysis, we follow the work of Ida[10] and use Eq. (2) as an approximation for the damping. The natural frequency of a bubble in isolation, ω_{oi}, is given by

$$\omega_{oi}^2 = \frac{3\gamma P_o}{\rho R_{oi}^2}, \tag{3}$$

where ρ is the density of the liquid, γ is the specific heat ratio, and P_o is the equilibrium pressure. Equation (1) can be written in matrix form as

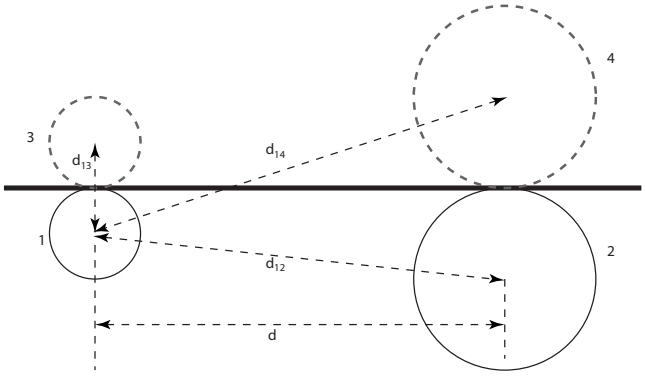


FIG. 1. Bubbles of different size close to a wall.

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = -\mathbf{S}\ddot{\mathbf{x}}, \tag{4}$$

where $\mathbf{x}$ is a time-dependent vector of all the $\delta_i(t)$, $\mathbf{M}$ is the inertia matrix, $\mathbf{C}$ is the damping matrix, $\mathbf{K}$ is the stiffness matrix, and $\mathbf{S}$ contains all the coupling terms of the bubble system. Equation (4) can be solved by assuming a solution of the form $\mathbf{x} = \mathbf{A}e^{\lambda t}$ which gives

$$[\lambda^2(\mathbf{M} + \mathbf{S}) + \lambda\mathbf{C} + \mathbf{K}]\mathbf{A} = 0. \tag{5}$$

In order to obtain a nontrivial solution for the right eigenvectors $\mathbf{A}$,

$$\det[\lambda^2(\mathbf{M} + \mathbf{S}) + \lambda\mathbf{C} + \mathbf{K}] = 0; \tag{6}$$

thus there are only distinct values of λ (eigenvalues) and associated $\mathbf{A}$ that can be solutions to Eq. (5). This is a quadratic eigenvalue problem and can be found to occur in many practical applications (see Ref. 25 for more examples). Equation (6) is a polynomial of order $2n$, where n is the matrix dimension or the number of bubbles in the system. Since all coefficients of the characteristic polynomial [Eq. (6)] are real numbers, the roots (eigenvalues) of Eq. (6) occur in n complex conjugate pairs. The natural frequencies of the system is given by the imaginary part of λ and the real part of λ (typically a negative constant) indicates the damping associated with the natural frequencies.

In general, closed form expressions for the eigenvalues and eigenvectors cannot be obtained. Numerical methods need to be employed to acquire the eigenvalues and eigenvectors and hence compute the solution to Eq. (4). However, insight into the solution can be obtained by examining the closed form expression for the eigenvalues of a simpler systems consisting of only two and three bubbles. First, consider the case where the system is made up of two bubbles with different equilibrium radii, R_{0i}, and the effects of the walls are modeled by the method of images (see Fig. 1). For such a system, Eq. (5) can be written explicitly as

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \ddot{\delta}_1(t) \\ \ddot{\delta}_2(t) \end{bmatrix} + \begin{bmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{bmatrix} \begin{bmatrix} \dot{\delta}_1(t) \\ \dot{\delta}_2(t) \end{bmatrix} + \begin{bmatrix} \omega_{o1}^2 & 0 \\ 0 & \omega_{o2}^2 \end{bmatrix} \begin{bmatrix} \delta_1(t) \\ \delta_2(t) \end{bmatrix}$$
$$= - \begin{bmatrix} \dfrac{R_{o1}}{d_{13}} & \dfrac{R_{o2}^2}{R_{o1}d_{12}} + \dfrac{R_{o2}^2}{R_{o1}d_{14}} \\ \dfrac{R_{o1}^2}{R_{o2}d_{23}} + \dfrac{R_{o1}^2}{R_{o2}d_{21}} & \dfrac{R_{o2}}{d_{24}} \end{bmatrix} \begin{bmatrix} \ddot{\delta}_1(t) \\ \ddot{\delta}_2(t) \end{bmatrix}, \quad (7)$$

where d_{12} is the distance between the two real bubbles and the other d_{ij} are the distances between the bubbles and their images. Following the discussion above, a nontrivial solution can be only obtained if

$$\det \begin{bmatrix} \lambda^2\left(1 + \dfrac{R_{01}}{d_{13}}\right) + \lambda\epsilon_1 + \omega_{01}^2 & \lambda^2\left(\dfrac{R_{02}^2}{R_{01}d_{14}} + \dfrac{R_{02}^2}{R_{01}d_{12}}\right) \\ \lambda^2\left(\dfrac{R_{01}^2}{R_{02}d_{21}} + \dfrac{R_{01}^2}{R_{02}d_{23}}\right) & \lambda^2\left(1 + \dfrac{R_{02}}{d_{24}}\right) + \lambda\epsilon_2 + \omega_{02}^2 \end{bmatrix} = 0. \quad (8)$$

Equation (8) is the characteristic polynomial for a system of two unequally sized bubble with images used to model the effects of the wall.

The analytical expression for roots of the polynomial specified by Eq. (8) above can get unwieldy and does not provide useful information. A further simplification can be made if we assume that the two bubbles are of the same size, i.e., $R_{01}=R_{02}=R_0$ and thus $\omega_{01}=\omega_{02}=\omega_0$. We will also ignore damping by setting ϵ_i (for millimeter sized bubbles, the damping is usually very small) and assume that the size of the bubbles, R_0, is small relative to the distance between the bubbles, hence $d_{14}\approx d_{12}\approx d$ (see Fig. 1). Making these assumptions, Eq. (8) simplifies to

$$\det \begin{bmatrix} \lambda^2\left(\dfrac{3}{2}\right) + \omega_0^2 & \lambda^2\left(2\dfrac{R_0}{d}\right) \\ \lambda^2\left(2\dfrac{R_0}{d}\right) & \lambda^2\left(\dfrac{3}{2}\right) + \omega_0^2 \end{bmatrix} = 0. \quad (9)$$

The eigenvalues corresponding to Eq. (9), λ, can be found to be

$$\pm\sqrt{\frac{2\omega_0^2}{-3-4/(d/R_0)}} \quad \text{and} \quad \pm\sqrt{\frac{2\omega_0^2}{-3+4/(d/R_0)}}, \quad (10)$$

with associated eigenvectors given by

$$\begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \quad \text{and} \quad \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}. \quad (11)$$

The first eigenvector has a lower natural frequency and it represent a regime where the two bubbles oscillate in phase relative to each other. This is the + (symmetric) mode. The second eigenvector is the − (antisymmetric) mode because it represents the state where the two bubbles oscillate 180° out of phase with each other phase with each other. The + mode of a two bubble system always has a lower frequency than the − mode of a two bubble system. The reason for this was explained by Feuillade.[26] When bubbles oscillate in phase with each other, they expand and contract simultaneously. Because the liquid in between the bubbles is assumed to be incompressible, the motion of the bubbles is retarded and leads to a reduction in natural frequency. The opposite is true for the − mode leading to a higher natural frequency. Note that in the limit where $d\to\infty$, the natural frequencies of the system approach

$$\sqrt{\frac{2}{3}}\omega_0 \approx 0.8165\omega_0, \quad (12)$$

which is the natural frequency of a single bubble attached to a wall. A plot of the natural frequencies of the system is shown in Fig. 2 which is similar to Fig. 3 of Feuillade[8] who studied the oscillations of bubbles in free space. The only difference here is that the asymptotic value of the natural frequencies is approximately $0.8165\omega_0$ (as oppose to just ω_0) which is the result of adding image bubbles in the system to model effects of the wall.

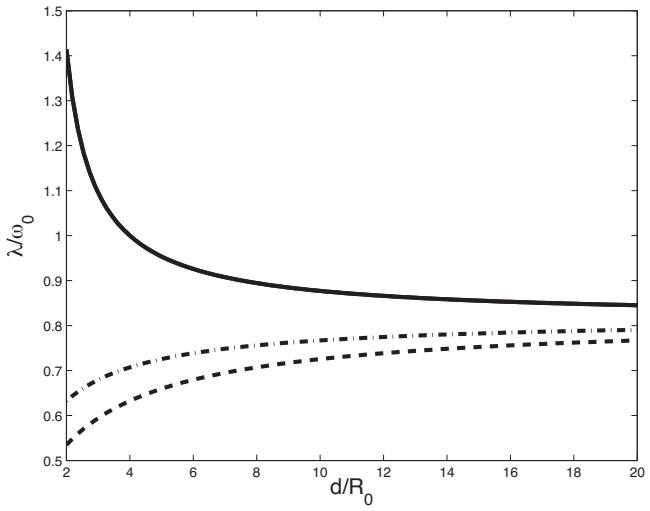


FIG. 2. Natural frequencies of the "−" mode for a two and three bubble system (—), natural frequencies for the "+" mode of a two bubble system (-·-), and natural frequency for the "+" mode of a three bubble system (---).

A similar analysis can be carried out for a system consisting of three bubbles arranged on the vertices of an equilateral triangle of length d attached to a wall. Again, the method of images will be used to model the effects of the wall and all bubbles will be assumed to be of equal size. The natural frequencies for such a system can be shown to be given by the following expressions:

$$\lambda = \pm\sqrt{\frac{2\omega_0^2}{-3-8/(d/R_0)}}, \quad \pm\sqrt{\frac{2\omega_0^2}{-3+4/(d/R_0)}}, \quad \pm\sqrt{\frac{2\omega_0^2}{-3+4/(d/R_0)}}.$$

Note that one of the eigenvalues is repeated. The corresponding eigenvectors are

$$\begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix}, \quad \begin{Bmatrix} -1 \\ 1 \\ 0 \end{Bmatrix}, \quad \begin{Bmatrix} -1 \\ 0 \\ 1 \end{Bmatrix}. \quad (13)$$

The first eigenvector is the + (symmetric) mode because it represents the state where all three bubbles oscillate in phase. The other two eigenvectors have the same natural frequen-

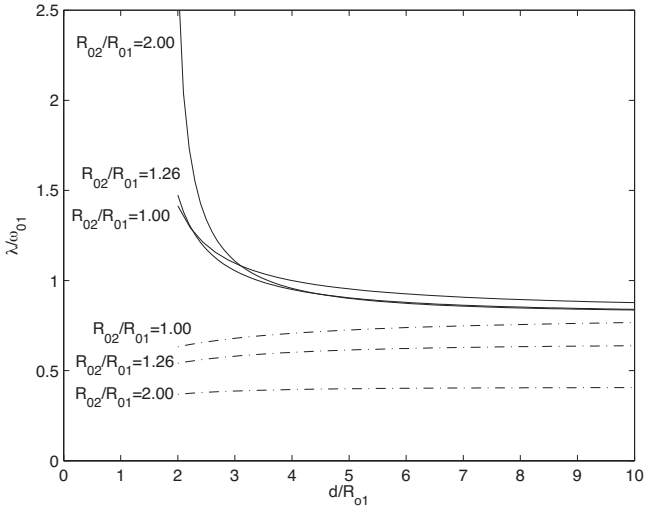


FIG. 3. Natural frequencies of a two bubble system close to a wall for different values of R_{02}/R_{01}. — Natural frequencies of the "−" mode, –·– Natural frequencies for the "+" mode.

cies and they represent a regime where two bubbles oscillate 180° out of phase with each other, while the third bubble remains in equilibrium (the − mode). Figure 2 also shows a plot of the natural frequencies of a three bubble system. The natural frequency associated with the − mode is the same as for the two bubble system. The frequency associated with the + mode is lower for the three bubble system than for the two bubble system.

The analysis above was formulated for bubbles that are of the same size. For a system consisting of bubbles that are of different sizes, the closed form solution can be unwieldy and not very informative. To study the effects of different bubble sizes, the eigenvalues for a two bubble system is obtained numerically by solving Eq. (8). Similar to the analysis carried out above, we will again assume that damping can be neglected and that the spacing between the bubbles are large compared to the size of the bubbles $d_{14} \approx d_{12}=d$. Computations were carried out with reference bubble size, $R_{01}=2.29$ mm and $R_{02}/R_{01}=1.00,1.29,2.00$. These bubble sizes were chosen because they are similar to the bubble size used in our experimental setup (see Sec. III). The resulting eigenvalues obtained are shown in Fig. 3 . The modes of oscillations can again be classified as either a symmetric (or +) mode where both bubbles oscillate in phase and antisymmetric (or −) mode where both bubbles oscillate 180° out of phase. As expected, and consistent with the reasons given by Feuillade[26] for the case with similar size bubbles, the frequency of the + mode is always less than the frequency of the − mode. From Fig. 3, it can be concluded that the − mode is affected by R_{02}/R_{01} only for small values of d/R_{01}. For $d/R_{01}>3$, there is very little variation in the natural frequency associated with the − mode. On the other hand, the + mode appears to be sensitive to the value of R_{02}/R_{01}. The natural frequency associated with the + mode appears to monotonically decrease with R_{02}/R_{01} for all values of d/R_{01}.

In the experiments, the bubble system was excited with an external source. So, strictly speaking, we will be measuring the resonance frequency of the bubble system. In systems where the damping is small, the resonance frequency is very similar to the natural frequency of the system. To investigate the behavior of system eigenmodal response to external excitation, we will take a closer look at the particular solution, $\boldsymbol{\delta}_p(t)$, of a two bubble system close to a wall [Eq. (7)] excited by a time harmonic force $\mathbf{q}(t)=\mathbf{q}_{\text{ext}}e^{i\omega_{\text{ext}}t}$. Following Tisseur and Meerbergen,[25] $\boldsymbol{\delta}_p(t)$ can be written as

$$\boldsymbol{\delta}_p(t) = e^{i\omega_{\text{ext}}t}\sum_{j=1}^{2n}\frac{\mathbf{B}_j^*\mathbf{q}_{\text{ext}}}{i\omega_{\text{ext}}-\lambda_j}\mathbf{A}_j \tag{14}$$

where $\mathbf{A}$, $\mathbf{B}$, and λ are right, left eigenvectors, and eigenvalues of the quadratic eigenvalue problem shown in Eq. (5). The contribution of the particular mode $\mathbf{A}_j$ to the particular solution is dependent on the jth coefficient

$$\phi_j = \left|\frac{\mathbf{B}_j^*\mathbf{q}_{\text{ext}}}{i\omega_{\text{ext}}-\lambda_j}\right|. \tag{15}$$

Thus ϕ is a measure of the contribution of the + or the − mode to the particular solution. It is clear that ϕ_j increases if we force the system at a frequency close to the eigenvalues associated with $\mathbf{A}_j$. A *resonance* condition in the system will occur if the frequency of the external excitation is very close to the natural frequencies (provided damping is small).

To study the effects of the external frequency excitation, the amplitude of the various modes, ϕ, is plotted as a function of excitation frequency. From Eq. (15) the peak of ϕ is expected to occur when the system is excited at the frequency similar to the natural frequency of the mode. If the same pressure amplitude were applied to both bubbles, then the response would only be made up by the + mode. However, if the pressure amplitudes applied to both both bubbles were different, then it could be expected that $\boldsymbol{\delta}_p(t)$ is made up of both the + and − modes. In the following graphs, results will be presented assuming that the external excitation is

$$\mathbf{q}_{\text{ext}} = W\begin{Bmatrix}1\\0.4\end{Bmatrix}. \tag{16}$$

These figures are selected because they are close to the amplitude of excitation that was applied to different bubbles in the experimental setup. In our experimental setup, all bubbles would experience different excitation amplitudes; hence $\boldsymbol{\delta}_p(t)$ will be made up of both the + and − modes. The magnitude of ϕ for both the + and − modes are shown Fig. 4. The effects of separation distance on ϕ is shown in Fig. 4(a). ϕ for the − mode is shown by the solid line and ϕ for the + mode is shown by the dashed line. As explained earlier, the peak of ϕ for the − mode will always occur at a frequency higher than the peak of ϕ for the + mode. It is clear that as the two bubbles are moved further apart, the − mode becomes the more dominant mode. The effects of bubble size ratio R_{02}/R_{01} on the amplitudes of the different modes is illustrated in Fig. 4(b). It is clear that increasing the size of the second bubble increases the amplitude of the + mode while preserving the peak corresponding to the − mode. Hence one can reasonably expect the + mode to be more dominant for larger values of R_{02}/R_{01}.

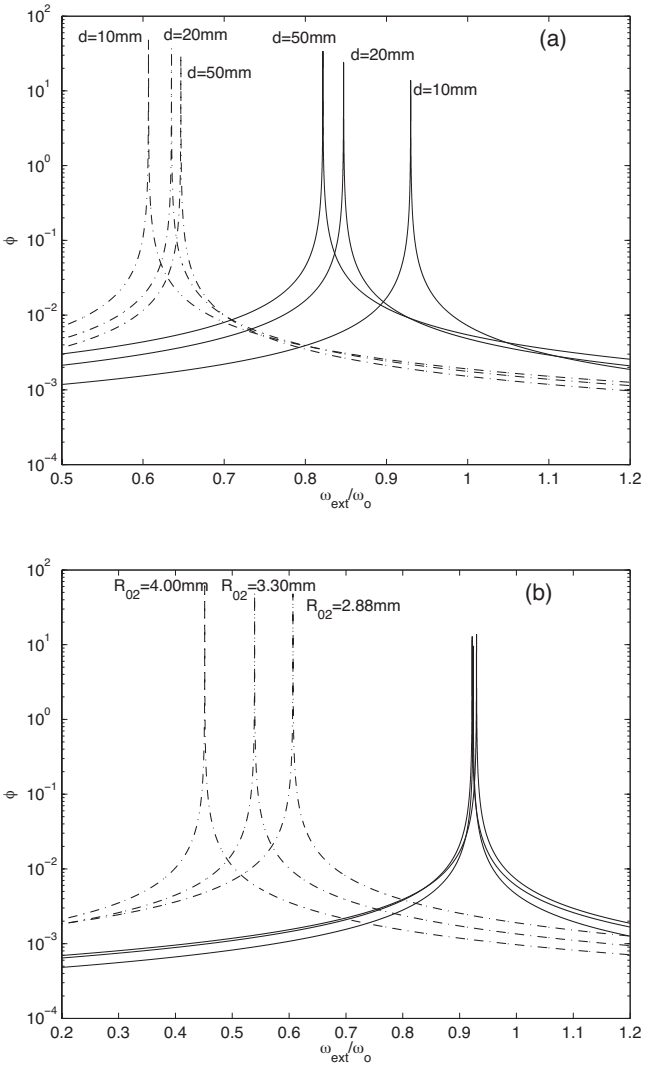


FIG. 4. Amplitude of "+" mode (–·–) and amplitude of "−" mode (—) as a function of excitation frequency ω_{ext}. (a) illustrates the effects of varying distance between the bubbles, d, which is indicated in the figure. R_{01} =2.29 mm and R_{02}=2.88 mm. The effects of varying R_{02} is shown in (b), where d=10 mm and R_{01}=2.29 mm. The values of R_{02} is shown in the figure.

III. EXPERIMENTAL SETUP

The principal purpose of the experimental apparatus is to transfer acoustic energy to the bubble system and to detect the system's response. The apparatus utilized is shown in Fig. 5. The rig and postprocessing methodology is very similar to that used by Payne *et al.*[20] The excitation signal is generated by a computer and is passed into a stereo amplifier. The amplified signal is then passed into the mechanical oscillatory unit which converts the electrical signal into oscillations in the liquid where the bubbles exist. The response of the bubble system is detected by a hydrophone and this signal is amplified by a charge amplifier unit. The bubble oscillations are logged using a high speed data acquisition system consisting of a data card, a data adapter, and a data logging computer.

The chamber is made from 12 mm thick acrylic and has a square base of length of 300 mm and height of 50 mm. A 30 mm diameter circular hole in the bottom of the cylinder

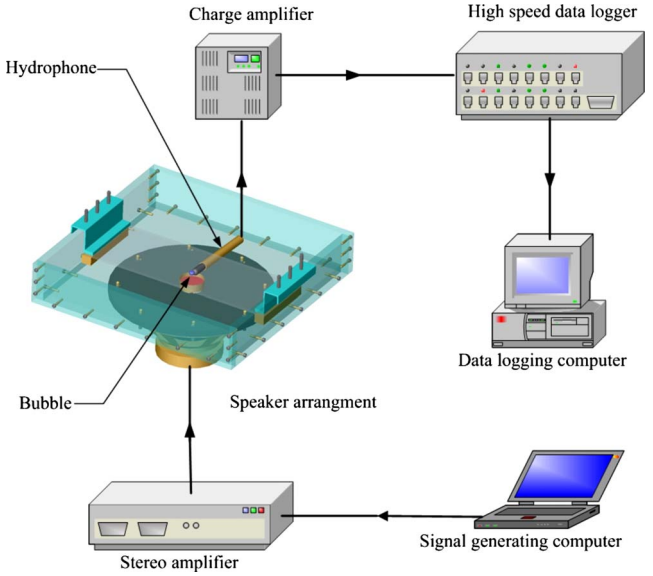
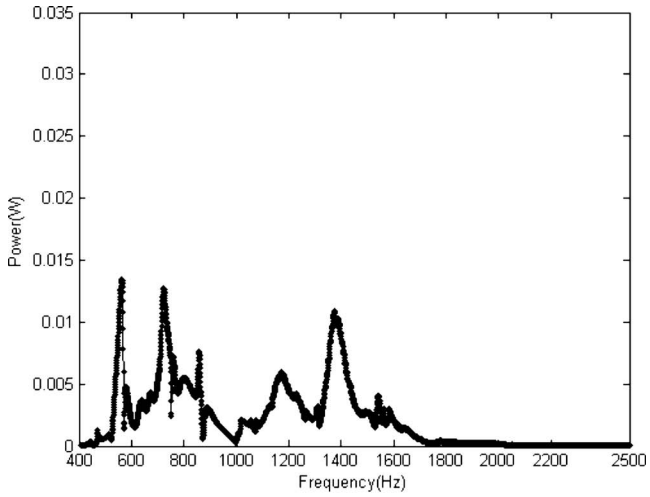


FIG. 5. (Color online) Experimental apparatus.

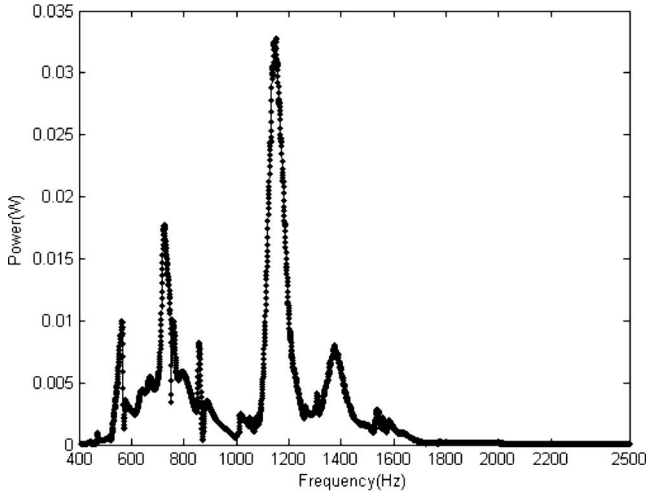
allows the acoustic waves to propagate through the liquid domain. To generate the mechanical oscillations, a common audio speaker was adopted which is similar to that used by Hsiao.[9] It was a 203.2 mm diameter, 8 Ω speaker and was modified by the attachment of an aluminum piston to its diaphragm. The piston in turn drove a piece of duct tape that sealed the hole in the chamber base.

The water in the tank was Melbourne tap water which had been filtered prior to filling the tank. The bubbles were introduced manually with a precision syringe (50 μl, Alltech Associates Australia, with a volumetric accuracy of ±5%) fitted with a needle, and were arranged as close as possible to the centerline of the piston as possible (±0.25 mm). The error in injected bubble volume corresponded to less than 1.7% of the bubble radii. Prior to this any small bubbles that have been introduced by the process of filling the tank were carefully removed by a combination of sweeping the plate surface with a wire and suction using a syringe and tube. Introduction of all the bubbles required generally took a few minutes and where bubble sizes were varied; this was done by injection of additional air. When a bubble was moved to a new location, it was swept with a wire. It is well-known that very small bubbles can increase their size over time, owing to rectified diffusion:[27] even under very low-amplitude forcing, bubbles with radii in the order of 100 μm can increase their radius by 50% after 1 h of continuous forcing.[28] However, since the bubbles of the present experiment were millimeter-sized bubbles (4000 times the volume of the just-cited experiments) and forcing pressures were less than 10^{-6} bar, for rectified diffusion or dissolution effects were negligible over the approximately 45–60 min required to perform a complete set of measurements. Careful observation indicated the bubble size had not changed measurably (to ±0.1 mm) before and after the experiments. Sound amplitudes were kept well below those at which bubbles moved or exhibited surface oscillations.

A chirp signal was used to drive the speaker; it had constant amplitude and increased in frequency over time.

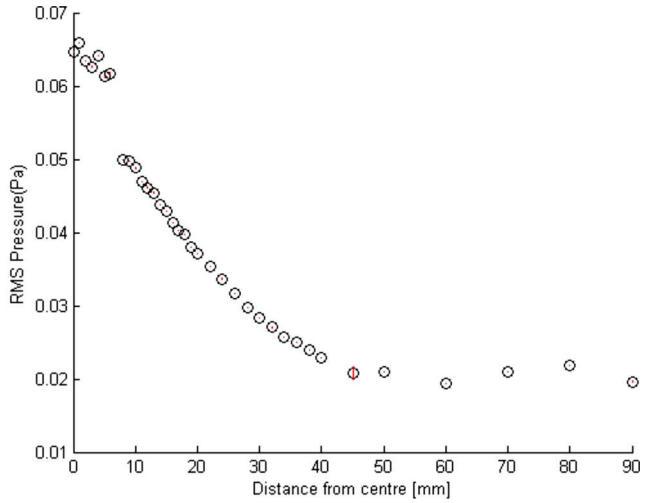


(a)

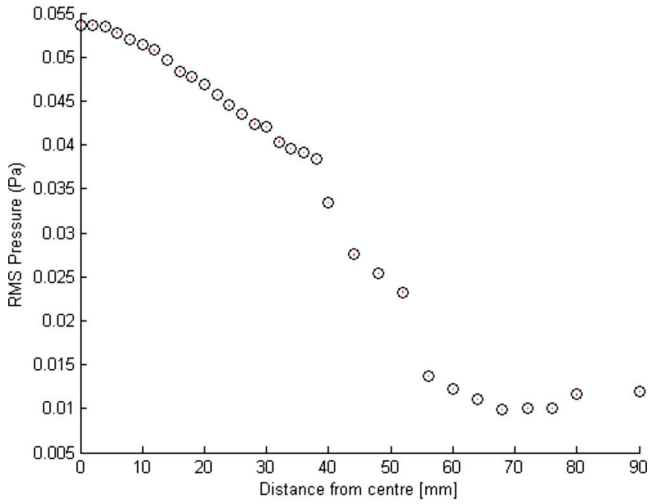


(b)

FIG. 6. Frequency domain response for an input chirp signal. (a) Without bubble response. (b) Response for a 50 μl bubble.



(a) Pressure distribution in the y direction



(b) Pressure distribution in the x direction

FIG. 7. (Color online) Pressure distribution in the x-y direction of the tank.

This signal was amplified before it was passed through to the speaker. The length of the chirp signal was kept constant at 480 ms; its frequency varied from 100 to 2500 Hz, covering the expected resonance of the bubble system in the tank. Chirps were run in rapid succession, 540 ms apart.

Thirty time domain chirp responses were detected via the hydrophone and were converted to the frequency domain via a fast Fourier transform. As a reference, and in order to determine the frequency response of the tank itself, data from the hydrophone were initially obtained for cases when no bubbles were present in the tank.[20] Typical frequency response is shown in Fig. 6(a). Experiments were subsequently carried out with bubble systems present in the tank. Typical frequency response when there is a single 50 μl bubble present in the tank is shown in Fig. 6(b). In order to obtain the resonance frequency of the bubble system, the data in Fig. 6(b) were scaled with the data in Fig. 6(a) and the resonance frequency of the bubble system is identified as the frequency when the scaled value is a maximum.

Experiments were conducted to detect pressure variations along the plate surface. The hydrophone was placed at the center of the acoustic source and directed away from the source in two directions normal to each other along the plate. As the pressure oscillations are cyclic, the rms value of the pressure is calculated for a given chirp length of 0.18 ms, with frequency varying from 800–1500 Hz. The rms pressure is calculated using the Euclidean length of the set of pressures given as

$$\|P_{ch}\| = \sqrt{\sum_{i=1}^{N_{cl}} P_{ch}(i)^2}, \tag{17}$$

where the $P_{ch}(i)$ is an element of the set of pressures which changes with time and N_{cl} is the length of the set of pressures. This Euclidean length is then converted to a rms value by considering

$$P_{\text{rms}} = \frac{\|P_{ch}\|}{\sqrt{N_{cl}}}. \tag{18}$$

This rms pressure value represented as a voltage value is shown in Fig. 7 for two orthogonal directions, considered in

the experiment. The hydrophone was placed as close to the plate as physically possible to detect the pressure oscillations at the plate.

Figure 7(a) depicts the pressure distribution along the axis denoted as x; the figure shows that pressure is highest where the hydrophone is directly above the acoustic source. As the distance from the acoustic source increases, the pressure decreases accordingly. The wavelengths of sound in the tank were at least 1 m, and the distance from the source to the tank wall was 0.15 m. The disk creating the oscillation in the tank was only 15 mm in radius, and therefore might be considered a point source relative to the wavelength. The sound field intensity falls off to 0.2 of its maximum value over only 5% of a wavelength, not because of standing-wave effects but simply because of the geometric spreading of the sound wave from the small source. Figure 7(b) depicts the pressure distribution along the axis denoted as y. Figures 7(b) and 7(a) are similar, yet differences between the plots exist. Note that the variations in the pressure was calculated for ten signals and the standard deviation is plotted which is barely visible. It could be considered that random errors due to signal generation and detection are insignificant.

In Secs. IV and V, when experimental data are compared with theoretical predictions, we will assume that the resonance frequency is similar to the natural frequency [obtained by solving Eq. (6)]. In the bubble system analyzed in this paper, the damping is usually very small. In general, it is well-known that the resonance frequency is very close to the natural frequency for a linear second order system with small damping. Justification of this assumption for a bubble system close to a wall can be found in Ref. 20.

IV. RESULTS AND DISCUSSION

As mentioned previously, this study will focus on the comparison of image-bubble theory with experimental data for a system consisting of bubbles of different sizes. As far as the authors are aware, this type of data has not been acquired before. Studies in the literature on equally sized bubbles (see Refs. 8, 10, 29, and 20) show that the coupled-oscillator model using image theory predicts resonant frequencies that agree well with experimental data. Practical applications usually consist of systems with bubbles that are of different sizes. Thus, it is important to investigate and prove if this agreement extends to data from a system made up of bubbles that are of different sizes. In order to keep the number of parameters manageable, only two and three bubble arrangements will be considered. For the three bubble system, the bubbles are arranged on the vertices of an equilateral triangle.

Experiments were conducted for two different-sized bubbles where the bubbles were positioned next to each other with a separation distance of 12 mm. One 50 μl bubble was left as the reference bubble next to a second bubble which varied in size. The hydrophone was placed at the center of the two bubbles to detect the oscillations of the two bubble system. The size of the second bubble was incrementally increased from 5 to 400 μl. At each increment the resonance frequency of the bubble system was measured. The

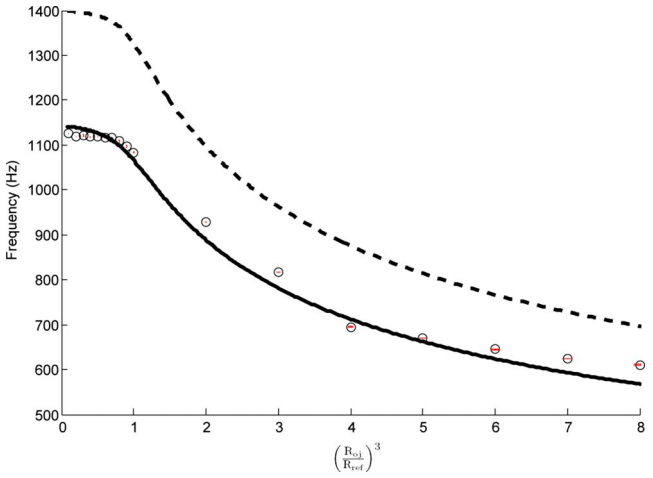


FIG. 8. (Color online) Resonance frequency vs ratio of bubble radius for two different-sized bubbles where the reference bubble remains at 50 μl ($R_{01}=R_{\mathrm{ref}}=2.29$ mm) and the other bubble incrementally increased size from 5 to 400 μl. (---) Symmetric mode eigenvalues from Eq. (5) without image bubbles, (—) with image bubbles. (○) The experimental results, with error bars shown by vertical lines.

results obtained for this two bubble system are shown in Fig. 8. The error bars on the figure are smaller than the symbols, showing that there was insignificant variation between the 30 chirps. The wavelength of acoustic pressure waves in the chamber was large relative to the distance between the bubbles. Hence, it could be assumed that the chirp signal excited the symmetric mode (where all bubbles oscillate in phase) of the system and the dominant frequency detected by the hydrophone would be the + mode frequency, especially if the bubbles were close. The theoretical + mode frequency using image bubbles to approximate the effects of the wall [i.e., the roots of Eq. (8)] is plotted against the ratio of the bubble sizes, namely, the ratio between the size of the volume incremented bubble and the reference bubble [i.e., $(R_{oj}/R_{\mathrm{ref}})^3$]. Note that for these set of results, $R_{\mathrm{ref}}=R_{01}=2.29$ mm. For comparison, the symmetric mode natural frequency for a bubble system in an unbounded domain (i.e., without image bubbles) is also shown in this figure. It is clear that the experimental data lie closer to the eigenvalues given by Eq. (8), validating the use of image bubbles to model wall effects.

Figure 8 shows that for small values of $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$, the system response is dominated by the larger 50 μl bubble on a wall. Using Eqs. (3) and (12) the + mode natural frequency of a 50 μl bubble on the wall is approximately 1163 Hz. This is very close to the resonance frequency measured by the hydrophone when $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$. As $(R_{oj}/R_{\mathrm{ref}})^3$ increases, the resonance requency decreases, consistent with the theoretical prediction shown in Fig. 3. When $(R_{oj}/R_{\mathrm{ref}})^3 > 6$, the experimental resonant frequency is larger than the theoretical prediction made using the mirror image. The bigger (dominant) bubble in the system is no longer the reference 50 μl but the other bubble in the system, which could be up to 400 μl in volume. This deviation from the experimental data could be partially explained using the argument presented by Strasberg.[21] When bubbles get bigger, buoyancy and surface tension forces change the shape of the

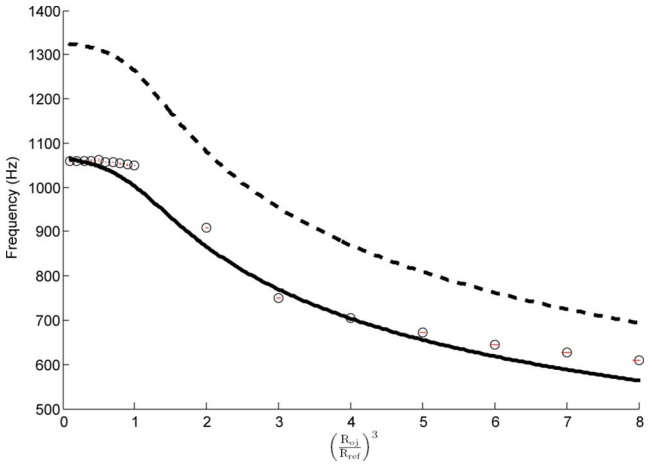


FIG. 9. (Color online) Resonance frequency vs ratio of bubble radius for three different-sized bubbles with two reference bubbles remains at 50 μl ($R_{01}=R_{\mathrm{ref}}=2.29$ mm) and the other bubbles incrementally increases from 5 to 400 μl for each experi ments. (---) Symmetric mode eigenvalues from Eq. (5) without image bubbles and (—) with image bubbles. (○) The experimental results, with error bars shown by vertical lines.

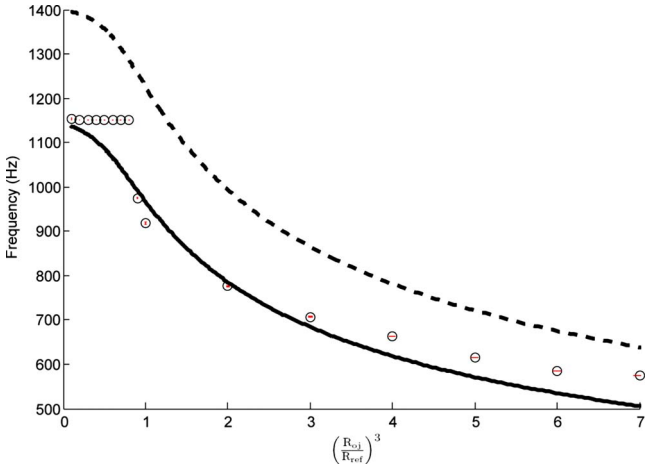


FIG. 10. (Color online) Resonance frequency vs ratio of bubble radius for three different-sized bubbles with one reference bubble at 50 μl ($R_{01}=R_{\mathrm{ref}}=2.29$ mm) and the other two bubbles incrementally increase size from 5 to 350 μl with each experiment. (---) Eigenvalues from Eq. (5) without image bubbles and (—) with image bubbles. (○) The experimental results, with error bars shown by vertical lines.

bubbles from a perfect sphere to an oblate spheroid. The theoretical analysis carried out by Strasberg[21] showed that this change in shape can cause a slight increase in the resonance frequency of bubbles in an unbounded domain. Even though the bubbles here are close to a wall, we argue that similar principles would apply. This slight increase in the measured frequency is also consistent with the results presented in Ref. 20.

Similar to the two different-sized bubble case, experiments were conducted for a system with three different-sized bubbles. Two different cases are possible. The first case is where a single bubble is incrementally increased in volume with each experiment, leaving two other bubbles as reference bubbles. In our experiments, the volume of the (two) reference bubbles were fixed at 50 μl, and the volume of the incrementally growing bubble was varied from 5 to 400 μl. The resonance frequencies measured by the experimental apparatus is shown in Fig. 9. The second case is where two bubbles incrementally increase in volume, leaving one bubble as the reference bubble. In our experiments, the volume of the reference bubble was kept constant at 50 μl and the size of the other two bubbles were varied from 5 to 350 μl. The results obtained are shown in Fig. 10. In both these cases the bubble separation distance was kept constant. Preliminary tests had showed that the positioning of the hydrophone had little effect on the measured + mode resonance frequency. Hence the hydrophone was placed close to the reference bubble, to be consistent with all the other experiments presented in this paper.

There are general similarities between data in Figs. 9 and 10. It is clear that there is a decrease in the symmetric mode resonance frequency for increasing values of $(R_{oj}/R_{\mathrm{ref}})^3$. The prediction using the mirror-image theory is plotted in these figures and, for reference, the + mode natural frequency of three bubbles in an unbounded domain (i.e., no image bubbles) is also shown in Figs. 9 and 10. Similar to the previous two bubble case, it is clear that the image theory provides a better match with the experimental data. In Fig. 9 and for $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$, the bigger (dominant) bubble in the system are the two reference 50 μl bubbles. It can reasonably be expected that the two 50 μl bubbles will dominate the response of the system for $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$. Thus one can simply ignore the third smaller bubble in the system and predict the resonance frequency by using Eq. (10). This gives the natural frequency of two 50 μl bubbles on a wall to be 1038 Hz, which is very close to the frequency measured experimentally for $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$ (see Fig. 9). When there is only one reference bubble in the three bubble system, the bigger (dominant) bubble in the system is the reference 50 μl when $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$. Figure 10 shows that for $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$, the measured frequency is approximately 1150 Hz. This is very close to the value of the natural frequency for a single 50 μl bubble attached to a wall which has a natural frequency of 1163 Hz. Hence, the effects of the two neighboring bubbles can be neglected provided that $(R_{oj}/R_{\mathrm{ref}})^3 \ll 1$.

There are, however, several differences worth mentioning when comparing data in Figs. 9 and 10. For $(R_{oj}/R_{\mathrm{ref}})^3 > 3$, it is clear that there is better agreement with theoretical predictions in Fig. 9 than in Fig. 10. In Fig. 10, there are two large bubbles in the system that are shaped like oblate spheroids. Since Strasberg[21] showed that the natural frequency of the system is increased for bubbles shaped like oblate spheroids, and since the measured frequencies are dominated by the larger bubbles in the system, the two larger bubbles shift the resonance frequency above the curve predicted by the mirror-image theory. In Fig. 9 with $(R_{oj}/R_{\mathrm{ref}})^3 > 3$, only one of the bubbles in the system was large (and shaped like an oblate spheroid); thus there is better agreement with predicted data. This shows that as the number of large bubbles increase, deviation from the theory developed here is more prominent. This is to be expected because the theory assumes that the bubbles are spherical, which is not the case when the bubbles are large.

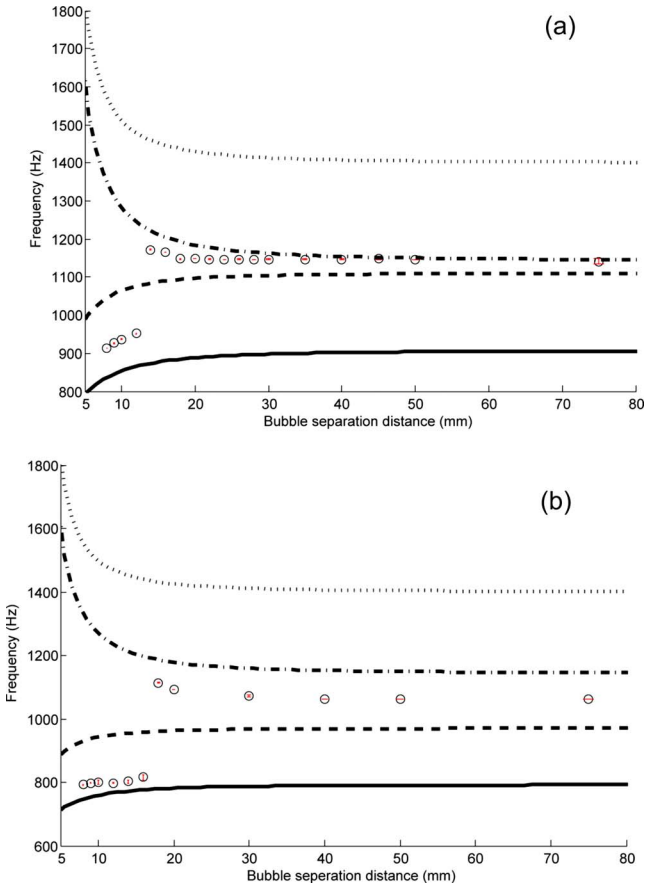


FIG. 11. (Color online) Resonance frequency vs bubble separation distance for two bubbles of 50 μl ($R_{01}=R_{\text{ref}}=2.29$ mm) and 100 μl ($R_{02}=2.88$ mm) (a) 150 μl ($R_{02}=3.30$ mm) (b). (---) symmetric and (···) antisymmetric mode eigenvalues for the mathematical model without any image bubbles. (—) symmetric and (–·–) antisymmetric mode eigenvalues using image bubbles to model the effects of the wall. (○) Experimental data, with error bars shown by vertical lines.

In the next set of experiments, the effect of bubble separation on the resonant frequency for a system of two different-sized bubbles is investigated. Experiments were conducted with two different-sized bubbles. The bubbles were initially placed close to each other and then one of the bubbles were gradually moved further away from the reference bubble. The first set of experiments were conducted with a 50 μl bubble placed next to a 100 μl ($r=R_{02}/R_{01}\approx 1.26$) bubble with a separation distance of 8 mm. The bubbles were then moved along the rigid boundary until the bubble separation distance reached 80 mm. The results obtained are shown in Fig. 11(a). To illustrate the effects of the wall, theoretical predictions were computed for cases with (to simulate wall effects) and without image bubbles in the model. In general, it is clear that the predicted data obtained using image bubbles tend to agree better with the experimental data. For small bubble separation, the dominant frequency is the + mode frequency. As the bubbles are moved further apart, the frequency response is dominated by the − mode. It can be seen that the + mode frequency predicted by the theoretical results deviate considerably from the experimental data. But as the distance between the bubbles increase, it seems that the − mode begins to dominate and there is better agreement with theoretical predictions. The occurrence of the

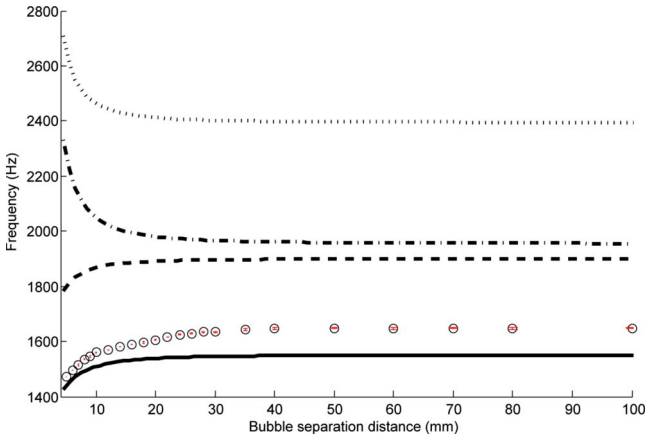


FIG. 12. (Color online) Resonance frequency vs bubble separation distance for two bubbles of 10 μl ($R_{01}=R_{\text{ref}}=1.34$ mm) and 20 μl ($R_{02}=1.68$ mm). (---) Symmetric and (···) antisymmetric mode eigenvalues for the mathematical model without any image bubbles. (—) Symmetric and (–·–) antisymmetric mode eigenvalues using image bubbles to model the effects of the wall. (○) Experimental data, with error bars shown by vertical lines.

− mode is likely to be due to the fact that the applied sound field was not perfectly uniform within the chamber. As shown previously in Fig. 7, the pressure is high at the center of the piston and decreases away from the piston center. As shown in Fig. 11(a), the theoretically predicted − mode shows closer agreement to the experimental data. The changeover from + to − mode occurs as the distance between the bubbles reaches approximately 14 mm, and the − mode appears to continue to dominate for bubble separation distances beyond 14 mm.

From the theoretical analysis described in Sec. II, it was conjectured that the − mode would be more difficult to excite at larger values of r. In order to prove this hypothesis, an experiment was conducted for a two bubble system with a larger value of r. This was achieved by keeping the reference bubble at 50 μl and increasing the size of the second bubble to 150 μl. This system would correspond to $r\approx 1.44$. Again, as in the previous case, the bubbles were moved apart to a maximum distance of 80 mm. The results obtained are shown in the Fig. 11(b). As in the previous case, for small values of d, the dominant frequency is the + mode frequency. The − mode becomes dominant when the distance between the bubble was increased. It is important to note that the transition of the dominant frequency from the + to the − mode occurs at bubble separation $d\approx 20$ mm. This is a larger transition value than for the prior set of experiment (when the transition occurs at $d\approx 14$ mm), indicating that the antisymmetric mode is harder to excite in this system (i.e., for $r\approx 1.44$) than in the previous system (for $r\approx 1.26$).

Lastly, the effects of smaller bubble size is investigated. Experiments were conducted with a bubble system consisting of 10 μl and a 20 μl bubbles. This case has $r\approx 1.26$ [similar to the case in Fig. 11(a)] but d/R_0 is much larger (due to the smaller size of the 10 μl reference bubble). The results obtained are shown in Fig. 12. It is clear that the dominant frequency obtained is the + mode resonance frequency. No − mode frequencies were detected by the microphone. The + mode resonance is closer to that of the theoretically predicted when the distance between the bubbles is

small. However, as the bubble separation distance increases, the experimental data deviate from the theoretical values. The + mode frequency increases gradually and asymptotes, similar to that predicted by the mathematical model. However, the experimental results and the theoretical results do not asymptote to the same value. It is interesting to note that for all separation distances, the experimentally measured resonance frequency did not pick up the − mode frequency.

V. CONCLUSION

The present results on bubble attached to a rigid boundary show that bubbles of different sizes exhibit similar trends in resonant acoustic frequency to bubbles of equal size. It has been illustrated that as in the case of systems consisting of bubbles that are of the similar size, the oscillations of a system of bubbles of different size attached to a rigid boundary can be modeled using coupled-oscillator theory with mirror-image bubbles approximating the effects of the wall. Both theory and experimental data show that as the ratio of the volume of the bubbles in the system increases, the resonance frequency of the system decreases. Results obtained using the coupled-oscillator model were generally in fair to good agreement with the experimental data. It was observed that when the distance between the bubbles is small, the symmetric mode is dominant and when the distance between the bubbles is large, the antisymmetric mode frequency becomes dominant. The antisymmetric mode is harder to detect when the ratio of the bubble sizes in the system is large.

[1]L. Rayleigh, "On the pressure developed in a liquid during the collapse of a spherical cavity," Philos. Mag. **34**, 94–98 (1917).
[2]M. S. Plesset, "The dynamics of cavitation bubbles," J. Appl. Mech. **2**, 277–282 (1949).
[3]E. A. Neppiras and B. E. Noltingk, "Cavitation produced by ultrasonics," Proc. Phys. Soc. London, Sect. B **63**, 674–685 (1950).
[4]M. Poritsky, "The collapse on growth of a spherical bubble on cavity in a viscous fluid," Proceedings of the First U.S. National Congress on Applied Mechanics, Illinoise Institute of Technology, June 11–16, 813–821 (1952).
[5]M. Minnaert, "On musical air bubbles and the sound of running water," Philos. Mag. **16**, 235–248 (1933).
[6]A. Ooi, A. Nikolovska, and R. Manasseh, "Analysis of time delay effects on a linear bubble chain system," J. Acoust. Soc. Am. **124**, 815–826 (2008).
[7]I. Tolstoy, "Superresonant systems of scatterers. I," J. Acoust. Soc. Am. **80**, 282–294 (1986).
[8]C. Feuillade and M. F. Werby, "Resonances of deformed gas bubbles in liquids," J. Acoust. Soc. Am. **96**, 3684–3692 (1994).
[9]P. Y. Hsiao, M. Devaud, and J. Bacri, "Acoustic coupling between two air bubbles in water," Eur. Phys. J. E **4**, 5–10 (2001).
[10]M. Ida, "A characteristic frequency of two mutually interacting gas bubbles in an acoustic field," Phys. Lett. A **297**, 210–217 (2002).
[11]R. Manasseh, A. Nikolovska, A. Ooi, and S. Yoshida, "Anisotropy in the sound field generated by a bubble chain," J. Sound Vib. **278**, 807–823 (2004).
[12]A. A. Doinikov, R. Manasseh, and A. Ooi, "On time delays in coupled multibubble systems," J. Acoust. Soc. Am. **117**, 47–50 (2005).
[13]V. Leroy, M. Devaud, T. Hocquet, and J.-C. Bacri, "The bubble cloud as an n-degree of freedom harmonic oscillator," Eur. Phys. J. E **17**, 189–198 (2005).
[14]A. Nikolovska, R. Manasseh, and A. Ooi, "On the propagation of acoustic energy in the vicinity of a bubble chain," J. Sound Vib. **306**, 507–523 (2007).
[15]V. Leroy, A. Strybulevych, M. Scanlon, and J. Page, "Transmission of ultrasound through a single layer of bubbles," Eur. Phys. J. E **29**, 123–130 (2009).
[16]R. Manasseh and A. Ooi, "The frequencies of acoustically interacting bubbles," Bubble Science, Engineering and Technology **82**, 58–74 (2009).
[17]R. Mettin, I. Akhatov, U. Parlitz, C.-D. Ohl, and W. Lauterborn, "Bjerknes forces between small cavitation bubbles in a strong acoustic field," Phys. Rev. E **56**, 2924–2931 (1997).
[18]C. Feuillade, "Acoustically coupled gas bubbles in fluids: Time-domain phenomena," J. Acoust. Soc. Am. **109**, 2606–2615 (2001).
[19]J. R. Blake and D. C. Gibson, "Cavitation bubbles near boundaries," Annu. Rev. Fluid Mech. **19**, 99–123 (1987).
[20]E. M. B. Payne, S. J. Illesinghe, A. Ooi, and R. Manasseh, "Symmetric mode resonance of bubbles attached to a rigid boundary," J. Acoust. Soc. Am. **118**, 2841–2849 (2005).
[21]M. Strasberg, "The pulsation frequency of nonspherical gas bubbles in liquids," J. Acoust. Soc. Am. **25**, 536–537 (1953).
[22]M. Ida, "Avoided crossings in three coupled oscillators as a model system of acoustic bubbles," Phys. Rev. E **72**, 036306 (2005).
[23]C. J. Devin, "Survey of thermal radiation, and viscous damping of pulsating air bubbles in water," J. Acoust. Soc. Am. **31**, 1654–1667 (1959).
[24]A. I. Eller, "Damping constants of pulsating bubbles," J. Acoust. Soc. Am. **47**, 1469–1470 (1959).
[25]F. Tisseur and K. Meerbergen, "The quadratic eigenvalue problem," SIAM Rev. **43**, 235–286 (2001).
[26]C. Feuillade, "Scattering from collective modes of air bubbles in water and the physical mechanism of superresonances," J. Acoust. Soc. Am. **98**, 1178–1190 (1995).
[27]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[28]P. Tho, R. Manasseh, and A. Ooi, "Cavitation microstreaming patterns in single and multiple bubble systems," J. Fluid Mech. **576**, 191–233 (2007).
[29]M. Ida, "Alternative interpretation of the sign reversal of secondary Bjerknes force acting between two pulsating gas bubbles," Phys. Rev. E **67**, 056617 (2003).

Angular and frequency spectral analysis of the ultrasonic backward beam displacement on a periodically grooved solid

Sarah W. Herbison[a] and Nico F. Declercq
George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 801 Ferst Drive, Atlanta, Georgia 30332-0405 and Laboratory for Ultrasonic Nondestructive Evaluation, Georgia Tech Lorraine, GT-CNRS UMI 2958, 2-3 rue Marconi, 57070 Metz, France

Mack A. Breazeale
National Center for Physical Acoustics, The University of Mississippi, Oxford, Mississippi 38677

(Received 28 November 2008; revised 10 September 2009; accepted 15 September 2009)

The ultrasonic backward beam displacement, which has been shown to occur when a bounded beam is incident upon a periodically corrugated liquid-solid interface, is studied experimentally. This effect has been previously studied on a periodic water-brass interface at one particular frequency (6 MHz) and one corresponding angle of incidence (22.5°), but the question has remained whether it would also exist at other frequency and angle combinations. The knowledge of whether this phenomenon is a coincidence or whether it will occur for other frequency and angle combinations contributes to a better understanding of the interaction of ultrasound with periodic structures and diffraction effects, in particular. Potential applications exist in the study of phononic crystals and in the non-destructive evaluation of materials. The present work reports results from recent experiments on the same periodically grooved brass sample that was employed in the first investigations of this phenomenon. Through the examination of frequency spectra in the form of angular and classical spectrograms, the experiments reported here show the backward beam displacement to occur for multiple angles of incidence and frequencies. Furthermore, evidence is shown as to the exact cause of the backward beam displacement, namely, a backward propagating Scholte–Stoneley wave. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3243467]

PACS number(s): 43.35.Bf, 43.30.Hw, 43.20.El, 43.35.Pt [RR] Pages: 2939–2948

I. INTRODUCTION

The ultrasonic backward beam displacement on a periodically corrugated surface, as shown in Fig. 1(a), has intrigued scientists since the 1970s when it was initially discovered by Breazeale and Torbett.[1] The motivation behind their work was to discover the acoustic phenomenon that would correspond to the optical phenomenon predicted by Tamir and Bertoni, where an interface with a periodic structure superimposed may cause a leaky wave to propagate in the backward direction, resulting in a backward displacement of the specularly reflected beam.[2] Using schlieren imaging and a bounded beam with a frequency of 6 MHz at an angle of incidence of 22.5°, Breazeale and Torbett did indeed observe the first ultrasonic backward beam displacement, as shown in Fig. 1(b).

Although the phenomenon was of physical interest from the time of its discovery, the backward beam displacement did not receive attention in the 1980s and 1990s when periodically corrugated surfaces were being used to transform bulk waves into surface waves for the non-destructive testing of surfaces of materials because at that time the focus was on the study of normal, and not oblique, incidence.[3] However, the interaction of sound with periodic structures, periodic surfaces being but one example, has attracted new interest in the past decade with the study of phononic crystals.[4–7] Due to the unique properties that stem from their periodicity (band gaps, etc.), these acoustic counterparts of photonic crystals in optics have potential applications including acoustic filtering and novel transducers.[8,9] In addition, periodic surfaces have recently been investigated in the context of Lamb wave propagation in corrugated plates (waveguides) with applications in the non-destructive evaluation of non-planar surfaces[10] and surface roughness characterization.[11]

Also in recent years, the backward beam displacement has received new attention with the proposal of a numerical technique by Declercq *et al.*[12] based on a combination of inhomogeneous wave theory and the Rayleigh theory of diffraction that enabled the simulation of the beam displacement. The same theoretical approach was then used to study the effect of the beam width and to reveal the cause of the effect, which turned out to be a leaky form of Scholte–Stoneley waves.[13] Shortly thereafter, experiments were conducted by Teklu *et al.*[14] that showed agreement with theoretical predictions concerning the influence of the beam width. Their experiments using schlieren photography also showed that the backward beam shift appeared when the negative first order diffracted sound beam was barely visible along the surface of the sample. This observation was considered a signal of the transition of the negative first order from a bulk wave to a Scholte–Stoneley wave with decreasing angle of

[a]Author to whom correspondence should be addressed. Electronic mail: sherbison@gatech.edu

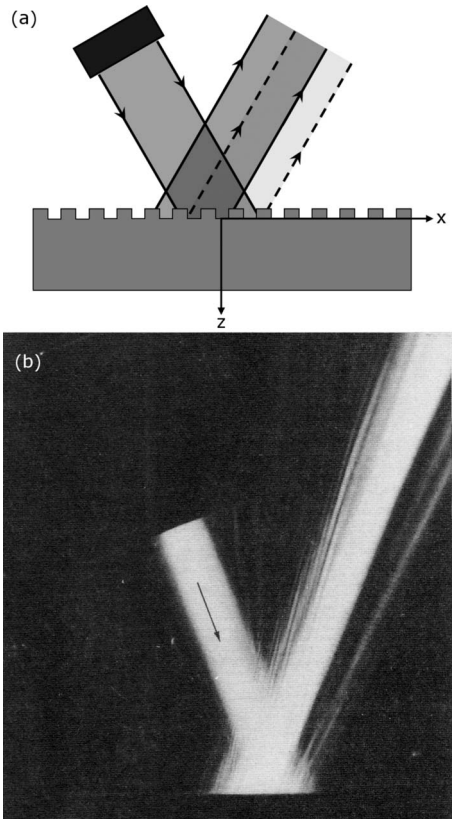


FIG. 1. (a) Diagram of the ultrasonic backward beam displacement on a periodically corrugated surface. The location of the specular beam predicted geometrically is shown by dashed lines. The backward shifted beam is shown in solid lines. The backward propagating leaky surface wave causing the beam shift propagates in the negative x-direction. Corrugation profile enlarged for illustration. (b) Schlieren image showing the backward displacement of a 6 MHz ultrasonic beam on a water-brass grating interface (Ref. 1).

incidence, but there was no hard evidence. (It should be noted that on homogeneous, planar solid surfaces, phase matching considerations prevent the coupling between bulk waves and Scholte–Stoneley waves. However, in the case of a homogeneous non-planar solid surface such as the periodically corrugated one studied in this work, a coupling between bulk waves and Scholte–Stoneley waves can exist.[15])

Prior to this experimental work by Teklu *et al.*, a numerical study on the behavior of Scholte–Stoneley waves when they encounter the corner of a solid plate[16] revealed that Scholte–Stoneley waves are scattered in the forward direction upon reaching the corner and do not propagate around the corner in the manner of leaky Rayleigh waves.[17,18] This phenomenon has also been shown experimentally.[19]

The present work has two objectives, both accomplished experimentally. The first objective is to quantitatively show that the backward beam displacement exists at additional frequencies and angles of incidence other than the single frequency and angle of incidence pair (6 MHz, 22.5°) studied in the past and to show that the classical grating equation can be used to theoretically predict potential frequency and angle pairs. The second objective is to show that the backward displacement is accompanied by a backward propagating Scholte–Stoneley wave, which results in the known transmission effect upon reaching the edge of the solid sample. This work realizes both of these objectives through the analysis of frequency spectra resulting from angular scans of the ultrasound fields in water using a newly developed polar/C-scan apparatus at Georgia Tech Lorraine. All the experiments reported here are performed on the original grooved brass sample that was used in 1976 when Breazeale and Torbett[1] first reported the ultrasonic backward beam displacement.

II. THEORETICAL CONSIDERATIONS

We consider a bounded beam incident from water (with a sound velocity v_{liq}) at an angle of incidence θ_i onto a grooved solid surface with spatial periodicity Λ. All angles of incidence are given from the normal to the surface, and the beam is considered to be time-harmonic with frequency f. From the classical diffraction grating equation,[2] an expression is formulated that gives the optimum angle of incidence for the generation of a diffracted backward propagating lateral wave (e.g., surface wave with velocity v_{surf}):

$$\sin(\theta_i) = v_{\mathrm{liq}}\left(\frac{1}{f\Lambda} - \frac{1}{v_{\mathrm{surf}}}\right). \tag{1}$$

This expression can be inverted to give the frequency of the backward propagating lateral wave given a known angle of incidence. [In the derivation of Eq. (1), the diffracted surface wave generated is considered to be of the negative first order since higher order modes will not propagate for the frequencies and surface periodicity examined in this work.] It was predicted by Tamir and Bertoni[2] that the presence of a backward propagating leaky wave would interfere with the specularly reflected beam and result in its lateral backward displacement. In the first investigation of this prediction, Breazeale and Torbett used a value of 2015 m/s, the leaky Rayleigh wave velocity on brass, as the velocity of the backward propagating leaky surface wave and a velocity of 1490 m/s for water. Their sample had a rectangular corrugation profile with a periodicity of 178 μm and a height of 25 μm. The length of the grooved portion of the sample was 25.4 mm. The result of their calculation was a prediction of 41° for the optimal angle of incidence to generate the backward displacement for an incident beam with a frequency of 6 MHz. However, their experiments using schlieren photography showed that the backward displacement occurred for their beam at an angle of incidence of $22.5° \pm 0.25°$, not 41° as predicted. This suggested that either the theory of Tamir and Bertoni did not accurately describe the backward beam displacement or that the surface wave responsible for the phenomenon was not, in fact, a leaky Rayleigh wave.

As discussed in the Introduction, it was later shown that the angle at which the backward beam displacement appears can be predicted by the knowledge that the effect is induced by a leaky type of backward propagating Scholte–Stoneley wave.[13] With this knowledge, calculations show that the surface wave responsible for the backward displacement of the 6 MHz beam incident at $22.5° \pm 0.25°$ on the grooved brass sample of Breazeale and Torbett would have a velocity between 1465.8 and 1477.6 m/s with an average of 1471.7 m/s.

Even though, in general, one may expect dispersion of Scholte–Stoneley waves on corrugated surfaces and a dependence of the velocity on the corrugation periodicity, this velocity can be used as a basis in order to predict the angle of incidence and frequency pairs at which one may expect the backward beam displacement to appear on the surface under study. Thus, the frequency and angle pairs predicted by Eq. (1), with a Scholte–Stoneley wave velocity of 1471.7 m/s used as v_{surf}, will be compared to the frequency and angle pairs observed experimentally in this work. In addition, the velocity of sound in the water used for the recent experiments reported here was measured to be 1479.5 m/s, and we will use this velocity as v_{liq} in Eq. (1) in order to be consistent with the actual experimental environment.

In addition to predicting the frequency and angle of incidence combinations that should result in the backward beam displacement and the generation of a backward propagating Scholte–Stoneley wave, the classical grating equation also provides information on the propagation of diffracted bulk modes that are not confined to the surface of the sample. In prior studies of the backward beam displacement, only single frequency beams (i.e., time-harmonic ultrasonic beams) have been employed, and so the presence of the backward propagating Scholte–Stoneley wave was examined without the possibility of other propagating bulk modes being present. However, in this study an ultrasonic pulse is employed so that many frequencies can be examined simultaneously. This results in the presence of additional propagating bulk modes in the fluid that are not confined to the vicinity of the surface of the sample; this occurs only for frequencies that are higher than the frequency of the backward Scholte–Stoneley wave generated at a given angle of incidence. The propagating bulk modes are also considered to be of the negative first order and their angles of propagation, θ_{pm}, can be predicted by the following expression, also derived from the classical grating equation:

$$\sin\,\theta_{pm} = \frac{v_{liq}}{f\Lambda} - \sin\,\theta_i. \tag{2}$$

Using Eq. (2) for an ultrasonic pulse at a given angle of incidence, the directions of the propagating modes for many frequencies can be predicted.

III. EXPERIMENTAL SETUP

A single experimental setup is used to accomplish both of the objectives of this work as described in the Introduction. All measurements are performed underwater in an ultrasonic immersion tank. The experimental setup consists of a pitch-catch arrangement with a stationary transducer emitting a pulse with a center frequency equal to approximately 5 MHz that is incident upon the grooved brass sample. The propagation distance between the incident transducer and the sample is 66 mm, and the beam width is approximately 12 mm. The beam used in the original experiments of Breazeale and Torbett was 10 mm in width, and it was shown by Teklu *et al.*[14] that the backward displacement is not necessarily visible for smaller beam widths. The incident pulse and its frequency spectrum are shown in Fig. 2. The diffracted fields

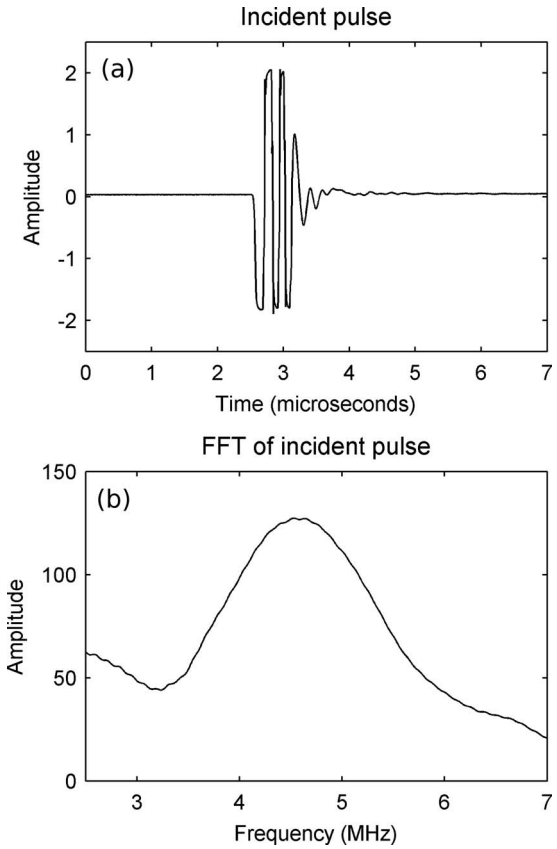


FIG. 2. (a) Time waveform of the incident pulse and (b) its frequency spectrum.

generated by the sample are scanned with a receiving transducer that rotates about the point of incidence on the sample surface. This transducer is mounted in the polar/C-scan apparatus which controls the rotation of the scan and the timing of the waveform acquisition. The distance between the sample and the receiving transducer is chosen to be 45 mm in order to be compatible with the rotation of the polar/C-scan equipment. The receiving transducer is continuously aimed at the point of incidence on the grooved sample and the propagation distance remains unchanged during rotation. In order to change the angle of incidence, the sample is mounted in such a manner that it can be rotated. A photograph of the experimental setup (before submersion in the tank) is shown in Fig. 3, with the sample rotated such that the beam from the emitting transducer would be normally incident to the sample surface.

Two types of scans are performed. The first scans the diffracted field in front of the emitting transducer (where the specularly reflected beam and any backward shifted frequencies would be present) in order to quantitatively observe the backward beam shift. A schematic of this scan is shown in Fig. 4(a). A scan is performed each time the angle of incidence is changed (clockwise rotation of the sample). The angular resolution is 0.25° and the angular range is 70° for all scans. The start and end points of each scan are determined by geometrical constraints and so their locations with respect to the polar/C-scan equipment are identical for all scans.

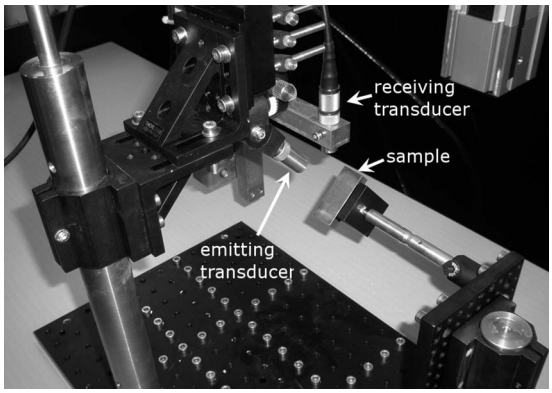


FIG. 3. Photo of the experimental setup with the sample rotated such that the beam from the emitting transducer is normally incident upon the sample. Photo was taken before submersion in the water tank.

The second type of scan measures the diffracted field behind the emitting transducer (near the surface of the sample where the backward propagating surface wave is predicted to be found) in order to quantitatively observe the backward propagating surface wave that is predicted to accompany the backward beam displacement. To scan this area of the diffracted field, the sample is rotated in the opposite (counter-clockwise) direction so that the receiving transducer is now in the field considered to be behind the emitting transducer. A schematic of this scan is shown in Fig. 4(b). A scan

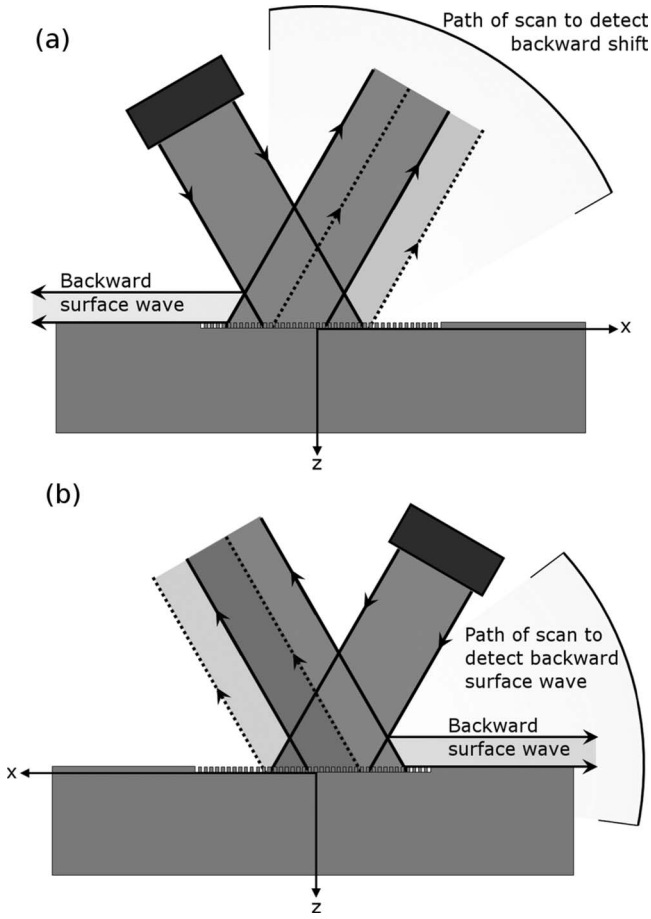


FIG. 4. (a) Scan path to detect the backward beam shift. (b) Scan path to detect the backward propagating Scholte–Stoneley wave.

is performed each time the sample is rotated to change the angle of incidence. The angular resolution and angular range of these scans are also 0.25° and 70°, respectively.

The measurements for the angles of incidence (i.e., angles of rotation of the sample) given in this work should be considered as accurate to ±1°. Before we report our experimental results, it is appropriate to discuss a consequence of the rotational nature of our scans, namely, their sensitivity to the magnitude of the backward displacement along the surface and to the angle of incidence (or rather specular reflection). The backward beam displacement is considered to be a lateral shift of a (time-harmonic) beam along the surface. However, in this work, as the beam employed is a pulse, we expect only a single or very narrow range of frequencies to be shifted backward while all other frequencies within the pulse will remain in the specularly reflected beam. For a constant angle of incidence, as the magnitude of the lateral shift of the backward displaced frequencies increases and moves away from the specularly reflected beam, their detection becomes more difficult since this backward shifted portion becomes less normal to the surface of the receiving transducer during the scan. For a constant magnitude of the lateral shift, the backward displaced portion of the beam at small angles of incidence will be less normal to the surface of the receiving transducer (and thus its detected amplitude will be lower than the reality) than it would be for larger angles of incidence as it rotates through the scan. In addition, this given lateral shift along the surface will also be more difficult to detect if it occurs for a too large angle of incidence. It should be noted that the receiving transducer does not detect the lateral shift of the backward displaced portion of the beam, but rather the projection of that lateral shift at the transducer surface. As the angle of incidence increases, the projection of the lateral shift and hence the distance between the backward displaced and the specularly reflected portions of the beam decreases at the surface of the transducer. The lateral shift then becomes more difficult to detect since it is seen as part of the specular beam.

The conclusion of this discussion is that using the rotational scans as described in this work, the backward beam displacement can be best detected within a middle range of angles of incidence. It should also be noted that even though these effects should be taken into account when using a rotational scan to measure the backward beam displacement, this type of scan is an ideal tool to measure both the backward propagating surface wave and the propagating modes that occur behind the emitting transducer, simultaneously.

IV. RESULTS

The earlier experiments reported by Breazeale and Torbett[1] and by Teklu *et al.*[14] studied the interaction of continuous time-harmonic waves with the corrugated brass sample using schlieren imaging. However, the experiments of this work employ a pulse in order to study many frequencies, all of which are simultaneously incident on the sample. To separate the different frequencies, a Fourier transformation is performed on the detected signals after their acquisition. First, the original observations of Breazeale and Torbett

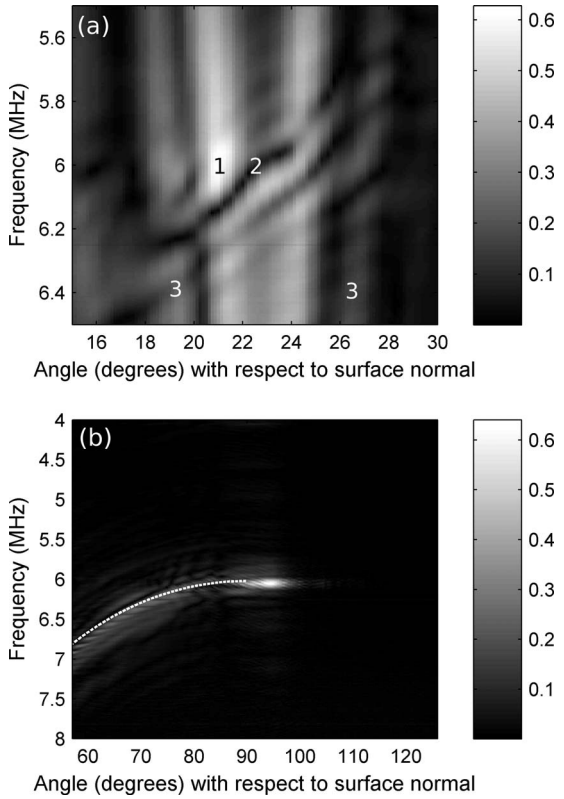


FIG. 5. Angular spectrograms confirming the results obtained by Breazeale and Torbett. For a θ_i of 22.5°, a spectrogram from the region of the specularly reflected beam (a) shows a null zone of frequencies in the range 5.95–6.16 MHz. The complementary spectrogram from a scan to detect the backward surface wave (b) shows it occurring at 6.05 MHz. Propagating bulk modes also detected in (b) and shown theoretically by dotted line.

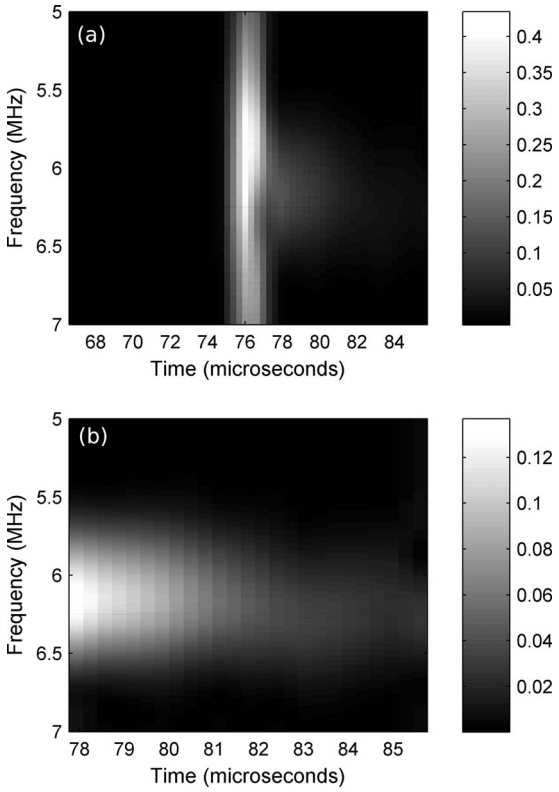


FIG. 6. (a) Classical spectrogram for a θ_i of 22.5° obtained from the time waveform captured at a receiver angle of 21°. (b) Close-up of the classical spectrogram in (a) to show the trailing frequencies arriving after the initial pulse.

are studied using the current experimental method. We next discuss the results obtained for additional angles of incidence and frequencies. Results are reported in the form of angular spectrograms, which show the amplitude of frequencies present in the waveform detected by the receiving transducer as a function of the angle of the transducer within the field. In addition, classical spectrograms, which show frequencies present in the time waveform detected by the receiving transducer as a function of time, are presented.

In what follows, all results are normalized with respect to the frequency spectrum of the incident beam. The angular spectrograms, then, are amplitude plots as a function of frequency and position. They are not schlieren-type images, but are measurements in the plane of interaction at a certain radius from the interaction spot.

A. Confirmation of the observations of Breazeale and Torbett

For a θ_i of 22.5°, a scan of the region of the specularly reflected beam results in the angular spectrogram shown in Fig. 5(a). The beam is centered at the angle of specular reflection, 22.5°, as expected. There are also three note-worthy features within this angular spectrogram, and they are numbered on Fig. 5(a).

First, there is a bright zone of frequencies with higher amplitudes, labeled as "1" in Fig. 5(a), centered approximately at an angular position of 21° and at a frequency of 6 MHz. If we examine the angular spectrogram closely, the maximum amplitude within this zone is found to be at an angular position of 21° and at a frequency of 5.98 MHz. Frequencies having high amplitudes (greater than 0.55) within the bright zone exist between 5.92 and 6.08 MHz, and frequencies within this range do not have such high amplitudes on the opposite side of the beam. The existence of this high-amplitude bright zone indicates that the energy from frequency components in this range is not distributed equally from left to right across the beam profile; these frequencies are disproportionately found on the left side. This result can be interpreted as evidence of these frequencies having been shifted backward. However, if this is indeed the case, these frequencies should be seen to have a later arrival time in a classical spectrogram.

Therefore, a classical spectrogram was obtained from the time waveform captured at a receiver position of 21°, and it is shown in Fig. 6(a). If the backward beam displacement occurs as it is currently understood, a frequency component that is shifted backward should arrive after the other non-shifted frequencies in the beam. This is due to finite nature of wave propagation speeds; any frequency that is shifted backward with respect to the specular direction must first travel backward along the surface at the speed of the backward propagating surface wave before re-radiating from the sample surface to the receiving transducer. When we examine this classical spectrogram, we see the arrival of the specular beam (a pulse) as a solid vertical band of frequencies. The time duration of the pulse as it is seen by the

receiving transducer is a function of the beam width, the angle of incidence, and the incident pulse duration. However, this specular beam arrival is followed by a faint trail of frequencies that can be seen more clearly in Fig. 6(b). In this close-up of Fig. 6(a), a band of "trailing" frequencies is seen to arrive after the initial pulse. The band of trailing frequencies is quite wide, and the frequency of maximum amplitude of the band is 6.17 MHz along many of the initial time windows of the spectrogram.

Returning to the angular spectrogram in Fig. 5(a), the second note-worthy feature is the presence of a null zone (with amplitudes less than 0.1), labeled as "2" in the figure, inside the specular beam. This null zone represents frequencies at positions between 20.75° and 23.75° that are detected as having very low amplitudes in the angular scan. The frequencies in the null zone range from 6.15 to 6.16 MHz at 20.75° and from 5.95 to 5.96 MHz at 23.75°. The existence of this null zone is most likely due to phase cancellation between specularly reflected sound and re-radiated (backward shifted) sound. The presence of null strips due to phase cancellation in the reflected fields generated by incident bounded beams, such as those seen in studies of the Schoch effect, are well documented.[20,21]

Finally, we note the presence of vertical bands showing many frequencies present to the left and right of the specular beam, namely, at angles of 19° and 26.5° and labeled as "3" in Fig. 5(a). The vertical bands to the left and right of the specular beam can be attributed to deformation of the beam that occurs upon reflection, which is not an uncommon occurrence especially in the case of periodically corrugated surfaces. Such beam deformation and beam widening, in particular, can be seen in several references. In particular, Figs. 5 and 6 in the work of Teklu *et al.*[14] show the deformation of an ultrasonic beam upon reflection from a periodically corrugated surface. Specular beam deformation is also seen in the original schlieren image of the ultrasonic backward beam displacement observed by Breazeale in 1976[1] that we have shown in Fig. 1.

The spectrogram of the complementary scan in the region of the backward propagating surface wave for the same angle of incidence, 22.5°, is shown in Fig. 5(b). Here a backward propagating surface wave with an amplitude maximum at a frequency of 6.05 MHz can be seen with a range of frequencies surrounding the amplitude maximum. This angular spectrogram shows that the backward propagating surface wave propagated directly off the surface of the sample and into the field of the receiving transducer, as one would expect from a Scholte–Stoneley wave. Higher frequency propagating bulk orders are also seen in the spectrogram of Fig. 5(b). The dotted line plotted on the spectrogram corresponds to the theoretical angles of the propagating bulk modes as a function of frequency, calculated from Eq. (2). The presence of faint additional frequencies at the same angular position as the backward propagating Scholte–Stoneley wave can be explained as being evanescent waves, negative first order for frequencies below the Scholte–Stoneley frequency or higher order for higher frequencies.

A summary of these results is the following. From the scan in the region of the specular beam, it is clear that there is an imbalance of frequencies across the beam profile from left to right for frequencies in the vicinity of 6 MHz (± 0.08 MHz). There is also a null zone consisting of a band of frequencies between 5.95 and 6.15 MHz depending on the angle of the receiver. From the classical spectrogram, it is seen that at a receiver angle of 21°, left of the center of the specular beam, a wide band of frequencies with a maximum at 6.17 MHz arrives after the specular beam. This band of trailing frequencies is sufficiently wide as to include frequencies in both the high-amplitude bright zone and the null zone in the angular spectrogram. From the scan in the region behind the incident beam, a backward propagating surface wave with an amplitude maximum at a frequency of 6.05 MHz was detected.

From this evidence, we are led to conclude that the null zone in the center of the specular beam is likely due to phase cancellation resulting from those frequencies having been shifted backward. This is based on the observation that many of the frequencies in the null zone can be located in the bright zone centered at 21° and that in the classical spectrogram taken from the time waveform arriving at 21°, the band of frequencies arriving after the specular beam contains all of the frequencies in the null zone. For those frequencies appearing in the null zone (located below the bright zone) at a receiver location of 21°, we suspect that they have been shifted backward as well, but since the sensitivity of the receiving transducer decreases dramatically for sound not normally incident to its surface, it is possible that this became an issue at the angles where these frequencies could have been detected.

All prior studies on the ultrasonic backward beam displacement have employed a time-harmonic beam, not a pulse. Therefore, when a pulse is employed, there exists the possibility that for a given angle of incidence there is actually a range of frequencies, and not just a single frequency, that can be shifted backward. We make this hypothesis based on the observation that in the classical spectrogram of Fig. 6(b), it is a fairly wide band of frequencies that arrives after the specular beam. This result could also be due to imperfections on the corrugated surface. One particularly interesting feature of this band of trailing frequencies is its duration: we can see for how long these frequencies continue to arrive at the receiving transducer, which provides insight into the time-dependent nature of the backward beam displacement, which is an area for future investigation.

We now wish to compare these experimental results with theory. The theoretical Scholte–Stoneley wave frequency, f_{SSt}, calculated for an angle of incidence of 22.5° is equal to 5.99 MHz. It can be seen in Fig. 5(a) that this frequency is within the bright high-amplitude zone at an angle of 21° and is in the null zone in the center of the specular beam. It is also a frequency contained within the trailing frequencies in the spectrogram of Fig. 6(b), and it is close to the maximum amplitude of the detected backward propagating surface wave. Therefore, we conclude that it is reasonable to assume that this frequency is displaced backward with respect to the specular beam and that its displacement is due to a backward propagating Scholte–Stoneley wave.

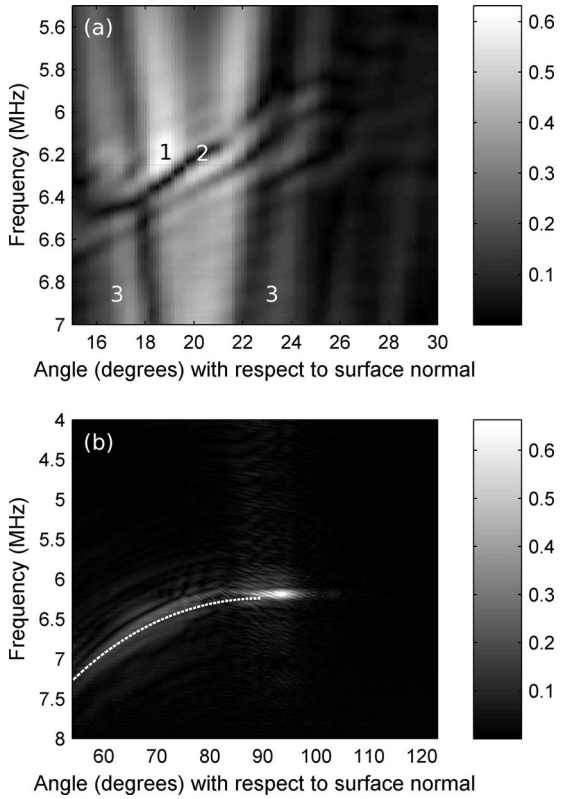


FIG. 7. (a) Angular spectrogram for a θ_i of 20° showing backward shifted frequencies in the region of specular reflection. (b) Complementary angular spectrogram showing the backward propagating Scholte–Stoneley wave and higher frequency propagating bulk modes (with theoretical locations shown by dotted line).

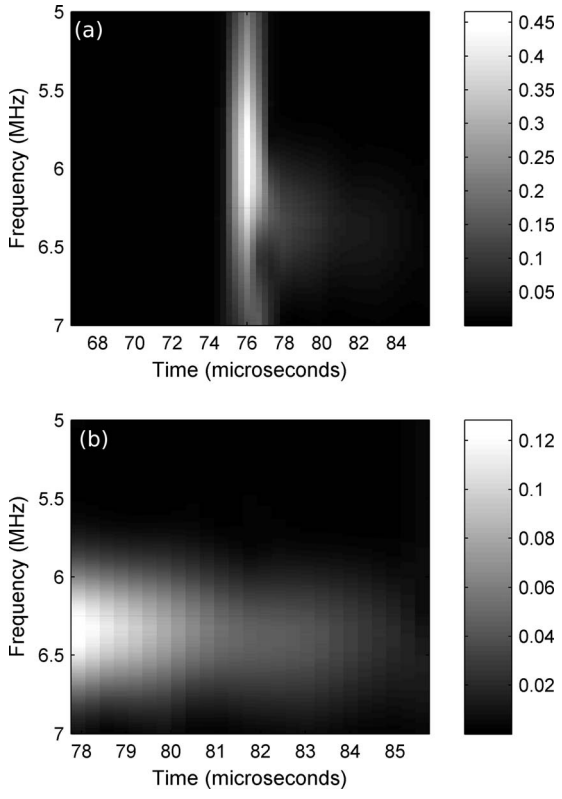


FIG. 8. (a) Classical spectrogram for a θ_i of 20° obtained from the time waveform captured at a receiver angle of 18.5°. (b) Close-up of the classical spectrogram in (a) to show the trailing frequencies arriving after the initial pulse.

B. Additional angles of incidence and associated frequencies

We now address results obtained for three additional angles of incidence, which are shown in Figs. 7–12 and are summarized in Table I along with the results previously discussed. For each case, Table I shows the appropriate figure numbers, the angle of incidence θ_i, the frequency of maximum amplitude found within the bright zone in each angular spectrogram, the range of frequencies found in the null zone of each angular spectrogram (amplitudes < 0.1), the frequency of maximum amplitude in the trailing frequency band in each classical spectrogram (in all cases this frequency was found to remain the maximum over many time windows), the frequency of maximum amplitude of the backward propagating surface wave found in the complementary spectrogram of the region behind the emitting transducer, and the theoretical Scholte–Stoneley wave frequency f_{SSt}.

Concerning the scans in the regions of specular reflection, Figs. 7(a) and 9(a) exhibit features similar to those of Fig. 5(a), and the figures have been labeled in a similar manner. First, a bright zone of higher amplitude frequencies exists on the left side of each specular beam. The bright zone in Fig. 11(a) does not occur to the left of the specular beam, but rather inside it. This is attributed to the fact that as the angles of incidence and specular reflection increase, the projection of a backward displacement appears closer to the center of the receiving transducer. Figures 7(a) and 9(a) also show null zones of frequencies that can be found in the higher amplitude bright zone in the left side of the specular beam. For Fig. 11(a), the null zone has a different form because higher frequencies drop off sharply in the center of the specular beam. This can be attributed to the efficiency with which higher frequency propagating modes may be generated for these frequencies at this angle of incidence, as well as possible destructive interference occurring between these propagating modes and the specular beam. Therefore, the null zone for Fig. 11(a) has been calculated to occur where the frequencies drop off with amplitudes less than 0.1.

Classical spectrograms obtained from time waveforms captured during the scans in the regions of the specular beams are shown in Figs. 8, 10, and 12. Each figure contains a classical spectrogram obtained from a time waveform captured at an angular position to the left of the specular beam with the exception of Fig. 12. In this case, it appears that the measurement of the angle of incidence was at the upper limit of its uncertainty and the classical spectrogram was obtained from a time waveform received at an angle of 31.5°. In all the classical spectrograms, the specular beam arrival is seen as a vertical band of frequencies followed by trailing frequencies that are considered to be backward shifted frequencies, since they arrive after the initial pulse. Close-ups of the spectrograms more clearly reveal the range of frequencies that trail the initial pulse and there is a definite trend: the trailing frequencies of maximum amplitude decrease with increasing angle of incidence, and this is consistent with the other experimental results and with the theoretical Scholte–Stoneley wave frequencies. The wide bands of trailing fre-

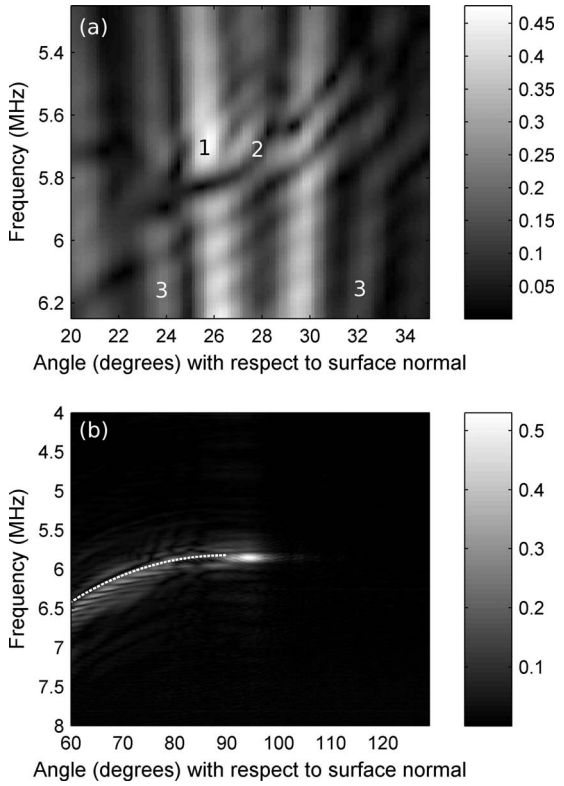
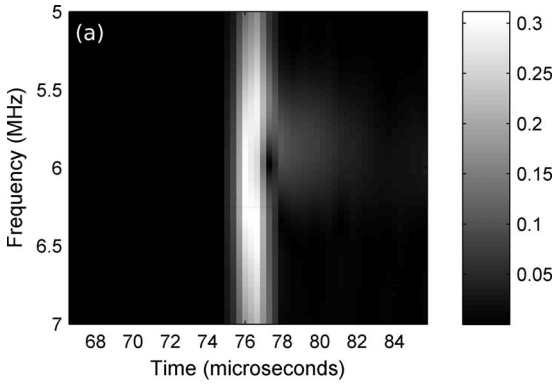


FIG. 9. (a) Angular spectrogram for a θ_i of 26.5° showing backward shifted frequencies in the region of specular reflection. (b) Complementary angular spectrogram showing the backward propagating Scholte–Stoneley wave and higher frequency propagating bulk modes (with theoretical locations shown by dotted line).



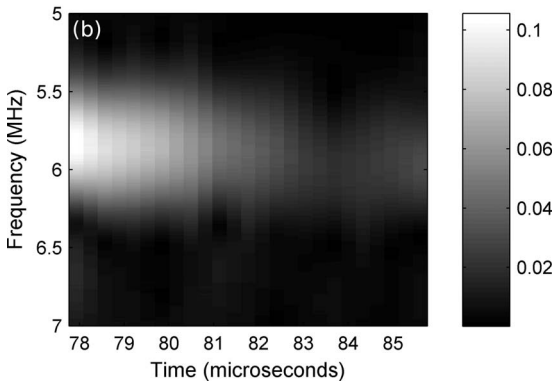


FIG. 10. (a) Classical spectrogram for a θ_i of 26.5° obtained from the time waveform captured at a receiver angle of 25.75°. (b) Close-up of the classical spectrogram in (a) to show the trailing frequencies arriving after the initial pulse.

quencies may suggest that when a pulse is incident on a periodically grooved surface, a range of backward shifted frequencies may be generated. One interesting feature found in all the classical spectrograms of Figs. 6(a), 8(a), 10(a), and 12(a) is a localized drop in amplitude for frequencies just below the trailing frequencies at the end of the specular beam arrival. The cause of this feature is unknown at this time, but it clearly has a time-dependent nature and is related to the diffraction occurring on the surface.

Concerning the complementary scans performed in the regions of the backward propagating surface wave, shown in Figs. 7(b), 9(b), and 11(b), a lateral wave is indeed observed at the surface of the sample and in all cases, and the wave has propagated off the edge of the sample into the field of the receiving transducer. This evidence, along with the fact that the experimentally observed frequencies of this surface wave are consistent with the other experimental evidence of backward shifted frequencies and the theoretical Scholte–Stoneley frequencies (f_{SSt}) for each angle of incidence, supports the claim that a backward propagating Scholte–Stoneley wave accompanies and is responsible for backward displaced frequencies. In addition to the backward propagating Scholte–Stoneley wave, higher order propagating bulk modes are observed for each angle of incidence, and their locations are well-described by theory, as shown by each dotted line.

In summary, it can be seen from the Figs. 5–12 and from Table I that in general, for all the angles of incidence studied here, experimental evidence has been observed that supports the existence of the backward beam displacement for multiple angles of incidence and frequencies and provides more insight as to the cause of that displacement. The frequencies of maximum amplitude found in the bright zones of the angular spectrograms, which, for three of the four angles studied, were observed to the left of the specular beam, are in reasonable agreement with the theoretically predicted backward Scholte–Stoneley frequencies, f_{SSt}. Also, the frequency ranges of all the null zones observed in the angular spectrograms contain the frequency of maximum amplitude observed in the bright zone and usually the appropriate f_{SSt} for each angle of incidence. The frequencies of maximum amplitude observed in the trailing frequencies in the classical spectrograms were toward the high end of the frequency ranges for the null zones and slightly higher than the theoretically predicted f_{SSt}. However, the classical spectrograms showed a wide range of potentially backward shifted frequencies which supports the existence of bright and null zones for a range of frequencies as was observed. Finally, all the frequencies of maximum amplitude of the observed backward propagating surface waves were consistent with the other experimental results and were close to the theoretically predicted f_{SSt}. We would like to make the comment that the ability of the theory to predict the experimental results could be improved with more research into the nature of the evolution of the Scholte–Stoneley wave velocity with frequency (dispersion) and the corrugation periodicity and form.

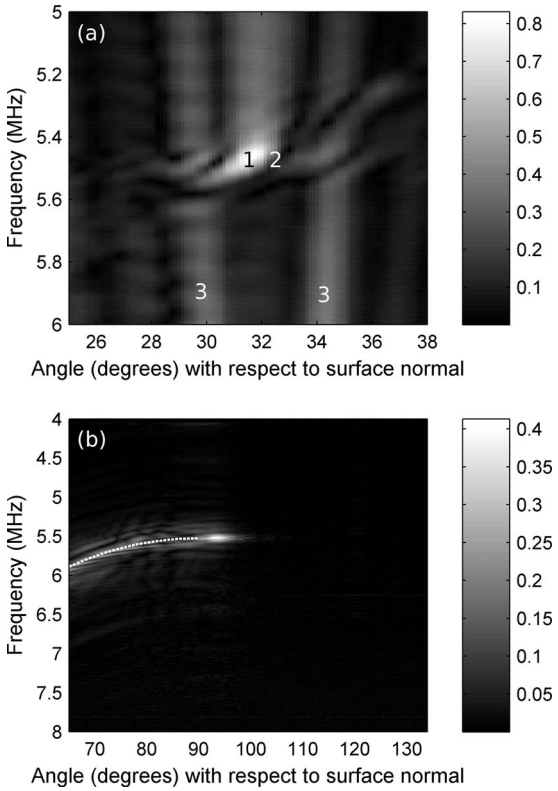


FIG. 11. (a) Angular spectrogram for a θ_i of 30.5° showing backward shifted frequencies in the region of specular reflection. (b) Complementary angular spectrogram showing the backward propagating Scholte–Stoneley wave and higher frequency propagating bulk modes (with theoretical locations shown by dotted line).

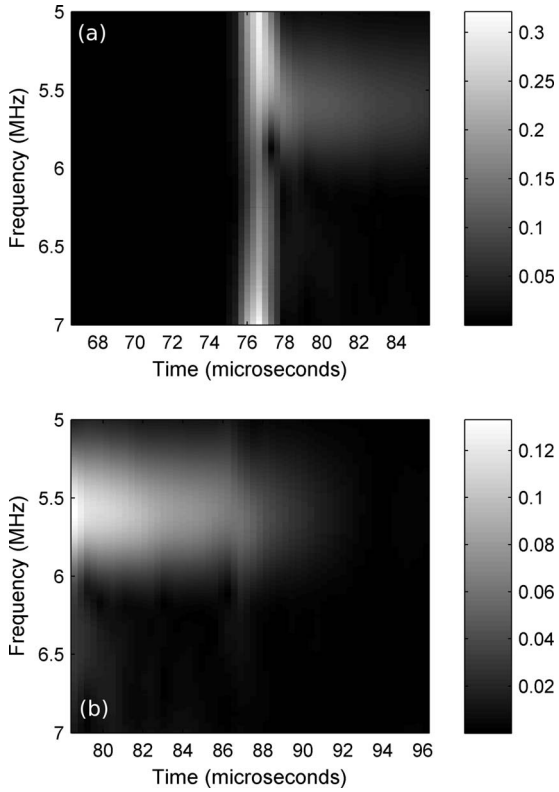


FIG. 12. (a) Classical spectrogram for a θ_i of 30.5° obtained from the time waveform captured at a receiver angle of 31.5°. (b) Close-up of the classical spectrogram in (a) to show the trailing frequencies arriving after the initial pulse.

These results lead us to conclude that the backward beam displacement can be observed for multiple angles of incidence and multiple frequencies, that the cause of the displacement is a backward propagating Scholte–Stoneley wave, and that the classical grating equation is a useful tool for predicting frequency and angle pairs that would result in the backward displacement.

V. CONCLUSIONS

New experiments concerning the ultrasonic backward beam displacement are reported in this work. Through the use of a pulse instead of a time-harmonic beam, it is possible to examine many frequencies at once. The two objectives of this work are to quantitatively show that the backward beam displacement exists at additional frequencies and angles of incidence other than the single frequency and angle of incidence pair (6 MHz, 22.5°) studied in the past and to show that the backward displacement is accompanied by a backward propagating Scholte–Stoneley wave. The experimental method employed in this work is first used to verify the original observations of Breazeale and Torbett and experimental results are then reported on additional angles of incidence where backward displaced frequencies and accompanying backward Scholte–Stoneley waves have been detected. The theory of Tamir and Bertoni[2] and the ability of the classical grating equation to predict the backward beam displacement as well as the propagation directions of bulk modes at multiple frequency and angle pairs are verified. It is possible, and highly likely, that the backward beam displacement is a continuous phenomenon that will occur for any frequency or angle of incidence on a periodically grooved solid, as long as the classical grating equation applies and is satisfied. These results have implications in the further study of periodic structures, for example, in the study of diffraction effects in phononic crystals.

TABLE I. Summary of results. Comparison of experimental observations and theoretical f_{SSt}.

Figures	Angle θ_i (deg)	Frequency of max amplitude in angular spectrogram (MHz)	Frequency range of null zone in angular spectrogram (MHz)	Trailing frequency of max amplitude in classical spectrogram (MHz)	Frequency of max amplitude of backward surface wave (MHz)	Theoretical f_{SSt} (MHz)
5 and 6	22.5	5.98	5.95–6.16	6.17	6.05	5.99
7 and 8	20	6.20	6.18–6.38	6.34	6.19	6.17
9 and 10	26.5	5.70	5.62–5.85	5.82	5.84	5.73
11 and 12	30.5	5.46	5.46–5.55	5.59	5.52	5.49

ACKNOWLEDGMENTS

The authors wish to acknowledge the French Centre Nationale de Recherche Scientifique (CNRS) and the Conseil Régional de Lorraine for their financial support of this work.

[1]M. A. Breazeale and M. A. Torbett, "Backward displacement of waves reflected from an interface having superimposed periodicity," Appl. Phys. Lett. **29**, 456–458 (1976).
[2]T. Tamir and H. L. Bertoni, "Lateral displacement of optical beams at multilayered and periodic structures," J. Opt. Soc. Am. **61**, 1397–1413 (1971).
[3]K. Mampaert and O. Leroy, "Reflection and transmission of normally incident ultrasonic waves on periodic solid-liquid interfaces," J. Acoust. Soc. Am. **83**, 1390–1398 (1988).
[4]J. V. Sánchez-Pérez, D. Caballero, R. Mártinez-Sala, C. Rubio, J. Sánchez-Dehesa, F. Meseguer, J. Llinares, and F. Gálvez, "Sound attenuation by a two-dimensional array of rigid cylinders," Phys. Rev. Lett. **80**, 5325–5328 (1998).
[5]V. Laude, M. Wilm, S. Benchabane, and A. Khelif, "Full band gap for surface acoustic waves in a piezoelectric phononic crystal," Phys. Rev. E **71**, 036607 (2005).
[6]B. Bonello, C. Charles, and F. Ganot, "Velocity of a SAW propagating in a 2D phononic crystal," Ultrasonics **44**, e1259–e1263 (2006).
[7]J. O. Vasseur, P. A. Deymier, G. Frantziskonis, G. Hong, B. Djafari-Rouhani, and L. Dobrzynski, "Experimental evidence for the existence of absolute acoustic band gaps in two-dimensional periodic composite media," J. Phys.: Condens. Matter **10**, 6051–6064 (1998).
[8]A. Khelif, A. Choujaa, S. Benchabane, B. Djafari-Rouhani, and V. Laude, "Guiding and bending of acoustic waves in highly confined phononic crystal waveguides," Appl. Phys. Lett. **84**, 4400–4402 (2004).
[9]C. Goffaux, F. Maseri, J. O. Vasseur, B. Djafari-Rouhani, and Ph. Lambin, "Measurements and calculations of the sound attenuation by a phononic band gap structure suitable for an insulating partition application," Appl. Phys. Lett. **83**, 281–283 (2003).
[10]T. Kundu, S. Banerjee, and K. V. Jata, "An experimental investigation of guided wave propagation in corrugated plates showing stop bands and pass bands," J. Acoust. Soc. Am. **120**, 1217–1226 (2006).
[11]D. Leduc, B. Morvan, A.-C. Hladky, P. Pareige, and J. L. Izbicki, "Lamb wave propagation in a plate with a grooved surface with several spatial periodicities," Ultrasonics **44**, e1359–e1363 (2006).
[12]N. F. Declercq, J. Degrieck, R. Briers, and O. Leroy, "Theoretical verification of the backward displacement of waves reflected from an interface having superimposed periodicity," Appl. Phys. Lett. **82**, 2533–2534 (2003).
[13]N. F. Declercq, J. Degrieck, R. Briers, and O. Leroy, "Theory of the backward beam displacement on periodically corrugated surfaces and its relation to leaky Scholte-Stoneley waves," J. Appl. Phys. **96**, 6869–6877 (2004).
[14]A. Teklu, M. A. Breazeale, N. F. Declercq, R. D. Hasse, and M. S. McPherson, "Backward displacement of ultrasonic waves reflected from a periodically corrugated interface," J. Appl. Phys. **97**, 084904 (2005).
[15]A. G. Every, R. E. Vines, and J. P. Wolfe, "Observation of Scholte-like waves on the liquid-loaded surfaces of periodic structures," Ultrasonics **38**, 761–766 (2000).
[16]R. Briers, O. Leroy, and G. N. Shkerdin, "Conversion of a Stoneley wave at the extremity of a fluid loaded plate," J. Acoust. Soc. Am. **101**, 1347–1357 (1997).
[17]E. Lamkanfi, N. F. Declercq, W. Van Paepegem, and J. Degrieck, "Finite element analysis of transmission of leaky Rayleigh waves at the extremity of a fluid-loaded thick plate," J. Appl. Phys. **101**, 114907 (2007).
[18]N. F. Declercq, A. Teklu, M. A. Breazeale, R. Briers, O. Leroy, J. Degrieck, and G. N. Shkerdin, "Study of the scattering of leaky Rayleigh waves at the extremity of a fluid loaded thick plate," J. Appl. Phys. **96**, 5836–5840 (2004).
[19]A. Tinel and J. Duclos, "Diffraction and conversion of the Scholte-Stoneley wave at the extremity of a solid," J. Acoust. Soc. Am. **95**, 13–20 (1994).
[20]N. F. Declercq, J. Degrieck, and O. Leroy, "The double-sided ultrasonic beam displacement," Appl. Phys. Lett. **85**, 4234–4236 (2004).
[21]W. G. Neubauer and L. R. Dragonette, "Measurement of Rayleigh phase velocity and estimates of shear speed by schlieren visualization," J. Appl. Phys. **45**, 618–622 (1974).

Resonant ultrasound studies of the layered perovskite system $Ca_{2-x}Sr_xRuO_4$

Yanbing Luan and Veerle Keppens[a)]
Department of Materials Science and Engineering, The University of Tennessee, Knoxville, Tennessee 37996-2200

Rongying Jin[b)]
Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

David Mandrus
Department of Materials Science and Engineering, The University of Tennessee, Knoxville, Tennessee 37996-2200 and Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

(Received 5 August 2009; revised 2 October 2009; accepted 6 October 2009)

The elastic response of the layered perovskite system $Ca_{2-x}Sr_xRuO_4$ ($0.2 \leq x \leq 2$) has been studied as a function of temperature and doping concentration x using resonant ultrasound spectroscopy. The elastic constants c_{11} and c_{44} have been obtained for three polycrystalline samples ($x=1.0$, 0.5, and 0.3) and show a softening trend with increasing Ca-content. In addition, the temperature-dependence of the elastic response of five single-crystals ($x=2.0$, 1.9, 0.5, 0.3, and 0.2) has been measured. For $2.0 \geq x \geq 0.5$, a dramatic softening over a wide temperature range is observed upon cooling, which is attributed to the rotational instability of RuO_6 octahedra (for $x=2.0$ and 1.9) and the static rotation of the octahedra (for $x=0.5$). For the Ca-rich samples ($x=0.3$ and 0.2), the softening occurs in a very narrow temperature range, corresponding to the structural phase transition from high-temperature tetragonal to low-temperature orthorhombic symmetry.
 [DOI: 10.1121/1.3257583]

PACS number(s): 43.35.Cg [RR] Pages: 2949–2953

I. INTRODUCTION

Sr_2RuO_4, the $n=1$ member of the Ruddlesden–Popper series $Sr_{n+1}Ru_nO_{3n+1}$, is the only copper-free superconductor that shares the layered perovskite structure with the high-T_c superconductor $La_{2-x}Ba_xCuO_4$. The discovery of exotic superconductivity in this compound by Maeno *et al.*[1] in 1994 has generated extensive research on all aspects of Sr_2RuO_4 (Refs. 2–7) and its doped varieties. The $Ca_{2-x}Sr_xRuO_4$ series, obtained by the isovalent substitution of Ca for Sr, has undoubtedly received the most attention. It is a Mott transition system, connecting the Mott insulator Ca_2RuO_4 with the spin-triplet superconductor Sr_2RuO_4 and exhibits a variety of physical properties for different x values. In 2000, the first experimental phase diagram of $Ca_{2-x}Sr_xRuO_4$ was reported by Nakatsuji *et al.*[8] and is schematically illustrated in Fig. 1. The phase diagram is known for its rich behavior, including an antiferromagnetic transition, a metal-insulator transition, a structural transition, and a ferromagnetic instability near $x \sim 0.5$. Unlike high-T_c cuprates, chemical doping of Sr_2RuO_4 eliminates superconductivity, as it is extremely sensitive to impurities.[3] Later, the magnetic phase diagram of $Ca_{2-x}Sr_xRuO_4$ (Ref. 9) was deduced from first-principles calculations and qualitatively explained the experimental phase diagram. Therefore, this is a prototype system for studying interplay between structural, electronic, and magnetic properties.

In an attempt to obtain a better understanding of the lattice dynamics of $Ca_{2-x}Sr_xRuO_4$, we have initiated a study of the elastic response of several members of the series. The elastic response is amongst a material's most important physical properties, with the elastic moduli directly related to the inter-atomic bonding and to quantities such as the velocity of sound. Measurements of the temperature-dependence of the elastic moduli can yield valuable information to fully understand the thermodynamics of a material, especially in the vicinity of phase transitions. So far, there are only a few reports[10–14] on the elastic properties of Sr_2RuO_4, and there are no reports on the elastic properties of other compositions in the $Ca_{2-x}Sr_xRuO_4$ series. The full elastic tensor of Sr_2RuO_4 single-crystal was measured at room temperature by Paglione *et al.*[12] Transverse ultrasonic measurements were performed on a single-crystal of Sr_2RuO_4 by Okuda *et al.*[14] across the superconducting transition temperature $T_c(\sim 1.40$ K), using the pulse-echo technique. Upon cooling, a jump-like decrease in T_c was found in the transverse elastic modulus c_{66}, which was attributed to the coupling between the strain and the two-dimensional superconducting order parameter (OP) with broken time-reversal symmetry.

In this paper, we present results, for the first time, on the elastic response of doped $Ca_{2-x}Sr_xRuO_4$ with $0.2 \leq x \leq 2$ as a function of temperature between 5 and 300 K. The elastic constants c_{11} and c_{44} have been obtained at room temperature

a)Author to whom correspondence should be addressed. Electronic mail: vkeppens@utk.edu
b)Present address: Louisiana State University, Baton Rouge, Louisiana 70803.

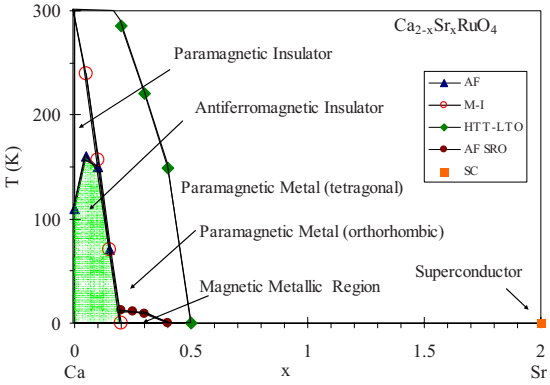


FIG. 1. (Color online) Phase diagram of $Ca_{2-x}Sr_xRuO_4$, based on data from Ref. 8.

for three polycrystalline samples ($x=1.0$, 0.5, and 0.3) and show a softening trend with increasing Ca-content. In addition, we have measured the temperature dependence of the elastic response of five single-crystals ($x=2.0$, 1.9, 0.5, 0.3, and 0.2), probing the high-temperature tetragonal to low-temperature orthorhombic phase transition.

II. EXPERIMENTAL DETAILS

Polycrystalline rods were prepared using a standard solid-state reaction with $CaCO_3$ (99.9995% Alfa Aesar), $SrCO_3$ (99.9995% Alfa Aesar), and RuO_2 (99.99% Alfa Aesar) powders as starting materials. These powders are mixed with the appropriate molar ratio, depending on the composition of the polycrystal to be made. After the mixture is well ground in a vibratory micro mill for about 1 h, the mixture is pressed into cylinder-shaped rods with a typical diameter of 5 mm and 120 mm length. The resulting polycrystalline rods are subsequently sintered at 950 °C in air for 24 h.

Single-crystals of $Ca_{2-x}Sr_xRuO_4$ are grown via the floating zone technique. A major advantage of this method is that it is crucible-free, avoiding contamination from the crucible material. A polycrystalline rod is used as the feed rod for the single-crystal growth in an NEC SC-M15HD image furnace. The seed rod can be a polycrystal or a single-crystal. During the growth, both the feed rod and the seed rod rotate in the opposite direction with the same rotation rate and the single crystal is grown in a mixed oxygen/argon atmosphere.

The current study focuses on the elastic response of the $Ca_{2-x}Sr_xRuO_4$ system, measured using resonant ultrasound spectroscopy (RUS).[15–18] RUS is a powerful method for determining the elastic moduli of a solid, based on measurements of the mechanical resonances of a sample. Whereas these "free vibrations" can easily be calculated for a sample with known elastic tensor and well-defined shape and dimensions, the RUS experiment works "backwards," and the sample's resonances are measured under nearly free boundary conditions. An iteration procedure is used to "match" the measured lines with the calculated spectrum. For samples with irregular shape or symmetry lower than orthorhombic, the procedure for calculating the elastic moduli from the resonances can be quite challenging, but RUS measurements can give important information even when it is not possible to obtain an absolute value for the elastic constants: any deviation from "normal" thermodynamic behavior will be reflected in the temperature-dependence of the resonant frequencies, and these can be measured regardless of sample shape or symmetry. In a RUS measurement, two transducers are used, one to excite the resonances and the other to detect the response of the sample. The frequency is swept within a typical range 200–2000 KHz, and a large response will be detected when the frequency corresponds to one of the sample's eigenfrequencies. Figure 2 shows a typical resonance spectrum measured using RUS and the inset shows a schematic drawing of the sample and transducer set-up.

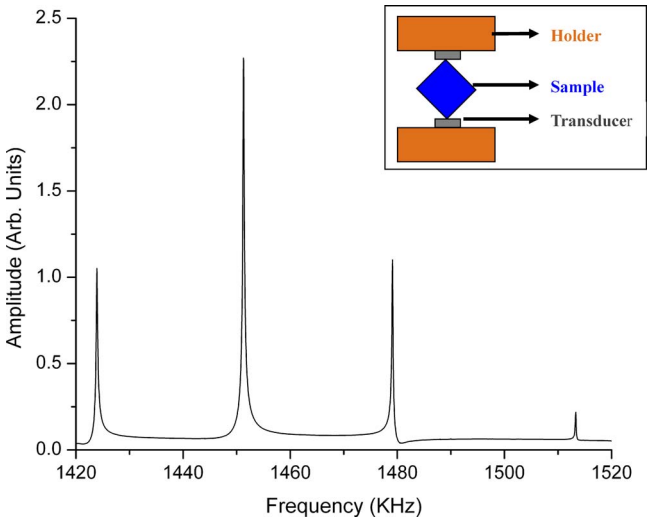


FIG. 2. (Color online) A typical resonance spectrum measured by RUS.

The RUS data reported here were carried out as a function of temperature (5–300 K) using a custom designed probe that was inserted in a commercial Quantum Design Physical Properties Measurement System (PPMS). For our study, three $Ca_{2-x}Sr_xRuO_4$ polycrystals with $x=1.0$, 0.5, and 0.3 were cut and polished into rectangular parallelepipeds (RPs) and their elastic constants were determined from measurements of the resonant frequencies. In addition, the temperature-dependence of the resonant frequencies were measured on $Ca_{2-x}Sr_xRuO_4$ single-crystals with five different compositions, i.e., $x=2.0$, 1.9, 0.5, 0.3, and 0.2. The shape and small size of these samples prevented crystallographic orientation, but valuable information can be obtained from the temperature-dependence of the frequencies, as explained below. RUS measurements were carried out upon both cooling and warming, and no difference was observed. The results presented here are based on the cooling data.

III. RESULTS AND DISCUSSION

A. $Ca_{2-x}Sr_xRuO_4$ polycrystals

Polycrystalline materials are elastically isotropic and have two independent elastic moduli: the longitudinal modulus c_{11} and the shear modulus c_{44}. Table I lists the experimental values for the elastic moduli obtained at room temperature using RUS for $Ca_{2-x}Sr_xRuO_4$ polycrystals with $x=1.0$, 0.5, and 0.3. Also listed in Table I are the calculated polycrystalline moduli for the parent compound Sr_2RuO_4 ($x=2$), using the Voigt approximation.[19] The latter allows

TABLE I. Elastic constants of $Ca_{2-x}Sr_xRuO_4$ polycrystals at room temperature. The calculated values are corrected for porosity, as explained in the text.

	C_{11} (10^{11} Pa)		C_{44} (10^{11} Pa)	
Composition	Exp.	Corrected	Exp.	Corrected
Sr_2RuO_4	2.1883[a]		0.6679[a]	
$CaSrRuO_4$	1.008	1.946	0.307	0.534
$Ca_{1.5}Sr_{0.5}RuO_4$	0.459	0.905	0.176	0.350
$Ca_{1.7}Sr_{0.3}RuO_4$	0.364	0.722	0.155	0.320

[a]Calculated by using single-crystal values in Ref. 12.

computation of the shear and bulk modulus of a polycrystalline solid based on the values of elastic moduli of the single-crystal, under the assumption that the stress is uniform everywhere within the sample. The general expressions for the Voigt approximation are as follows.

Bulk modulus,

$$B = \frac{1}{9}(C_{11} + C_{22} + C_{33}) + \frac{2}{9}(c_{12} + c_{23} + c_{13}). \quad (1)$$

Shear modulus:

$$G = \frac{1}{15}(c_{11} + c_{22} + c_{33}) - \frac{1}{15}(c_{12} + c_{23} + c_{13}) + \frac{1}{5}(c_{44} + c_{55} + c_{66}). \quad (2)$$

The c_{11} and c_{44} values of Sr_2RuO_4 polycrystal are calculated from the data reported in Ref. 12, using the above Voigt equations together with the polycrystalline relations $c_{44}=G$ and $c_{11}=B+(4G/3)$.

Before comparing the elastic moduli found for the various compositions, we need to point out that the density of the polycrystals is only about 70% of the theoretical density. As the elastic moduli of a solid depend on its density, a meaningful comparison of the experimental values requires that they are corrected to zero porosity. Following the model of Ledbetter and Datta[20] for spherical inclusions in a matrix and Chandra Sekhar *et al.*'s[21] treatment of a porous ceramic as a composite material containing spherical voids, the moduli for zero porosity $Ca_{2-x}Sr_xRuO_4$ polycrystals are estimated, and included in Table I. Figure 3 illustrates how the values of the elastic constants decrease with decreasing Sr-content.

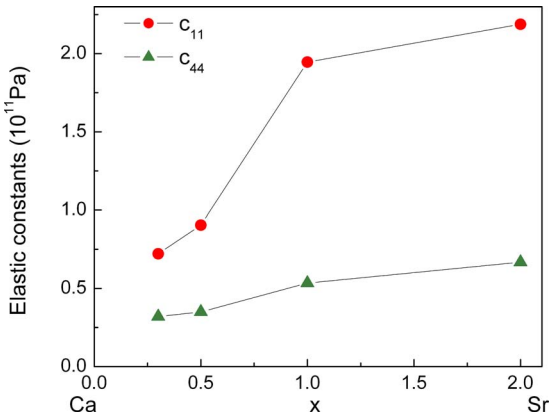


FIG. 3. (Color online) Elastic constants (corrected for porosity) versus composition x in $Ca_{2-x}Sr_xRuO_4$ polycrystals. The error for c_{11} is less than 1.5% and less than 0.3% for c_{44}.

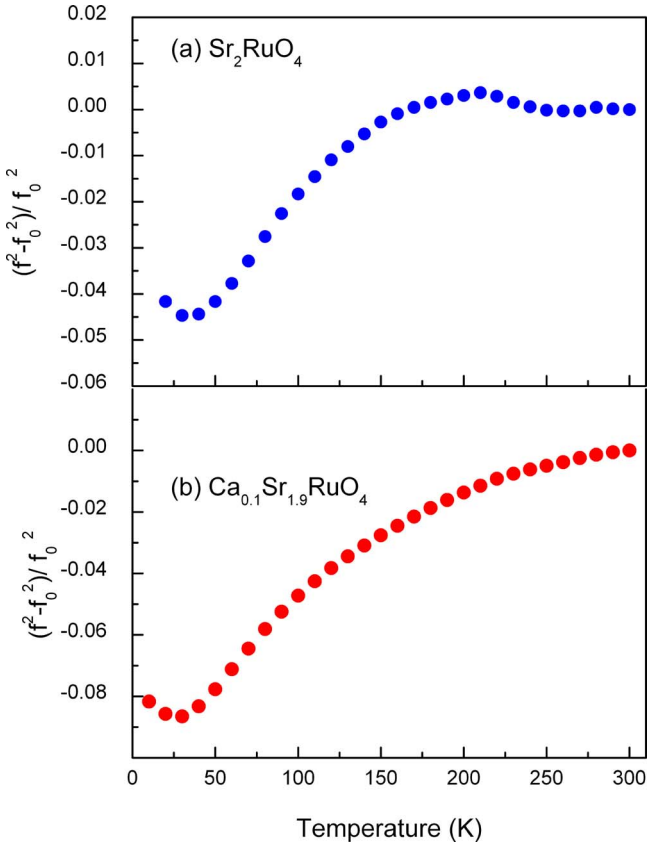


FIG. 4. (Color online) Relative change in representative squared resonant frequencies vs temperature for Sr_2RuO_4 (a) and $Ca_{0.1}Sr_{1.9}RuO_4$ (b).

Since Ca atoms are smaller than Sr atoms, the replacement of Sr by Ca atoms in Sr_2RuO_4 will cause both randomness in SrO layers and lattice distortion in RuO_2 layers, and thus induce a decrease in elastic constants.

B. $Ca_{2-x}Sr_xRuO_4$ single-crystals

The resonant frequencies of five $Ca_{2-x}Sr_xRuO_4$ single crystals (x=2.0, 1.9, 0.5, 0.3, and 0.2) were measured between 5 and 300 K (the x=0.2 composition was measured up to 350 K). The squared resonant frequencies are directly proportional to the elastic moduli[16,17] and any irregularity in the elastic response is therefore reflected in the temperature-dependence of the resonant frequencies.

The temperature dependence of representative squared resonant frequencies for the five single-crystals is plotted in Figs. 4 and 5. A significant softening is apparent in the behavior of all five crystals, but whereas this softening is rather gradual and spans more than 100 K in the compounds with $x>0.5$, it is more abrupt in the samples with high Ca-content ($x\leq 0.5$).

Starting with the parent compound, Sr_2RuO_4, a gradual softening of about 5% is observed upon cooling, reaching a minimum around 40 K, as shown in Fig. 4(a). While it may be tempting to attribute this softening to the superconductivity in this material, the two are most likely unrelated, since a similar—albeit more significant—softening is observed in the slightly Ca-doped (and non-superconducting) $Ca_{0.1}Sr_{1.9}RuO_4$ [Fig. 4(b)]. In contrast to La_2CuO_4, where

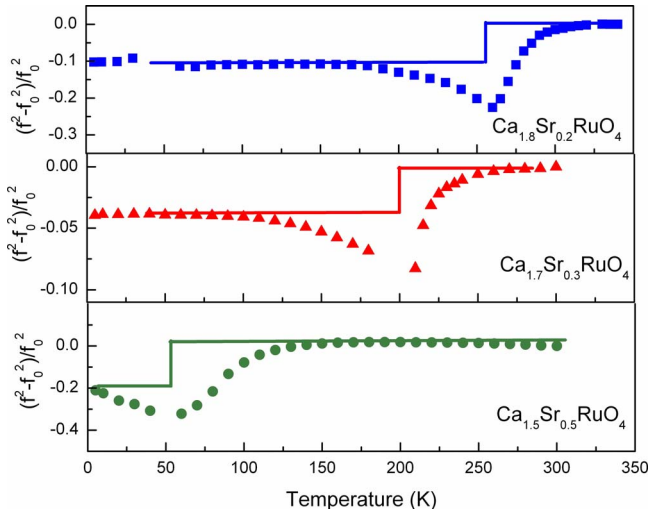


FIG. 5. (Color online) Relative change in representative squared resonant frequencies vs temperature for $Ca_{1.5}Sr_{0.5}RuO_4$ (a), $Ca_{1.7}Sr_{0.3}RuO_4$ (b), and $Ca_{1.8}Sr_{0.2}RuO_4$ (c).

tilting of the Cu–O octahedra leads to a tetragonal-to-orthorhombic phase transition, neutron diffraction on single-crystals of Sr_2RuO_4 confirms that its K_2NiF_4-type structure remains undistorted.[22] Thus the observed softening of resonant frequencies in Sr_2RuO_4 upon cooling cannot be attributed to a structural transition. Instead, we believe that the softening is due to a rotational instability of RuO_6 octahedra around the *c*-axis. Even though neither tilting nor rotation of RuO_6 is revealed in Sr_2RuO_4, large atomic displacement parameters[22] (ADPs) indicate that Sr_2RuO_4 is close to a rotational instability. Such instability has indeed been observed in the phonon dispersion relation of Sr_2RuO_4, using inelastic neutron scattering.[23] The rotation of the RuO_6 octahedra corresponds to a Σ_3 mode, whose second optical branch along [110] direction is found to soften continuously in the Brillouin zone.[23] This softening interacts with the first optical branch and the acoustic branch of Σ_3, leading to a sharp drop in the acoustic branch near the zone boundary. This steep drop in the dispersion relation of the Σ_3 mode is an indication of a rotational instability, which we believe to be the cause of the observed softening of the resonant frequencies in Sr_2RuO_4 and $Ca_{0.1}Sr_{1.9}RuO_4$ upon cooling. With increasing Ca-concentrations, the instability develops into a static rotation of RuO_6 that distorts the $I4/mmm$ symmetry and results in $I4_1/acd$ symmetry when the Sr-content decreases below $x=1.5$.[24]

Plotted in Fig. 5(a) is the temperature-dependence of a representative squared resonant frequency for $Ca_{1.5}Sr_{0.5}RuO_4$ ($x=0.5$). The softening in this sample is more pronounced, and happens in a narrower temperature region than the softening observed in Sr_2RuO_4 and $Ca_{0.1}Sr_{1.9}RuO_4$. This indicates that a different mechanism is most likely responsible for the sudden drop in resonant frequency, observed below 150 K in $Ca_{1.5}Sr_{0.5}RuO_4$. $Ca_{1.5}Sr_{0.5}RuO_4$ is known to be at the verge of a structural phase transition, from the tetragonal $I4_1/acd$ symmetry to orthorhombic $Pbca$ symmetry. As the Ca-content increases, a tilt of RuO_6 octahedra develops in addition to the already present rotation. This tilt around an

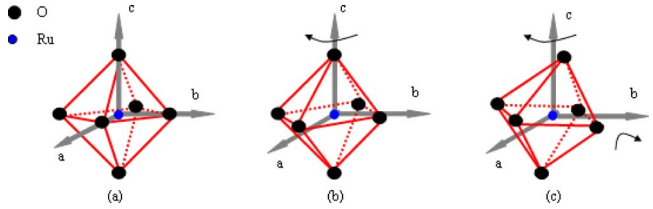


FIG. 6. (Color online) The RuO_6 octahedron in different tilt and rotational distorted configurations: (a) no distortion, (b) rotation around *c*-axis, and (c) combined tilt and rotational distortion.

axis parallel to the *ab* plane deepens the lattice distortion and leads to the structural phase transition in the Ca-rich compounds ($x<0.5$). Figure 6 illustrates the RuO_6 octahedron in different tilt and rotational configurations.

For $x=0.5$, the transition is believed to have quantum-critical character.[8] A quantum-critical phase transition takes place at the absolute zero of temperature, where crossing the phase boundary means that the quantum ground state of the system changes in some fundamental way.[25] As shown in the phase diagram (Fig. 1), the temperature of the transition increases rapidly with increasing Ca-content. Figures 5(a)–5(c) illustrate how the gradual softening below 100 K, indicating the proximity of a phase-transition in $Ca_{1.5}Sr_{0.5}RuO_4$, develops into a sharp decrease in a very narrow temperature range in $Ca_{1.7}Sr_{0.3}RuO_4$ ($x=0.3$) and $Ca_{1.8}Sr_{0.2}RuO_4$ ($x=0.2$). The minimum in the resonant frequencies corresponds to a transition temperature T_c of 50 K for $Ca_{1.5}Sr_{0.5}RuO_4$, 210 K for $Ca_{1.7}Sr_{0.3}RuO_4$, and 260 K for $Ca_{1.8}Sr_{0.2}RuO_4$, respectively. Note that $T_c>0$ K for $Ca_{1.5}Sr_{0.5}RuO_4$, indicating that the actual Ca content in our sample is slightly greater than 1.5. Nevertheless, these transition temperatures agree well with the phase diagram of $Ca_{2-x}Sr_xRuO_4$ and illustrate how RUS measurements probe structural phase transitions. In addition to the softening, a significant broadening of the resonance peaks is observed in the vicinity of the transition, indicating considerable ultrasonic attenuation.

The sharp softening of resonant frequencies across the structural transition is due to the coupling between the elastic strain and the order parameter, i.e., the tilt angle of the RuO_6 octahedra. The Landau theory of structural phase transitions predicts a step-like decrease in the elastic moduli when approaching T_c from high temperatures, as observed in $Ca_{1.7}Sr_{0.3}RuO_4$ and $Ca_{1.8}Sr_{0.2}RuO_4$, when the coupling is linear in strain and quadratic in order parameter.[26] The deviation from true step-like behavior can be attributed to the thermal fluctuations or some imperfections in the crystals. A similar behavior of the resonant frequencies was also revealed across the tetragonal-to-orthorhombic transition in the cuprates $La_{2-x}Sr_xCuO_4$ (Ref. 27) and in the pyrochlore $Cd_2Re_2O_7$.[28]

Whereas the above results clearly illustrate the ability of RUS to qualitatively probe instabilities and phase transitions in a crystal lattice, a full quantitative analysis requires measurements on oriented single-crystals. Since the resonant frequencies represent a mixture of the elastic moduli, i.e., a given frequency typically depends on a combination of elastic constants, it is not possible to evaluate the temperature-dependence of individual moduli at this time. Efforts to grow

larger single-crystals that can be oriented are currently under way, and a quantitative study of the elastic response of these crystals is planned for the future.

IV. CONCLUSIONS

In summary, we report the elastic constants c_{11} and c_{44} for three $Ca_{2-x}Sr_xRuO_4$ polycrystals ($x=1.0$, 0.5, and 0.3). It is found that the replacement of Sr by smaller Ca atoms causes a decrease in elastic constants. The temperature-dependence of the resonant frequencies for five $Ca_{2-x}Sr_xRuO_4$ single-crystals ($x=2.0$, 1.9, 0.5, 0.3, and 0.2) all show softening, due to rotation and/or tilts of the Ru–O tetrahedral, and clearly illustrate the ability of RUS to qualitatively probe instabilities and phase transitions in a crystal lattice. The gradual softening of resonant frequencies over a large temperature span in single-crystals Sr_2RuO_4 ($x=2$) and $Ca_{0.1}Sr_{1.9}RuO_4$ ($x=1.9$) is caused by a rotational instability in the lattice. The sharp softening of resonant frequencies over a very narrow temperature range for $Ca_{1.7}Sr_{0.3}RuO_4$ ($x=0.3$) and $Ca_{1.8}Sr_{0.2}RuO_4$ ($x=0.2$) corresponds to the tetragonal to orthorhombic structural phase transition, which involves both rotation and tilting of RuO_6 octahedra.

ACKNOWLEDGMENTS

We wish to thank the National Science Foundation (NSF-DMR-0804719) for the financial support for this research. Work at Oak Ridge National Laboratory was supported by the Department of Energy, BES Division of Materials Science and Engineering.

[1]Y. Maeno, H. Hashimoto, K. Yoshida, S. Nishizaki, T. Fugita, J. G. Bednorz, and F. Lichtenberg, "Superconductivity in a layered perovskite without copper," Nature (London) **372**, 532–534 (1994).
[2]T. M. Rice and M. Sigrist, "Sr_2RuO_4: An electronic analogue of ^{3}He?," J. Phys.: Condens. Matter **7**, L643–L648 (1995).
[3]A. P. Mackenzie, R. K. W. Haselwimmer, A. W. Tyler, G. G. Lonzarich, Y. Mori, S. Nishizaki, and Y. Maeno, "Extremely strong dependence of superconductivity on disorder in Sr_2RuO_4," Phys. Rev. Lett. **80**, 161–164 (1998).
[4]G. M. Luke, Y. Fudamoto, K. M. Kojima, M. I. Larkin, J. Merrin, B. Nachumi, Y. J. Uemura, Y. Maeno, Z. Q. Mao, Y. Mori, H. Nakamura, and M. Sigrist, "Time-reversal symmetry-breaking superconductivity in Sr_2RuO_4," Nature (London) **394**, 558–561 (1998).
[5]Y. Sidis, M. Braden, P. Bourges, B. Hennion, S. NishiZaki, Y. Maeno, and Y. Mori, "Inelastic neutron scattering studies on Sr_2RuO_4," Physica B **281-282**, 967–968 (2000).
[6]K. Ishida, H. Mukuda, Y. Kitaoka, Z. Q. Mao, H. Fukazawa, and Y. Maeno, "Ru NMR probe of spin susceptibility in the superconducting state of Sr_2RuO_4," Phys. Rev. B **63**, 060507 (2001).
[7]N. Kikugawa, K. Deguchi, and Y. Maeno, "Superconducting order parameter of Sr_2RuO_4," Physica C **388–389**, 483–486 (2003).
[8]S. Nakatsuji and Y. Maeno, "Switching of magnetic coupling by a structural symmetry change near the Mott transition in $Ca_{2-x}Sr_xRuO_4$," Phys. Rev. B **62**, 6458–6466 (2000).
[9]Z. Fang and K. Terakura, "Magnetic phase diagram of $Ca_{2-x}Sr_xRuO_4$ governed by structural distortions," Phys. Rev. B **64**, 020509 (2001).
[10]H. Matsui, M. Yamaguchi, Y. Yoshida, A. Mukai, R. Settai, Ya. Onuki, H. Takei, and N. Toyota, "Elastic anomalies and acoustic de Haas-van Alphen effects in Sr_2RuO_4," J. Phys. Soc. Jpn. **67**, 3687–3690 (1998).
[11]H. Matsui, Y. Yoshida, A. Mukai, R. Settai, Y. Onuki, H. Takei, N. Kimura, H. Aoki, and N. Toyota, "Acoustic de Haas-van Alphen effect and elastic anomaly of transverse mode $(C_{11}-C_{12})/2$ in Sr_2RuO_4," J. Phys. Soc. Jpn. **69**, 3769–3772 (2000).
[12]J. P. Paglione, C. Lupien, W. A. MacFarlane, J. M. Perz, L. Taillefer, Z. Q. Mao, and Y. Maeno, "Elastic tensor of Sr_2RuO_4," Phys. Rev. B **65**, 220506 (2002).
[13]N. Okuda, T. Suzuki, Z. Mao, Y. Maeno, and T. Fujita, "Elastic moduli of unconventional superconductor Sr_2RuO_4," Physica B **312–313**, 800–802 (2002).
[14]N. Okuda, T. Suzuki, Z. Mao, Y. Maeno, and T. Fujita, "Transverse elastic moduli in spin-triplet superconductor Sr_2RuO_4," Physica C **388–389**, 497–498 (2003).
[15]J. D. Maynard, "The use of piezoelectric film and ultrasound resonance to determine the complete elastic tensor in one measurement," J. Acoust. Soc. Am. **91**, 1754–1762 (1992).
[16]A. Migliori, J. L. Sarrao, W. M. Visscher, T. M. Bell, M. Lei, Z. Fisk, and R. G. Leisure, "Resonant ultrasound spectroscopic techniques for measurement of the elastic moduli of solids," Physica B **183**, 1–24 (1993).
[17]A. Migliori and J. L. Sarrao, *Resonant Ultrasound Spectroscopy*, (Wiley, New York, 1997).
[18]A. Migliori and J. D. Maynard, "Implementation of a modern resonant ultrasound spectroscopy system for the measurement of the elastic moduli of small solid specimens," Rev. Sci. Instrum. **76**, 121301 (2005).
[19]E. Schreiber, O. L. Anderson, and N. Soga, *Elastic Constants and Their Measurement* (McGraw-Hill, New York, 1973).
[20]H. M. Ledbetter and S. K. Datta, "Effective wave speeds in an SiC-particle-reinforced Al composite," J. Acoust. Soc. Am. **79**, 239–248 (1986).
[21]M. Chandra Sekhar, K. Padmavathi, J. G. Park, and P. Venugopal Reddy, "Elastic behavior of $YMnO_3$ and $ErMnO_3$ manganites," Mod. Phys. Lett. B **17**, 1119–1125 (2003).
[22]M. Braden, A. H. Moudden, S. Nishizaki, Y. Maeno, and T. Fujita, "Structural analysis of Sr_2RuO_4," Physica C **273**, 248–254 (1997).
[23]M. Braden, W. Reichardt, S. Nishizaki, Y. Mori, and Y. Maeno, "Structural stability of Sr_2RuO_4," Phys. Rev. B **57**, 1236–1240 (1998).
[24]O. Friedt, M. Braden, G. André, P. Adelmann, S. Nakatsuji, and Y. Maeno, "Structural and magnetic aspects of the metal-insulator transition in $Ca_{2-x}Sr_xRuO_4$," Phys. Rev. B **63**, 174432 (2001).
[25]S. L. Sondhi, S. M. Girvin, J. P. Carini, and D. Shahar, "Continuous quantum phase transitions," Rev. Mod. Phys. **69**, 315–333 (1997).
[26]W. Rehwald, "The study of structural phase transitions by means of ultrasonic experiments," Adv. Phys. **22**, 721–755 (1973).
[27]J. L. Sarrao, D. Mandrus, A. Migliori, Z. Fisk, I. Tanaka, H. Kojima, P. C. Canfield, and P. D. Kodali, "Complete elastic moduli of $La_{2-x}Sr_xCuO_4$ ($x=0.00$ and 0.14) near the tetragonal-orthorhombic structure phase transition," Phys. Rev. B **50**, 13125 (1994).
[28]I. A. Sergienko, V. Keppens, M. McGuire, R. Jin, J. He, S. H. Curnoe, B. C. Sales, P. Blaha, D. J. Singh, K. Schwarz, and D. Mandrus, "Metallic 'ferroelectricity' in the pyrochlore $Cd_2Re_2O_7$," Phys. Rev. Lett. **92**, 065501 (2004).

Estimating the shell parameters of SonoVue® microbubbles using light scattering

Juan Tu
Key Laboratory of Modern Acoustics of Nanjing University, Ministry of Education, Nanjing, Jiangsu 210093, People's Republic of China

Jingfeng Guan
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Yuanyuan Qiu
Key Laboratory of Modern Acoustics of Nanjing University, Ministry of Education, Nanjing, Jiangsu 210093, People's Republic of China

Thomas J. Matula[a)]
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

(Received 26 May 2009; revised 24 August 2009; accepted 6 September 2009)

Experiments were performed to measure the dynamical response of individual SonoVue® microbubbles subjected to pulsed ultrasound. Three commonly used bubble dynamic models (i.e., Hoff's, Sarkar's, and linearized Marmottant's models) were compared to determine the most appropriate model for fitting to the experimental data. The models were evaluated against published optical microscopy data. The comparison suggests that it is difficult to rank these models for lipid-shelled microbubbles undergoing small-amplitude oscillations, because under these conditions the shell parameters in these models are closely related. A linearized version of the Marmottant model was used to estimate the shell parameters (i.e., shear modulus and shear viscosity) of SonoVue® microbubbles from the experimental light scattering data, as a function of ambient microbubble radius. The SonoVue® microbubble shell elasticity and dilatational viscosity increase with ambient bubble radius, in agreement with previously published data. The results suggest that light scattering, used in conjunction with one of several popular bubble dynamics models, is effective at characterizing microbubble response and evaluating shell parameters.
 [DOI: 10.1121/1.3242346]

PACS number(s): 43.35.Ei, 43.80.Vj [CCC] Pages: 2954–2962

I. INTRODUCTION

The sensitivity and specificity of diagnostic ultrasound (US) imaging can be improved by IV injection of gas microbubbles as contrast agents.[1] These agents can enhance the acoustic backscattering from blood and produce enhancement of both Doppler flow signals and gray-scale *B*-mode images, due to the large acoustic impedance difference between the gas and the surrounding blood/tissue. Most US contrast agents (UCAs) are comprised of suspensions of micron-sized bubbles ("microbubbles"), with a stabilizing shell (e.g., albumin or lipid) and a relatively insoluble gas (e.g., perfluorocarbons). The UCA studied in this paper is SonoVue® (Bracco Diagnostics Inc., Geneva, Switzerland), one of the more commonly used new generation UCAs.[2,3] SonoVue® microbubbles, coated with a very thin lipid monolayer membrane shell encapsulating the gas SF_6, have a mean radius of around 1.5 μm and a concentration of $(2-5)\times 10^8$ bubbles/ml.

The encapsulating shell, which is the additional material present at the gas-liquid interface to stabilize UCA microbubbles against dissolution and coalescence,[4] affects the dynamical response of microbubbles. A better understanding of the physical interaction of US with UCAs will provide new insights and enable new approaches in both diagnostic and therapeutic US applications; e.g., the nonlinear behavior of oscillating microbubbles has led to the development of contrast harmonic imaging, whereas observations of microbubble destruction have led to the development of high mechanical index imaging techniques, and novel therapeutic applications such as high intensity focused ultrasound (HIFU) and targeted UCA drug/gene delivery.[5,6] Understanding shell properties is thus an important area of current investigations.

May *et al.*[7] measured the shell thickness of MRX552 microbubbles (ImaRx, Co., Tucson, AZ), a therapeutic US agent with triacetin shell, based on SEM fracture measurements. The precondition of their scanning electron microscopy (SEM) study is that MRX552 microbubble has a thick measurable fluid shell (~500 nm) with known viscosity. However, for most diagnostic UCAs with thin-shell materials

a)Author to whom correspondence should be addressed. Electronic mail: matula@apl.washington.edu

(e.g., albumin or lipid), it is impossible to employ SEM analysis. In fact, there is no method to directly measure UCA shell properties. Thus, the shell properties of UCAs can only be investigated by comparing experimental dynamical responses [e.g., $R(t)$ curve or scattering cross-section] to simulation results obtained from numerical models. This approach, measurements coupled to numerical simulations, is the focus of this paper.

To date, about a half dozen numerical models have been proposed to describe the dynamical response of shelled UCA microbubbles, with parameters for shell characteristics such as shell thickness, viscosity, stiffness, friction parameter, and surface tension.[8–13] These numerical models are usually based on approximations and scaling arguments from Rayleigh–Plesset (RP)-like equations that describe the nonlinear spherical oscillations of bubbles exposed to an external acoustic field.[14,15] The early version of these models was based on semi-empirical observations[8] by treating the shell as surface layers of elastic solids and introducing elasticity and shell friction parameters. The model was validated and the shell parameters were estimated by fitting the acoustic transmission measurements of the scattering and extinction cross-section of Albunex® microbubbles with the numerical model.[9] Church[10] regarded the shell as a continuous layer of incompressible solid elastic material by accounting for shell elasticity, in terms of a shear modulus and viscous dissipation of the shell. His theoretical derivations agreed with the experimental work of de Jong and Hoff[16] on the scattering cross-section and attenuation of Albunex®. According to Hoff's report,[11] Albunex® microbubbles have an albumin shell with the thickness of ~15 nm. However, SonoVue® microbubbles are encapsulated with a lipid monolayer membrane (~4 nm, in Hoff's report[11]), which is thinner than the albumin shell. In order to study the dynamical motion of lipid-shelled UCA microbubbles, thin-shell assumptions have been incorporated by Hoff[11] and Chatterjee and Sarkar,[12] and Morgan, *et al.*[17] All the above models regard the microbubble shell as infinitesimal or constant in thickness with a fixed surface tension coefficient. However, an oscillating UCA microbubble may express variations in surface tension during its oscillation. Marmottant *et al.*[13] proposed an improved model that described the shell elasticity with a radius-dependent surface tension over a certain range of microbubble radii (elastic range), below which the microbubble would buckle and above which the microbubble shell would break. They claimed that this model should be more precise and suitable for UCA microbubbles with thin lipid monolayers or elastic solid shells. However, the question remains whether or not the various models can be ranked in order of accuracy. Such a ranking is attempted in this paper in order to determine which model is most appropriate for fitting to the data.

Since the shell parameters cannot be measured directly, some indirect experimental approaches, including backscatter detection[18–20] and high-speed photography,[21–23] have been applied to characterize UCA shell properties and quantify the responses of UCA microbubbles. However, backscattering signals are often machine or concentration dependent;[24] and high-speed cameras are expensive and the amount of data collection is small. Furthermore, even though high-speed cameras offer a "direct" measurement of bubble size via visual inspection of the images, identifying the indistinct bubble edge generates inaccuracies in the measurement. In the present work, light scattering was used to measure the instantaneous dynamical response of individual SonoVue® microbubbles exposed to pulsed US from an *M*-mode diagnostic US system. Compared to other modalities, light scattering offers some advantages such as collecting real-time data over nearly unlimited successive acoustic pulses to an individual microbubble (to study, e.g., microbubble evolution), and to a nearly unlimited number of microbubbles (e.g., to characterize a large sample of microbubbles), and low-cost implementation. It can also be incorporated easily into an experimental setup. For example, it should be possible to analyze thousands of microbubbles in a few minutes (or even seconds), whereas a high-speed camera can take days to image and analyze just a few microbubbles.

Two major studies were preformed here: (1) A comparative study was performed to evaluate three typical shelled microbubble dynamics models by fitting the models to two types of experimental data: In one case, the experimental data were performed on BR14 microbubbles using high-speed microscopy, and come from previously published work;[23] the other case is a SonoVue® microbubble, whose dynamical response [i.e., $R(t)$ curve] was measured with light scattering. (2) The shell parameters of SonoVue® microbubbles were quantified as a function of initial ambient radius using a best-fit approach to the measured light-scattered data. The results suggest that (1) the models generate similar results, and, thus, when comparing a single data set, cannot be ranked; (2) under assumptions of thin shell and linear oscillations, the shell parameters in each of the different models are related to each other; (3) the shell parameters increase with bubble size, in agreement with previously published data; and (4) light scattering is a very useful technique for characterizing UCA microbubbles.

II. SHELLED UCA MODELS

Modeling the response of UCAs to pulsed US is nontrivial due to the addition of a coated shell. In the present work, three models (i.e., the Hoff, Sarkar, and Marmottant models) were compared to determine if a "best" model can be selected for microbubbles encapsulated with lipid monolayer membranes. While not inclusive, the models are representative. All these models are modified from a RP-like equation, which describes the response of a spherical free gas bubble to a time-varying acoustic pressure field in an incompressible liquid. A general RP-like (unshelled) equation goes like

$$\rho_L R\ddot{R} + \frac{3}{2}\rho_L \dot{R}^2 = P_g\left(\frac{R_0}{R}\right)^{3\gamma} + P_v - P_0 - \frac{2\sigma}{R} - \delta\omega\rho_L R\dot{R} - P_a\cos(\omega t), \quad (1)$$

where R_0 is the ambient bubble radius, ρ_L is the density of a Newtonian liquid, P_0 is the ambient pressure, P_v is the vapor pressure, σ is the surface tension, γ is the polytropic exponent of the gas, δ is the damping coefficient, P_a is the am-

plitude of the incident acoustic pressure, ω is the angular frequency of driving signal, and P_g is the gas pressure inside the bubble ($P_g = P_0 - P_v + 2\sigma/R_0$).

The assumptions for this equation are as follows: (1) The motion of the bubble is symmetric; (2) the wavelength of US is much larger than the bubble radius; (3) no rectified diffusion occurs; and (4) the bubble contains gas or vapor, which is compressed and behaves like an ideal gas with a constant polytropic exponent.

A. Hoff model

A classical three-region RP-like equation [Eq. (2)] describing the dynamics of shelled gas microbubbles was derived by Church.[10] In this model a continuous layer of incompressible, solid elastic shell with damping separates the gas microbubble from the bulk Newtonian liquid. The elastic surface layer stabilizes the microbubble against dissolution by supporting a strain that counters the Laplace pressure. Viscous damping is considered in this model.

$$\begin{aligned}
&\rho_s R_1 \ddot{R}_1 \left[1 + \left(\frac{\rho_L - \rho_s}{\rho_s}\right)\frac{R_1}{R_2}\right] \\
&\quad + \rho_s \dot{R}_1^2 \left\{\frac{3}{2} + \left(\frac{\rho_L - \rho_s}{\rho_s}\right)\left[\frac{4R_2^3 - R_1^3}{2R_2^3}\right]\frac{R_1}{R_2}\right\} \\
&\quad = P_{G,\mathrm{eq}}\left(\frac{R_{01}}{R_1}\right)^{3\gamma} - P_\infty(t) - \frac{2\sigma_1}{R_1} - \frac{2\sigma_2}{R_2} \\
&\quad - 4\frac{\dot{R}_1}{R_1}\left[\frac{V_s\eta_s + R_1^3\eta_L}{R_2^3}\right] - 4\frac{V_s G_s}{R_2^3}\left(1 - \frac{R_{e1}}{R_1}\right),
\end{aligned} \tag{2}$$

where R_1 and R_2 are the inner and outer radii of the elastic shell, R_{01} are R_{02} are corresponding ambient radii, ρ_s is the shell density, σ_1 is the surface tension of the gas-shell interface, σ_2 is the surface tension of the shell-liquid interface, $P_{G,\mathrm{eq}} = P_0$ for the surface layer permeable to gas, η_s and η_L are shell and liquid shear viscosities, respectively, and G_s is the shell shear modulus

$$P_\infty(t) = P_0 - P_a \sin \omega t,$$

$$V_s = R_{02}^3 - R_{01}^3,$$

and

$$R_{e1} = R_{01}\left[1 + \frac{\left(\frac{2\sigma_1}{R_{01}} - \frac{2\sigma_2}{R_{02}}\right)\frac{R_{02}^3}{V_s}}{4G_s}\right]. \tag{3}$$

Hoff[11] simplified Church's equation for the case of a bubble whose shell thickness is thin, $d_s(t) \ll R_2$,

$$\begin{aligned}
\rho_L R\ddot{R} + \frac{3}{2}\rho_L \dot{R}^2 &= P_0\left(\frac{R_0}{R}\right)^{3\gamma} - 4\eta_L\frac{\dot{R}}{R} - 12\eta_s\frac{d_s R_0^2}{R^3}\frac{\dot{R}}{R} \\
&\quad - 12G_s\frac{d_s R_0^2}{R^3}\left(1 - \frac{R_0}{R}\right) - P_0 - P_{\mathrm{drive}}(t).
\end{aligned} \tag{4}$$

Throughout this paper, Eq. (4) is referred to as the "Hoff" model. For microbubbles coated with a lipid monolayer shell, we will set $d_s = 4$ nm according to Hoff's report.[11]

B. Sarkar model

Chatterjee and Sarkar[12] developed a new model for encapsulated contrast agent microbubbles

$$\begin{aligned}
\rho_L\left(R\ddot{R} + \frac{3}{2}\dot{R}^2\right) &= \left(P_0 + 2\frac{\sigma_i}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\gamma} - 4\eta_L\frac{\dot{R}}{R} - 2\frac{\sigma_i}{R} \\
&\quad - 4\frac{\kappa_s\dot{R}}{R^2} - P_0 - P_{\mathrm{drive}}(t).
\end{aligned} \tag{5}$$

They assumed the encapsulation of a contrast agent to be an interface of infinitesimal thickness with complex interface rheological properties, where the interfacial tension (σ_i) and dilatational viscosity (κ_s) are unknown interface and shell parameters. The other symbols used are the same as above. For simplicity, we refer to Eq. (5) as the "Sarkar" model.

C. Marmottant model

Most shelled UCA models assume constant surface tension coefficients and small deformations of the microbubble surface. However, for phospholipid monolayer coatings, the surface area available per phospholipid molecule apparently varies as the microbubble oscillates. Thus, Marmottant *et al.*[13] derived an improved model specifically for microbubbles with lipid monolayer coatings. The model considers the microbubble shell as a two-dimensional viscoelastic medium and suggests that the shell elasticity can be modeled with a radius-dependent surface tension that takes into account shell buckling and rupture. For our purposes, we use the linearized version (no rupture or buckling), from Van der Meer *et al.*[23] There are two parameters to model the shell properties: the shell elastic compression modulus χ and a shell dilatational viscosity κ_s.

$$\begin{aligned}
\rho_L\left(R\ddot{R} + \frac{3}{2}\dot{R}^2\right) &= \left(P_0 + 2\frac{\sigma_w}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\gamma}\left(1 - \frac{3\gamma}{c}\dot{R}\right) - 2\frac{\sigma_w}{R} \\
&\quad - 4\chi\left(\frac{1}{R_0} - \frac{1}{R}\right) - 4\eta_L\frac{\dot{R}}{R} - 4\frac{\kappa_s\dot{R}}{R^2} - P_0 \\
&\quad - P_{\mathrm{drive}}(t).
\end{aligned} \tag{6}$$

Again, for simplicity, Eq. (6) is referred to throughout this paper as the linearized "Marmottant" model. The model has been applied very successfully to SonoVue® and BR14, from Bracco Diagnostics.

III. COMPARISON OF THREE SHELLED BUBBLE DYNAMIC MODELS

In Sec. V, it will be necessary to use a specific bubble dynamics model to fit the measured dynamical response of SonoVue® microbubbles from our light scattering system. We do not know *a priori* which model is most appropriate. The three models described above (i.e., the Hoff, Sarkar, and Marmottant models) provide a means for comparison. It will be useful to first compare these models by fitting some experimental data obtained from literature published by Van der Meer *et al.*[23] As described in their paper, high-speed

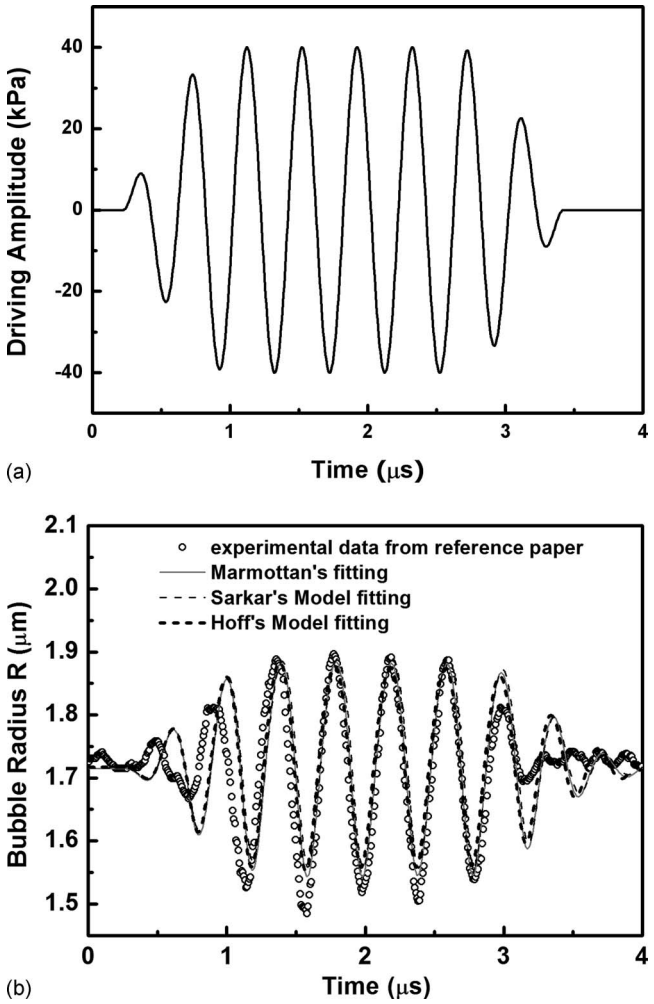


FIG. 1. (a) A Gaussian-tapered, 8-cycle, 2.5-MHz, 40-kPa forcing function is used to generate simulated acoustic microbubble responses. (b) The best-fit simulated responses for the three models are compared to each other and to the data ($R_0=1.7$ μm) available in Ref. 23. Specific values for the best fit shell parameters are given in Table I. Only slight differences between the models can be observed.

optical imaging was used to record the $R(t)$ curves for individual lipid-shelled UCA microbubbles (BR14, Bracco Diagnostics) up to 25×10^6 frames/s. Optical imaging allows direct measurement of the bubble radius, which can be an advantage over other indirect experimental methods such as acoustic backscattering and light scattering, for which a secondary means of calibrating the size is required. However, even optical imaging suffers from inaccuracies in identifying a bubble's indistinct edge. Still, it is the gold standard for measuring sizes and, thus, provides a good framework for our purposes of comparing bubble dynamics models.

For the comparison, we chose a specific $R(t)$ curve from Fig. 4 in Van der Meer *et al.*'s paper,[23] with an ambient radius of 1.7 μm. To simulate a forcing function, we assume (as they did) an 8-cycle acoustic pulse [Fig. 1(a)], tapered with a Gaussian envelope. The amplitude of the driving signal was 40 kPa with the frequency of 2.5 MHz, again as was done in Ref. 23. Since R_0 was measured to be 1.7 μm, the two-parameter fitting sets used here were (χ,κ_s), (σ_i,κ_s), and (η_s,G_s) for the linearized Marmottant, Sarkar, and Hoff models, respectively. Minimum standard deviation (STD) evaluation is applied to determine the best fitting parameters. The STD value is defined as

$$\mathrm{STD}=\frac{\sqrt{\sum_{i=t_{\mathrm{start}}}^{t_{\mathrm{end}}}(R_{\mathrm{exp},i}-R_{\mathrm{cal},i})^2}}{N}, \tag{7}$$

where t_{start} and t_{end} are the starting and ending points of the fitting region, $R_{\mathrm{exp},i}$ and $R_{\mathrm{cal},i}$ are the ith measured and simulated microbubble radius along the time span, respectively, and N is the total number of the fitting points. The following are some physical constants used for the simulation: $P_0=1.013\times10^5$ Pa, $P_v=2.33$ kPa, $\rho_g=1.161$ kg/m^3, $\sigma_w=0.072$ N/m, $\gamma=1.07$, $\rho_L=10^3$ kg/m^3, $\rho_s=1100$ kg/m^3, $c=1500$ m/s, and $\eta_L=2\times10^{-3}$ Pa s (effective liquid viscosity accounts for thermal damping[23]). The best fit results are shown in Fig. 1(b) for all three models. Table I lists the best fit shell parameters and the minimum STD values given by these three models.

The results shown in Fig. 1(b) suggest that all three models perform equally well in describing the experimental data in the central region, and all models show deviations from the experimental data at the beginning and end stages. The minimum STD values are similar for all three sets of shell parameters. The relative equality between the models suggests that any of these models can be used to fit our light scattering data.

IV. MEASUREMENTS OF SONOVUE® MICROBUBBLE DYNAMICS USING LIGHT SCATTERING

Having determined that any of these three models can be used successfully to characterize a thin lipid-shelled microbubble from *direct* optical imaging data, we attempted to determine if an identical fitting analysis can be applied to light scattering data, an *indirect* measure of a bubble's oscillation that must be converted to an $R(t)$ curve using Mie scattering theory. A modification of Mie scattering theory, applicable to a coated sphere, was first developed by Aden and Kerker,[25] and later applied to coated bubbles by Marston *et al.*[26,27] Our implementation of the model for coated SonoVue® microbubbles is based on the equations in the book by Bohren and Huffman.[28]

TABLE I. Best fit shell parameters and minimum STD for the data from Ref. 23.

UCA model	Shell elasticity	Shell viscosity	Minimum STD
Marmottant	Shell elasticity $\chi=0.25$ N/m	Dilatational viscosity $\kappa_s=4\times10^{-9}$ kg/s	0.054
Sarkar	Interfacial tension $\sigma_i=0.32$ N/m	Dilatational viscosity $\kappa_s=4\times10^{-9}$ kg/s	0.054
Hoff	Shear modulus $G_s=23$ MPa	Shear viscosity $\eta_s=0.5$ Pa s	0.059

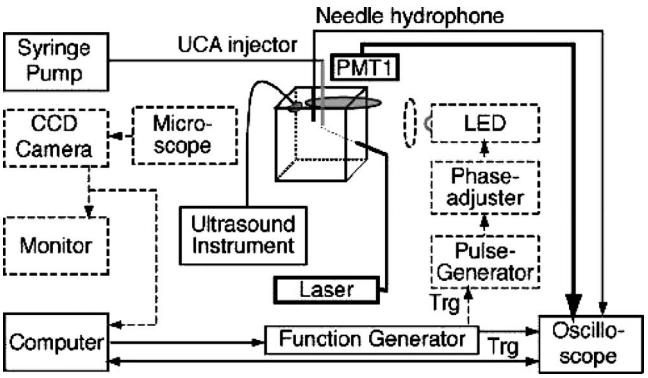


FIG. 2. Diagram of the light scattering system that was used to measure SonoVue® microbubble responses to pulsed US from a diagnostic US system. Dark solid lines refer to laser probe, whereas dashed lines refer to light scattering calibration components.

A. Experimental setup

Figure 2 illustrates the experimental setup. Details of the design and methods have been described elsewhere.[29] In brief, highly diluted SonoVue® suspensions were injected into the region of interest (ROI) (a small volume within a water tank where the US and laser beams intersect with SonoVue® microbubbles) using a syringe pump (74900 series, Cole-Parmer Instrument Co., Vernon Hills, IL) at a rate of 10 ml/h with a 0.5 mm inner-diameter tube. Acoustic interrogation pulses were sent from a probe of a diagnostic US instrument (Ultramark 4 Plus, ATL-Philips, Andover, MA) operated in *M*-mode at 1-kHz pulse-repetition-frequency and monitored using a calibrated needle hydrophone (NTR Systems Inc., Seattle, WA). Although the Ultramark 4 Plus is an older instrument, its output does more closely resemble a clinical pulse than does a research transducer.

A HeNe laser (Melles Griot, Carlsbad, CA) was used as the light scattering probe. The waist of laser beam was focused to less than 100 μm at the ROI. The scattered light signals from the microbubbles in the ROI were collected and focused onto a photo-multiplier tube (PMT) detector (Hamamatsu, model 2027). The output signals from the PMT and the hydrophone were recorded using a high-speed digital oscilloscope (LeCroy, Chestnut Ridge, NY) in sequence mode, and then transferred to a computer and post-processed using MATLAB (Mathworks Inc., Natick, MA).

Although optical imaging has the advantage of direct measurements of microbubble radius, an ultrahigh-speed camera is very expensive and not commonly found in laboratories. Previously Guan and Matula[29] showed that UCA dynamics could be measured with light scattering, and that these signals could be interpreted as volume mode oscillations of shelled microbubbles using bubble dynamics equations such as those described above. The question remains, and we will try to answer, whether or not one can use these models to accurately (or consistently) interpret the light-scattered signal to estimate shell parameters.

B. Shell properties estimation of a sample microbubble

Individual SonoVue® microbubbles were exposed to diagnostic US pulses with pressure amplitude 150 kPa and

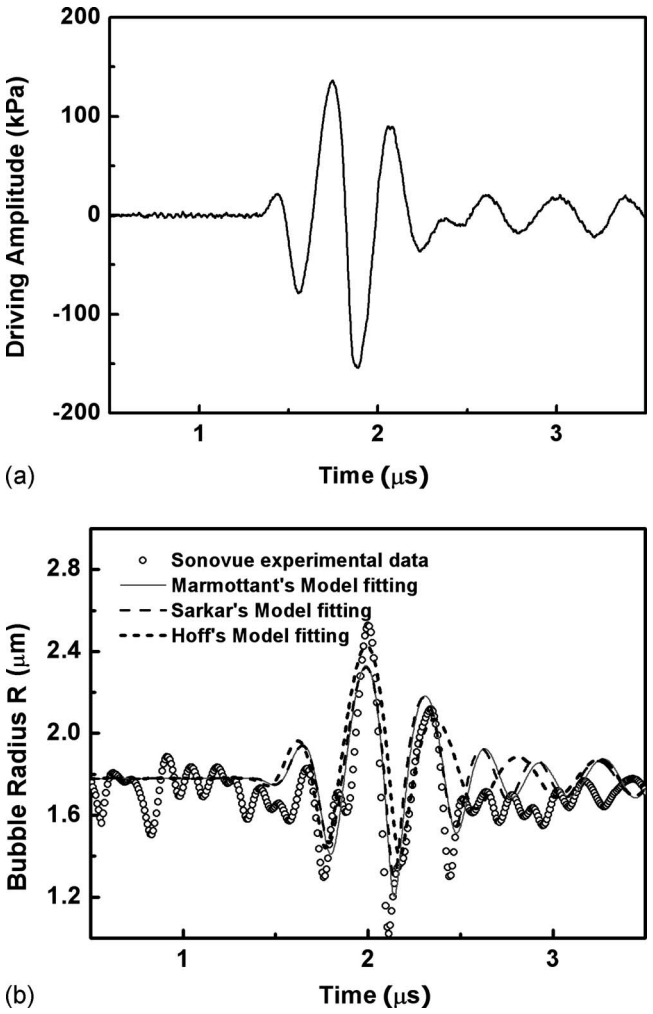


FIG. 3. (a) A 150 kPa, 2.5-MHz *M*-mode pulse from an Ultramark 4 Plus diagnostic US system is used to excite SonoVue® oscillations. (b) The best-fit simulated responses for the three models are compared to each other and to the light scattering data (R_0=1.78 μm). Specific values for the best fit shell parameters are given in Table II.

frequency 2.5 MHz [Fig. 3(a)]. The microbubble response [i.e., $R(t)$ curve] was measured with light scattering. Backlit optical imaging was used to calibrate the light scattering amplitude,[29] and for these data, $R_0 \approx 1.78$ μm. A relatively low signal/noise may affect the overall calibration constant, which will be discussed further in Sec. VI. As before, the parameter sets used to fit the data were (χ, κ_s), (σ_i, κ_s), and (G_s, η_s) for the linearized Marmottant, Sarkar, and Hoff models, respectively. The best fit $R(t)$ curves for these models are shown in Fig. 3(b). Again, we assume a constant shell thickness of 4 nm, according to the report by Hoff.[11] The corresponding best-fit shell parameters and minimum STD values are listed in Table II.

Figure 3(b) and the minimum STD values listed in Table II show that SonoVue® microbubble $R(t)$ curves can be fitted reasonably well by each of the three models. Overall, the best fit shell parameters from the light scattering data are quite close to those values obtained for BR14 microbubbles reported in Van der Meer *et al.*'s work.[23]

Comparing Fig. 1(b) with Fig. 3(b), and the minimum STDs in Tables I and II, it is clear that better fits were ob-

TABLE II. Best fit shell parameters and minimum STDs for SonoVue® microbubbles (calibrated $R_0 = 1.78\ \mu$m) measured with light scattering.

UCA model	Shell elasticity	Shell viscosity	Minimum STD
Marmottant	Shell elasticity $\chi = 0.3$ N/m	Dilatational viscosity $\kappa_s = 3.2 \times 10^{-9}$ kg/s	0.132
Sarkar	Interfacial tension $\sigma_i = 0.4$ N/m	Dilatational viscosity $\kappa_s = 4 \times 10^{-9}$ kg/s	0.134
Hoff	Shear modulus $G_s = 20$ MPa	Shear viscosity $\eta_s = 0.6$ Pa s	0.136

tained with Van der Meer's data. One possible reason for this disparity may lie in the microbubble oscillation amplitude. In Van der Meer's experiment, the maximum driving amplitude is 40 kPa. The present experiment used a relatively higher driving amplitude (150 kPa), and nonlinear effects might negate some of the assumptions (such as constant shell thickness).

V. SONOVUE® SHELL PARAMETERS VERSUS AMBIENT MICROBUBBLE SIZE

Having shown that light scattering can be used with any of the three models to estimate the shell parameters for a microbubble, we now describe light scattering experiments performed on a series of SonoVue® microbubbles, whose ambient radii were determined based on the calibration constant for the received relative scattered light intensity.[29] In this experiment, the bubbles were injected into the ROI via a syringe pump, while continuously taking data. Although we could have chosen to use any of the three models for our fitting algorithms, we used the linearized version of Marmottant's model for consistency and for later comparison (Sec. VI describes the relationship between the shell parameters for the various models so that one can convert between the various parameter sets). As before, the Ultramark 4Plus was used to excite microbubble oscillations at 150 kPa and 2.5 MHz. A total of 50 Sonovue® microbubbles were examined, although in principal, because the acquisition speed is rapid, the sample number can be increased significantly. The results, shown in Fig. 4, illustrate that both the shell elastic compression modulus χ and shell dilatational viscosity κ_s increase with increasing bubble ambient radius (R_0).

In Fig. 4(b), we add the data from Van der Meer *et al.*[23] to our own data. The comparison is quite good. The two independent methods for measuring UCA characteristics suggest that the variability in dilatational viscosity is real. These results agree with previous experimental observations on other lipid-shelled microbubbles.[17,23,30] Van der Meer *et al.*[23] suggested that the physical significance of the viscosity variation involves rheological thinning processes.

VI. DISCUSSION

A. Shell parameters

In our comparative study, a different set of shell parameters exists for each model, (χ, κ_s), (σ_i, κ_s), and (η_s, G_s), for the linearized Marmottant, Sarkar, and Hoff models, respectively. However, it is natural to ask whether or not these parameters are related. Comparing the shell parameter terms in Sarkar's model [Eq. (5)] with those in Hoff's model [Eq. (4)], we have

$$-4\kappa_s \frac{\dot{R}}{R^2} = -12\eta_s d_s \frac{R_0^2}{R^2}\frac{\dot{R}}{R^2}, \tag{8}$$

$$\frac{2\sigma_i}{R_0}\left(\frac{R_0}{R}\right)^{3\gamma} - \frac{2\sigma_i}{R} = -12G_s d_s \frac{R_0^2}{R^3}\left(1 - \frac{R_0}{R}\right). \tag{9}$$

In order to make progress we note that under small-amplitude oscillation, the shell volume can be assumed constant, in which case $V_s \approx 4\pi R^2 d_s = 4\pi R_0^2 d_{s0}$, where d_{s0} is the initial shell thickness. Using this in Eq. (8) gives

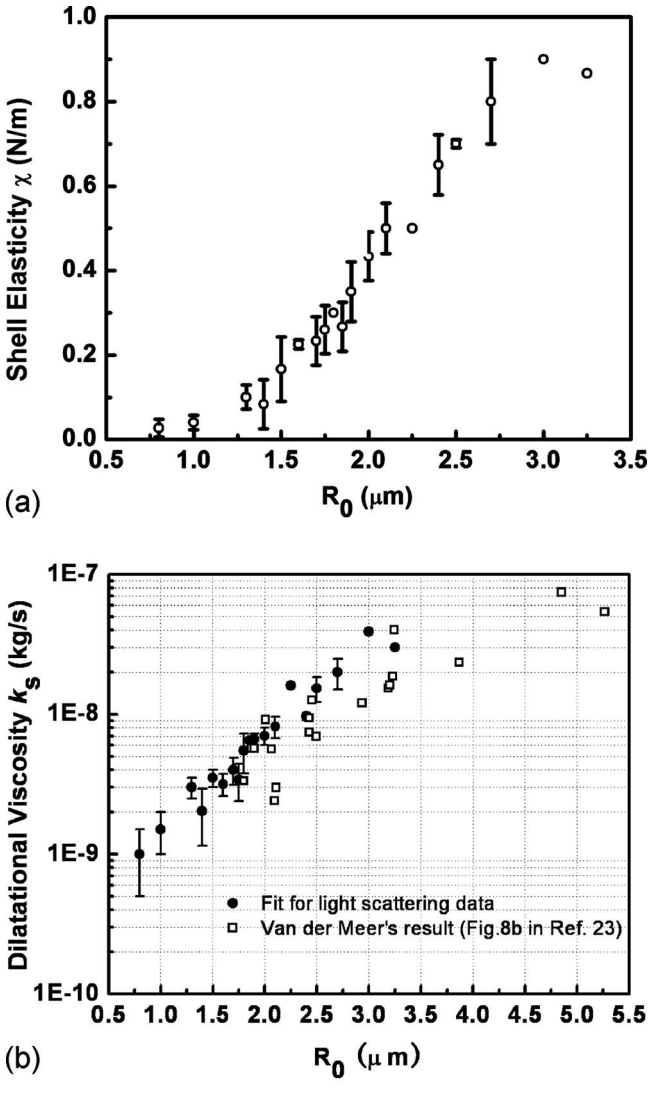


FIG. 4. SonoVue® shell parameters estimated from fitting the experimental light scattering data with the linearized Marmottant model. (a) The relationship between shell elasticity and R_0. (b) The relationship between shell dilatational viscosity and R_0. Standard deviation error bars are used when multiple bubbles are "fitted" to the same radius.

TABLE III. Best fit shell parameters and minimum STDs for SonoVue® microbubbles (best-fitted R_0 =1.72 μm) measured with light scattering.

UCA model	Shell elasticity	Shell viscosity	Minimum STD
Marmottant	Shell elasticity χ=0.22 N/m	Dilatational viscosity $\kappa_s=2.5\times10^{-9}$ kg/s	0.131
Sarkar	Interfacial tension σ_i=0.3 N/m	Dilatational viscosity $\kappa_s=3.5\times10^{-9}$ kg/s	0.132
Hoff	Shear modulus G_s=19 MPa	Shear viscosity η_s=0.4 Pa s	0.135

$$\kappa_s \approx 3\eta_s d_s (\Delta R \ll 1; d_s = d_{s0} = 4 \text{ nm}). \quad (10)$$

The fits in Table I show good agreement with this analysis.

Now, looking at Eq. (9), and allowing $\gamma=1.07\approx1$ in the left hand side, we have

$$\frac{2\sigma_i}{R_0}\left(\frac{R_0}{R}\right)^{3\gamma} - \frac{2\sigma_i}{R} \approx -2\sigma_i\left(\frac{1}{R} - \frac{R_0^2}{R^3}\right). \quad (11)$$

The right hand side of Eq. (9) is analyzed as follows: The variable χ is used in Marmottant's model to describe the shell elasticity. This elastic compression modulus is related to the shear modulus under thin-shell elasticity theory by[31]

$$\chi = 3G_s d_s. \quad (12)$$

Substituting this into the right hand side of Eq. (9) gives

$$-12G_s d_s \frac{R_0^2}{R^3}\left(1 - \frac{R_0}{R}\right) = -4\chi\left(\frac{R_0^2}{R^3} - \frac{R_0^3}{R^4}\right). \quad (13)$$

Substituting Eqs. (11) and (12) into Eq. (9) gives

$$-2\sigma_i\left(\frac{1}{R} - \frac{R_0^2}{R^3}\right) \approx -4\chi\left(\frac{R_0^2}{R^2}\frac{1}{R} - \frac{R_0}{R}\frac{R_0^2}{R^3}\right), \quad (14)$$

which, upon using the small-amplitude assumption, suggests that σ_i should be of the same magnitude as χ. Table I suggests a good agreement between the values of G_s in Hoff's model and χ in Marmottant's model (assuming a shell thickness of 4 nm) when using thin plate elasticity theory. The best-fit value of χ is also on the same order of magnitude as previously reported for SonoVue® microbubbles.[32]

The best fit shell dilatational viscosities (κ_s) for both Van de Meer's and Sarkar's model are 4 nm Pa s (viz., 4×10^{-9} kg/s), which also agrees with the results of Van de Meer *et al.*[23] Finally, we note that when applying Eq. (10), there is a reasonable agreement between the best-fit η_s in Hoff's model (0.5 Pa s) to the κ_s values found in Marmottant's and Sarkar's models (4×10^{-9} kg/s).

The above analysis suggests that although different shell parameters are applied in each model, they appear to be related to each other under the assumption of small-amplitude oscillation and constant (yet thin) shell thickness. The results are consistent with optical imaging data (Table I), as well as our light scattering data (Table II). We have not, however, performed an analysis on the shell thickness parameter, which is assumed constant at 4 nm. We hypothesize that other types of bubbles would show a similar agreement between models. This analysis is restricted to small-amplitude behavior, and thus "degeneracy" in the models might be expected. Nonlinearities introduced with stronger forcing conditions may lead to a "splitting" in the degeneracy in which the various models might show differences. Indeed, Doinikov *et al.*[33] suggested that the increase in shell shear modus and viscosity of SonoVue® bubbles with radius might result from material rheological properties of both shear-thinning and strain-softening processes. Due to complicated rheological changes in the lipid shell, SonoVue® microbubbles may generate stronger nonlinear responses when insonified by stronger US pulses, making these bubbles an ideal choice to explore nonlinear dynamical shell effects. One model ideally suited for such exploration is the full Marmottant model, which includes rupture and buckling effects—an elastic parameter that depends on the bubble radius.

B. Light scattering

Our implementation of light scattering is straightforward, and shows great promise for characterizing shelled microbubbles. The additional imaging system used for calibration (see dashed lines in Fig. 2) is somewhat inconvenient. For comparative purposes, we re-analyzed the light scattering data by including a third fitting parameter, R_0, into the routine instead of assuming the measured calibration constant. For that case, the best-fit value for R_0 is 1.72 μm, and the corresponding best-fit shell parameters and minimum STD are listed in Table III. The $R(t)$ curve using this slightly lower value for R_0 is shown in Fig. 5. Comparing to the $R(t)$ curve using the calibrated value R_0=1.78 μm [Fig. 3(b)], we note that the addition of R_0 as a fitting parameter does not appreciably change the fits to the data. The corresponding shell parameters are slightly lower (see Table III), but the trend shown in Fig. 4 does not change.

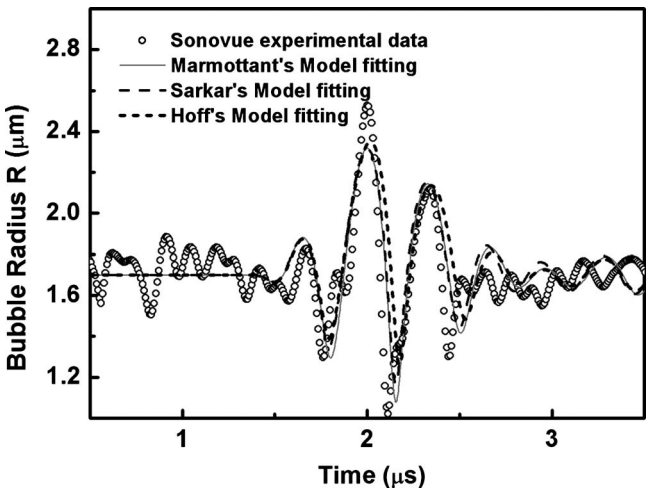


FIG. 5. The data in Fig. 3(b) are re-analyzed by adding a third fitting parameter, R_0. The best fit R_0=1.72 μm. The best fit shell parameters are listed in Table III.

There is relatively close agreement between our calibrated and fitted data. Indeed, professional multi-beam light scattering systems are already used to size microbubbles.[34,35] That a simple light scattering system can estimate microbubble size and shell parameters suggests that it should be possible to implement a light scattering system for microbubble characterization in the laboratory or industrially, for both sizing and shell characterization. In particular, with the addition of an US driver, an off-the-shelf particle sizer could be modified to characterize both size and shell properties. Such an effort is already being undertaken in our laboratory, with promising results.

VII. CONCLUSION

Three shelled UCA bubble dynamic models (i.e., the Hoff, Sarkar, and linearized Marmottant models) were compared by fitting to experimental data obtained from previously published literature. The comparison results show that it is difficult to rank these models for lipid monolayer shelled bubbles undergoing small-amplitude oscillations. Although different shell parameters are used in each of the models, they appear to be related to each other under the assumptions of small-amplitude oscillation and constant shell thickness. SonoVue® bubble shell parameters were quantified by fitting the measured $R(t)$ curves generated with light scattering to a linearized version of the Marmottant model. The results show that the shell elastic modulus χ and shell dilatational viscosity κ_s increase with increasing ambient radius. Finally, as an industrial method for characterizing microbubbles, light scattering appears to have an advantage of low cost, high speed of data acquisition and analysis, and high throughput.

ACKNOWLEDGMENTS

The authors wish to thank Peter Frinking and Bracco Imaging for supplying the SonoVue® microbubbles. We also wish to thank Michel Versluis for providing the MATLAB program for simulating the 8-cycle Gaussian windowed acoustic pulse used in Ref. 23 and shown in Fig. 2 of this paper. This work was funded in part by NIH Grant No. 5R01EB000350 (NIBIB), the National Natural Science Foundation of China (Grant Nos. 10434070 and 10704037), the Young Scholar Technological Innovation Projects of Jiangsu Province (China) Grant No. BK2007569, and the Research Fund for the Doctoral Program (for new scholar) of Higher Education of China (Grant No. 20070284070).

[1]B. B. Goldberg, J. S. Raichlen, and F. Forsberg, *Ultrasound Contrast Agents—Basic Principles and Clinical Applications*, 2nd ed. (Martin Dunitz, London, 2001).
[2]P. A. Dijkmans, L. J. M. Juffermans, R. J. P. Musters, A. van Wamel, F. J. ten Cate, W. van Gilst, C. A. Visser, N de Jong, and O. Kamp, "Microbubbles and ultrasound: from diagnosis to therapy," Eur. J. Echocardiogr. **5**, 245–256 (2004).
[3]G. Regine, M. Atzori, V. Miele, V. V. Buffa, M. Galluzzo, M. Luzietti, and L. Adami, "Second-generation sonographic contrast agents in the evaluation of renal trauma," Radiol. Med. (Torino) **112**, 581–587 (2007).
[4]A. Kabalnov, D. Klein, T. Pelura, E. Schutt, and J. Weers, "Dissoluton of multicomponent microbubbles in the bloodstream: 1. Theory," Ultrasound Med. Biol. **24**, 739–749 (1998).
[5]S. Tinkov, R. Bekeredjian, G. Winter, and C. Coester, "Microbubbles as ultrasound triggered drug carriers," J. Pharm. Sci. **98**, 1935–1961 (2009).
[6]J. E. Kennedy, "High-intensity focused ultrasound in the treatment of solid tumours," Nat. Rev. Cancer **5**, 321–327 (2005).
[7]D. J. May, J. S. Allen, and K. W. Ferrara, "Dynamics and fragmentation of thick-shelled microbubbles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1400–1410 (2002).
[8]N. de Jong, R. Cornet, and C. T. Lancée "Higher harmonics of vibrating gas-filled microspheres. Part two: Measurements," Ultrasonics **32**, 455–459 (1994).
[9]N. de Jong, R. Cornet, and C. T. Lancee, "Higher harmonics of vibrating gas-filled microspheres. Part one: Simulation," Ultrasonics **32**, 447–453 (1994).
[10]C. C. Church, "The effects of an elastic solid surface layer on the radial pulsations of gas bubbles," J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[11]L. Hoff, *Acoustic Characterization of Contrast Agents for Medical Ultrasound Imaging* (Kluwer, Dordrecht, 2001).
[12]D. Chatterjee and K. Sarkar, "A Newtonian rheological model for the interface of microbubble contrast agents," Ultrasound Med. Biol. **29**, 1749–1757 (2003).
[13]P. Marmottant, S. van der Meer, M. Emmer, M. Versluis, N. de Jong, S. Hilgenfeldt, and D. A. Lohse, "A model for large amplitude oscillations of coated bubbles accounting for buckling and rupture," J. Acoust. Soc. Am. **118**, 3499–3505 (2005).
[14]L. Rayleigh, "On the pressure developed in a liquid during the collapse of a spherical cavity," Philos. Mag. **34**, 94–98 (1917).
[15]M. S. Plesset, "The dynamics of cavitation bubbles," ASME J. Appl. Mech. **16**, 277–282 (1949).
[16]N. de Jong and L. Hoff, "Ultrasound scattering properties of Albunex microspheres," Ultrasonics **31**, 175–181 (1993).
[17]K. E. Morgan, J. S. Allen, P. A. Dayton, J. E. Chomas, A. L. Klibaov, and K. W. Ferrara, "Experimental and theoretical evaluation of microbubble behavior: Effect of transmitted phase and bubble size," IEEE Trans. Ultrason. Ferroelectr. Freq. Control. **47**, 1494–1509 (2000).
[18]R. Basude and M. A. Wheatley, "Generation of ultraharmonics in surfactant based ultrasound contrast agents: Use and advantages," Ultrasonics **39**, 437–444 (2001).
[19]P. H. Chang, K. K. Shun, and H. B. Levene, "Second harmonic imaging and harmonic Doppler measurements with Albunex," IEEE Trans. Ultrason. Ferroelectr. Freq. Control. **42**, 1020–1027 (1995).
[20]W. T. Shi, F. Forsberg, A. L. Hall, R. Y. Chia, J. B. Liu, S. Miller, K. E. Thomenius, M. A. Wheatley, and B. B. Goldberg, "Subharmonic imaging with microbubble contrast agents: Initial results," Ultrason. Imaging **21**, 79–94 (1999).
[21]P. A. Dayton, K. E. Morgan, A. L. Klibanov, G. Brandenburger, and K. W. Ferrara, "Optical and acoustical observation of the effects of ultrasound on contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control. **46**, 220–232 (1999).
[22]N. de Jong, P. J. A. Frinking, A. Bouakaz, M. Goorden, T. Schourmans, X. Jingping, and F. Mastik, "Optical imaging of contrast agent microbubbles in an ultrasound field with a 100-MHz camera," Ultrasound Med. Biol. **26**, 487–492 (2000).
[23]S. M. Van der Meer, B. Dollet, M. M. Voormolen, C. T. Chin, A. Bouakaz, N. de Jong, M. Versluis, and D. Lohse, "Microbubble spectroscopy of ultrasound contrast agents," J. Acoust. Soc. Am. **121**, 648–656 (2007).
[24]V. Sboros, K. V. Ramnarine, C. M. Moran, S. D. Pye, and W. N. McDicken, "Understanding the limitations of ultrasonic backscattering measurements from microbubble populations," Phys. Med. Biol. **47**, 4287–4299 (2002).
[25]A. L. Aden and M. Kerker, "Scattering of electromagnetic waves from two concentric spheres," J. Appl. Phys. **22**, 1242–1246 (1951).
[26]P. L. Marston, S. C. Billette, and C. E. Dean, "Scattering of light by a coated bubble in water near the critical and Brewster scattering angles," Ocean Optics IX, edited by M. A. Blizard, Proc. SPIE **925**, 308–316 (1988).
[27]P. L. Marston, "Colors observed when sunlight is scattered by bubble clouds in sea water," Appl. Opt. **30**, 3479–3484 (1991).
[28]C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1998).
[29]J. Guan and T. J. Matula, "Using light-scattering to measure the response of individual ultrasound contrast microbubbles subjected to pulsed ultrasound *in vitro*," J. Acoust. Soc. Am. **116**, 2832–2842 (2004); T. J. Matula and J. Guan, "Using light-scattering techniques to better understand the interaction of pulsed ultrasound with contrast microbubbles," Proc. SPIE **5750**, 117–126 (2005).

[30]A. A. Doinikov and P. A. Dayton, "Nonlinear dynamics of lipid-shelled ultrasound microbubble contrast agents," *Computational Methods in Multiphase Flow IV*, edited by A. A. Mammoli and C. A. Brebbia (WIT, Southampton, United Kingdom, 2007), pp. 261–270.
[31]D. Boal, *Mechanics of the Cell* (Cambridge University Press, Cambridge, 2002).
[32]M. Gorce, M. Arditi, and M. Schneider, "Influence of bubble size distribution on the echogenicity of ultrasound contrast agents. A study of SonoVue," Invest. Radiol. **35**, 661–671 (2000).
[33]A. A. Doinikov, J. F. Haac, and P. A. Dayton, "Modeling of nonlinear viscous stress in encapsulating shells of lipid-coated contrast agent microbubbles," Ultrasonics **49**, 269–275 (2009).
[34]B. E. Oeffinger and M. A. Wheatley, "Development and characterization of a nano-scale contrast agent," Ultrasonics **42**, 343–347 (2004).
[35]S. H. Bloch, R. E. Short, K. W. Ferrara, and E. R. Wisner, "The effect of size on the acoustic response of polymer-shelled contrast agents," Ultrasound Med. Biol. **31**, 439–444 (2005).

The natural frequencies of microbubble oscillation in elastic vessels

Sergey Martynov,[a] Eleanor Stride, and Nader Saffari
Department of Mechanical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom

(Received 23 March 2009; revised 10 September 2009; accepted 11 September 2009)

A theoretical model for the dynamics of a bubble in an elastic blood vessel is applied to study numerically the effect of confinement on the free oscillations of a bubble. The vessel wall deformations are described using a lumped-parameter membrane-type model, which is coupled to the Navier–Stokes equations for the fluid motion inside the vessel. It is shown that the bubble oscillations in a finite-length vessel are characterized by a spectrum of frequencies, with distinguishable high-frequency and low-frequency modes. The frequency of the high-frequency mode increases with the vessel elastic modulus and, for a thin-wall vessel, can be higher than the natural frequency of bubble oscillations in an unconfined liquid. In the limiting case of an infinitely stiff vessel wall, the frequency of the low-frequency mode approaches the well-known solution for a bubble confined in a rigid vessel. In order to interpret the results, a simple two-degree-of-freedom model is applied. The results suggest that in order to maximize deposition of acoustic energy, a bubble confined in a long elastic vessel has to be excited at frequencies higher than the natural frequency of the equivalent unconfined bubble.
 [DOI: 10.1121/1.3243292]

PACS number(s): 43.35.Pt, 43.20.Ks, 43.80.Qf, 43.80.Sh [CCC] Pages: 2963–2972

I. INTRODUCTION

A. Background and motivation

Microbubbles have become well established as contrast agents for medical ultrasound imaging (Stride and Saffari, 2003; Burns, 1996; Forsberg *et al.*, 2000; Cosgrove, 2006). They have also been investigated as vehicles for targeted drug and gene delivery, transient disruption of the blood brain barrier, thrombolysis, and high-intensity focused ultrasound therapy (Bull, 2007; Bao *et al.*, 1997; Hynynen, 2008; Tachibana and Tachibana, 1995; Clement, 2004). Their efficacy in these applications is due to the highly non-linear response of bubbles to ultrasound, which results in the appearance of subharmonic and harmonic frequency components in the scattered pressure field, and allows the echoes from the microbubbles and tissue to be better distinguished. Under resonant excitation, the most favorable conditions for acoustic energy transfer to the bubble can be achieved. For bubbles located in large blood vessels far from the vessel walls, the effect of confinement on the bubble response is negligibly small, and the resonance frequency can be estimated based on the results of studies of spherical oscillations of coated and uncoated bubbles in a large volume of liquid. However, the effect of confinement cannot be neglected for bubbles located in narrow capillaries and near walls of larger vessels. Experimental studies have shown that the presence of a confining surface can significantly change the flow field around the deforming bubble (Lauterborn and Bolle, 1975; Brujan *et al.*, 2001) as well as the bubble's acoustic response (Zhao *et al.*, 2006). Recent numerical studies also have suggested that the effects of vessel size and type of vasculature have to be considered when choosing the transmission frequency in ultrasound imaging and drug delivery with microbubbles (Sassaroli and Hynynen, 2005; Qin and Ferrara, 2007). In the dynamic system composed of a bubble and a compliant wall as two elastic structures, more than one resonance mode can be expected. To the best of the authors' knowledge, the effect of wall stiffness on the resonance modes of bubble oscillations in elastic vessels has not been quantified to this date.

B. Literature review

Miller (1979) examined the effect of wall elasticity on the radial oscillations of long cylindrical gaseous inclusions in plants, neglecting their axial dimension. It was shown that the resonance frequency of the gas inclusions increases with increasing stiffness of the vessel wall. Leighton *et al.* (1995) developed a theory for free vibrations of a one-dimensional (1D) bubble occupying the whole cross-section of a rigid tube having a divergent section at some distance away from the bubble. In order to describe more accurately the non-spherical deformations of small (millimeter-size) bubbles near confining surfaces, the boundary integral method has been used, assuming inviscid flow (Blake *et al.*, 1987, 1986; Krasovitski and Kimmel, 2001; Oguz and Prosperetti, 1998; Best and Kucera, 1992). Assuming that a blood vessel can be approximated as a rigid tube, and following the model developed by Oguz and Prosperetti (1998), Sassaroli and Hynynen (2005) analyzed the resonance frequency of a microbubble, taking into account the effects of viscous and thermal damping. The study confirmed the conclusions of Oguz and Pros-

[a]Author to whom correspondence should be addressed. Electronic mail: s.martynov@ucl.ac.uk

peretti (1998) and Leighton *et al.* (1995) that the natural frequency of a bubble in a rigid tube is reduced as a result of confinement and decreases with the vessel length.

The boundary element method has also been applied to study the potential mechanism of vessel damage by oscillations of a microbubble in a high-intensity ultrasound field (Miao *et al.*, 2008). Moderate amplitude oscillations of microbubbles can be more accurately described using the Navier–Stokes equations, which can be solved numerically using finite element or finite volume methods (Qin and Ferrara, 2007, 2006; Gao *et al.*, 2007).

The effect of vessel compliance has been studied mainly for bubbles confined in capillaries, which are the smallest vessels in the body enabling interchange of substances between the blood and body tissues. They have a simpler structure than large arteries and veins and, also, can be more easily affected by the oscillations of a single bubble. Since the elasticity of capillaries depends largely on the nature of the surrounding tissue (Fung *et al.*, 1966), several different modeling approaches have been developed to describe the structure and stiffness of the capillary wall.

Qin and Ferrara (2006) applied a lumped-parameter model from Humphrey and Na (2002) to describe the dynamics of a capillary wall as a membrane-type anisotropic elastic structure (Belardinelli and Cavalcanti, 1992) embedded in a small volume of surrounding tissue. They confirmed the results of experimental *ex vivo* observations by Caskey *et al.* (2006) that bubble oscillations in long rat cecum capillaries are less affected by the capillary length than they are in rigid vessels. Miao *et al.* (2008) assumed that the vessel wall is a finite-thickness structure possessing properties of a homogeneous isotropic elastic material. They adopted the algorithm proposed by Duncan *et al.* (1996) to couple the boundary element method solution in the fluid domain and the finite element method (FEM) solution for the structural deformations. For short vessels with lengths less than 10 vessel diameters, the bubble oscillations were found to be affected by the vessel length. It was found, as in the study by Leighton *et al.* (1995), that narrow blood vessels can be damaged by excessive wall stresses generated by collapsing bubbles. Gao *et al.* (2007) applied the pseudoplastic vessel elasticity model by Humphrey and Na (2002), taking into account the orthotropic nature of the vessel wall as a finite-thickness structure. In agreement with Qin and Ferrara (2007), they concluded that higher stiffness of the vessel wall causes a larger stress amplitude and a higher frequency for the wall oscillations. While the results of Duncan *et al.* (1996) indicate that the oscillations of a bubble near an elastic structure contain two frequency components, this effect has not been observed for bubbles confined in elastic vessels. The lumped-parameter model developed by Qin and Ferrara (2006) has been applied to study only the special case of bubble oscillations in capillaries where both thickness and elasticity change simultaneously with tumor development (Qin and Ferrara, 2006, 2007). Although elastic modulus and thickness can be considered as two independent parameters of a blood vessel, their effect on the bubble oscillations had not been quantified up to this point.

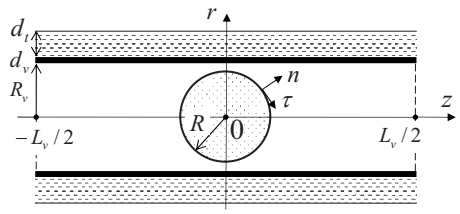


FIG. 1. Geometry and dimensions of a bubble located in the middle of a cylindrical blood vessel. r-z are the global coordinates, and n-τ are the local coordinates on the bubble surface.

C. Objectives

The objective of the present study is to characterize the frequency spectrum of the natural oscillations of a bubble confined in a vessel of a finite length and variable stiffness (elastic modulus and thickness). First, the continuum mechanics model for asymmetrical oscillations of a bubble in an elastic vessel is described. An alternative simplified model is then developed for a bubble in a vessel as a two-degree-of-freedom system. Subsequently, the results for the effect of bubble confinement in an elastic vessel on the frequency of free small-amplitude oscillations in the bubble-vessel system are presented and compared with results of analysis of normal modes using the simplified model.

II. MATHEMATICAL MODEL

A. Fluid flow

For the purpose of the present study, the oscillations of an initially spherical bubble located at the midpoint on the axis of a cylindrical blood vessel (Fig. 1) are considered. Advection of the bubble by the blood stream is neglected on the basis that the relevant time scale is much longer than that for the bubble oscillations. Assuming that in cylindrical coordinates (Fig. 1) the solution is independent of the angular coordinate, the problem is solved in two dimensions (r,z). The liquid content of the vessel (blood plasma) is modeled as a homogeneous single-phase Newtonian fluid. Assuming that the acoustic pressure wavelength is much larger than the vessel length, and liquid velocities are small compared with the speed of sound, the liquid is considered as incompressible. The fluid motion is described using the Navier–Stokes equations

$$\nabla \cdot \mathbf{u} = 0,$$

$$\rho_l \left(\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} \right) = \nabla \cdot (-p\mathbf{I} + \mu_l(\nabla \mathbf{u} + \nabla \mathbf{u}^T)), \tag{1}$$

where $\mathbf{u}$ is the velocity, p is the pressure, $\mathbf{I}$ is the identity tensor, ρ_l is the liquid density, and μ_l is the dynamic coefficient of viscosity, which in the present study are set to $\rho_l = 10^3$ kg/m^3 and $\mu_l = 10^{-3}$ Pa s.

B. Bubble motion

Since the gas in a bubble is much less dense than the surrounding liquid, and amplitudes of oscillations are assumed to be small, the inertia of the gas is neglected. The gas diffusion from the bubble is neglected as a relatively slow

process. It is also assumed that the gas is ideal and behaves isothermally ($\gamma=1$), which is an acceptable approximation for the relatively small bubble deformations expected for diagnostic ultrasound applications (Stride and Saffari, 2003). Then, the gas pressure p_g is related to the bubble volume as $p_g=p_{go}(V_o/V)^{\gamma}$, where $p_{go}=p_{lo}+2\sigma/R_o$ is the initial gas pressure inside the bubble, p_{lo} is the ambient pressure in liquid, σ is the surface tension coefficient ($\sigma=0.0643$ N/m), V is the instantaneous gas volume, and $V_o=\frac{4}{3}\pi R_o^3$ is the volume of an initially spherical bubble with radius R_o. For an arbitrary bubble shape, the volume V is calculated by applying the Gauss divergence theorem.

C. Vessel wall deformations

Following the study by Qin and Ferrara (2006) the vessel wall is described as a membrane-type elastic structure composed of a thin layer of endothelial cells embedded in tissue. The linearized form of the equation for the deformations of the vessel wall with equilibrium radius R_v, adopted in the present study, is

$$(\rho_v d_v+\rho_t d_t)\ddot{\tilde{R}}_v=p_l-p_a-\frac{E_v d_v}{(1-\nu^2)\tilde{R}_v^2}(\tilde{R}_v-R_v), \tag{2}$$

where d_v+d_t is the effective thickness of the vessel wall (the membrane wall and embedding tissue), $\tilde{R}_v$ is the instantaneous local vessel radius, ρ_v and ρ_t are the densities of the wall and embedding tissue, p_l is the pressure in the liquid, p_a is the pressure in the periphery of the vessel wall, the third term on the right-hand side describes the elastic response from the membrane wall, E_v is Young's modulus of the wall, and ν is Poisson's ratio.

Assuming equal densities of the embedding tissue and the vessel-filling liquid, the natural frequency of oscillations of a bubble in a vessel of infinitesimal stiffness is expected to match with the theoretical limit of unconfined bubble. In order to describe this effect the thin-wall vessel is modeled as surrounded by liquid with an effective thickness d_t. This extends the model by Qin and Ferrara (2007) originally proposed for vessels embedded in a small amount of embedding tissue. In order to make a comparison with the results by Qin and Ferrara (2007), the present study is performed for the equilibrium bubble radius of $R_{eq}=4$ μm and the following set of constants in Eq. (2): $R_v=5$ μm, $E_v=1$ MPa, $d_v=1$ μm, $d_t=1$ μm, $\rho_v=\rho_t=900$ kg/m^3, and $\nu=0.5$. At rest, it is assumed that $p_a=p_{lo}=0.1$ MPa so that the transmural pressure p_l-p_a is zero, and the vessel is unstrained.

In the human body, blood capillaries connect larger blood vessel (arterioles and venules) and are typically less than 1 mm in length. For bubbles confined in finite-length stiff tubes with opened ends, the mass of liquid in the tube represents a major contribution to the radiation mass of a bubble (Leighton *et al.*, 1995; Oguz and Prosperetti, 1998). The study by Caskey *et al.* (2006) has indicated that in compliant blood capillaries, the vessel length has less effect on the bubble-vessel oscillations than in rigid tubes. Qin and Ferrara (2006) showed that bubble oscillations in sufficiently long and soft vessels are not affected by the vessel length. In the present study the effect of vessel length is clarified for short vessels with walls of an arbitrary stiffness.

D. Boundary conditions

Equation (1) is closed by the boundary conditions at the vessel wall, vessel symmetry axis ($u_r=\partial u_z/\partial r=0$), vessel ends (where the ambient pressure p_{lo} is set to 0.1 MPa), and the bubble free surface. The velocity is continuous across the vessel and bubble walls. At the vessel wall, the normal stress in the liquid is balanced with the pressure p_l in Eq. (2). The dynamic boundary condition at the bubble surface states the balances for normal stress components:

$$((p_g-p_l)\mathbf{I}+\mu_l(\nabla\mathbf{u}+\nabla\mathbf{u}^T))\cdot\mathbf{n}=2\sigma\kappa\mathbf{n},$$

where $\mathbf{n}$ is the unit normal vector at the bubble surface (Fig. 1), and κ is the local mean curvature of the surface, defined as

$$\kappa=\frac{1}{2}\left(\kappa'+\frac{n_r}{r}\right).$$

Here n_r is the radial component of the unit normal vector $\mathbf{n}$, r is the local radial coordinate of the bubble surface, and κ' is the local curvature of the bubble surface in the r-z plane:

$$\kappa'=\frac{d^2r}{dz^2}\Big/\left[1+\left(\frac{dr}{dz}\right)^2\right]^{3/2}. \tag{3}$$

E. Method of solution

The governing equations subject to the above boundary conditions are solved using the FEM package COMSOL MULTIPHYSICS 3.4 (COMSOL AB, Palo Alto, CA). With the finite element method, the solution domain is covered by a triangular non-uniform mesh with higher resolution applied at the bubble surface and near the vessel wall. Motion of the free bubble surface and deformations of the vessel wall are described using a moving mesh arbitrary Lagrangian–Eulerian (ALE) algorithm. The ALE method smoothly deforms the mesh in the fluid domain by solving the Laplace equation for the mesh displacements.

In order to stabilize the non-physical pressure oscillations in the solution of the Navier–Stokes equations, the second-order elements for the velocity components and linear elements for the pressure are used. The local curvature κ' was calculated numerically using the finite-difference approximation of Eq. (3), with 120 points located on the bubble interface. It was found that doubling the number of discretization points did not change noticeably the calculated shape and the mean radius of the bubble.

For the time integration, an implicit backward-differencing second-order method was used. The time step integration was completed when the desired level of accuracy (0.1%) was achieved for the flow variables.

In order to verify the numerical model, several tests have been developed based on the available theoretical results. The results of the model validation using the solution of the Rayleigh–Plesset equation and free oscillations of a bubble in a rigid vessel (Oguz and Prosperetti, 1998) and in

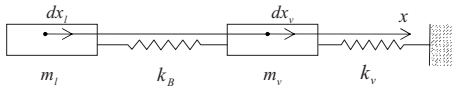


FIG. 2. A spring-mass system for analysis of the bubble oscillations in an elastic vessel. The x axis is arranged in such a way that its direction coincides with the direction of the r axis and is opposite to the direction of the z axis in Fig. 1.

a compliant vessel (Qin and Ferrara, 2007) are summarized in the Appendix.

III. APPROXIMATE MODEL WITH TWO DEGREES OF FREEDOM

A. Formulation of the model

Bubble oscillations in a vessel can also be described using a simplified approach originally proposed by Oguz and Prosperetti (1998) to model the dynamics of a one-dimensional bubble in a rigid vessel. This method is extended here to a bubble in an elastic vessel. In this case, the liquid motion in the radial direction is restricted by the elasticity and inertia of the vessel wall as well as the inertia of the liquid in the gap between the bubble and vessel wall. The system is approximately described using a model with two degrees of freedom (Fig. 2), associated with the motion of the liquid in the vessel and the vessel wall, both dynamically interacting with a massless bubble. The effect of viscous damping on the oscillations is neglected.

Following the concept introduced by Oguz and Prosperetti (1998), a bubble is considered as a massless and compressible (isothermal) cylindrical slice having the same volume as the equivalent spherical bubble $V=\frac{4}{3}\pi R^3$ and occupying the whole cross-section of a vessel of length $\Delta L = V/\pi R_v^2$. Furthermore, it is assumed that the vessel wall deforms only locally in a section occupied by a bubble of height ΔL, while away from the bubble location, the wall remains undeformed. On each side of the vessel cross-section at $z=0$, the liquid mass $m_l=\frac{1}{2}\rho_l A_v(L_v-\Delta L)$ is in contact with the cylindrical bubble slice over the area $A_v=\pi R_v^2$. The mass of the deforming section of the wall with an area $A_s=2\pi R_v \Delta L$ is $m_v=\rho_v d_v A_s$. These masses are linked to the spring elements representing a massless bubble (k_B) and an elastic vessel wall (k_v) (Fig. 2). The dynamics of this system are described by the equations

$$m_l \ddot{x}_l = -k_{B,z}(x_l - \tilde{x}_v), \tag{4}$$

$$m_v \ddot{x}_v = -k_{B,r}(x_v - \tilde{x}_l) - k_v x_v, \tag{5}$$

where x_l and x_v describe the relative displacements of masses m_l and m_v, $k_{B,r}$ and $k_{B,z}$ are the corresponding bubble stiffness coefficients defined below, and $\tilde{x}_v$ and $\tilde{x}_l$ are the relative displacements caused by the motion of masses m_v and m_l. These are expressed using the mass conservation relation for the bubble at a fixed pressure ($dV=2A_v dz+A_s dr=0$):

$$\tilde{x}_l = x_l C, \tag{6}$$

$$\tilde{x}_v = x_v / C, \tag{7}$$

where

$$C = -\frac{dr}{dz} = \frac{2A_v}{A_s} \tag{8}$$

and dr and dz are the magnitudes of the bubble deformation in r- and z-directions.

In order to derive an expression for $k_{B,z}$, the axial deformations of the bubble slice are considered assuming that the radial deformations are restricted. These deformations are associated with the axial motion of liquid with the mass m_l in contact with the bubble over the area A_v. The stiffness of a bubble in this case can be defined as $k_{B,z}=-A_v dp_i/dz$, and p_i is pressure inside the bubble (Oguz and Prosperetti, 1998). Taking into account that $A_v dz = dV/2$,

$$k_{B,z} = -2A_v^2 \frac{dp_i}{dV}. \tag{9}$$

Assuming that the volumetric stiffness of an ideal gas inside the bubble $dp_i/dV=-\gamma p/V$ is the same as for an unconstrained bubble (Oguz and Prosperetti, 1998), Eq. (9) can be expressed as

$$k_{B,z} = 2(\pi R_v^2)^2 \frac{\gamma p}{\frac{4}{3}\pi R^3} = \gamma \pi p R_v \frac{3}{2}\left(\frac{R_v}{R}\right)^3.$$

Disregarding the liquid motion in the axial direction, the system can deform along the radial coordinate. These deformations are restricted by the inertia of the vessel wall m_v and combined stiffness of the bubble $k_{B,r}$ and the vessel wall k_v. The bubble stiffness $k_{B,r}=-A_s dp_i/dR_v$ is affected by the surface area of the bubble in contact with the vessel wall A_s. Taking into account that $dV=A_s dR_v$ and using Eq. (9),

$$k_{B,r} = \frac{A_s^2}{2A_v^2} k_{B,z}. \tag{10}$$

The coefficient in the last term on the right-hand side of Eq. (2) multiplied by A_s is equal to the wall stiffness $k_v = (8/3)\pi E_v d_v (R/R_v)^3/(1-\nu^2)$.

Substitution of Eqs. (6)–(8) and (10) into Eqs. (4) and (5) gives

$$m_l \ddot{x}_l = -k_{B,r}\frac{C^2}{2}\left(x_l - \frac{x_v}{C}\right), \tag{11}$$

$$m_v \ddot{x}_v = -k_{B,r}(x_v - C x_l) - k_v x_v. \tag{12}$$

The oscillation frequency ω of the natural modes for the coupled equations (11) and (12) satisfies the characteristic equation

$$\begin{vmatrix} m_l\omega^2 - k_{B,r}\frac{C^2}{2} & k_{B,r}\frac{C}{2} \\ k_{B,r}C & m_v\omega^2 - (k_{B,r}+k_v) \end{vmatrix} = 0, \tag{13}$$

which can be solved for the two normal mode frequencies $\omega_{\text{high}}(m_l, m_v, k_{B,r}, k_v, C)$ and $\omega_{\text{low}}(m_l, m_v, k_{B,r}, k_v, C)$.

B. Limiting cases

The solution of Eq. (13) can be analyzed in the limiting cases of confinement of a bubble in infinitely stiff and infinitely soft vessels.

In the limiting case of bubble confinement in a rigid vessel ($E_v \to \infty, r \to 0$) Eq. (12) vanishes, and the characteristic frequency of the low-frequency mode is

$$\omega_{\text{rigid}}^2 = \frac{k_{B,z}}{m_l}, \tag{14}$$

which can be expressed as

$$\left(\frac{\omega_{\text{rigid}}}{\omega_\infty}\right)^2 = \frac{R_v^2}{R(L_v - \Delta L)}, \tag{15}$$

where $\omega_\infty = \sqrt{-4\pi R/\rho_l dp_i/dV}$ is the natural frequency of an unconfined bubble with negligible surface tension. This result is consistent with the approximate equation derived by Oguz and Prosperetti (1998) for a bubble located in the middle of a vessel:

$$\left(\frac{\omega_{\text{rigid}}}{\omega_\infty}\right)^2 = \frac{R_v^2}{RL_v^e}, \tag{16}$$

where $L_v^e = L_v - (4/3)R^3/R_v^2 + 4R_v$ is the equivalent length of the liquid columns on both sides of the bubble.

In another limiting case when $k_v \to 0$, the eigenfrequency is

$$\omega_{\text{soft}}^2 = \frac{k_{B,r}(m_l + m_v C^2/2)}{m_l m_v}.$$

From the practical point of view it is also interesting to consider a limit $m_l \to \infty$ ($m_l \gg m_v, z \ll r$) when the eigenfrequency of the high-frequency mode is

$$\omega_{\text{high}}^2 = \frac{k_v + k_{B,r}}{m_v}. \tag{17}$$

This gives for $k_v \to 0$ (bubble stiffness controlled oscillations):

$$\left(\frac{\omega_{\text{high},B}}{\omega_\infty}\right)^2 = \frac{1}{6}\frac{\rho_l}{\rho_v}\frac{R^2}{R_v d_v}, \tag{18}$$

and for $k_{B,r} \to 0$ (vessel stiffness controlled oscillations):

$$\left(\frac{\omega_{\text{high},v}}{\omega_\infty}\right)^2 = \frac{E_v}{3(1-\nu^2)\gamma p}\frac{\rho_l}{\rho_v}\left(\frac{R}{R_v}\right)^2. \tag{19}$$

Equation (18) shows that oscillations of bubbles with $R_v \approx R$ in thin-walled vessels ($d_v < R$) with $\rho_v \approx \rho_l$ are characterized by a natural frequency that is larger than ω_∞. It is interesting to note that Eq. (19) gives the natural frequency of the vessel wall oscillations, as described by Eq. (2).

IV. RESULTS

As previously stated in the Introduction, for the practically important case of bubble confinement in a narrow elastic vessel, the effects of individual vessel parameters on the fundamental frequency of bubble oscillations have not been investigated. This section describes the results of studies of the influence of the vessel parameters on the free response of a microbubble.

In order to obtain the natural response of the system (bubble in a vessel), a microbubble with an equilibrium radius of 4 μm was initially spherically perturbed by $\Delta R = 0.1$ μm, and then released for free oscillations. From the results of the computations, the time traces for the mean radius $R = (3V/4\pi)^{1/3}$ of an equivalent spherical bubble of volume V were obtained. A fast Fourier transform (FFT) of the radius time curves was then performed to calculate the frequency spectrum for the volumetric oscillations of the bubble.

In order to quantify the effect of vessel stiffness and confinement on the frequency spectrum of bubble oscillations, Young's modulus and the radius of the vessel wall as well as the thickness of the embedding tissue were varied. The values of Young's modulus of vascular tissues reported in the literature vary from about 0.98 MPa (or lower) to 9.6 MPa (and higher) (Duck, 1990; Yamada, 1970; Rowe *et al.*, 2003; Snowhill and Silver, 2005). Analysis of measurements of elasticity of capillaries in frog mesentery (Fung *et al.*, 1966) has shown that Young's modulus of the endothelium, considered as a membrane in a liquid, is less than 0.68 MPa. In order to cover this range, the wall elastic modulus E_v has been varied from 0.1 to 50 MPa. The thickness of the embedding tissue d_t has been varied from 0 to 4 μm. In order to quantify the effect of vessel size, the computations were performed for a vessel wall radius of 5–10 μm and a vessel length of 40–500 μm.

A. Effect of vessel elasticity and length

In an unbounded liquid, an uncoated bubble can be characterized by its radiation mass, which is a function of the bubble equilibrium size, liquid density, and stiffness, which are in turn functions of the bubble size, gas pressure, and surface tension coefficient. For a bubble in an elastic vessel, the inertia of the liquid and stiffness of the vessel wall also affect the bubble oscillations. In fact, the vessel length was shown to be a major factor influencing the period of bubble oscillations in finite-length stiff vessels (Oguz and Prosperetti, 1998; Sassaroli and Hynynen, 2004; Qin and Ferrara, 2006; Miao *et al.*, 2008). On the other hand, the numerical study by Qin and Ferrara (2006) has shown that the bubble oscillations in long and compliant vessels are not affected by the vessel length. In order to clarify the effect of vessel length on the bubble oscillations in an elastic vessel, this section compares the results obtained for relatively short ($L_v = 40$ μm) and relatively long ($L_v = 200$ μm) vessels.

Figure 3(a) shows the radial oscillation curves for a bubble of equilibrium radius $R_{\text{eq}} = 4$ μm in vessels with radius $R_v = 5$ μm, wall thicknesses $d_v = 1$ μm and $d_t = 0$, and different elasticities and lengths. It may be seen that bubble oscillations in a short and stiff vessel contain more than one mode. The high-frequency mode varies with the vessel wall stiffness, while the low-frequency mode is more affected by the vessel length.

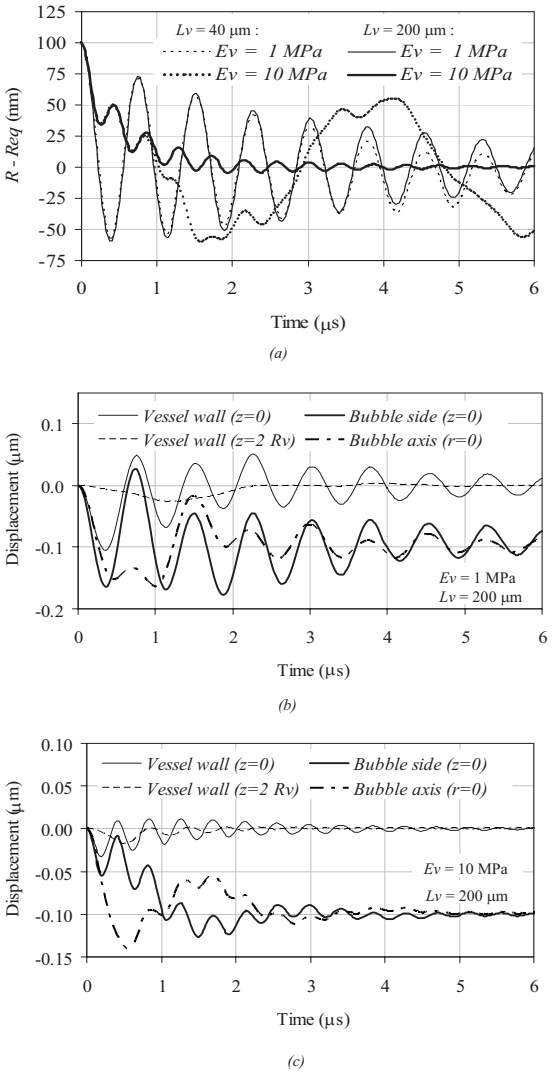


FIG. 3. (a) The effect of vessel length and wall stiffness on the volumetric oscillations of a bubble. [(b) and (c)] The radial displacements of the vessel wall and the radial and axial deformations of the bubble ($R_{eq}=4$ μm, $R_v=5$ μm, and $d_v=1$ μm).

Figures 3(b) and 3(c) show that the natural oscillations of a bubble in a narrow elastic vessel are non-spherical and induce vessel wall deformations. In the plane $z=0$ the high-frequency bubble oscillations are in phase with the vessel wall oscillations. The frequency of these oscillations is seen to increase with an increase in the vessel wall stiffness. In a stiffer vessel [Fig. 3(c)] the amplitude of the high-frequency oscillations of a bubble is reduced and the low-frequency oscillations become more pronounced. Up to about 1 μs, the radial and axial displacements of the bubble are out of phase, indicating the possible excitation of a shape mode of the bubble oscillations. The time traces of the vessel wall displacement at $z=2R_v$ show that the wall deforms non-locally. The phenomenon of wave propagation in the elastic vessel can affect the bubble oscillations and will be addressed in future studies.

Analysis of the spectral frequency content of the radius time traces plotted in Fig. 3(a) was obtained using the FFT algorithm. Figure 4 shows the spectra of bubble oscillations in a short vessel ($L_v=40$ μm). The spectrum of a bubble in

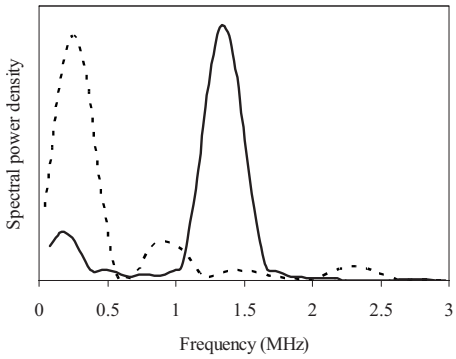


FIG. 4. The effect of the vessel wall stiffness on the spectrum for the radial oscillations of a bubble of equilibrium radius $R_{eq}=4$ μm in a vessel with $R_v=5$ μm, $d_v=1$ μm, and $L_v=40$ μm. Solid curve, $E_v=1$ MPa; dotted curve, $E_v=10$ MPa.

a soft vessel ($E_v=1$ MPa) has two local peaks (characteristic frequencies or modes): one at about 0.17 MHz and another at about 1.33 MHz. The amplitude of the high-frequency mode is higher than that of the low-frequency mode. For a stiff vessel ($E_v=10$ MPa), the low-frequency mode has the highest amplitude, while the lower-amplitude high-frequency content of the spectra contains three local maxima. The highest-frequency mode appears at about 2.3 MHz. In short vessels, the low-frequency oscillations become more pronounced with increasing vessel stiffness, while in long vessels the low-frequency mode can be significantly suppressed by viscous damping in the liquid. This is clearly seen from a comparison of the curves $L_v=40$ μm and 200 μm for $E_v=10$ MPa in Fig. 3(a). Figure 4 also shows the presence of another, third, natural frequency mode at about 0.9 MHz, which may be a nonlinear harmonic produced by interference of the low-frequency second harmonic and subharmonic of the high-frequency mode. The results based on the FEM calculations for $d_t=0$ show that the high-frequency mode is almost unaffected by the vessel length. At the same time the low-frequency mode can be over-damped in long compliant vessels ($L_v=200$ μm) [see also Fig. 3(a)].

From the frequency spectra, calculated for various combinations of the vessel parameters, the natural frequencies have been identified. Figure 5 shows frequencies for the high and low natural frequency modes as a function of the vessel wall elastic modulus. It can be seen that the frequency of the high-frequency mode increases with the vessel stiffness and is larger than the natural frequency of an unconfined bubble:

$$f_\infty = \frac{1}{2\pi R}\sqrt{\frac{3\gamma p_{lo} + 2(3\gamma - 1)\sigma/R}{\rho_l}}, \tag{20}$$

which for a bubble with radius $R=4$ μm gives $f_\infty \approx 0.76$ MHz.

B. Effect of the vessel radius

As discussed earlier, the low-frequency component of natural oscillations is more pronounced for stiffer and shorter vessels, while in soft elastic vessels the high-frequency oscillation mode is dominant. For rigid vessels the variation in the natural frequency with the vessel diameter and length has already been studied by Oguz and Prosperetti (1998). For

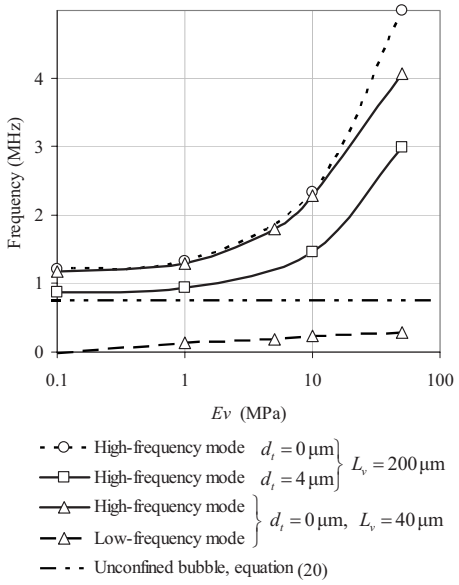


FIG. 5. The effect of the vessel wall elastic modulus, thickness of the embedding tissue, and vessel length on the natural frequency of the modes of bubble oscillation (R_{eq}=4 μm, R_v=5 μm, and d_v=1 μm).

vessels of finite stiffness, the effect of vessel size on the high-frequency natural oscillations remains unclear and is addressed in this section. For this purpose the vessel radius has been varied from 5 to 10 μm (Fig. 6). It is seen from Fig. 6 that for relatively soft vessels with $E_v < 1$ MPa, the natural frequency is affected more by the vessel diameter than the elastic modulus E_v. The natural frequency increases by less than 8% for a tenfold increase in E_v, by about 20% when the vessel radius is reduced from infinity to 10 μm, and by about 35% when the vessel radius is further reduced from 10 to 5 μm. For stiffer vessels, with $E_v > 1$ MPa, the vessel wall Young's modulus and diameter both have comparable effects on the natural frequency. Thus, in a vessel with radius R_v=7 μm, an increase in the vessel elastic modulus from 1 to 10 MPa gives an increase of approxi-

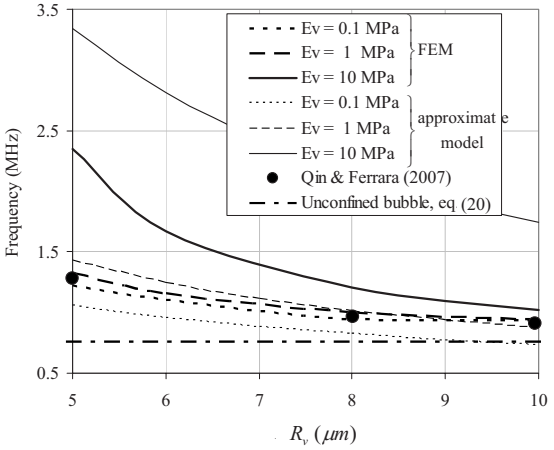


FIG. 6. The effect of the vessel radius and elastic modulus on the natural high-frequency mode of bubble oscillations in a long vessel (R_{eq}=4 μm, d_v=1 μm, and L_v=500 μm). Comparison of the results of computations using the FEM and 1D models, and the results from Fig. 5 in the paper by Qin and Ferrara (2007).

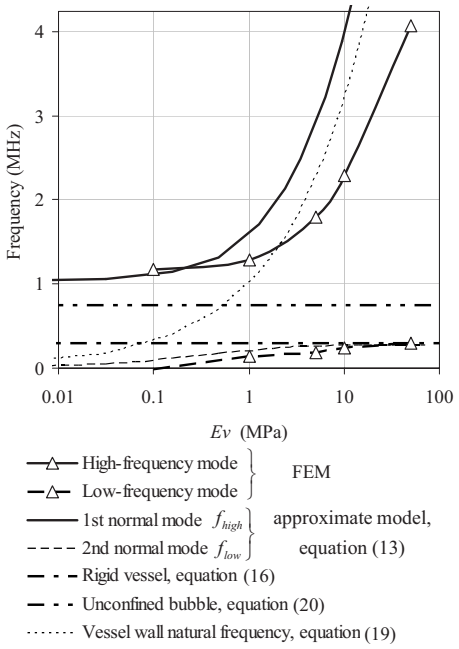


FIG. 7. The effect of the vessel wall elastic modulus on the natural frequency of the modes of bubble oscillation (R_{eq}=4 μm, R_v=5 μm, L_v =40 μm, and d_v=1 μm).

mately 20% in the natural frequency; a similar effect can be achieved by a reduction in the vessel radius from 7 to 5 μm when E_v=1 MPa.

C. Effect of thickness of the embedding tissue

The effect of the surrounding tissue on the bubble natural frequencies was studied by varying the effective thickness of the embedding tissue d_t from 0 to 4 μm. Figure 5 shows that with an increasing amount of embedding tissue (d_t), the frequency of the high-frequency mode decreases by about 25%–30%. A large amount of embedding material can cause a significant reduction in the frequency of oscillations due to a large mass loading term $\rho_t d_t$ in Eq. (2). To correctly model the case when the vessel is embedded in a liquid ($\rho_t = \rho_l$) it is essential to choose the thickness d_t that would guarantee that in the limiting case of an infinitely soft vessel ($E_v \to 0$) the natural frequency tends to f_∞. For the conditions in Fig. 5, this maximum thickness d_t is close to 4 μm.

D. Comparison with the approximate model

Figure 7 shows the natural frequencies calculated from Eq. (13) for a bubble with equilibrium radius R_{eq}=4 μm and a vessel with R_v=5 μm, L_v=40 μm, and elastic modulus of the vessel wall E_v in the range from 0.01 to 100 MPa. It is seen that the high-frequency branch follows qualitatively the trend predicted from the FEM computations. For large E_v Eq. (19) provides a good approximation for the natural frequency mode f_{high} described by Eq. (13). For small E_v, the larger difference between the f_{high} modes described by Eqs. (13) and (19) can be attributed to the fact that the contribution of the bubble stiffness was neglected in approximation (19). In the limit when $E_v \to 0$, the FEM results predict the high-frequency mode in agreement with Eq. (18).

For a finite vessel wall stiffness, Eqs. (14) and (18) overestimate the natural frequency. This can partially be explained by the fact that the inertia of the liquid in the gap between the bubble and the vessel wall was neglected in Eq. (5). Also, Eq. (14) predicts the undamped natural frequency, while analysis based on Eq. (1) gives the frequency damped by the liquid viscosity effect.

Figure 6 shows that in agreement with the FEM results, the approximate model predicts ω_{high} decreases as the vessel radius increases. This trend is also described by Eqs. (18) and (19).

In Fig. 7, the low-frequency mode f_{low} as predicted by Eq. (13) agrees with the results of FEM computations, and, at large E_v, both solutions converge to the limit given by Eq. (16).

The two modes of oscillation described by the two-degree-of-freedom model correspond to the in-phase and out-of phase motions of the masses m_l and m_v. The model takes into account the bubble axial and radial motions, but does not differentiate between the shape and volume oscillations of a bubble. For a bubble confined in a vessel with infinitely soft and absolutely rigid walls the model adequately captures the bubble shape deformations. Thus, in an ideally soft vessel in vacuum (no embedding tissue) the bubble deforms into an oblate spheroid under high-frequency oscillations in phase with the vessel wall. In a rigid vessel, the low-frequency mode characterizes periodical elongations of a bubble in the axial direction.

V. DISCUSSION

Earlier studies have shown that bubble oscillations in an unbounded liquid and in rigid vessels are characterized by a single natural frequency. The results obtained in the present paper indicate that bubble oscillations in a finite-length elastic vessel have a spectrum of frequencies with distinguishable high-frequency and low-frequency modes. Section IV D explains that these frequencies appear as a result of the interaction of a compressible bubble with an elastic vessel wall and liquid column in a vessel. The natural frequencies are shown to increase with an increase in the stiffness of the vessel wall. In contrast with the high-frequency mode, the low-frequency mode of the bubble oscillations in a compliant vessel was shown to be significantly affected by the vessel length. The low-frequency oscillations require a longer period of observation to be detected in softer vessels and can be significantly damped in long vessels.

The FEM results also showed the presence of mid-frequency components, which may result from the non-linear interaction of the harmonics and sub-harmonics of the first two modes. The harmonic and sub-harmonic components could be further suppressed if the effect of viscous damping is considered for the vessel wall. Since there is a lack of data in the literature about the viscous properties of blood vessels, the viscous damping in the vessel wall has been ignored in the present study. These effects have to be addressed in future studies.

In the present consideration it was assumed that the bubble is located in the midpoint of the vessel axis. Breaking this symmetry would induce other high-frequency translational bubble oscillations, as indicated in studies by Sassaroli and Hynynen (2004) and Duncan *et al.* (1996).

The effect of a bubble coating on the oscillations of confined bubbles is the subject of current research by the authors. For coated microbubbles, the higher stiffness of the coating increases the natural frequency of oscillations of an unconfined bubble, and a similar proportional increase in resonance frequency is expected for confined bubbles.

VI. CONCLUSIONS

It has been shown that the oscillations of a bubble in a finite-length elastic vessel are characterized by a spectrum of frequencies with two distinguishable natural frequencies. The frequency of the high-frequency mode increases with the vessel elastic modulus and, for a thin-wall vessel, can be higher than the natural frequency of bubble oscillations in an unconfined liquid. In the limiting case of an infinitely stiff vessel wall, the frequency of the low-frequency mode approaches the well-known solution for bubbles confined in a rigid vessel. In order to interpret the results, a simple model has been applied, which assumes that oscillations can be approximately described using a system with two degrees of freedom.

The results suggest that in order to maximize the amplitude of bubble oscillations in an elastic vessel, it has to be excited at a frequency higher than the natural frequency of an unconfined bubble.

ACKNOWLEDGMENTS

The authors would like to thank the Engineering and Physical Sciences Research Council for supporting this work through Grant No. EP/E029310/1, Professor A. Prosperetti for stimulating discussions, Dr. D. Sinden for useful comments on the manuscript, and Dr. G. ter Haar and her team at the Institute of Cancer Research.

APPENDIX: MODEL VALIDATION

1. Spherical oscillations of an unconfined bubble

For the purposes of validation, initially the spherically symmetric solution for an unconfined bubble was obtained using the Rayleigh–Plesset equation (Leighton, 1994)

$$\rho_l\left(R\ddot{R}+\frac{3}{2}\dot{R}^2\right)=p_{go}\left(\frac{R_o}{R}\right)^{3\gamma}-p_{lo}-4\mu_l\frac{\dot{R}}{R}. \tag{A1}$$

Equation (A1) was solved using MATLAB® R2006b for a bubble with an equilibrium radius of $R_{\text{eq}}=4$ μm, zero surface tension, and equilibrium liquid pressure of 0.1 MPa. In COMSOL, in order to minimize the influence of a finite size of the liquid domain on the bubble oscillations, the radius of the external boundary was specified to be 100 times larger than the bubble radius (Popinet and Zaleski, 2002).

Figure 8 shows the time variations in the bubble radius as predicted by the finite element model in comparison with the solution of Eq. (A1). From Fig. 8 it is seen that the finite

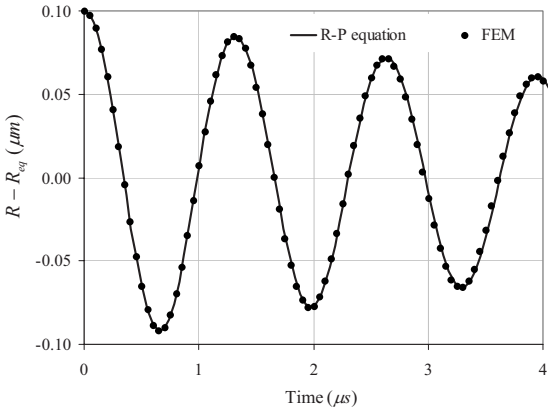


FIG. 8. Radius vs. time curves for free oscillations of a spherical microbubble of equilibrium radius $R_{eq}=4$ μm based on the solution using the FEM model (points) and the Rayleigh–Plesset equation (A1) (solid line).

element model predicts the bubble radius curves in good agreement with Eq. (A1). The discrepancy between the two solutions was found to be less than 5%.

2. Bubble oscillations in a rigid vessel

Oguz and Prosperetti (1998) [and earlier Devin (1961)] showed that the natural frequency of a bubble located on the axis of a rigid cylindrical tube decreases with the tube length, as described by Eq. (16). Equation (16) has been shown to be valid for relatively large bubbles with a zero surface tension and radii larger than $0.2R_v$. Figure 9 shows the variation in the bubble natural frequency with the vessel length according to Eq. (19) in comparison with the calculations based on the FEM results. Computations were performed for a bubble with initial radius of 3 μm, vessel radius of 4 μm, and relative vessel length of $2L_v/R_v=75$, 150, and 300. It is seen that the finite element model predicts a natural frequency, which is in good agreement with that obtained from the theory by Oguz and Prosperetti (1998).

3. Bubble oscillations in an elastic vessel

Figure 6 compares the results from this study with the data reported by Qin and Ferrara (2007) for "compliant" capillaries unaffected by the tumor. Comparison is performed for a bubble with radius of 4 μm and $\sigma=0.0643$ (Pa m) placed in a vessel with elastic modulus $E_v=1$ MPa and

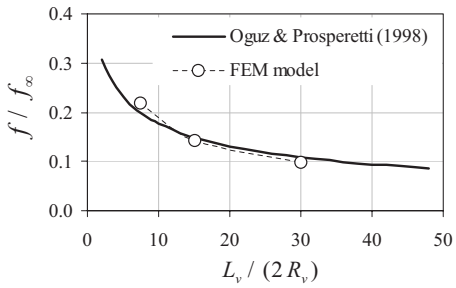


FIG. 9. The effect of vessel length on the natural frequency of a bubble confined in a rigid vessel. Comparison of the frequency obtained from the finite element solution (points) and the theory by Oguz and Prosperetti (1998) (solid line).

length of 500 μm. It is seen that the present results are in very good agreement with those predicted by Qin and Ferrara (2007).

Bao, S. P., Thrall, B. D., and Miller, D. L. (**1997**). "Transfection of a reporter plasmid into cultured cells by sonoporation in vitro," Ultrasound Med. Biol. **23**, 953–959.

Belardinelli, E., and Cavalcanti, S. (**1992**). "Theoretical-analysis of pressure pulse-propagation in arterial vessels," J. Biomech. **25**, 1337–1349.

Best, J. P., and Kucera, A. (**1992**). "A numerical investigation of nonspherical rebounding bubbles," J. Fluid Mech. **245**, 137–154.

Blake, J. R., Taib, B. B., and Doherty, G. (**1986**). "Transient cavities near boundaries. 1. Rigid boundary," J. Fluid Mech. **170**, 479–497.

Blake, J. R., Taib, B. B., and Doherty, G. (**1987**). "Transient cavities near boundaries. 2. Free-surface," J. Fluid Mech. **181**, 197–212.

Brujan, E. A., Nahen, K., Schmidt, P., and Vogel, A. (**2001**). "Dynamics of laser-induced cavitation bubbles near elastic boundaries: Influence of the elastic modulus," J. Fluid Mech. **433**, 283–314.

Bull, J. L. (**2007**). "The application of microbubbles for targeted drug delivery," Expert Opinion on Drug Delivery **4**, 475–493.

Burns, P. N. (**1996**). "Harmonic imaging with ultrasound contrast agents," Clin. Radiol. **51**, 50–55.

Caskey, C. F., Kruse, D. E., Dayton, P. A., Kitano, T. K., and Ferrara, K. W. (**2006**). "Microbubble oscillation in tubes with diameters of 12, 25, and 195 microns," Appl. Phys. Lett. **88**, 033902.

Clement, G. T. (**2004**). "Perspectives in clinical uses of high-intensity focused ultrasound," Ultrasonics **42**, 1087–1093.

Cosgrove, D. (**2006**). "Ultrasound contrast agents: An overview," Eur. J. Radiol. **60**, 324–330.

Devin, C. (**1961**). "Resonant frequencies of pulsating air bubbles generated in short, open-ended pipes," David Taylor Model Basin, Hydromechanics Laboratory Technical Report No. 1522.

Duck, F. A. (**1990**). *Physical Properties of Tissue* (Academic, London).

Duncan, J. H., Milligan, C. D., and Zhang, S. (**1996**). "On the interaction between a bubble & a submerged compliant structure," J. Sound Vib. **197**, 17–44.

Forsberg, F., Shi, W. T., and Goldberg, B. B. (**2000**). "Subharmonic imaging of contrast agents," Ultrasonics **38**, 93–98.

Fung, Y. C., Zweifach, B. W., and Intaglie, M. (**1966**). "Elastic environment of capillary bed," Circ. Res. **19**, 441–461.

Gao, F. R., Hu, Y. T., and Hu, H. P. (**2007**). "Asymmetrical oscillation of a bubble confined inside a micro pseudoelastic blood vessel and the corresponding vessel wall stresses," Int. J. Solids Struct. **44**, 7197–7212.

Humphrey, J. D., and Na, S. (**2002**). "Elastodynamics and arterial wall stress," Ann. Biomed. Eng. **30**, 509–523.

Hynynen, K. (**2008**). "Ultrasound for drug and gene delivery to the brain," Adv. Drug Delivery Rev. **60**, 1209–1217.

Krasovitski, B., and Kimmel, E. (**2001**). "Gas bubble pulsation in a semi-confined space subjected to ultrasound," J. Acoust. Soc. Am. **109**, 891–898.

Lauterborn, W., and Bolle, H. (**1975**). "Experimental investigations of cavitation-bubble collapse in neighborhood of a solid boundary," J. Fluid Mech. **72**, 391–399.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic Press, London).

Leighton, T. G., White, P. R., and Marsden, M. A. (**1995**). "Applications of one-dimensional bubbles to lithotripsy, and to diver response to low frequency sound," Acta Acust. **3**, 517–529.

Miao, H., Gracewski, S. M., and Dalecki, D. (**2008**). "Ultrasonic excitation of a bubble inside a deformable tube: Implications for ultrasonically induced hemorrhage," J. Acoust. Soc. Am. **124**, 2374–2384.

Miller, D. L. (**1979**). "Cylindrical-bubble model for the response of plant-tissue gas bodies to ultrasound," J. Acoust. Soc. Am. **65**, 1313–1321.

Oguz, H. N., and Prosperetti, A. (**1998**). "The natural frequency of oscillation of gas bubbles in tubes," J. Acoust. Soc. Am. **103**, 3301–3308.

Popinet, S., and Zaleski, S. (**2002**). "Bubble collapse near a solid boundary: A numerical study of the influence of viscosity," J. Fluid Mech. **464**, 137–163.

Qin, S., and Ferrara, K. W. (**2006**). "Acoustic response of compliable microvessels containing ultrasound contrast agents," Phys. Med. Biol. **51**, 5065–5088.

Qin, S., and Ferrara, K. W. (**2007**). "The natural frequency of nonlinear oscillation of ultrasound contrast agents in microvessels," Ultrasound Med. Biol. **33**, 1140–1148.

Rowe, A. J., Finlay, H. M., and Canham, P. B. (**2003**). "Collagen biomechanics in cerebral arteries and bifurcations assessed by polarizing microscopy," J. Vasc. Res. **40**, 406–415.
Sassaroli, E., and Hynynen, K. (**2004**). "Forced linear oscillation's of microbubbles in blood capillaries," J. Acoust. Soc. Am. **115**, 3235–3243.
Sassaroli, E., and Hynynen, K. (**2005**). "Resonance frequency of microbubbles in small blood vessels: A numerical study," Phys. Med. Biol. **50**, 5293–5305.
Snowhill, P., and Silver, F. (**2005**). "A mechanical model of porcine vascular tissues—Part II: Stress–strain and mechanical properties of juvenile porcine blood vessels," Cardiovasc. Eng. **5**, 157–169.
Stride, E., and Saffari, N. (**2003**). "Microbubble ultrasound contrast agents: A review," Proc. Inst. Mech. Eng., Part H: J. Eng. Med. **217**, 429–447.
Tachibana, K., and Tachibana, S. (**1995**). "Albumin microbubble echocontrast material as an enhancer for ultrasound accelerated thrombolysis," Circulation **92**, 1148–1150.
Yamada, H. (**1970**). *Strength of Biological Materials* (Williams & Wilkins, Baltimore, MD).
Zhao, S., Kruse, D. E., Ferrara, K. W., and Dayton, P. A. (**2006**). "Acoustic response from adherent targeted contrast agents," J. Acoust. Soc. Am. **120**, EL63–EL69.

Post-processing of guided wave array data for high resolution pipe inspection

Alexander Velichko[a] and Paul D. Wilcox
Department of Mechanical Engineering, University of Bristol, Bristol BS8 1TR, United Kingdom

(Received 10 December 2008; revised 7 October 2009; accepted 9 October 2009)

This paper describes a method for processing data from a guided wave transducer array on a pipe. The raw data set from such an array contains the full matrix of time-domain signals from each transmitter-receiver combination. It is shown that for certain configurations of an array, the total focusing method can be applied, which allows the array to be focused at every point on a pipe in both transmission and reception. The effect of array configuration parameters on the sensitivity of the proposed method to random and coherent noise is discussed. Experimental results are presented using electromagnetic acoustic transducers for exciting and detecting the S_0 Lamb wave mode in a 12-in. diameter steel pipe at 200 kHz excitation frequency. The results show that using the imaging algorithm, a 2-mm (0.08 wavelength) diameter half-thickness hole can be detected.
 [DOI: 10.1121/1.3257604]

PACS number(s): 43.35.Zc, 43.20.Mv, 43.60.Fg, 43.60.Jn [JGM] Pages: 2973–2982

I. INTRODUCTION

Ultrasonic guided waves provide a highly efficient technique for rapid pipe inspection.[1,2] Guided wave inspection is mainly used for long-range pipe screening over 10 s of meters at relatively low frequencies. This technique allows the presence and axial location of defects to be obtained. The leading commercial guided wave instruments for pipeline inspection also have data processing tools which allow maps of unrolled pipe circumference to be plotted. Typically, this is achieved by focusing on either transmission or reception only, and for this reason circumferential resolution is limited. Therefore, the problems of improving circumferential defect location, characterization of the defects, and accurate defect sizing still remain.

The defect detection and sizing capability of guided wave inspection can be improved using array techniques, which are well known in ultrasonic bulk wave imaging.[3] The complete raw data set from a transducer array contains the full matrix of time-domain signals from each transmitter-receiver combination. If such a matrix is collected, then all imaging can be performed in post-processing.

This approach for pipe imaging was used in Refs. 4–6. In Ref. 5, three different synthetic focusing algorithms were investigated: the common source method (CSM), the synthetic aperture focusing technique (SAFT), and the total focusing method (TFM). Although the processing technique is the same for all three methods, only the TFM uses the full transmit-receive matrix of data. The SAFT requires pulse-echo data and the CSM corresponds to the case when all transmitter elements are fired simultaneously with focusing only performed on reception. In other words, only the TFM uses the full focusing capability of the array. As a result, the TFM gives significantly reduced sidelobe level relative to the SAFT and the CSM. However, for the considered array configuration, the authors showed that SAFT and TFM suffer from the coherent noise in the image and for imaging only, the CSM can be used.

In this paper, it is shown that for certain configurations of the array on a pipe, the TFM can be applied, which allows the array to be focused at every point on the pipe in both transmission and reception. The sensitivity of the proposed method to random and coherent noise is investigated. Experimental results are presented using electromagnetic acoustic transducers (EMATs) for exciting and detecting the S_0 Lamb wave mode in a 12-in. diameter steel pipe at 200 kHz excitation frequency.

II. PRELIMINARIES

A. Excitation and scattering of guided waves in a pipe

A hollow cylinder of outer radius R and wall thickness h is considered with a cylindrical coordinate system θ, y, z defined with the y axis normal to the pipe surface. The guided waves are excited by an ultrasonic transducer, which is located at the outer surface of the cylinder and produces a time harmonic load $\mathbf{q}e^{-i\omega t}$. The total harmonic displacement field due to the surface load can be represented as the superposition of normal modes of the cylinder:

$$\mathbf{u} = \begin{cases} \mathbf{u}^+, & z > \dfrac{d}{2}, \\ \mathbf{u}^-, & z < -\dfrac{d}{2}, \end{cases}$$

$$\mathbf{u}^{\pm} = \sum_{M,m} f_m^{M\pm} \mathbf{w}_m^M(y) e^{im\theta} e^{\pm i k_m^M z}, \tag{1}$$

where $\mathbf{u} = \{u_\theta, u_y, u_z\}^T$ is the displacement field, d is the axial size of the transducer, and $\mathbf{w}_m^M$ and k_m^M are the power normalized displacements and wavenumber of the M-th mode of m-th circumferential order, respectively. The indices "+" and

[a]Author to whom correspondence should be addressed. Electronic mail: a.velichko@bristol.ac.uk

"−" correspond to wave propagation in the positive and negative axial directions. The mode amplitudes $f_m^{M\pm}$ are related to the load $\mathbf{q}$ as

$$f_m^{M\pm} = \frac{i\omega}{4}\int_0^{2\pi} Rd\theta \int_{-d/2}^{d/2} dz[\mathbf{w}_m^{M*}(R)\mathbf{q}(\theta,z)e^{-im\theta}e^{\mp ik_m^M z}]. \tag{2}$$

The guided waves $\mathbf{u}^+$ and $\mathbf{u}^-$ propagate in positive and negative axial directions, interact with the scatterers, and generate scattered wave fields $\mathbf{u}_{sc}^+$ and $\mathbf{u}_{sc}^-$, consisted of all possible combinations of mode-converted guided waves. The scattered waves propagating back toward the transmitter can be written in the form

$$\mathbf{u}_{sc} = \mathbf{u}_{sc}^+ + \mathbf{u}_{sc}^-, \quad \mathbf{u}_{sc}^{\pm} = \sum_{n,N}\left(\sum_{m,M} f_m^{M\pm} S_{mn}^{MN}\right)\mathbf{w}_n^N(y)e^{in\theta}e^{\mp ik_n^N z}, \tag{3}$$

where S_{mn}^{MN} is the scattered modal amplitude.

B. Approximate theory for large diameter pipes

When the wall thickness h and the wavelength λ are small compared with the pipe radius R, the effect of the curvature of the pipe wall on the wave propagation becomes negligibly small. In this case, the propagation and scattering of guided waves in a pipe are described by an approximate theory.[7] The pipe can be regarded as a flat isotropic plate of the same thickness that is periodic in the circumferential direction. Such an approximation leads to the simple analogy between guided wave modes in pipe and plate. Index M defines the set of guided wave modes in a pipe which corresponds to a particular guided wave mode in a plate. For example, the longitudinal and flexural modes in a pipe correspond to Lamb modes in a plate, and torsional modes in a pipe correspond to shear horizontal modes in a plate. If the index M is fixed, then the circumferential order m of a pipe mode corresponds to specific directions of propagation $\varphi_m^{M\pm}$ of plate mode. The angle $\varphi_m^{M\pm}$ is defined by

$$k_M \cos\varphi_m^{M\pm} = \pm k_m^M, \quad k_M \sin\varphi_m^{M\pm} = \frac{m}{R}, \tag{4}$$

so the zero angle corresponds to the axisymmetric mode $m=0$. The axial wavenumber of a pipe mode k_m^M is the projection of the plate wavenumber k_M into the axial direction z:

$$k_m^M = \sqrt{k_M^2 - \frac{m^2}{R^2}}. \tag{5}$$

The power normalized displacements $\mathbf{w}_m^n$ for the pipe mode are given by

$$\mathbf{w}_m^M = \frac{\mathbf{u}_m^M}{\sqrt{P_m^M}}, \quad P_m^M = 2\pi\,\mathrm{Re}\int_{R_{\mathrm{in}}}^{R_{\mathrm{out}}} \mathbf{P}_m^M \mathbf{e}_z y dy, \tag{6}$$

where $\mathbf{u}_m^n$ are pipe mode displacements and $\mathbf{P}_\mathbf{m}^\mathbf{M}$ is the complex Poynting vector.[8] If the pipe curvature can be neglected, then the Poynting vector is in the direction of propagation of the equivalent plate mode, and from Eq. (6), it follows

$$\mathbf{w}_m^M \approx \left(\frac{1}{2\pi R|\cos\varphi_m^M|}\right)^{1/2}\mathbf{w}_M(\varphi_m^M), \tag{7}$$

where $\mathbf{w}_M(\varphi_m^M)$ is the power normalized displacements for the plate mode propagating in the φ_m^M direction.

Using Eq. (7) and results obtained in Ref. 7, it is possible to relate the modal amplitudes f_m^M and S_{mn}^{MN} to the directivity function of the transmitter f_M and the scattering matrix of the defect in a plate S_{MN}, respectively:

$$f_m^M = e^{i\pi/4}\left(\frac{1}{Rk_m^M}\right)^{1/2} f_M(\varphi_m^M), \tag{8}$$

$$S_{mn}^{MN} = e^{i\pi/4}\frac{1}{R}\left(\frac{k_M}{2\pi k_m^M k_n^N}\right)^{1/2} S_{MN}(\varphi_m^M, \varphi_n^N). \tag{9}$$

At relatively low frequency-thickness values (approximately less than 1 MHz mm), the normal stress corresponding to the S_0 mode is small compared with the in-plane stresses and can be neglected. Note that for the SH_0 mode, the normal stress is identically zero. Therefore, the propagation of these modes in a plate can be modeled using the plane stress approximation.[9] The plate is replaced by a two-dimensional (2D) infinite elastic media, and the longitudinal and shear bulk waves in this media correspond to the S_0 and SH_0 modes in the plate. A similar approach has been used before for finite element modeling of the S_0 mode in a plate.[10]

It should be stressed that the pipe imaging method, described in this paper, does not depend on any approximation and is valid for any pipe. Nevertheless, the approximate theory allows better understanding of propagation of guided waves in a pipe and their interaction with defects.

C. Configuration of transducer array on a pipe

An ultrasonic array system consists of set of transmitter elements and set of receiver elements. In order to perform circumferential mode decomposition of transmitted and reflected wave fields, all transmitters and all receivers have to be arranged in circumferential ring configurations around the pipe.

For the data processing, it is desirable that only one type of guided wave mode (i.e., equivalent to a particular plate mode such as S_0) is excited and detected. This can be achieved at the transducer element level. For example, omnidirectional EMAT transducer[11] allows excitation and detection of only the S_0 mode. In Refs. 5 and 6, PZT shear transducers were used, which excite two modes: SH_0 and S_0. The S_0 mode is transmitted predominantly in the circumferential direction around the pipe, while the SH_0 mode propagates in the axial direction along the pipe and, therefore, can be formally regarded as the only mode that is transmitted and detected by the transducers. However, the array configuration considered in Refs. 5 and 6 had transmitter and receiver elements with the same axial positions on a pipe. In this case, the received signals contain periodically repeated pure circumferential S_0 wave packets. On the other hand, if all transmitters are fired simultaneously (the CSM), the excited wave is an axisymmetric SH_0 mode, and there is no generation of

any circumferentially propagating waves. This limits the focusing capability of such an array to reception only but avoids the artifacts in an image caused by the unwanted circumferentially traveling waves.

An alternative pipe array configuration consists of separate transmitter and receiver rings of transducers. In this situation, the transmitter-receiver time-traces contain signals reflected from features and defects in the positive and negative axial direction and a direct transmitted signal. It is supposed that the positive axial direction is the direction of imaging. This means that waves reflected from features in the negative direction, as well as the direct transmit-receive signal, must be suppressed in post-processing. The appropriate array configuration has two transmitter rings at $z=z_1$, $z_2>0$, and one receiver ring at $z=0$, and, hence, the collected data can be considered as two full transmit-receive matrices, one corresponding to each transmitter ring. These data sets will be denoted by the indices "1" and "2," respectively.

Because it is assumed that the transducers are sensitive to only one type of mode, the indices M and N, which refer to the mode type in the wave field representation, can now be omitted. The signal transmitted by the transducer at $(\theta_t, z_{1,2})$ position on the pipe surface and recorded by the same type of transducer at $(\theta_r, z=0)$ position can be written as

$$u^{(1,2)}(\omega,\theta_t,\theta_r) = u^{(1,2)+}(\omega,\theta_t,\theta_r) + u^{(1,2)-}(\omega,\theta_t,\theta_r) + u^{(1,2)0}(\omega,\theta_t,\theta_r), \tag{10}$$

where the positive $u^{(1,2)+}$ and negative $u^{(1,2)-}$ reflected waves are

$$u^{(1,2)\pm} = \sum_{n,m} a^{\pm}_{mn} e^{\mp ik_m z_{1,2}} e^{i(-m\theta_t + n\theta_r)}, \quad a^{\pm}_{mn} = f^{\pm}_m S_{mn} f^{\mp}_n, \tag{11}$$

and the direct transmit-receive signal $u^{(1,2)0}$ is given by

$$u^{(1,2)0} = \sum_m a_m e^{ik_m z_{1,2}} e^{i(-m\theta_t + m\theta_r)}, \quad a_m = f^-_m f^+_m. \tag{12}$$

III. TOTAL FOCUSING METHOD FOR PIPE IMAGING

The TFM for bulk wave imaging is usually formulated in the time domain as a delay-and-sum algorithm.[12] Because of the pipe geometry, this simple approach cannot be applied directly to pipe imaging. However, a delay in the time domain is equivalent to a phase shift in the temporal frequency domain, and the TFM can also be formulated as a phase compensation in the frequency domain. This form of the TFM has the advantage that the dispersive properties of guided waves can be easily taken into account. For example, this method has been used for the guided wave inspection of plate structures.[13]

In the case of a cylindrical waveguide, the phase compensation approach is applied to circumferential modal amplitudes instead of transmit-receive signals.[5] The conversion of the raw transmit-receive data into the angular wavenumber domain is performed by 2D Fourier transform with respect to the circumferential positions θ_t and θ_r of transmitter and receiver elements.[4] In order to satisfy the Nyquist criteria, the angular interval between transducers $\Delta\theta$ must be less than $\pi/m_{\max}$, where $m_{\max}$ is the maximum angular mode order. From Eq. (5), it follows that $m_{\max}=kR$ and, therefore, $\Delta x<\lambda_{\min}/2$, where $\Delta x=\Delta\theta R$ is the transducer pitch and $\lambda_{\min}$ is the shortest wavelength of guided wave mode within the frequency range of the transmitted signal. Using Eqs. (10)–(12), the extracted angular wavenumber spectra can be written as

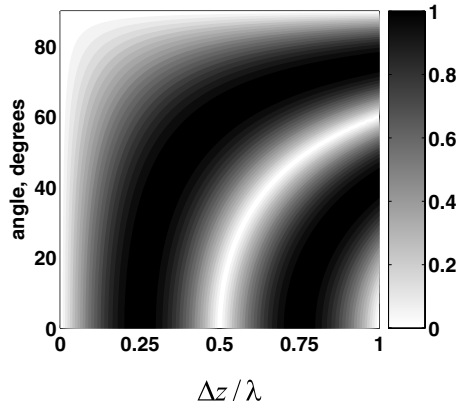


FIG. 1. The isolines of the function $\sin(2\pi\Delta z/\lambda\cos\varphi)$.

$$U^{(1,2)}_{mn} = U^{(1,2)+}_{mn} + U^{(1,2)-}_{mn} + U^{(1,2)0}_{mn},$$

$$U^{(1,2)\pm}_{mn} = a^{\pm}_{m,-n} e^{\mp ik_m z_{1,2}}, \quad U^{(1,2)0}_{mn} = \delta_{m,-n} a_m e^{ik_m z_{1,2}}. \tag{13}$$

The chosen configuration of the array (two transmitter rings and one receiver ring) allows the simulation in post-processing of the situation when only angular modes in the positive axial direction are excited. This can be achieved by the linear combination of the angular modes obtained from the two transmit-receive data sets. The amplitude of the resulting angular mode c_{mn} is given by

$$c_{mn} = \frac{1}{2i}\left(u^{(1)}_{mn} e^{ik_m z_2} - u^{(2)}_{mn} e^{ik_m z_1}\right). \tag{14}$$

If there is no noise in the system and all parameters are known exactly, then direct and "negative" reflected signals will be totally suppressed. Using expression (5) for the angular wavenumber k_m, the modal amplitude c_{mn} can be written in the form

$$c_{mn} = a^{+}_{m,-n} \sin k_m \Delta z, \tag{15}$$

where $\Delta z=z_2-z_1$ is the axial interval between the transmitter rings.

In the framework of the approximate theory [formula (4)], the axial wavenumber k_m is equal to $2\pi/\lambda\cos\varphi_m$. Figure 1 shows the isolines of the function $\sin(2\pi\Delta z/\lambda\cos\varphi)$ as a function of the angle φ and the distance between transmitters. Small values mean that, for given parameters, the "positive" and negative modes cannot be separated. It can be seen that the optimal value of Δz is equal to $\lambda/4$ and in this case, almost all angular modes can be separated. Note that for the CSM, the only incident mode is axisymmteric (i.e., incident angle is 0°), and the distance between transmitters can be taken equal to $(n+1/2)\lambda/2$ with any integer n.

The modal amplitudes c_{mn} are subject to the phase shift $\Phi_{mn}=-(k_m+k_n)z-(m+n)\theta$. Then the phase shifted wavenumber-frequency domain spectra are summed over all angular wavenumbers and frequencies to get an image I:

$$I(\theta,z)=\sum_{m,n}\left(\sum_{\omega}c_{mn}f_{mn}e^{-i(k_m+k_n)z}\right)e^{-i(m+n)\theta}, \tag{16}$$

where the coefficients f_{mn} represent an angular mode filter. This filter can be taken as a function of mode angles, $f_{mn}=f_{mn}(\varphi_m,\varphi_n)$, for example, in the form of the Hanning filter $f_{mn}=\cos\varphi_m\cos\varphi_n$. It should be stressed that the filter f_{mn} is optional and has the effect of reducing sidelobes in an image at the expense of broadening the main lobe.

The sum over frequencies represents the back-propagation of each angular mode and can be efficiently implemented as a Fourier transform in the (k_m+k_n)-domain using a dispersion compensation algorithm.[14] Note that the velocity of each angular mode in the axial direction depends on frequency, and the mode becomes more dispersive for bigger angular orders m and n. The sum over m,n can be considered as the main diagonal of the inverse 2D Fourier transform of the back-propagated modal amplitudes.

The complete procedure for the TFM for pipe imaging is as follows.

- 2D Fourier transform of transmit-receive data matrices to extract the angular mode amplitudes.
- Linear superposition of the angular modes obtained from the two transmit-receive data sets in order to suppress signals transmitted in the negative axial direction.
- Back-propagation of each angular mode (to obtain axial resolution).
- 2D inverse Fourier transform of back-propagated angular modes at each axial position. The circumferential defect distribution (angular resolution) is given by the main diagonal of the output matrix of the inverse Fourier transform.

IV. NOISE SENSITIVITY OF THE IMAGING METHOD

In practice, it is important for the imaging methods to be stable with respect to errors in initial data. Stability means that small errors in the initial data cause small errors in the image.

In the analysis given below, the pipe end is used as a reference reflector. If the incident mode is S_0 or SH_0, then at the relatively low frequencies, the plane stress approximation can be used. In this case, the reflection of angular modes from the pipe end is equivalent to the reflection of plane waves from a free boundary. The detected amplitudes of angular modes $a_{mn}^{\pm}$ can be written in the form

$$a_{mn}^{\pm}=\delta_{mn}f_m^{\pm}s(\varphi_m^{\pm})f_m^{\mp}e^{\pm 2ik_{m^{z_\pm}}}, \tag{17}$$

where $s(\varphi)$ is the reflection coefficient of a plane wave incident at angle φ at a free boundary and $z_\pm$ is location of the positive and negative pipe ends. Using relationships (7) and (8), this expression can be rewritten as

$$a_{mn}^{\pm}=\delta_{mn}s_m^{\pm}e^{\pm 2ik_{m^{z_\pm}}}, \tag{18}$$

where coefficient $s_m^{\pm}$ is given by

$$s_m^{\pm}=\frac{i}{kR|\cos\varphi_m^{\pm}|}f(\varphi_m^{\pm})s(\varphi_m^{\pm})f(\varphi_m^{\mp}). \tag{19}$$

A. Random noise sensitivity

Consider the true modal amplitudes $U_{mn}^{(1,2)+}$ with amplitude and phase errors:

$$\tilde{U}_{mn}^{(1,2)+}=U_{mn}^{(1,2)+}(1+q_{mn}^{(1,2)})e^{i\epsilon_{mn}^{(1,2)}}, \tag{20}$$

where $\tilde{U}_{mn}^{(1,2)+}$ is the modal amplitude corrupted by the amplitude and phase errors $q_{mn}^{(1,2)}$ and $\epsilon_{mn}^{(1,2)}$. It is assumed that these errors are independent and have normal distributions with standard deviations σ_q and σ_ϵ, respectively.

To characterize the stability of the imaging method, the standard deviation σ_I of the image can be taken.[15,16] Using a method of calculation similar to that given in Ref. 15, it is possible to show that for the function

$$G=\sum_{m,n}g_{mn}(1+q_{mn})e^{i\epsilon_{mn}}, \tag{21}$$

the standard deviation is given by

$$\sigma_G=\sqrt{(\sigma_q^2+1-e^{-\sigma_\epsilon^2})\sum_{m,n}|g_{mn}|^2}. \tag{22}$$

It is assumed that the maximum image amplitude $I_{\max}$ is reached at the nominal center of the reflector (θ_0,z_0), i.e., $I_{\max}=I(\theta_0,z_0)$. From Eq. (16), it follows

$$I_{\max}=\left|\sum_{\omega}I(\omega)\right|,$$

$$I(\omega)=\sum_{m,n}a_{m,-n}^{+}f_{mn}\sin(k_m\Delta z)e^{-i(k_m+k_n)z_0}e^{-i(m+n)\theta_0}. \tag{23}$$

Using formula (22) and relationship (14), the standard deviation of the image at a particular frequency can be written as

$$\sigma_I(\omega)=\sqrt{(\sigma_q^2+1-e^{-\sigma_\epsilon^2})\sum_{m,n}|a_{m,-n}^{+}f_{mn}|^2}. \tag{24}$$

Since the majority of energy is concentrated in the vicinity of the center frequency ω_0 of the transmitted signal, the relative noise level δ_I can be estimated by the ratio

$$\delta_I=\frac{\sigma_I(\omega_0)}{|I(\omega_0)|}. \tag{25}$$

To estimate the random noise level for a real reflector, scattering from the pipe end wall is considered. It is assumed that transducers are omni-directional and the Hanning filter f_{mn} is applied. Using formulas (18) and (19), and supposing that $\sigma_q=\sigma_\epsilon\equiv\sigma\ll 1$, the following expression for the δ_I can be obtained:

$$\delta_I=\sigma d(\omega,\Delta z,R),$$

$$d(\omega,\Delta z,R)=\sqrt{2}\frac{\left(\sum_m |s(\varphi_m)\cos\varphi_m|^2\right)^{1/2}}{\left|\sum_m s(\varphi_m)\cos\varphi_m \sin\left(2\pi\frac{\Delta z}{\lambda}\cos\varphi_m\right)\right|}. \quad (26)$$

Note that the reflection coefficient $s(\varphi)$ does not depend on frequency. From Eq. (26), it follows that for a constant value of $\Delta z/\lambda$, the function $d\sim N^{-1/2}$, where N is the number of circumferential modes. Expression (5) shows that $N=2kR=2Rv^{-1}\omega$, where v is the mode velocity and, therefore, $d\sim(\omega R)^{-1/2}$.

The calculations show that at each frequency, the function d has a minimum at $\Delta z/\lambda=0.25$. For example, for 12-in. diameter steel pipe, S_0 incident mode at 200 kHz, and $\Delta z/\lambda=0.25$, the function $d\approx 0.2$. Therefore, if the random noise amplitude $\sigma\leq 0.05$, then the relative noise due to the random errors is less then -40 dB.

Note that one source of the random errors is uncertainty of the signal strength from each transducer. The estimation of the noise amplitude in the image due to the coupling errors for the CSM has been performed by Davies and Cawley.[6] It has been shown that the CSM is very robust against the most extreme levels of coupling errors likely to be encountered in practice ($\pm 50\%$). However, for the array configuration considered in the current paper, coupling errors can be compensated for in post-processing. Indeed, the first signal detected by each receiver is the direct transmit-receive signal [Eq. (12)]. The amplitude of this signal can be obtained theoretically by either an exact model[17] or the approximate theory for large diameter pipes.[7] Then, taking any transmit-receive signal as a reference, the amplitudes of all other transmit-receive time-traces can be corrected based on the model data.

B. Coherent noise sensitivity

Apart from random noise, there is also coherent noise in the image caused by the waves propagating in the negative axial direction. The suppression of the negative signals depends on the exact knowledge of the parameters of the system such as the interval between transmitters and the wave velocity. It is also assumed that the transmitter ring is not tilted. In practice, all these parameters are known only with some fixed precision. As a result, negative waves are not totally suppressed and appear as artifacts in the image. The biggest amplitude of the artifacts corresponds to the biggest reflector located in the negative direction. An estimation of the coherent noise can be obtained using the approximate theory for thin pipes and assuming that the biggest reflector is a pipe end wall.

1. Interval between transmitters

If the error in measurement of the interval between transmitters is δz, then the modal amplitudes c_{mn}, defined by expression (14), have additional terms caused by the uncompensated negative waves. For the negative reflected wave, this term is given by

$$\delta c_{mn}=\frac{1}{2i}a^-_{m,-n}(e^{ik_m(z_1+\tilde{z}_2)}-e^{ik_m(z_2+\tilde{z}_1)}), \quad (27)$$

where $z_{1,2}$ and $\tilde{z}_{1,2}$ are the exact and approximate axial positions of the transmitter rings, respectively,

$$z_2=z_1+\Delta z,\quad \tilde{z}_2=\tilde{z}_1+\Delta z+\delta z. \quad (28)$$

Expression (27) becomes

$$\delta c_{mn}=a^-_{m,-n}e^{ik_m(z_1+\tilde{z}_1+\Delta z)}\delta_m,\quad \delta_m=\frac{1}{2i}(e^{ik_m\delta z}-1). \quad (29)$$

In the case of scattering from the pipe end, the modal amplitudes a^-_{mn} are given by expression (18). It can be seen that the negative pipe end appears in the image at $z_0\approx -z_-+z_1+0.5\Delta z$. For a small error $\delta z\ll 1$, the relative noise amplitude can be approximately estimated as

$$\delta\approx \pi\frac{\delta z}{\lambda}\frac{1}{\sin\left(2\pi\frac{\Delta z}{\lambda}\right)}. \quad (30)$$

For example, if $\Delta z/\lambda=0.25$, then, to keep noise level below -40 dB, the error in axial position has to be less than 0.0032λ. If mode of interest is S_0, then at 200 kHz, the wavelength is 26 mm and $0.0032\lambda=0.08$ mm. It is impossible to achieve such precision in practice. However, this error can be removed fairly easily in post-processing by iteratively adjusting the value of Δz to minimize the amplitude of the artifacts caused by features in the negative axial direction.

2. Tilting of transducer ring

If a ring of transmitters or receivers is tilted, then the axial position of each transducer depends on its circumferential location θ. If the nominal axial position of the transducer ring is z, then the axial transducer coordinate in a tilted ring is $z+\epsilon(\theta)$, where $\epsilon=\epsilon_0\cos(\theta-\theta_0)$. In this case, the Fourier transform of the transmit-receive data gives corrupted angular wavenumber spectra. Each extracted amplitude of a particular angular mode is represented by the superposition of all angular modes. For example, the extracted amplitudes corresponding to the positive and negative reflected waves are given by

$$U^{(1,2)\pm}_{mn}=a^{\pm}_{m,-n}e^{\mp ik_m z_{1,2}}+\delta U^{(1,2)\pm}_{mn}. \quad (31)$$

The additional terms $\delta U^{(1,2)\pm}_{mn}$ are caused by the tilting and are given by

$$\delta u^{(1,2)\pm}_{mn}=\sum_{m'} a^{\pm}_{m',-n}e^{\mp ik_{m'}z_{1,2}}f^{\pm}_{m'm},\quad f^{\pm}_{mm'}$$

$$=\frac{1}{N}\sum_{\theta_t} e^{i(m-m')\theta_t}(e^{\pm ik_m\epsilon}-1), \quad (32)$$

where N is the number of transducers around the pipe.

Using expression (18) for the pipe end reflection, the terms in the modal amplitudes c_{mn} corresponding to $\delta U^{(1,2)\pm}_{mn}$ can be written as

$$\delta c_{mn}^{\pm} = \frac{1}{2i} s_{-n} f_{-n,m}^{\pm} e^{i(k_m \mp k_n)z_1} (e^{ik_m \Delta z} - e^{\mp k_n \Delta z}) e^{\pm 2ik_m z_{\pm}}. \tag{33}$$

The term δc_{mn}^{+} results in blurring of the positive image, while the term δc_{mn}^{-} gives an artifact of the negative pipe end wall at about $z_0 \approx -z_- + z_1$.

In order to get a quantitative estimation of the relative noise level, the calculations were performed for the S_0 mode transmitted and detected by omni-directional transducers. The transmitted signal was a 5 cycle Hanning windowed toneburst with a center frequency in the range from 50 to 200 kHz. The amplitude of the artifact depends on a number of parameters: error amplitude ϵ_0, interval between transmitters Δz, pipe radius, location of the end wall z_-, and center frequency of the transmitted signal.

For small errors ($\epsilon_0 \lesssim 10$ mm), the amplitude of the noise is directly proportional to ϵ_0. It was found that in the frequency interval considered and with $\Delta z/\lambda$ held constant, the noise amplitude is also nearly constant with frequency. Figure 2(a) shows the noise amplitude as a function of $\Delta z/\lambda$ for a 12-in. diameter pipe, $\epsilon_0 = 1$ mm, and $z_- = -1$ m. It is seen that the maximum noise corresponds to $\Delta z = 0.5\lambda$. The dependences of the noise on the pipe radius and negative pipe end location for $\Delta z/\lambda = 0.25$ are shown in Figs. 2(b) and 2(c). The noise level slightly varies with the pipe radius and rapidly decreases with increasing distance between the negative pipe end and the array.

3. Wave velocity

The artifacts caused by the uncompensated negative waves also appear in an image if incorrect material properties are used to compute the dispersion curves. In this case, the additional term in the mode amplitudes c_{mn}, which corresponds to the negative reflected waves, is given by

$$\delta c_{mn} = \frac{1}{2i} a_{m,-n}^{-} e^{(k_m + \tilde{k}_m)z_1} (e^{i\tilde{k}_m \widetilde{\Delta z}} - e^{ik_m \Delta z}), \tag{34}$$

where $\tilde{k}_m$ and $\widetilde{\Delta z}$ are the theoretical mode wavenumber and interval between transmitters used in the imaging method.

In the framework of the plane stress approximation, the phase velocities of the S_0 and SH_0 modes are constant. If the relative velocity error of the mode of interest is ϵ, then $\tilde{k} = k(1+\epsilon)^{-1}$. From Eq. (34), it follows that the negative axisymmetric mode can be suppressed (i.e., $\delta c_{00} = 0$) if the value $\widetilde{\Delta z} = (1+\epsilon)\Delta z$ is taken. Below, it is assumed that $\widetilde{\Delta z}$ is first adjusted (in processing) to minimize the amplitude of the artifacts caused by features in the negative axial direction.

To estimate the noise amplitude, the numerical calculations for the pipe end reflection were performed with the same parameters as in the previous paragraph. For small errors ($\epsilon < 0.01$) the noise amplitude is directly proportional to ϵ. If $\Delta z/\lambda$ is held constant, then the noise amplitude is almost independent of frequency and the location of the negative pipe end. The dependence of the noise on the interval between transmitters exhibits behavior similar to the noise caused by the tilting error [Fig. 2(a)].

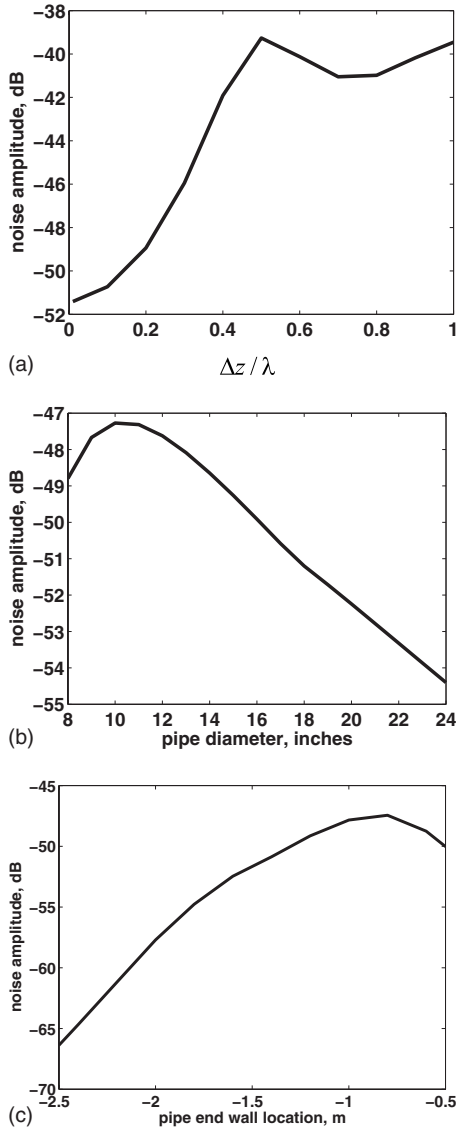


FIG. 2. The dependence of the noise amplitude caused by the tilting of transmitter ring on (a) interval between transmitter rings, (b) pipe radius, and (c) location of the negative pipe end wall. Amplitude is in decibels relative to the image amplitude of the positive pipe end wall.

The above analysis of the sensitivity of the imaging algorithm to noise shows that the relative noise level in an image critically depends on the distance Δz between transmitter rings. The optimal Δz is always less than a quarter of the wavelength $\Delta z/\lambda \leq 0.25$. In the interval $0 \leq \Delta z/\lambda \leq 0.25$, the random noise decreases with the $\Delta z/\lambda$, while the coherent noise increases. Thus, the best value of $\Delta z/\lambda$ depends on the relative size of different errors and the random noise level. However, in practice, the value $\Delta z/\lambda = 0.25$ can be recommended. It has been shown that in this case for typical errors, it is possible to keep the total noise level below −40 dB relative to an end wall reflection.

V. DETECTABILITY AND SIZING OF DEFECTS

A. Detectability of imaging method

It has been shown in Sec. IV that it is possible to achieve −40 dB noise level relative to an end wall reflection. This

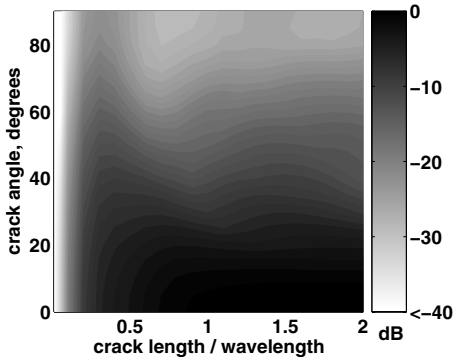


FIG. 3. The image amplitude (relative to the image amplitude of the pipe end wall) for a through thickness crack as a function of crack orientation and crack length.

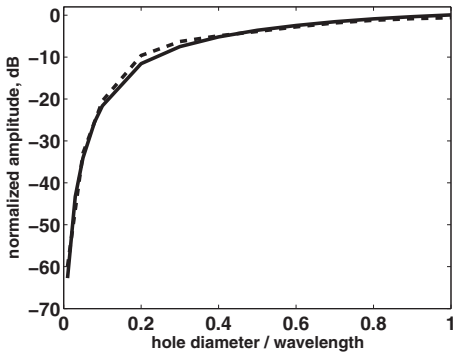


FIG. 4. The image amplitude for a through thickness hole (solid line) and a circumferential crack with the length equal to the hole diameter (dotted line). Amplitudes are in decibels relative to the image amplitude of the pipe end wall.

means that any defect with an image amplitude greater than −40 dB relative to a pipe end can be detected.

To illustrate the minimum defect size that can be detected, a crack-like defect is considered. The general problem of diffraction of guided waves by a crack in a pipe can be solved only by direct numerical methods (e.g., finite element, boundary element, or finite difference methods). However, for a through thickness defect with a constant profile through the plate thickness, the plane stress approximation described in Sec. II A can be used. Such a defect is fully described by its 2D shape at the pipe surface, and mode conversion occurs between the S_0 and SH_0 modes only. Therefore, the three-dimensional problem of diffraction of guided waves reduces to the problem of bulk wave scattering in an infinite 2D media. For example, in the case of a circular hole, the known analytical solution can be used.[18] If the defect is a crack, then the diffraction problem can be solved by the fast semi-analytical technique, developed in Ref. 19, which is based on a boundary integral equation method.

The detectability of the imaging method is characterized by a detectability map. Such a map represents the amplitude of the defect in an image relative to the biggest reflector as a function of defect parameters. For a through thickness crack, these parameters are crack length and crack orientation. As a reference reflector, a pipe end is taken. Note that in an ideal pipe, angular modes propagate along the pipe without attenuation and beam spreading and, hence, the detectabilty is independent of defect location

Figure 3 shows the predicted image amplitude for a through thickness crack as a function of crack orientation and crack length. The calculations were performed for a 12-in. nominal diameter steel pipe and omni-directional transducers, which excite and detect the S_0 mode at the 200 kHz center frequency. The distance between transmitter rings is taken to be equal to $\lambda/4$, where λ is the S_0 wavelength at the center frequency. A crack angle of 0° corresponds to a circumferential crack orientation. Assuming a defect threshold of −40 dB, it can be seen that the TFM allows the detection of through thickness cracks of any orientation over 0.05λ (1.3 mm) length.

Using analytical formulas for the scattering of bulk waves from a cylindrical hole,[18] the calculation of image amplitude for through thickness holes of different diameters was performed; the result is shown in Fig. 4. For comparison, the image amplitude of a circumferential crack with a length equal to the hole diameter is also shown in the same graph. It is seen that the two curves are almost identical. Note that if the diameter of the hole is less than 0.3λ, the image amplitude very rapidly decreases.

B. Defect sizing

The size of a crack can be characterized directly from the image by the full width at half maximum (FWHM).[5] In the case of a circumferential crack, this parameter is defined as the circumferential width of the image of the crack measured between the positions where the image is half of its peak value. Figure 5(a) shows the images of a 1λ length circumferential crack and a crack at 40° orientation obtained from the modeling data with the same system parameters as above. The location of the crack center for the circumferential crack is given by the position of the maximum of image amplitude. If the crack is inclined at some nonzero angle, then the maximum of the image does not coincide with the crack center. For example, the image of the crack inclined at 40° has two local maxima, which correspond to reflections from the crack's tips.

The circumferential profiles of the cracks at the position of the maximum of the images of the cracks are shown in Fig. 5(b). It can be seen that the FWHM for the crack at 40° orientation is smaller than for the circumferential crack. So, the FWHM defined in such a way depends on the crack orientation.

Figure 6 shows the FWHM as a function of orientation and length of crack. It can be seen that the FWHM gives an estimate of crack length for cracks over approximately λ length and inclined at up to 20°. Although the FWHM is not a linear function of crack length for subwavelength cracks, it is still monotonically related to the crack length and can still be used for sizing of cracks over 0.5λ length.

In the case of a through thickness hole, the calculations show that the FWHM does not depend on the hole diameter and is always about 0.7λ.

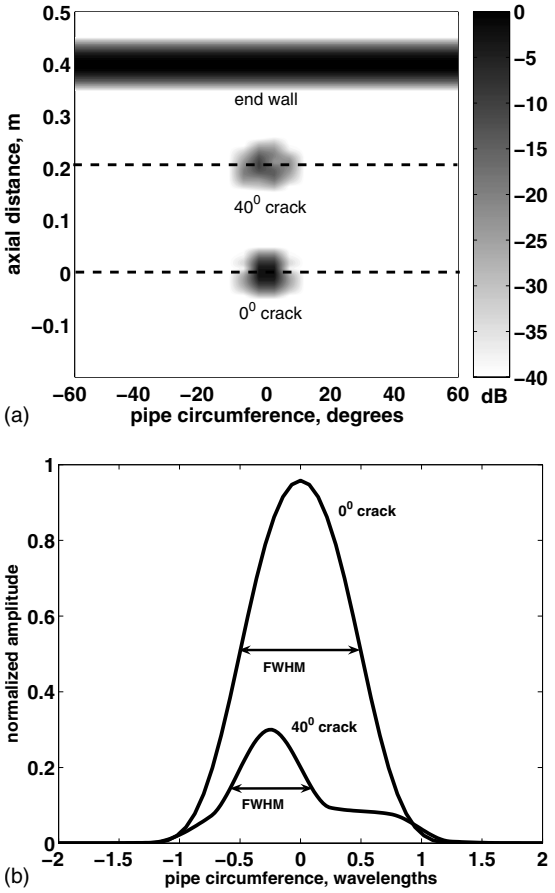


FIG. 5. (*a*) Image of 1λ length through thickness cracks in 12-in. diameter steel pipe inclined at 0° and 40°. Amplitude is in decibels relative to the image amplitude of the pipe end wall (shown in the image). (b) Circumferential profiles of the cracks at the position of the maximum image amplitude [dotted lines in (a)].

VI. EXPERIMENTAL RESULTS

A. Apparatus

Instead of a full transmitter-receiver array, a transducer manipulator manufactured by Phoenix Inspection Systems Ltd. (United Kingdom) was used. The manipulator allows the synthesis of the performance of any array by mechanical scanning of two transducers. A schematic diagram of the experimental set-up is shown in Fig. 7. The manipulator for each transducer consists of a mounting ring, circumferential and axial motors, and a transducer mounting. This provides

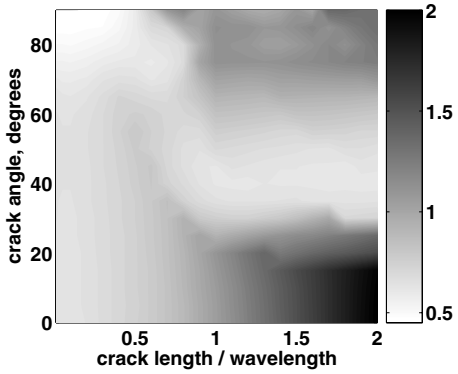


FIG. 6. FWHM as a function of crack length and crack orientation.

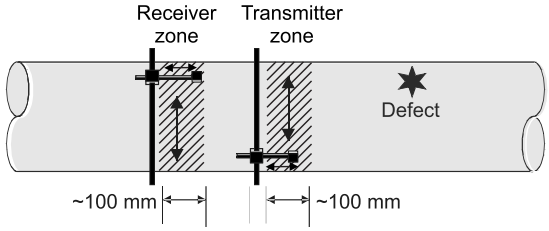


FIG. 7. The schematic diagram of pipe scanner configuration.

each transducer with 360° circumferential movement and about 100 mm of axial movement. Using one transmitter and receiver, any array configuration can be simulated. The manipulator was operated using a personal computer (PC) controller. The acquired data were exported to a PC and all post-processing was performed using MATLAB (The MathWorks, Inc., Natick, MA).

All experiments were carried out on a 12-in. diameter schedule 10 s (4.57-mm wall thickness) stainless steel pipe. The transducers were omni-directional EMATs designed to excite and detect the S_0 Lamb wave mode in metallic plate up to 10-mm thick.[11] The transmitted signal used was a 5 cycle Hanning windowed toneburst with a center frequency of 200 kHz. At this frequency, the wavelength of the S_0 mode is 26 mm. The circumferential interval between transducer positions was taken equal to half of the wavelength at the central frequency, which resulted in 80 circumferential transducer positions. The pipe was 3 m in length, the distance between receiver ring and the negative end wall was 0.9 m, and the distance between receiver and transmitter rings was 0.4 m.

B. TFM results

To illustrate the sensitivity of the imaging method to noise, two experiments were carried out. In the first experiment, a through thickness 9-mm-diameter (0.3λ) hole was drilled, and the axial distance between transmitter rings was set to 0.75λ. In the second experiment, three more defects were made: a 2-mm-diameter (0.08λ) through thickness hole, a 2-mm-diameter half-thickness hole, and a 6-mm (0.23λ) through thickness slot inclined at 30° to the circumferential direction. The axial distance between transmitter rings in this case was equal to 0.25λ. The TFM images are shown in Figs. 8(a) and 8(b). The axial distance is measured from the location of the second transmitter ring at $z=z_2$ as the area of interest of the pipe is $z>z_2$.

The image of the hole has the same amplitude in both cases. However, as was predicted by the noise sensitivity study, in the case of the 0.75λ transmitter interval, there is coherent noise caused by the unsuppressed negative waves. The artifact caused by the reflection from the negative pipe end is located at about $z=0.9$ m (1.3 m from the receiver ring) as predicted and has an amplitude of about −25 dB relative to the positive end wall. The second artifact is caused by the direct transmit-receive signal. It is located in the vicinity of $z=0$ and has an amplitude of about −15 dB. If the interval between transmitters is 0.25λ, then the total

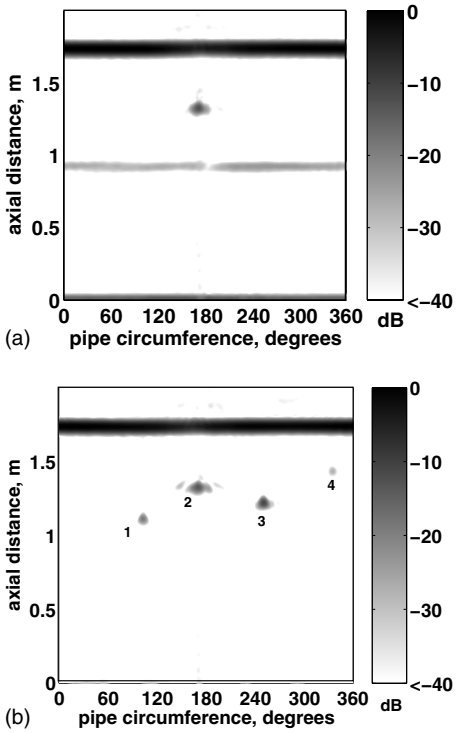


FIG. 8. (a) TFM image for the pipe with 9-mm-diameter through hole and 0.75λ axial transmitter interval. (b) TFM image for the pipe with multiple defects and 0.25λ axial transmitter interval. Signals 1–4 correspond to 2-mm-diameter through hole (signal 1), 9-mm-diameter through hole (signal 2), 6 mm, 30° slot (signal 3), and 2-mm-diameter half-thickness hole (signal 4).

noise level is below −40 dB relative to the end wall reflection, and all four defects are clearly visible in the image.

For the 2-mm-diameter through thickness hole, the experimental image amplitude is −20 dB and the FWHM is 0.8λ. The corresponding calculated values are −24 dB and 0.7λ. However, as mentioned before, for small holes, the image amplitude changes very fast with the hole diameter (Fig. 4). Therefore, even a small error in the modeling parameters (such as material properties) may cause a big discrepancy between modeling and experimental results.

In the case of a 9-mm-diameter hole, the experimental image amplitude and the FWHM are −13 dB and 1.2λ, respectively. On the other hand, the model gives −9 dB for the image amplitude and 0.7λ for the FWHM. This difference between experimental and theoretical results can be explained by the fact that for a hole, a significant contribution to the image amplitude is given by the higher order circumferential modes. However, higher order modes are increasingly dispersive. In practice, the material properties are known only approximately and focusing of these modes based on the dispersion compensation algorithm becomes worse for reflectors located far from the array.[14] This results in spreading of the processed signal in space, i.e., in reduction in the image amplitude and increasing of the size of the image.

The measured image amplitude and FWHM parameter for the 6-mm slot are approximately −11 dB and 0.9λ, respectively. This is in reasonable agreement with the theoretical prediction, which gives values of −11 dB for the image amplitude and 0.7λ for the FWHM. However, as the slot is less than 1 wavelength long, its sizing cannot be performed directly from the image based on the FWHM. In this case in order to estimate its size, more advanced methods are needed (for example, the vector total focusing method[20]).

The image amplitude for the 2-mm-diameter half-thickness hole is −28 dB and, hence, this defect is close to the sensitivity limit.

VII. CONCLUSION

The total focusing method, which allows an array to be focused at every point on a pipe surface in both transmission and reception, has been applied to pipe imaging. The technique is based on circumferential mode decomposition of the full matrix of transmit-receive data and linear superposition of extracted circumferential modes in the frequency domain. The sensitivity of the method to random and coherent noise has been studied.

The imaging algorithm has been experimentally tested on 12-in. diameter steel pipe using S_0 EMATs at 200 kHz excitation frequency. The results show that a 2-mm-diameter (0.08 wavelength) half-thickness hole can be detected.

ACKNOWLEDGMENTS

This work was supported by the United Kingdom Engineering and Physical Sciences Research Council (EPSRC) through the United Kingdom Research Centre in NDE and by BNFL, Nexia Solutions, DSTL, and British Energy.

[1] D. N. Alleyne, B. Pavlakovic, M. J. S. Lowe, and P. Cawley, "Rapid long-range inspection of chemical plant pipework using guided waves," Insight **43**, 93–96 (2001).
[2] J. L. Rose, J. J. Ditri, A. Pilarski, K. Rajana, and F. Carr, "A guided wave inspection technique for nuclear steam generator tubing," NDT Int. **27**, 307–310 (1994).
[3] B. W. Drinkwater and P. D. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," NDT Int. **39**, 525–541 (2006).
[4] T. Hayashi and M. Murase, "Defect imaging with guided waves in a pipe," J. Acoust. Soc. Am. **117**, 2134–2140 (2005).
[5] J. Davies and P. Cawley, "The application of synthetically focused imaging techniques for high resolution guided wave pipe inspection," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2007), Vol. **26A**, pp. 681–688.
[6] J. Davies and P. Cawley, "Synthetic focusing for high resolution guided wave pipe inspection: further results and robustness studies," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, NY, 2008), Vol. **27A**, pp. 139–146.
[7] A. Velichko and P. D. Wilcox, "Excitation and scattering of guided waves: Relationships between solutions for plates and pipes," J. Acoust. Soc. Am. **125**, 3623–3631 (2009).
[8] B. A. Auld, *Acoustic Fields and Waves in Solid*, 2nd ed. (Krieger, Malabar, FL, 1990).
[9] J. D. Achenbach, *Wave Propagation in Elastic Solids* (Elsevier, Amsterdam, 1999).
[10] O. Diligent, T. Grahn, A. Bostrom, P. Cawley, and M. J. S. Lowe, "The low-frequency reflection and scattering of the S_0 Lamb mode from a circular through-thickness hole in a plate: Finite element, analytical and experimental studies," J. Acoust. Soc. Am. **112**, 2589–2601 (2002).
[11] P. D. Wilcox, M. J. S. Lowe, and P. Cawley, "Omnidirectional guided wave inspection of large metallic plate structures using an emat array," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 653–665 (2005).
[12] C. Holmes, B. Drinkwater, and P. Wilcox, "Post-processing of the full matrix of ultrasonic transmit-receive array data for non-destructive evaluation," NDT Int. **38**, 701–711 (2005).

[13]P. D. Wilcox, "Omni-directional guided wave transducer arrays for the rapid inspection of large areas of plate structures," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 699–709 (2003).
[14]P. D. Wilcox, "A rapid signal processing technique to remove the effect of dispersion from guided wave signals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 419–427 (2003).
[15]R. J. Mailloux, *Phased Array Antenna Handbook* (Artech House, Boston, 1994).
[16]A. Velichko and P. D. Wilcox, "Guided wave arrays for high resolution inspection," J. Acoust. Soc. Am. **123**, 186–196 (2008).
[17]J. J. Ditri and J. L. Rose, "Excitation of guided elastic wave modes in hollow cylinders by applied surface tractions," J. Appl. Phys. **72**, 2589–2597 (1992).
[18]R. J. Brind, J. D. Achenbach, and J. E. Gubernatis, "High-frequency scattering of elastic waves from cylindrical cavities," Wave Motion **6**, 41–60 (1984).
[19]E. Glushkov, N. Glushkova, A. Ekhlakov, and E. Shapar, "An analytically based computer model for surface measurements in ultrasonic crack detection," Wave Motion **43**, 458–473 (2006).
[20]P. D. Wilcox, C. Holmes, and B. W. Drinkwater, "Advanced reflector characterization with ultrasonic phased arrays in NDE applications," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1541–1550 (2007).

Analytical modeling of piezoelectric ceramic transducers based on coupled vibration analysis with application to rectangular thickness poled plates[a)]

Boris S. Aronov,[b)] Corey L. Bachand, and David A. Brown
BTech Acoustics, LLC and Electroacoustics 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 16 March 2009; revised 4 September 2009; accepted 4 September 2009)

The energy method for analyzing piezoelectric ceramic transducers [B. S. Aronov, J. Acoust. Soc. Am. **117**, 210–220 (2005)] is applied to the treatment of transducers with mechanical systems that can be considered as two-dimensional. Analysis is made following the general outline of the theory of coupled vibration in two degrees-of-freedom systems and its extension in calculating the resonance frequencies of elastic bodies, as suggested by Giebe and Blechschmidt [Ann. Physik, Ser. 5 **18**, 417–485 (1933)]. The approach to the problem is illustrated with examples of piezoceramic rectangular thickness poled plates. The resonance frequencies and effective coupling coefficients are presented as functions of the plates' aspect ratio.
 [DOI: 10.1121/1.3242342]

PACS number(s): 43.38.Ar, 43.38.Fx, 43.38.Pf, 43.30.Yj [AJZ] Pages: 2983–2990

I. INTRODUCTION

Mechanical systems of piezoelectric ceramic transducers are composed of elastic bodies of finite size. A rigorous general analysis of the vibration of finite-size elastic bodies for the design of transducers is difficult. Although computer simulations of the mechanical and electrical behavior of the transducers have been made (see, for example, Ref. 1), the trial and error approach to the development of such transducers based on intuitive conceptions remains widespread. Therefore, having a simple approximate method that enables one to determine transducer parameters with a reasonable level of accuracy and gives a physical clear insight into the process of optimizing the parameters would be helpful.

The objective of this paper is to provide a simple means for calculating vibration of the mechanical system of piezoelectric transducers as two-dimensional and for estimating to what extent one-dimensional approximations are applicable within a prescribed accuracy.

Such a convenient approach in designing transducers can be developed based on the energy method for analyzing piezoelectric ceramic transducers[2] and following the general outline of the theory of coupled vibrations in two degrees-of-freedom systems and its extension in calculating resonance frequencies of elastic bodies, as suggested by Giebe and Blechschmidt.[3]

A brief outline of theory of coupled vibration in two degrees-of-freedom systems can be presented as follows. In general the energy state of a system that can be represented as a combination of two partial systems undergoing coupled vibrations can be characterized by the set of equations

$$W_{\text{pot}} = \tfrac{1}{2}(K_{11}\xi_1^2 + 2K_{12}\xi_1\xi_2 + K_{22}\xi_2^2),$$

$$W_{\text{kin}} = \tfrac{1}{2}(M_{11}\dot{\xi}_1^2 + 2M_{12}\dot{\xi}_1\dot{\xi}_2 + M_{22}\dot{\xi}_2^2), \tag{1}$$

where ξ_i are the generalized coordinates (usually displacements in the case of mechanical systems), K_{ii} and M_{ii} are the equivalent rigidities and masses of the partial systems that remain if one of the generalized coordinates is set equal to zero, and K_{12} and M_{12} are the parameters that characterize the elastic and inertial coupling mechanism between the partial systems. If the coupling is assumed to be purely elastic, then $M_{12}=0$.

The frequency equation for the system can be derived from Eq. (1) using Lagrange's equations as

$$(f_{1p}^2 - f^2)(f_{2p}^2 - f^2) - (\gamma_k^2 f_{1p}^2 f_{2p}^2 + \gamma_m^2 f^4) = 0, \tag{2}$$

where $f_{ip}=(1/2\pi)\sqrt{K_{ii}/M_{ii}}$ are the partial resonance frequencies and $\gamma_k=K_{12}/\sqrt{K_{11}K_{22}}$ and $\gamma_m=M_{12}/\sqrt{M_{11}M_{22}}$ are the coefficients of elastic and inertia coupling between the partial systems (coupling factors). Thus, the resonance frequencies of the system can be found from Eq. (2) so far as the assumptions are made regarding the partial mechanical systems and the coupling factors between them.

It is instructive to illustrate this technique with a typical example of two degrees-of-freedom mechanical system, shown in Fig. 1, having lumped parameters. Masses M_1 and M_2 vibrate on the springs with the stiffness constants K_1 and K_2 in the vertical direction. The configuration of the system is completely defined by displacements ξ_1 and ξ_2 of the masses. The partial systems are as follows: at $\xi_2=0$ the mass M_1 vibrates between springs K_1 and K_2 having a combined stiffness K_1+K_2 and at $\xi_1=0$ the mass M_2 vibrates on the spring K_2. The partial resonance frequencies are $f_{1p}=(1/2\pi)\sqrt{(K_1+K_2)/M_1}$ and $f_{2p}=(1/2\pi)\sqrt{K_2/M_2}$, respec-

[a)]The results of this work were reported in part at the 156th Meeting of the Acoustical Society of America [J. Acoust. Soc. Am. 124, 2551(A) (2008)].
[b)]Author to whom correspondence should be addressed. Electronic mail: baronov@comcast.net

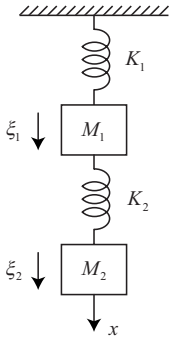


FIG. 1. Illustration of two degrees-of-freedom mechanical system with lumped parameters.

tively. After calculating the potential and kinetic energies of the system it is found that $M_{12}=0$ and $K_{12}=-K_2$. Thus, the coupling between the partial systems is elastic, and the coupling factor is $\gamma_k=-\sqrt{K_2/(K_1+K_2)}$. Now the resonance frequencies of the mechanical system can be found from Eq. (2). It is noteworthy that the solution obtained is exact to the extent that the masses and springs can be considered as elements with lumped parameters, i.e., that they hold only kinetic and potential energies in the course of vibration, respectively.

Giebe and Blechschmidt[3] suggested applying the coupling theory to determine the resonance frequencies of passive isotropic elastic bodies of certain configurations (rectangular plates and rectangular rods, as well as solid and hollow cylinders). Their approach to the problem is based on the assumptions that (a) vibration of a real elastic body may be represented as the coupled vibration of the partial one-dimensional systems, to which the real body approaches at the extreme values of its aspect ratios, and (b) the mode coupling factors may be selected in such a way as to yield the known resonance frequencies for the extreme one-dimensional systems (so called "hypothesis of Giebe and Blechschmidt"). Thus, for example, the extensional vibration of a rectangular plate was considered as a coupled vibration of two bars, as shown in Fig. 2, with inertia coupling between them. For this case the partial resonance frequencies are $f_{1p}=(1/2L_1)\sqrt{Y/\rho}$ and $f_{2p}=(1/2L_2)\sqrt{Y/\rho}$ (only the first resonance modes of vibration are presented for simplicity), and the coefficient of coupling is chosen as $\gamma_m=\sigma$ (here Y, σ, and ρ are Young's modulus, Poisson's ratio, and density). It is easy to verify that with the coupling factor chosen in this way, the resonance frequencies of the plate obtained from Eq. (2) at $\gamma_k=0$ are $f_r=f_{2p}$ and $f_r=f_{2p}/\sqrt{1-\sigma^2}$ for the extreme cases of $L_1\to 0$ (thin bar of the length L_2) and $L_1\to\infty$ (infinitely long strip vibrating through its width L_2).

The analogous approach was used by Onoe and Tiersten[4] to determine the resonance frequencies for piezoelectric ceramic bodies having the same geometry. While their approach is only approximate, it does give simple expressions for the resonance frequencies vs aspect ratio of elastic bodies considered in Refs. 3 and 4 (L_2/L_1 in the case of an isotropic rectangular plate), which prove to be in good agreement with experimental data reported in Ref. 4 at least for the lowest branch of the frequency spectrum.

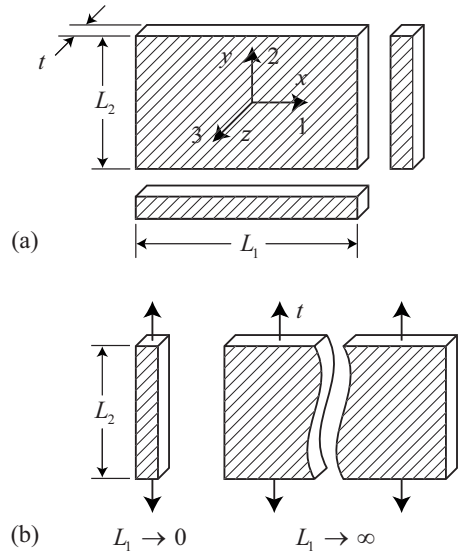


FIG. 2. Drawing of a rectangular plate (a) considered as a coupled vibration of two rods and (b) shown at extreme aspect ratios.

Although the Onoe and Tiersten results are sufficiently accurate for determining the resonance frequencies of piezoelectric bodies, the technique used was not and cannot be directly applied in calculating the electromechanical parameters of transducers. Moreover, it cannot be applied to treatment of transducers as electroacoustical, i.e., under the action of acoustical loads, because this requires knowing how the modes of vibration of piezoelement change vs their aspect ratios. For this purpose the coupling theory approach needs to be modified. This can be done by employing the energy method for the treatment of electroacoustical transducers.[2] For the application of the energy method it is crucial that the assumed modes of displacement in the mechanical system of the transducer are chosen appropriately. If the modes of displacement are assumed to be the same as those suggested for the partial systems, then all of the transducer parameters can be determined using the energy method, as previously described in Ref. 2.

This approach to the treatment of the coupled vibrations in the piezoelectric ceramic transducers is illustrated with an example of the electromechanical transducers in the shape of thin rectangular piezoceramic plates poled through their thickness, which can be regarded as a two-dimensional problem.

II. THICKNESS POLED THIN PLATE (STRIP)

Consider the vibration of a rectangular ceramic plate. The geometry of the plate is shown in Fig. 2(a). The z (3) axis is in the direction of polarization and is perpendicular to the fully electroded surfaces. We assume that the thickness of the plate t is much smaller than its lateral dimensions L_1 and L_2, and therefore the boundary condition can be imposed that the stress $T_3=0$. We are interested in the dependence of the plate's parameters on its aspect ratio, L_1/L_2. In the coupling theory approach[3] the partial systems for a rectangular plate were chosen in the shape of thin rods having lengths L_1 and L_2. Thus, if to take into consideration only the fundamental

modes of the rod vibration, which is the case in this paper, the partial distributions of displacement in the plate may be represented as

$$\xi_x = \xi_1 \sin(\pi x/L_1), \quad \xi_y = \xi_2 \sin(\pi y/L_2). \tag{3}$$

The corresponding distribution of displacement and strain in the plate will be

$$\xi(x,y) = \xi_1 \sin(\pi x/L_1) + \xi_2 \sin(\pi y/L_2), \tag{4}$$

$$S_1 = \xi_1(\pi/L_1)\cos(\pi x/L_1), \quad S_2 = \xi_2(\pi/L_2)\cos(\pi y/L_2). \tag{5}$$

For the boundary condition $T_3=0$ it is convenient to use the piezoelectric equations specified in the form[5]

$$\begin{aligned} S_1 &= s_{11}^E T_1 + s_{12}^E T_2 + d_{31}E_3, \\ S_2 &= s_{12}^E T_1 + s_{11}^E T_2 + d_{31}E_3, \\ S_3 &= s_{13}^E T_1 + s_{13}^E T_2 + d_{33}E_3, \\ D_3 &= d_{31}(T_1 + T_2) + \varepsilon_{33}^T E_3. \end{aligned} \tag{6}$$

After converting Eq. (6) to the form with the strains as independent variables, the following expressions will be obtained.

For the stress at $E_3=0$,

$$\begin{aligned} T_1^E &= \frac{Y_1}{1-\sigma_1^2}(S_1 + \sigma_1 S_2), \\ T_2^E &= \frac{Y_1}{1-\sigma_1^2}(\sigma_1 S_1 + S_2), \end{aligned} \tag{7}$$

where the notations $Y_1 = 1/s_{11}^E$ and $\sigma_1 = -(s_{12}^E/s_{11}^E)$ are introduced.

For the strain in the z axis direction at $E_3=0$ we obtain

$$S_3 = -\frac{\sigma_1}{1-\sigma_1}(S_1 + S_2). \tag{8}$$

And for the charge density we obtain

$$D_3 = D_3^E(S_1,S_2) + \varepsilon_{33}^{S_{1,2}} E_3, \tag{9}$$

where

$$D_3^E(S_1,S_2) = \frac{Y_1 d_{31}}{1-\sigma_1}(S_1 + S_2), \tag{10}$$

$$\varepsilon_{33}^{S_{1,2}} = \varepsilon_{33}^T(1-k_p^2). \tag{11}$$

In Eq. (11) $\varepsilon_{33}^{S_{1,2}}$ is the dielectric constant at $S_1=S_2=0$ and k_p is the planar coupling coefficient of the piezoceramic material.

Following the procedure of employing the energy method,[2] the energies associated with plate vibration should be determined.

The *potential (mechanical) energy* calculated at $E_3=0$ is

$$\begin{aligned} W_m^E &= \frac{1}{2}\int_{\tilde{V}} (T_1^E S_1 + T_2^E S_2) d\tilde{V} \\ &= \frac{1}{2}\frac{Y_1}{1-\sigma_1^2}\int_{\tilde{V}} (S_1^2 + 2\sigma_1 S_1 S_2 + S_2^2) d\tilde{V}. \end{aligned} \tag{12}$$

Here $\tilde{V}$ is the volume of mechanical system. After substituting expressions (5) for the strain and integrating over the volume of the plate we obtain

$$W_m^E = \tfrac{1}{2}(K_{11}\xi_1^2 + 2K_{12}\xi_1\xi_2 + K_{22}^E\xi_2^2), \tag{13}$$

where

$$K_{11} = \frac{\pi^2 Y_1 t}{2(1-\sigma_1^2)}\frac{L_2}{L_1}, \quad K_{22} = \frac{\pi^2 Y_1 t}{2(1-\sigma_1^2)}\frac{L_1}{L_2}, \quad K_{12} = \frac{4Y_1\sigma_1 t}{1-\sigma_1^2}. \tag{14}$$

The *kinetic energy* of a vibrating plate is

$$W_{\text{kin}} = \frac{\rho}{2}\int_{\tilde{V}} (\dot{\xi}_x^2 + \dot{\xi}_y^2 + \dot{\xi}_z^2) d\tilde{V}, \tag{15}$$

where ρ is the plate's density and $\dot{\xi}_x$, $\dot{\xi}_y$, and $\dot{\xi}_z$ are the velocity distributions along the coordinate axes. Velocity $\dot{\xi}_z$ can be found from Eq. (8) as $\dot{\xi}_z = \dot{\xi}_3 = \dot{S}_3 z$. Thus, the velocity of the plate's surface vibration is

$$\dot{\xi}_{z\Sigma} = -\frac{\sigma_1 t}{2(1-\sigma_1)}(\dot{S}_1 + \dot{S}_2). \tag{16}$$

Note that the velocity $\dot{\xi}_{z\Sigma}$, being relatively small compared with velocities in lateral directions, is important for measuring the mode shapes of the plate vibration. This is the quantity that determines the acoustic field in the case that a laterally vibrating plate is used for radiating in the z direction.

After integrating over the volume of the plate, the kinetic energy can be represented as

$$W_{\text{kin}} = \tfrac{1}{2}M_{11}\dot{\xi}_1^2 + M_{12}\dot{\xi}_1\dot{\xi}_2 + \tfrac{1}{2}M_{22}\dot{\xi}_2^2, \tag{17}$$

where

$$\begin{aligned} M_{11} &= \frac{M}{2}\left[1 + \frac{\pi^2\sigma_1^2 t^2}{12(1-\sigma_1^2)L_1^2}\right], \\ M_{22} &= \frac{M}{2}\left[1 + \frac{\pi^2\sigma_1^2 t^2}{12(1-\sigma_1^2)L_2^2}\right], \\ M_{12} &= M\frac{\sigma_1^2 t^2}{3(1-\sigma_1^2)L_1 L_2}. \end{aligned} \tag{18}$$

In these relations, $M = L_1 L_2 t\rho$ is the mass of the plate. So far as $t \ll L_1, L_2$, the contribution of the terms with factors t^2/L^2 is negligible and

$$M_{11} = M_{22} = M/2, \quad M_{12} \cong 0. \tag{19}$$

Comparing expressions for the potential energy W_m^E by formulas (13) together with Eq. (14) and for the kinetic energy W_{kin} by formulas (17) together with Eq. (18) and the general expressions (1), we arrive at the following conclusions. In the case under consideration the partial system, de-

termined by the condition that $\xi_1=0$ or $S_1=0$, as shown in Fig. 2(b) at $L_1\to\infty$, constitutes a strip that is infinite in the x direction and vibrating along its width L_2. Similarly, another partial system (at $\xi_2=0$) is the strip that is infinite in the y direction and vibrating along dimension L_1. [Note that a bar, which represents another extreme case at $L_1\to 0$ shown in Fig. 2(b), does not coincide with the partial system at $\xi_2=0$.]

The partial resonance frequencies are

$$f_{ip}=\frac{1}{2\pi}\sqrt{\frac{2K_{ii}}{M}}=\frac{1}{2L_i}\sqrt{\frac{Y_1}{\rho(1-\sigma_1^2)}},\quad i=1,2. \tag{20}$$

The coupling between the partial systems is elastic with the coupling factor

$$\gamma_k=\frac{K_{12}}{\sqrt{K_{11}K_{22}}}=\frac{8\sigma_1}{\pi^2} \tag{21}$$

(the subscript k will be further omitted). For PZT-4 $\sigma_1=0.33$ and $\gamma=0.27$.

The *electrical energy* of a transducer at $S_1=S_2=0$ is

$$W_{\mathrm{el}}^{S_{1,2}}=\frac{1}{2}\int_{\tilde{V}}\varepsilon_{33}^{S_{1,2}}E_{33}^2 d\tilde{V}=\frac{1}{2}C_{\mathrm{el}}^{S_{1,2}}V^2, \tag{22}$$

where $C_{\mathrm{el}}^{S_{1,2}}=\varepsilon_{33}^{S_{1,2}}L_1L_2/t$ is the capacitance of a transducer at $S_1=S_2=0$, and $\varepsilon_{33}^{S_{1,2}}$ is given by formula (11).

The *electromechanical energy* of a transducer is

$$W_{\mathrm{em}}=\frac{1}{2}\int_{\tilde{V}}D_3^E(S_1,S_2)E_3 d\tilde{V}. \tag{23}$$

After substituting expression (10) for D_3^E and integrating over the volume, it will be obtained as

$$W_{\mathrm{em}}=\tfrac{1}{2}V(n_1\xi_1+n_2\xi_2), \tag{24}$$

where

$$n_1=\frac{2Y_1d_{31}L_2}{1-\sigma_1},\quad n_2=\frac{2Y_1d_{31}L_1}{1-\sigma_1}. \tag{25}$$

After all the energies associated with the functioning of the transducer as an electromechanical device are determined, Lagrange's equations, which describe its vibration, can be obtained, as was done in Ref. 2. Lagrange's equations can be represented in the following form without taking into account the energy losses and external loads:

$$(j\omega M_{11}+K_{11}/j\omega)U_1+(K_{12}/j\omega)U_2=n_1V,$$

$$(K_{12}/j\omega)U_1+(j\omega M_{22}+K_{22}/j\omega)U_2=n_2V. \tag{26}$$

Here $U_i=j\omega\xi_i$ and the time dependent factor $e^{j\omega t}$ is omitted.

It is not essential to consider the losses or the external loads for illustrating the coupled vibration analysis. When necessary, they can be introduced in Lagrange's equations in exactly the same general way, as it is described in Ref. 2.

In order to determine the *resonance frequencies* and the *resonance mode shapes* of the transducer, we first consider its free vibration. After some manipulations of Eq. (26) at $V=0$, we obtain

$$(f_{1p}^2-f^2)U_1+(K_{12}/K_{11})f_{1p}^2U_2=0,$$

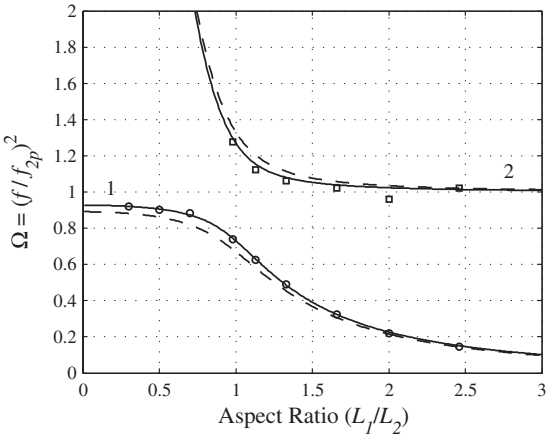


FIG. 3. Resonance frequency dependence on aspect ratio for lower (1) and upper (2) frequency branches (solid, calculated; dashed, Refs. 3 and 4; and markers, measured).

$$(K_{12}/K_{22})f_{2p}^2U_1+(f_{2p}^2-f^2)U_2=0. \tag{27}$$

The frequency equation follows in the form of Eq. (2), from the set of Eq. (27) where $\gamma_m=0$, and the elastic coupling factor is defined by Eq. (21).

The dependence of parameters of a transducer on the ratio of the plate dimensions may then be determined. We assume for definiteness that the dimension L_2 (i.e., f_{2p}) is kept constant and L_1 changes. After denoting $f^2/f_{2p}^2=\Omega$ as the normalized non-dimensional resonance frequency factor and $(f_{2p}/f_{1p})=\beta$ as ratio of the partial frequencies, Eq. (2) may be transformed to

$$\Omega^2-(1+1/\beta^2)\Omega+(1-\gamma^2)/\beta^2=0. \tag{28}$$

From this equation two resonance frequency branches corresponding to the solutions Ω_1 and Ω_2 may be found as functions of the aspect ratio for the plate (for a uniform plate $\beta=L_1/L_2$). We will use the convention that $\Omega_1(\beta)$ forms the lower and $\Omega_2(\beta)$ the upper frequency branch. For PZT-4 $\gamma=(8/\pi^2)\sigma_1=0.27$. The frequency dependency calculated at this value of the coupling factor is shown in Fig. 3 as solid lines. It is clear from Fig. 3 that $\Omega_1<1$. From Eq. (28) at $\beta=1$ it follows, in particular, that

$$\Omega_1=1-\gamma,\quad \Omega_2=1+\gamma,\quad \frac{f_2^2}{f_1^2}=\frac{\Omega_2}{\Omega_1}=\frac{1+\gamma}{1-\gamma}. \tag{29}$$

After the resonance frequencies are determined, the corresponding mode shape factors, which will be defined as

$$\mathrm{ms}_i=(U_1/U_2)_{\mathrm{at}\ \Omega=\Omega_i}=(\xi_1/\xi_2)_{\mathrm{at}\ \Omega=\Omega_i},\quad i=1,2, \tag{30}$$

may be found from the set of Eq. (27). Namely,

$$\mathrm{ms}_i=-\frac{K_{12}}{K_{11}}\frac{1}{1-\Omega_i\beta^2}. \tag{31}$$

It follows from expressions (21) and (14) that $(K_{12}/K_{11})=\gamma\beta$. Thus, Eq. (31) can be represented as

$$\mathrm{ms}_i=-\frac{\gamma\beta}{1-\Omega_i\beta^2}. \tag{32}$$

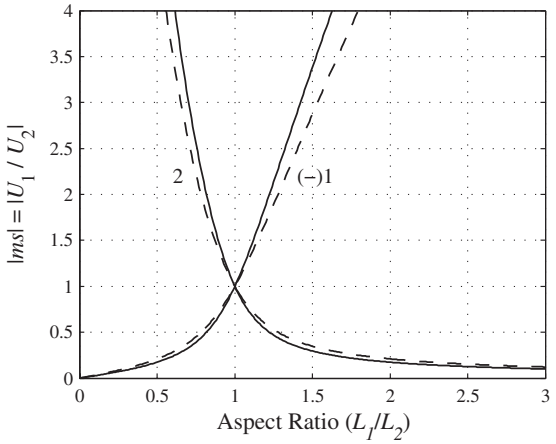


FIG. 4. Dependence of the mode shape factors on aspect ratio for lower (1) and upper (2) frequency branches (solid, calculated; dashed, calculated using $\gamma=\sigma_1$ following Refs. 3 and 4).

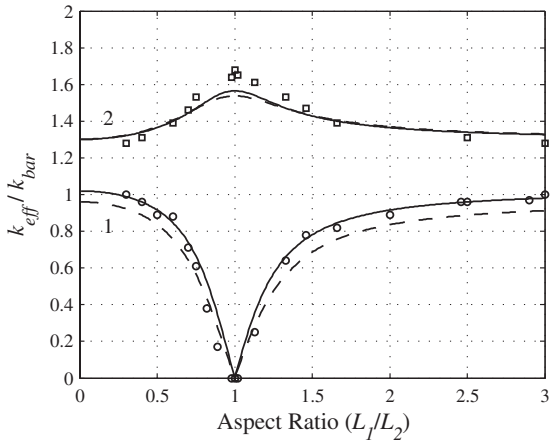


FIG. 5. Dependence of the effective coupling coefficients on aspect ratio for the lower (1) and upper (2) resonance frequency branches (solid, calculated; dashed, calculated using $\gamma=\sigma_1$ following Refs. 3 and 4; and markers, measured). Values are normalized to the effective coupling coefficient for a bar.

The dependence of the mode shape factors on the aspect ratio β is shown in Fig. 4. It follows from the figure that, at frequencies pertaining to the lower branch, the velocities U_1 and U_2 are in anti-phase. In particular, it follows from expressions (32) and (29) at $\beta=1$ that $\mathrm{ms}_1=-1$ and $\mathrm{ms}_2=1$. This means that at higher and lower resonance frequencies the velocities have the same magnitude but are in-phase at the higher frequency and are anti-phase at the lower frequency.

The *effective electromechanical coupling coefficient* of a transducer can be found from the relation[6]

$$\frac{k_{\mathrm{eff}}^2}{1-k_{\mathrm{eff}}^2}=\frac{W_{\mathrm{em}}^2}{W_{\mathrm{el}}^{S1,2}W_m^E}, \tag{33}$$

where the expressions for the energies W_m^E, $W_{\mathrm{el}}^{S1,2}$, and W_{em} are given by formulas (13), (22), and (24), respectively. The expressions for the energies W_m^E and W_{em} can be modified as follows by using the mode shape coefficients:

$$W_{mi}^E=\frac{1}{2}K_{22}\xi_2^2\left(1+\frac{2K_{12}}{K_{22}}\mathrm{ms}_i+\frac{K_{11}}{K_{22}}\mathrm{ms}_i^2\right)$$
$$=\frac{1}{2}\frac{K_{22}\xi_2^2}{\beta^2}(\beta^2+2\gamma\beta\mathrm{ms}_i+\mathrm{ms}_i^2), \tag{34}$$

$$W_{\mathrm{em}}=\frac{1}{2}Vn_2\xi_2\left(\frac{n_1}{n_2}\mathrm{ms}_i+1\right)=\frac{1}{2}V\frac{n_2\xi_2}{\beta}(\mathrm{ms}_i+\beta). \tag{35}$$

Substituting expressions (22), (34), and (35) into relation (33) results in

$$\frac{k_{\mathrm{eff}\,i}^2}{1-k_{\mathrm{eff}\,i}^2}=\frac{n_2^2}{C_{\mathrm{el}}^{S1,2}K_{22}}\frac{(\mathrm{ms}_i+\beta)^2}{\mathrm{ms}_i^2+2\gamma\beta\mathrm{ms}_i+\beta^2}=A_{\mathrm{em}}A_i(\beta). \tag{36}$$

The right side of relation (36), for a given coupling factor γ, can be represented as a product of two factors, one of which, $A_{\mathrm{em}}=n_2^2/C_{\mathrm{el}}^{S1,2}K_{22}$, depends on the electromechanical properties of a transducer only and the other, $A_i(\beta)$, is a function of the aspect ratio

$$A_i(\beta)=\frac{(\mathrm{ms}_i+\beta)^2}{\mathrm{ms}_i^2+2\gamma\beta\mathrm{ms}_i+\beta^2}. \tag{37}$$

After substituting expressions for the parameters K_{22}, $C_{\mathrm{el}}^{S1,2}$, and n_2 given by formulas (14), (22), and (25), A_{em} becomes

$$A_{\mathrm{em}}=\frac{n_2^2}{C_{\mathrm{el}}^{S1,2}K_{22}}=\frac{4}{\pi^2}(1+\sigma_1)\frac{k_p^2}{1-k_p^2}. \tag{38}$$

For a transducer made from PZT-4 the factor $A_{\mathrm{em}}=0.27$.

Dependences of the effective coupling coefficients on the aspect ratio for the upper and lower resonance frequency branches are shown in Fig. 5. The coupling coefficients are normalized to the effective coupling coefficient of a thin longitudinally vibrating side electroded bar, $k_{\mathrm{eff\,bar}}$, that represents the limiting case of the plate at $\beta\to 0$ and L_2 constant, which is known[6] as

$$k_{\mathrm{eff\,bar}}^2=\frac{1}{1+(\pi^2/8)(1-k_{31}^2)/k_{31}^2}. \tag{39}$$

It is interesting to estimate the relation of the obtained solutions for the resonance frequencies and the coupling coefficients with known results for the limiting one-dimensional configurations of the plate, which are a long bar (at $\beta\to 0$ and L_2 constant) and a long thin strip (at $\beta\to\infty$ and L_2 constant). It follows from Eq. (28) that

$$\Omega=\frac{1+\beta^2}{2\beta^2}\left[1\pm\sqrt{1-\frac{4\beta^2(1-\gamma^2)}{(1+\beta^2)^2}}\right], \tag{40}$$

and for $\beta\to 0$ $\Omega\approx(1/2\beta^2)(1+\beta^2)\{1\pm[1-2\beta^2(1-\gamma^2)]\}$ and

$$\Omega_1\approx(1-\gamma^2),\quad \Omega_2\to 1/\beta^2. \tag{41}$$

Thus at $\beta\to 0$

$$f_1\to f_{2p}\sqrt{1-\gamma^2}\quad\text{and}\quad f_2\to(f_{2p}/\beta)=f_{1p}.$$

Taking into account expression (20) for the partial frequencies f_{ip}, we arrive at the resonance frequency for an infinitely thin bar of the length L_2,

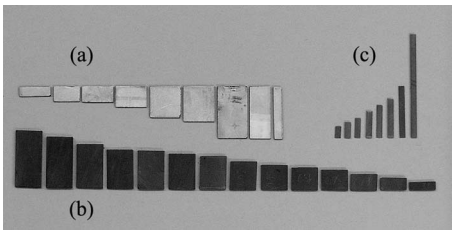


FIG. 6. Photograph of the sets of plates with different aspect ratios used in the experimental investigation of the paper. Set (a): $t=1.1$ mm, $L_2=15$ mm, $\beta=2.0-0.3$ and PZT-4. Set (b): $t=1.8$ mm, $L_2=12.7$ mm, $\beta=2.0-0.3$, and PZT-4. Set (c): $t=1.0$ mm, $L_2=2.8$ mm, $\beta=18-2.0$, and PZT-8.

$$f_1 = \frac{1}{2L_2}\sqrt{\frac{Y_1(1-\gamma^2)}{\rho(1-\sigma_1^2)}}, \tag{42}$$

and for an infinitely long strip of the width L_1,

$$f_2 = \frac{1}{2L_1}\sqrt{\frac{Y_1}{\rho(1-\sigma_1^2)}}. \tag{43}$$

Given that the exact value for the resonance frequency of the bar should be $f_{\text{bar}}=(1/2L_2)\sqrt{Y_1/\rho}$ and that $\gamma=0.81\sigma_1$, for PZT-4 ($\sigma_1=0.33$, $\gamma=0.27$) we obtain $f_1=1.02f_{\text{bar}}$. Thus, the error of the current approach can be considered as negligible for the limiting case that $\beta\to 0$. For another extreme case (long strip) the value of resonance frequency obtained by formula (43) is exact. It is noteworthy that the exact value of $f_1=f_{\text{bar}}$ would be achieved at $\gamma=\sigma_1$. Such a value of the coupling factor was adopted in Ref. 3, though under the assumption of inertial coupling.

In terms of the effective coupling coefficients obtained from the two-dimensional and one-dimensional approximations, comparison of the values for the limiting cases of the transducers in the shape of a bar and of a strip gives the following results. Using expressions (32) for the mode shape coefficients and Eq. (41) for the normalized frequencies Ω_i at $\beta\to 0$, we arrive at $\text{ms}_1\to-\gamma\beta$ and $\text{ms}_2\to 1/\gamma\beta$. From formula (37) for $A_i(\beta)$ we obtain $A_1(\beta)_{\beta\to 0}\to(1-\gamma)/(1+\gamma)$ and $A_2(\beta)_{\beta\to 0}\to 1$, respectively. Given that for PZT-4 $A_{\text{em}}=0.27$ and $\gamma=0.27$, it follows from Eq. (36) that for a bar $k_{\text{eff }1}=0.32$ and for a strip $k_{\text{eff }2}=0.45$. Using the one-dimensional approximation we obtain the results for a bar $k_{\text{eff bar}}=0.30$ [by formula (39)] and for a strip $k_{\text{eff strip}}=0.45$. The effective coupling coefficient for a strip is obtained from formula (39) after substituting k_{31} by k'_{31}, which is known from Ref. 5 as the coupling coefficient at the boundary conditions that are experienced by a ceramic material in the thin strip geometry $(T_3=0, S_2=0)$. According to Ref. 5 $k'_{31}=(k_{31}/\sqrt{1-k_{31}^2})\times(\sqrt{(1+\sigma_1)/(1-\sigma_1)}$, and for PZT-4 $k'_{31}=0.49$.

An experimental investigation was made in order to verify the results of calculations presented in Figs. 3 and 5 in a broad range of aspect ratios.

III. RESULTS OF EXPERIMENTAL INVESTIGATION

An experimental investigation was fulfilled with a variety of thin $(t/L<10)$ plates made from PZT-4 and PZT-8. The sets of plates with different aspect ratios, which were used in the investigation, are displayed in Fig. 6.

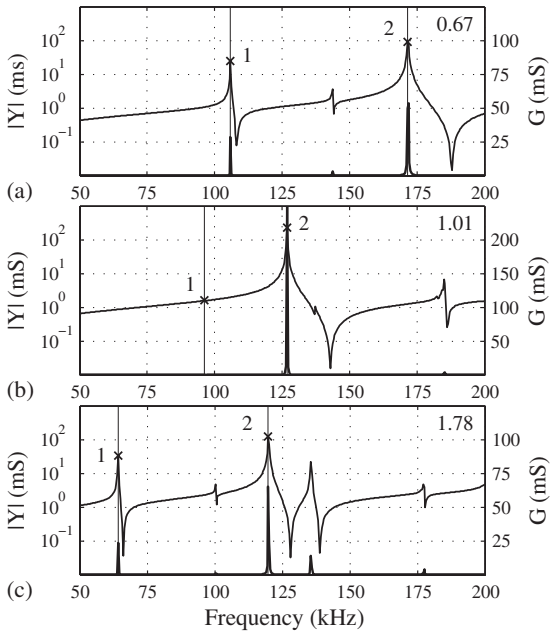


FIG. 7. Plot of admittance (magnitude) and conductance curves for plates with different aspect ratios: (a) $L_1/L_2=0.67$, (b) 1.01, and (c) 1.78.

The resonance-antiresonance method was used for determining the resonance frequency and the effective coupling coefficient dependencies on the aspect ratio of the plates. The measurements were made using an impedance analyzer HP 4194A and in accordance with IEEE Standard on Piezoelectricity (IEEE Standard 176-1987).

Peculiarities exist in the experimental investigation of coupled vibrations in mechanical systems in general, and in plates, in particular, due to the fact that these systems are inherently multi-resonance. A number of "unwanted" resonances may exist in their frequency spectra that are related to overtones of vibrations and their coupling with the fundamental modes. Examples of the admittance plots and corresponding plots of the conductance for the plates with different aspect ratios are shown in Fig. 7. The frequencies belonging to the frequency branches under investigation are marked in these plots by an x. A separation between the "targeted" fundamental resonance frequencies can be small depending on the aspect ratio, especially in the vicinity of strongest coupling. Under these circumstances, difficulties arise in identifying the targeted resonance frequencies. Therefore, the accuracy becomes questionable in application of the resonance-antiresonance method for determining the effective coupling coefficients for the corresponding modes of vibration since, strictly speaking, this method is valid for electromechanical systems with one degree-of-freedom.

In order to overcome these difficulties, an experimental investigation was performed on a large number of the sample plates (examples of which are depicted in Fig. 6), having small intervals between their aspect ratios with one dimension kept constant. Consequently, a reliable trend was obtained for the experimentally measured resonance frequencies and their dependence on aspect ratios as well as for the effective coupling coefficients for the corresponding modes of vibration, which were determined by using the resonance-antiresonance method. In order to increase reliability of the

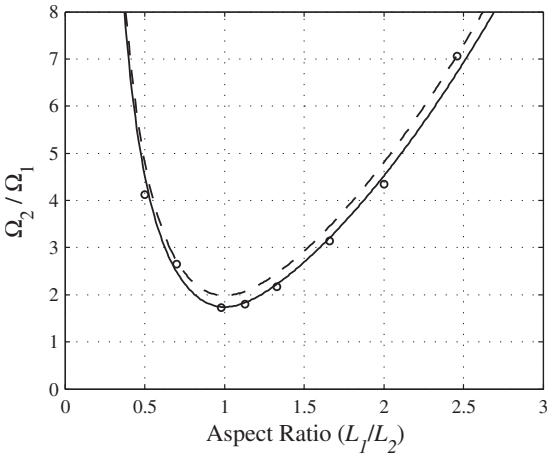


FIG. 8. Ratio of resonance frequencies of upper (2) to lower (1) branches vs aspect ratio (solid, calculated; dashed, following Refs. 3 and 4; and markers, measured).

data obtained for particular aspect ratios, the measurements were performed with several sets of plates (not shown). The trend lines of the experimental data obtained for the resonance frequencies and coupling coefficients are presented in Figs. 3 and 5 along with the theoretically predicted dependency. The plots are normalized to values at the limiting case of $\beta \ll 1$ in order to avoid differences between ceramic compositions and absolute values of dimensions.

The agreement between the experimental and theoretical results is very good. Based on these results, the conclusion can be made that both experimental technique and theoretical analysis are accurate at least within the discrepancy between the experimental and calculated data.

It was noted before that, for determining the resonance frequencies of the thickness poled piezoceramic plates, the different partial systems (a long bar and strip) and the coupling factor $\gamma=\sigma_1$ between them (which makes $\gamma=0.33$ for PZT-4) were suggested in Ref. 4. As it follows from formula (16) at $\gamma=\sigma_1$, the exact value of resonance frequency for the limiting case of a bar would be obtained from the current solution. This result is expected because this value of the coupling factor was chosen in Ref. 4 exactly from the condition of satisfying the values for the resonance frequencies at the extreme aspect ratios.

It is interesting to compare the accuracy of calculating the resonance frequencies in a broad range of aspect ratios by the two approaches: one used in Ref. 4 and the other used in the current paper. For this purpose the frequency curves calculated at $\gamma=0.33$ are also presented in Figs. 3 and 8. It is noteworthy that for the lower branch the results are in better agreement with experiment than for the upper branch, although less accurate than with the current approach (in Ref. 4 it was also noted that good agreement with experiment was obtained only for the lower branch). In order to make the comparison between the results obtained by these two approaches more quantitatively expressive, Fig. 8 shows the ratio of resonance frequencies of upper to lower branches vs aspect ratio. This is advantageous in terms of comparing the accuracy of calculations than the estimation of the absolute values since the properties of ceramics do not matter in this case. It can be seen that the results of the current treatment are more accurate in the range around the point of the strongest coupling, which is of great interest. In particular, for $\beta=1$ according to formula (29), $f_2/f_1=1.32$ for $\gamma=0.27$ and $f_2/f_1=1.41$ for $\gamma=0.33$. The experimentally measured value is $f_2/f_1=1.31$. (The measurement was made at $\beta=0.98$. At $\beta=1$ it is hard to determine f_1 accurately because the effective electromechanical coupling coefficient goes to zero for the lower frequency branch at this point, as it is shown in Fig. 5.)

IV. DISCUSSION OF THE RESULTS

The results presented in this paper illustrate with an example of the thickness poled rectangular plates that the coupled vibration analysis technique considered provides a simple, physically clear, and sufficiently accurate for practical applications means for treatment of piezoceramic transducers.

Good agreement between the calculated and measured resonance frequencies for both low and high frequency branches of the fundamental modes of vibration and for the effective coupling coefficients related to these modes is obtained.

It is noteworthy that the results for the effective coupling coefficients were not previously reported in the literature, yet these data are particularly important for designing transducers. These results are in accord with the general conclusion for the coupled systems that at low frequency branch vibrations in directions of the generalized coordinates take place in anti-phase, whereas at high frequency branch they are in-phase. The effective electromechanical coupling coefficient drops to zero for a mode of vibration related to the low frequency branch and it reaches a maximum for a mode related to the high branch at the point of the strongest coupling (at $\beta=1$). Analogous results were reported in Ref. 7 for thin walled radial poled cylindrical piezoelements, in which case the point of the strongest coupling was at the aspect ratio $\beta=1.57$.

In terms of determining the resonance frequencies the proposed modified coupling theory approach is, in practice, as accurate for the extreme case of a large aspect ratio as the approach of the coupling theory, described in Ref. 4, and is more accurate for the aspect ratios in vicinity of the point of the strongest coupling. The difference becomes negligible at aspect ratios close to the extreme case at $\beta=0$, and in vicinity of the extreme case the results obtained from Ref. 4 become slightly more accurate. However, the most appropriate way to consider the vibration of the mechanical systems that correspond to the extreme cases for a thin plate is to treat them as one-dimensional in the first place.

Since the plots in Figs. 3 and 5 are normalized to the values that correspond to the one-dimensional approximation, they allow the estimation of results up to values of the aspect ratios where the one-dimensional approximation can be used instead of two-dimensional within certain accuracy. Thus, at aspect ratio $\beta=0.5$ the resonance frequency drops

by 1.2% compared with those for infinitely thin bar. The effective electromechanical coupling coefficient drops by 2.5% at aspect ratio $\beta=0.3$.

Based on the results discussed in this paper it can be concluded that the analytical approach and the experimental technique used proved to be appropriate for treating bar and plate transducers, which can be considered as two-dimensional.

ACKNOWLEDGMENT

This work was funded by BTech Acoustics, LLC.

[1]R. Lerch, "Simulation of piezoelectric devices by two- and three-dimensional finite elements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 233–247 (1990).

[2]B. S. Aronov, "The energy method for analyzing the piezoelectric electroacoustic transducers," J. Acoust. Soc. Am. **117**, 210–220 (2005).

[3]E. Giebe and E. Blechschmidt, "Experimental and theoretical studies of extensional vibrations of rods and tubes," Ann. Physik, Ser. 5 **18**, 417–485 (1933).

[4]M. Onoe and H. F. Tiersten, "Resonant frequencies of finite piezoelectric vibrators with high electromechanical coupling," IEEE Trans. Ultrason. Eng. **10**, 32–39 (1963).

[5]D. A. Berlincourt, D. R. Curran, and H. Jaffe, in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1964), Vol. **IA**.

[6]B. S. Aronov, "On the optimization of the effective electromechanical coupling coefficients of a piezoelectric body," J. Acoust. Soc. Am. **114**, 792–800 (2003). Note that there is a misprint in the referred formula (38) for the effective coupling coefficient of the side electroded longitudinally vibrating bar. It should read as presented by formula (40) of the current paper.

[7]B. S. Aronov, D. A. Brown, and S. Regmi, "Experimental investigation of coupled vibrations in piezoelectric cylindrical shells," J. Acoust. Soc. Am. **120**, 1374–1380 (2006).

Estimation of vehicle speed using wayside sound pressure onset rate

Jeffrey A. Zapfe,[a] Eric W. Wood, and Marc S. Newmark
Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138-1118

(Received 30 April 2009; revised 10 September 2009; accepted 8 October 2009)

Transportation-related environmental noise studies, particularly those associated with rail traffic, often require long term measurements near existing rights of way. In addition to the sound produced by passing vehicles, it is also useful to know their speed. Previously, speed measurements have generally required an additional piece of instrumentation, such as a second microphone, video camera, or a radar gun—resulting in added measurement cost and complexity. The present study shows how estimations of vehicle speed can be obtained using a single wayside microphone. The method is based on the rate of pressure rise as the vehicle approaches, specifically the maximum onset rate. This paper shows how the maximum onset rate depends only on the vehicle speed, the microphone distance, and the speed of sound. In order to demonstrate the feasibility of the approach field test data are presented for train passages ranging in speed from 33.5 to 55.6 m/s, as measured by a microphone located 28.65 m from the tracks. Speed estimates derived from the onset rate were generally accurate to within 8% of the independently measured speed of the train.
 [DOI: 10.1121/1.3257599]

PACS number(s): 43.50.Lj, 43.50.Yw [AJZ] Pages: 2991–2997

I. INTRODUCTION

An environmental impact study (EIS) is usually conducted as part of the planning phase for a new transportation project. The analysis of wayside noise is often a critical component of the EIS, particularly those associated with highway and rail projects. The Federal Transit Administration (FTA),[1] for example, provides specific criteria for acceptable levels of wayside noise for new transit projects.

In most cases, the environmental assessment requires the characterization of the existing pre-project noise exposure. If the planned project area already includes related noise sources, such as an existing rail line or highway, then it may be necessary to identify and measure specific noise events. For example, if a new light rail system was being planned for an existing right of way used by freight trains, then it would be important to characterize the noise produced by the freight operations so the impact of the light rail system could be assessed fairly.

When measuring the sound produced by a moving noise source, it may be necessary to know the vehicle's speed at the time of the measurement. Previously, if a single microphone was used to monitor the sound, then additional equipment or additional processing (Doppler analysis) was needed to obtain the speed information. The usual methods employed to measure speed involve either a radar gun or a timed gate (time to travel a known distance). Direct feedback or telemetry from the vehicle can also be used, although this is usually limited to cases where dedicated test vehicles are involved.

[a]Author to whom correspondence should be addressed. Electronic mail: jzapfe@acentech.com

A method that could estimate the speed of a vehicle using just the microphone would be highly desirable because it would eliminate the need for supplemental equipment. This is the basis of the present work.

Considerable work has been done to characterize the acoustic sources associated with trains, automobiles, and trucks in order to predict wayside noise. Guidance manuals in the United States published by the Federal Railroad Administration,[2] the Federal Transit Administration,[1] and the Federal Highway Administration[3] provide general methods to estimate wayside noise from highway and rail operations. However, to date wayside noise prediction methods have not addressed the reverse problem, namely, "Can the wayside pressure be used to estimate the vehicle's speed?" Hanson[4] presented a graphical relationship between onset rate and speed for MAGLEV and high speed rail systems. Hanson's data show that onset rate increases as vehicle speed increases; however, he did not present an analytical or empirical prediction model in his paper. The present paper develops an analytical rationale and formal method to predict vehicle speed that was implied, but not explored in detail by Hanson.

II. ANALYTICAL DEVELOPMENT

A. Assumptions

The underlying assumptions of the analytical development are as follows:

(1) the vehicle is traveling in a straight line, at a constant speed;
(2) the sound field in front of the vehicle can be modeled as a monopole source;
(3) the vehicle is subsonic; and
(4) the environmental conditions are calm and steady.

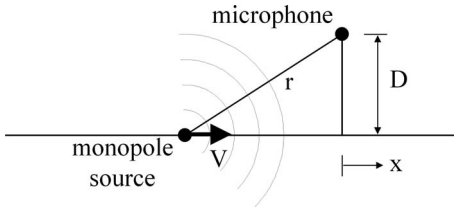


FIG. 1. Problem setup. Monopole source traveling in a straight line, at constant speed V, past a microphone. The distance to the closest point of approach is D, and the distance from the source to the microphone is r.

Unless stated otherwise, all distances are in meters, all speeds are in m/s, and all pressures are in pascals. Sound pressure level is referenced to 20 μPa.

B. Sound pressure measured by a wayside microphone

In investigating the effect of speed on the directivity of rail noise sources, Zhang[5] presented basic relationships for the pressure field produced by moving sound sources. While the work of Zhang is not directly applicable to the present problem of estimating vehicle speed from wayside sound pressure, the work does provide a basis for the development of the theory.

Consider a monopole sound source moving in a straight line past a stationary microphone, as illustrated in Fig. 1. The microphone is located a distance, D, from the path of travel and the source position, x, is referenced to the closest point of approach. If the source speed, V, is constant, then at any time, t, the distance from the source to the microphone, r, is given by

$$r = (x^2 + D^2)^{1/2}, \tag{1}$$

where x and r are functions of time. Zero time is referenced to the point when the source passes the microphone ($x=0$); consequently, when the source is approaching the microphone ($x<0$), the time is also negative.

Assuming a monopole radiation pattern, the sound pressure level, Lp, measured at the microphone will be given by

$$\begin{aligned} Lp &= Lp_0 + 20\log\left(\frac{R_0}{r}\right) \\ &= Lp_0 + 20\log(R_0) - 10\log(x^2 + D^2), \end{aligned} \tag{2}$$

where the monopole source is defined by a reference sound pressure level, Lp_0, measured at a reference distance, R_0. By using Eq. (2), it is implicitly assumed that the source is sufficiently close to the ground, and the wavelengths are sufficiently long that associated ground reflections can be ignored.

The sound pressure at the microphone given by Eq. (2) is not observed at the same instant that it is produced by the source; there is a time delay owing to the propagation limitation imposed by the speed of sound. The time delay, Δt, is given by

$$\Delta t = \frac{r}{c} = \frac{(x^2 + D^2)^{1/2}}{c}, \tag{3}$$

where c is the speed of sound. Equation (3) leads to the following relationship between "source time," t, and "microphone time," t_{mic}

$$t_{\text{mic}} = t + \Delta t = \frac{x}{V} + \frac{(x^2 + D^2)^{1/2}}{c}. \tag{4}$$

Onset rate is defined[2] as the rate of change in sound pressure level with time. However, the wayside microphone measures pressure in terms of microphone time and, in general, would have no knowledge of the source position or source time. Using the chain rule of differentiation, the onset rate at the microphone, dLp/dt_{mic} can be expressed as

$$\frac{dLp}{dt_{\text{mic}}} = \frac{dLp}{dx}\frac{dx}{dt_{\text{mic}}}. \tag{5}$$

The first derivative, dLp/dx, can be found by differentiating Eq. (2) resulting in the following expression:

$$\frac{dLp}{dx} = -\frac{20}{\ln(10)}\frac{x}{(x^2 + D^2)}. \tag{6}$$

The second derivative, dx/dt_{mic}, can be found by taking the differential of both sides of Eq. (4), resulting in the expression

$$\frac{dx}{dt_{\text{mic}}} = V\frac{(x^2 + D^2)^{1/2}}{\left[\frac{V}{c}x + (x^2 + D^2)^{1/2}\right]}. \tag{7}$$

Equation (7) is essentially the apparent speed of the source in terms of microphone time.

Analogous to the Doppler effect, the apparent speed of the approaching source ($x<0$) exceeds its actual speed, while the apparent speed of the receding source ($x>0$) is slower. Similarly, if the speed of sound was infinite, there would be no distinction between source time and microphone time and the apparent speed would simply be equal to V. Combining Eqs. (6) and (7) yields the following expression for the onset rate measured at the microphone

$$\frac{dLp}{dt_{\text{mic}}} = -\frac{20}{\ln(10)}\frac{1}{(x^2 + D^2)^{1/2}}\frac{Vx}{\left[\frac{V}{c}x + (x^2 + D^2)^{1/2}\right]}. \tag{8}$$

According to Eq. (8), as expected, the onset rate is positive for an approaching source and negative for a receding source. The onset rate is also proportional to the source speed, its position, and the microphone distance.

Unfortunately, Eq. (8) contains a dependence on the source/vehicle position which would generally not be known during a field measurement. In order to remove the dependence on x, consider the point where the onset rate is maximum, found by differentiating Eq. (8) with respect to x and setting the resulting quantity equal to zero. The point where the onset rate is greatest is given by the solution

$$\frac{V}{c}\left(\frac{x}{D}\right)^3 - \left[1-\left(\frac{x}{D}\right)^2\right]\left[1+\left(\frac{x}{D}\right)^2\right]^{1/2} = 0. \quad (9)$$

Equation (9) has two solutions, one for the approaching source ($x<0$) and one for the receding source ($x>0$). For very low speed sources, the limiting solution to Eq. (9) (as $V/c \to 0$) is

$$\frac{x}{D} = \pm 1, \quad (10)$$

which corresponds to a 45° look angle to the source. To find a solution to Eq. (9) that is not limited to low speeds, assume the following parametric form for an approaching source ($x/D<0$):

$$\frac{x}{D} = -1-\beta, \quad (11)$$

where $\beta \ll 1$. Substitution of Eq. (11) into Eq. (9), while discarding higher order powers of the parametric quantity, β, results in the approximate solution to Eq. (9) for approaching sources,

$$\left.\frac{x}{D}\right|_{\text{onset rate}}^{\max} = -1-\frac{V}{c}\frac{1}{\sqrt{8}-3\frac{V}{c}}. \quad (12)$$

The error in using Eq. (12) to estimate the source position, where the maximum onset rate occurs, increases with increased source speed. For example, at Mach 0.2 the position error would be less than 1% and at Mach 0.3 the error would be less than 2%.

To find the maximum onset rate, Eq. (12) or Eq. (10) can be substituted into Eq. (8). Using the simpler low-Mach number approximation ($x/D=-1$) yields the following expression for the maximum onset rate, $\text{OR}_{\max}$, as measured at the microphone:

$$\text{OR}_{\max} = \left.\frac{dLp}{dt_{\text{mic}}}\right|^{\max} = \frac{20}{\ln(10)}\frac{V}{D}\frac{1}{2-\sqrt{2}\frac{V}{c}}. \quad (13)$$

An alternative form of Eq. (13) is given by the following:

$$D \times \text{OR}_{\max} = \frac{20}{\ln(10)}\frac{V}{2-\sqrt{2}\frac{V}{c}}, \quad (14)$$

where now all of the quantities on the left hand side are either known or measured, and the source speed is the only unknown quantity on the right hand side of the equation.

C. Numerical simulation

A numerical simulation was used to examine the implications of the low-Mach number approximation used to generate Eq. (14). The numerical model was based on a monopole source with a reference sound pressure level of 94 dB (re 20 μPa) measured at a distance of 5 m. The source speed was 90 m/s (Mach 0.265 with $c=340$ m/s). The pressure time history at two microphone positions, $D=20$ m and $D=30$ m, was calculated numerically using Eq. (2) to calculate the sound pressure level history and Eq. (4) to calculate microphone-referenced time. Figure 2 shows the results of the simulation in terms of the microphone sound pressure level as a function of source position. The calculated sound pressure level is greater at the closer microphone, as expected. The onset rate was calculated by differentiating the numerical data using a first order finite difference. The maximum onset rate was then determined by inspection of the differentiated numerical data. For each microphone, the maximum onset rate occurs at the same source position ($x/D=-1.111$) which is slightly before the low-Mach number solution ($x/D=-1$). The corresponding angle difference from 45° is about 3°.

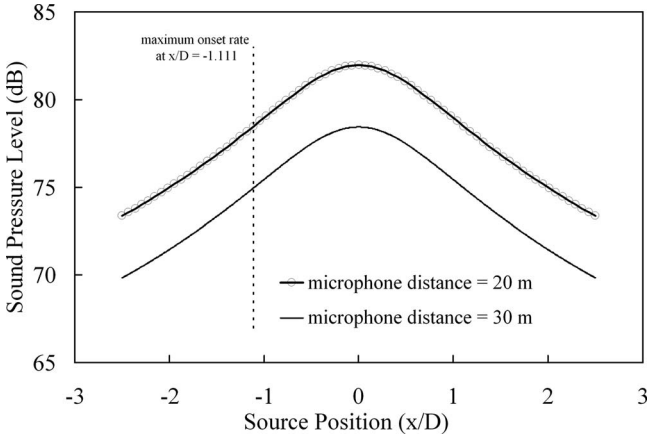


FIG. 2. Numerical simulation—sound pressure history for monopole source moving past two microphone positions. Source parameters: speed=90 m/s, source=94 dB (re 20 μPa) at 5 m, speed of sound=340 m/s.

The numerical simulation was also used to investigate the accuracy of the low-Mach number approximation in estimating source speed. Figure 3 plots the low-Mach number speed estimate, given by Eq. (12), along with the results from a series of numerical simulations where Eq. (2) was used to calculate the sound pressure level history and Eq. (4) was used to calculate microphone-referenced time. The maximum onset rate was then determined by inspection of the differentiated (first order finite difference) numerical

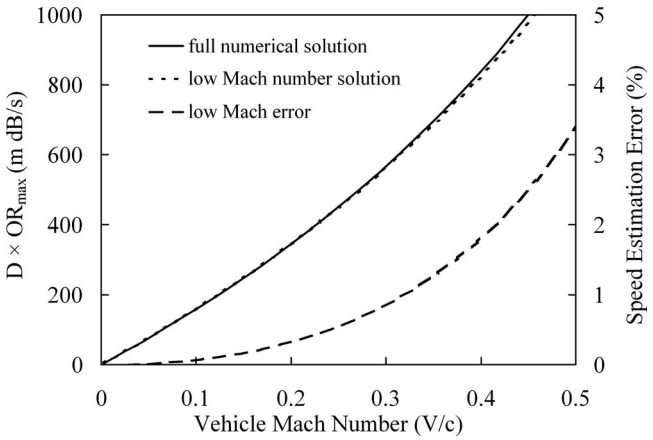


FIG. 3. Dependence of distance-adjusted maximum onset rate ($D \times \text{OR}_{\max}$) on vehicle Mach number using full numerical solution and low-Mach number approximation. Speed estimate error resulting from low-Mach number approximation.

simulation data. As Fig. 3 shows, the low Mach number approximation is quite accurate at low speed, and is still reasonable up to Mach numbers as high as M 0.5. For example, at Mach 0.5, the expected speed estimation error incurred by using the low-Mach number approximation is only about 3.4%.

As Fig. 3 shows, Eq. (14) provides a simple and reasonable means to estimate source speed. However, Eq. (14) can also be expressed in a polynomial form which will become useful later when exponential averaging algorithms are considered. If Eq. (14) is recast with source speed as the unknown, it becomes

$$V = \frac{\frac{\ln(10)}{10} D \times \mathrm{OR}_{\max}}{1 + \frac{\ln(10)}{10\sqrt{2}} \frac{D \times \mathrm{OR}_{\max}}{c}}. \tag{15}$$

Equation (15) can also be expanded as a geometric series of the form

$$V = \frac{\ln(10)}{10} D \times \mathrm{OR}_{\max} \left[1 - \left(\frac{\ln(10)}{10\sqrt{2}} \frac{D \times \mathrm{OR}_{\max}}{c} \right) + \left(\frac{\ln(10)}{10\sqrt{2}} \frac{D \times \mathrm{OR}_{\max}}{c} \right)^2 - + \cdots \right]. \tag{16}$$

Since the low-Mach number approximation for source speed effectively becomes a polynomial expansion in powers of the maximum onset rate, $\mathrm{OR}_{\max}$, and D, it seems reasonable to curve-fit a polynomial to the full numerical solution to avoid being limited by the low-Mach number approximation. This will effectively eliminate even the small errors associated with Eq. (14). A series of numerical simulations was performed for Mach numbers ranging from 0.01 to 0.5. The maximum onset rate was calculated for each speed and a cubic polynomial was curve fitted to the resulting data. Note that the dependence on microphone distance, D, can be removed by curve fitting to the parameter $D \times \mathrm{OR}_{\max}$. The resulting curve-fit polynomial for source speed is

$$V_{\mathrm{est}}|_{C=340} = 0.228(D \times \mathrm{OR}_{\max}) - 9.71 \times 10^{-5}(D \times \mathrm{OR}_{\max})^2 + 2.2 \times 10^{-8}(D \times \mathrm{OR}_{\max})^3, \tag{17}$$

where the units are m/s for V_{est}, m for D, and dB/s for $\mathrm{OR}_{\max}$. It should be noted that the coefficients in Eq. (17) are appropriate for a sea level standard atmosphere ($c=340$ m/s). The corresponding curve fit for a standard altitude of 1524 m (5000 ft), where the speed of sound is 334 m/s, is

$$V_{\mathrm{est}}|_{C=334} = 0.228(D \times \mathrm{OR}_{\max}) - 9.885 \times 10^{-5}(D \times \mathrm{OR}_{\max})^2 + 2.284 \times 10^{-8}(D \times \mathrm{OR}_{\max})^3, \tag{18}$$

The difference between Eqs. (17) and (18) is quite small. For example, if the $C=340$ m/s (sea level standard atmosphere) relation, Eq. (17), was used to estimate the speed of a source when the speed of sound was actually 334 m/s; the errors would only be 0.3% at Mach 0.25 and 0.6% at Mach 0.5.

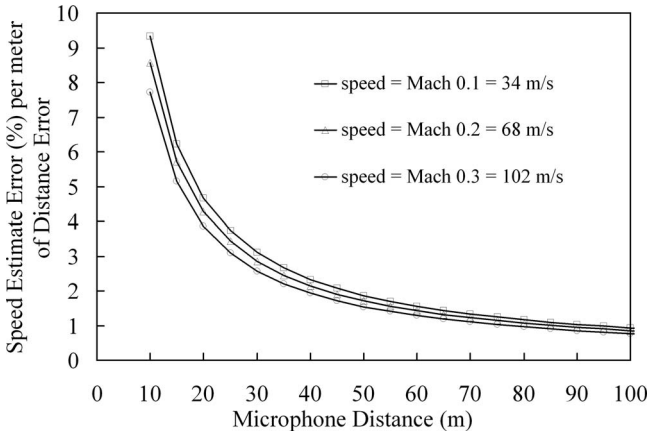


FIG. 4. Speed estimate error produced per meter of error in microphone distance.

D. Sensitivity to microphone distance

The microphone distance is a key parameter in determining the speed of the passing source. For actual vehicles with real physical dimensions, there could be a question as to where the microphone distance should be measured. For a train, for example, should the distance be measured to the center of the tracks, or the near rail, or the side of the vehicle? In order to investigate the sensitivity of the speed estimate to microphone distance, Eq. (17) can be differentiated to give

$$\left| \frac{dV_{\mathrm{est}}}{dD} \right| = 0.228(\mathrm{OR}_{\max}) - 1.942 \times 10^{-4}(D \times \mathrm{OR}_{\max}^2) + 6.6 \times 10^{-8}(D^2 \times \mathrm{OR}_{\max}^3). \tag{19}$$

Figure 4 shows how the speed estimate is affected by a distance error for speeds ranging from 34 to 102 m/s. The speed estimate error is shown as a percentage per meter of distance error. For example, for a vehicle traveling at 102 m/s, and a microphone distance of 50 m, one could expect a 1.5% error in the speed estimate for every meter of distance error. The speed error per meter decreases as the microphone distance increases. Also, the speed error decreases as the vehicle speed increases. This implies that more distant microphone positions are more tolerant of uncertainties in microphone distance. However, one would also expect that more distant microphone positions may be more affected by parts of the vehicle that trail the "leading edge." Recall that one of the assumptions of the method was that the leading edge of the vehicle could be modeled as a monopole source.

E. Exponential averaging

Commercial sound level meters typically use an exponentially weighted average in the calculation of the sound pressure level. The exponentially weighted sound pressure, p_a, at any time, t, is given by

$$p_a(t) = \left[\frac{1}{\tau} \int_{-\infty}^{t} p(t)^2 \exp\left(-\frac{t-\zeta}{\tau} \right) d\zeta \right]^{1/2}, \tag{20}$$

where τ is the exponential time constant. Common time constants are $\tau=1$ s, which is known as a slow detector, and $\tau=0.125$ s, which is known as a fast detector.

The exponential average tends to slow down the apparent pressure rise, resulting in a lower onset rate. An equation similar in form to Eq. (17) can be generated for the exponentially averaged onset rate. While similar in form to Eq. (17), the exponential average imposes a more complex relationship between microphone distance, D, and the polynomial coefficients. The general form of the exponentially averaged speed estimate is

$$V_{\text{est}} = a_1(D)(\text{OR}_{\text{max}}) + a_2(D)(\text{OR}_{\text{max}})^2 + a_3(D)(\text{OR}_{\text{max}})^3, \tag{21}$$

where the coefficients a_1, a_2, and a_3 are all functions of the microphone distance, D. The coefficient functions were calculated from curve fits to numerical simulation data for various source speeds and microphone distances. For each distance/speed combination, the un-averaged sound pressure level time history was calculated at the microphone position in terms of microphone time using Eqs. (2) and (4). Equation (20) was then used to operate on the un-averaged pressure data to generate the exponentially averaged sound pressure level time history. The maximum onset rate was then determined from the exponentially averaged time history. For each exponential averaging setting (fast, slow), the maximum onset rate was tabulated for each distance/speed combination. Numerical curve fits to the tabulated data were then used to determine the $a_i(D)$ coefficient functions that appear in Eq. (21). For example, for a fast detector ($\tau=0.125$ s), the numerically derived coefficient expressions are

$$a_1(D) = 0.044\,58 + 0.2242D + 1.386 \times 10^{-4}D^2 - 1.187 \times 10^{-6}D^3,$$

$$a_2(D) = 3.072 \times 10^{-4}D - 8.534 \times 10^{-5}D^2 - 6.438 \times 10^{-7}D^3 + 5.836 \times 10^{-9}D^4,$$

$$a_3(D) = -4.449 \times 10^{-7}D^2 + 3.374 \times 10^{-8}D^3 + 6.786 \times 10^{-11}D^4 - 1.107 \times 10^{-12}D^5. \tag{22}$$

For a slow detector ($\tau=1.0$ s), the coefficient expressions are

$$a_1(D) = 0.14 + 0.2451D - 5.914 \times 10^{-4}D^2 + 5.612 \times 10^{-6}D^3,$$

$$a_2(D) = -3.421 \times 10^{-4}D + 7.124 \times 10^{-7}D^2 - 7.164 \times 10^{-7}D^3 - 6.326 \times 10^{-9}D^4 + 9.185 \times 10^{-11}D^5,$$

$$a_3(D) = 2.847 \times 10^{-7}D^2 - 1.156 \times 10^{-7}D^3 + 2.725 \times 10^{-9}D^4 - 2.705 \times 10^{-11}D^5 + 1.172 \times 10^{-13}D^6. \tag{23}$$

The coefficient expressions in Eqs. (22) and (23) were obtained from curve fits to numerical simulation data with $C=340$ m/s (sea level standard atmosphere) and with source velocities ranging from 5 to 120 m/s and microphone distances ranging from 5 to 50 m. The expressions should not be used for conditions outside these ranges.

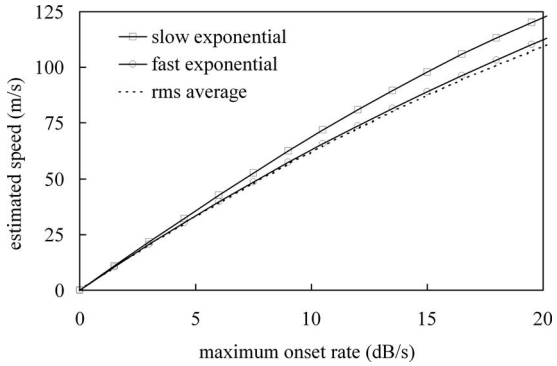


FIG. 5. Speed estimation curves based on maximum onset rate; for a microphone distance of 30 m at sea level; using rms averaging, fast exponential averaging ($\tau=0.125$ s), and slow exponential averaging ($\tau=1.0$ s).

Figure 5 compares the exponentially averaged speed estimation models for a microphone located at 30 m with a speed of sound equal to 340 m/s. As the figure shows, the onset rate reduction effect of the exponential average increases with increased speed. For example, at a source speed of 75 m/s, a microphone at 30 m without exponential averaging would measure a 12.5 dB/s maximum onset rate. If a sound level meter with a fast detector was used to measure the passby, and Eq. (15) (the expression for no exponential averaging) was mistakenly used to estimate the speed, the error would only be about 1.7% (1.3 m/s). However, if the meter was set to a slow detector, and Eq. (15) was used to estimate the speed, the error would be about 9% (7.5 m/s). In both cases, the estimated speed would have been too low.

III. FIELD TEST DATA

A. Test results

Passby sound pressure measurements were made in Sharon, MA next to the northbound track of the Northeast Corridor. Three train passbys were measured using a microphone that was located 28.65 m (94 ft) from the near rail (42.096 988° N, 71.193229 ° W, 79 m above sea level). The microphone was placed 1.5 m above the ground. The measurements were made using a B&K model 4189 microphone with a PCB model 426E01 preamplifier and a PCB model 480E09 ICP signal conditioner. The data were recorded in the field using a Rion model DA20 instrumentation recorder at a rate of 12 000 samples/s. Prior to testing, a 94 dB, 1 kHz calibration tone was applied using a B&K model 4231 acoustical calibrator.

The passby data were analyzed using MATLAB®, operating directly on the WAV files that were recorded with the DA20. After the sensor scale factor was applied, the data were filtered using a fourth order, 10–1000 Hz, Butterworth bandpass filter. The equivalent sound pressure level, *L*eq, was then calculated from the filtered data at 0.1 s intervals. In addition to the basic root-mean-square (rms) *L*eq, the data were also exponentially averaged before the *L*eq was calcu-

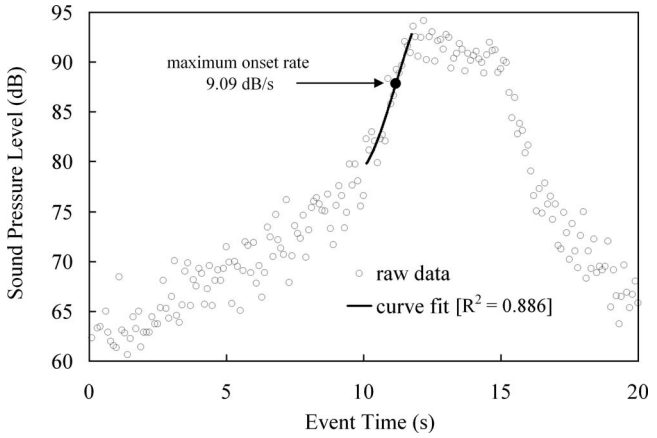


FIG. 6. Sound pressure level time history measured during passby of Amtrak Acela train at 55.6 m/s (124 m/h). Microphone located 28.65 m from near rail. rms averaged *L*eq levels (0.1 s) shown along with polynomial curve fit. Point where maximum onset rate occurs shown as a solid dot.

lated using a fast (τ=0.125 s) and a slow (τ=1.0 s) time constant.

The measurement site was generally flat although the approaching trains were partially obscured visually by a wooded area. The measurements were made on March 27, 2008. The air temperature was 11 °C and the winds were calm (<2.5 m/s). At the time of the measurements, the ground was soft and the surrounding deciduous trees were bare. The train speeds were measured independently using a radar gun and a timed gate (time for train of known length to pass a fixed point). The radar speed and timed gate speed agreed to within 2 m/s.

The passby sound pressure was measured for three trains; a Massachusetts Bay Transportation Authority (MBTA) commuter train, an Amtrak Regional train, and an Amtrak Acela express train. The MBTA commuter train was powered by a leading diesel-electric locomotive; the AMTRAK trains were both electric, drawing power from an overhead catenary. The observed train speeds varied from 33.5 m/s (75 m/h) to 55.6 m/s (124 m/h).

Reasonable estimates of onset rate were not possible because of the scatter in the raw 0.1 s *L*eq data. In order to obtain a smoothed time history that could be differentiated reliably, a cubic polynomial curve fit was used to approximate the raw *L*eq data. The onset rate was then calculated by differentiating the curve-fit polynomial. Because the area of interest was restricted to the portion of the passby where the pressure was rising, the curve fit was confined to the leading edge of the pressure time history, extending approximately 5 dB before and after the point of maximum onset rate. An *R*-square goodness of fit statistic was calculated for each curve according to the formula

$$R^2 = 1 - \frac{\sum (L\mathrm{eq}_{\mathrm{data}} - L\mathrm{eq}_{\mathrm{fit}})^2}{\sum (L\mathrm{eq}_{\mathrm{data}} - L\mathrm{eq}_{\mathrm{data\_mean}})^2}. \quad (24)$$

where $L\mathrm{eq}_{\mathrm{data}}$ refers to the raw *L*eq data at each time point in the curve-fit region, $L\mathrm{eq}_{\mathrm{fit}}$ are the *L*eq levels calculated from the curve-fit polynomial, and $L\mathrm{eq}_{\mathrm{data\_mean}}$ is the mean value of $L\mathrm{eq}_{\mathrm{data}}$ in the curve-fit region. The summations are done over all of the data points within the curve-fit region.

Figure 6 shows the rms-averaged sound pressure time history that was measured during an Amtrak Acela passby at 55.6 m/s. The curve-fit polynomial and associated R^2 value are also shown on the figure. The maximum onset rate, calculated from the curve fit, is 9.09 dB/s. According to Eq. (17), for a microphone distance of 28.65 m, this corresponds to an estimated speed of 53.2 m/s which is only 4% lower than the measured train speed.

Speed estimates were made for all three observed train passages and processed using (1) a rms average (no exponential), (2) a fast (0.125 s) exponential average, and (3) a slow (1.0 s) exponential average. The speed estimates are summarized in Table I. For the slowest train (MBTA commuter rail), the best estimates of the speed were produced using the fast and slow exponential averages. The best estimates for the faster Amtrak trains were generally obtained using the fast exponential average which produced speed estimates within 3% of the measured speed. In general the best curve fits (highest R^2) were obtained with the exponential averages because of the data smoothing that the exponentials provide.

B. Discussion of test results

The primary goal of the field test program was to show that it was possible to obtain reasonable estimates of vehicle speed based solely on the microphone position and the onset rate. This goal was achieved, with some speed estimates within a few percent of the measured speed. For the present paper, no extensive effort was made to investigate why some estimates were better than others. The authors plan to explore such issues in more detail as part of an ongoing research effort to investigate ways to improve the estimates using alternative filtering, curve fitting, or averaging schemes. The effect of microphone distance also has to be investigated (do closer microphones produce better estimates?).

TABLE I. Measured and estimated speeds for train passbys measured using a microphone located 28.65 m from near rail.

Train type	Measured speed (m/s)	Estimated speed (% error) (R^2) rms average (m/s)	Fast exponential (m/s)	Slow exponential (m/s)
MBTA commuter rail	33.5	32.7 (−2%) [0.957]	33.6 (0%) [0.986]	33.5 (0%) [0.997]
Amtrak regional	52.8	49.9 (−5%) [0.918]	53.5 (+1%) [0.975]	48.8 (−8%) (0.994)
Amtrak acela	55.6	53.2 (−4%) [0.886]	57.2 (+3%) [0.968]	57.0 (+2%) [0.998]

IV. CONCLUSIONS

For the case of a subsonic monopole sound source moving past a stationary microphone, a simple theoretical relationship exists between the microphone distance, the vehicle speed, speed of sound, and the maximum onset rate. Knowing the microphone position, this relationship allows one to estimate the speed of a passing vehicle based solely on the measured onset rate. While slightly more complicated versions of the relationship can also be derived to account for exponential averaging, the only quantities required to estimate speed are the maximum onset rate, the speed of sound, and the microphone distance.

The speed estimation method was evaluated using passby data from three train passages (one diesel, two electric) ranging in speed from 33.5 to 55.6 m/s. All speed estimates were within 8% of the actual measured speed with the best estimates within a few percent. The test results suggest that it is possible to estimate the speed of a passing vehicle using a wayside sound pressure measurement provided the sound field in front of the vehicle is similar to the sound field produced by a monopole source.

[1]United States Department of Transportation, Federal Transit Administration, "Transit noise and vibration impact assessment," FTA-VA-90-1003-06 (2006).
[2]United States Department of Transportation, Federal Railroad Administration, "High-speed ground transportation noise and vibration impact assessment" (2005).
[3]United States Department of Transportation, Federal Highway Administration, "FHWA traffic noise model users guide (version 2.5 addendum)," Final Report (2004).
[4]C. E. Hanson, "Proposed new noise impact criteria for passenger rail systems in the United States," J. Sound Vib. **193**, 29–34 (1996).
[5]X. Zhang, "On the directivity of railway noise sources," Ph.D. thesis, Chalmers University of Technology, Goteborg, Sweden (2007).

Determination of elastic constants of generally anisotropic inclined lamellar structure using line-focus acoustic microscopy

Jin-Yeon Kim[a] and Stanislav I. Rokhlin[b]
Department of Materials Science and Engineering, Edison Joining Technology Center, The Ohio State University, 1248 Arthur E. Adams Drive, Columbus, Ohio 43221

(Received 2 December 2008; revised 4 September 2009; accepted 17 September 2009)

A methodology for measuring elastic constants of different phases in materials with lamellar microstructure by line-focus acoustic microscopy is developed. The material microstructure investigated is modeled by generally anisotropic multilayers arbitrarily inclined to the sample surface on which acoustic microscopy measurements are performed. To calculate surface acoustic wave (SAW) propagation in such structures quasi-static effective elastic constants are determined and compared with calculated frequency-dependent constants. As a model material, practically important, Ti–6Al–2Sn–4Zr–2Mo alloy is selected. Time-resolved line-focus acoustic microscopy experiments are performed on a Ti-6242 α/β single colony (Ti–6Al–2Sn–4Zr–2Mo alloy) and on a Ti–6Al α-phase single crystal for which elastic constants of different phases are determined using inversion of measured SAW velocities. To validate the experimental methodology, SAW velocities in an *X*-cut quartz crystal are measured as a function of sample orientation angle and compared with predictions based on the known elastic moduli of quartz.
 [DOI: 10.1121/1.3245032]

PACS number(s): 43.35.Sx [ADP] Pages: 2998–3007

I. INTRODUCTION

Acoustic microscopy (AM) has been established as an effective tool in measuring elastic properties of anisotropic and multilayered materials.[1–5] Its advantage manifests in high resolution local measurements on actual material microstructures. Kim *et al.*[2] calculated the effective elastic constants of superlattice films from elastic constants of the constituting single crystal layers measured by high frequency AM. AM for measuring elastic constants of thin films on an anisotropic substrate was performed by Achenbach *et al.*,[3] and Wang *et al.*[4] and Wang and Rokhlin[5] measured the elastic constants of multilayered composites using low frequency time-resolved AM. In most previous work, AM property measurements were performed on multilayers constructed by periodically stacking anisotropic layers in the principal material axes. Moreover, the measurements were performed in the principal planes of the constituting layers with those planes parallel to the supporting substrate. A very limited number of studies have been performed on layered structures oriented perpendicular to surface wave propagation. In this regard Chimenti[6] analyzed dispersion of surface acoustic waves (SAWs) propagating perpendicular to the layering direction in a vertically laminated composite semispace. In all these cases, the structure of the anisotropic layer is well defined and thus the analysis of SAWs on multilayered anisotropic media[7] can be directly applied.

Modern metal alloys have complicated, often multiphase, microstructures. Elastic properties of different phases in these materials are usually poorly known, since most available data are for single crystals. However, knowledge of elastic properties can be essential for understanding of mechanical behavior of these types of materials. This paper's objective is to develop an experimental methodology and a property reconstruction model for determination of elastic properties of different phases in multiphase materials. The model material for this study is Ti-6242 α/β alloy with complex microstructure formed by quasi-random anisotropic layers of α- and β-phases inclined arbitrarily to the sample surface on which measurements are performed.

Due to their high specific strength, high service temperature, and good fracture toughness, commercial titanium alloys are employed widely in aerospace, energy, and biomedical applications.[8,9] The strength of these alloys is manipulated and maximized by controlling the microstructure during thermomechanical processing. The elastic constants of titanium alloy vary considerably depending on alloying and heat-treating history and on composition. Not much data on the elastic constants of titanium alloy crystals are available. Fisher and Renken[10] determined the elastic constants of an α-phase Ti single crystal using bulk ultrasonic waves propagating in principal directions of the crystal. Fisher and Dever[11] measured elastic constants of β-phase Ti–Cr alloy for varying amounts of chromium. The knowledge of elastic constants of different phases and crystalline effective properties of realistic Ti alloys is essential for modeling their mechanical properties and ultrasonic nondestructive evaluation.

[a]Present address: GWW School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332.
[b]Author to whom correspondence should be addressed. Electronic mail: rokhlin.2@osu.edu

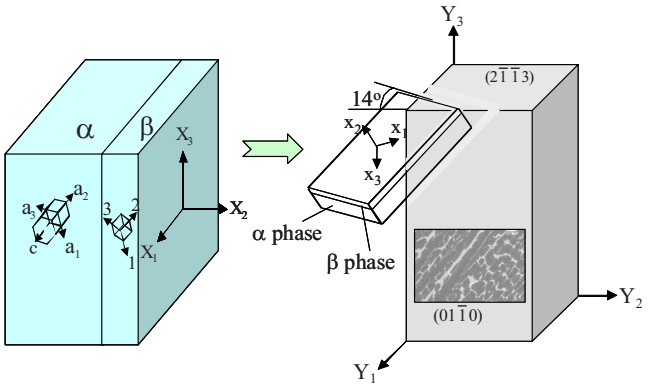


FIG. 1. (Color online) Schematics of the microstructure of Ti-6242 α/β single colony showing relative orientations of α (cubic)- and β (hexagonal)-phase crystal layers (left) and sample used in the AM [it depicts how layers are inclined to the sample surface and SEM micrograph embedded showing two phases α (dark) and β (white)].

In this paper, the elastic constants of α and β Ti phases are measured on a Ti-6242 α/β single colony and a Ti-6 α-phase crystal by means of time-resolved line-focus AM. It is performed on a non-principal plane of the Ti-6242 α/β single colony sample, measuring SAW velocity as a function of orientation angle. The ultrasonic wave propagation model was developed, which has combined morphological characteristics of microstructure and elastic properties of α- and β-phases to obtain the effective elastic constants of the anisotropic multiphase and relate them to velocity of the SAW. As a result by using the model inversion the elastic properties of both the effective and each of α- and β-phases were obtained from AM measurements.

II. PROBLEM STATEMENT

A. Microstructure of Ti-6242 α/β single colony

The microstructure of the Ti–6Al–2Sn–4Zr–2Mo α/β single colony is characterized as a composite crystalline structure of α- and β-phase single crystals. The α-phase crystal has a hexagonal close packed (hcp) structure with the shape factor $c/a=1.587$,[12] so that the direction $[2\bar{1}\bar{1}3]_\alpha$ is tilted up 57.8° from the basal plane. The β-phase having a body centered cubic structure is present in the form of platelets with a maximum thickness near 10 μm. The β-phase platelets are distributed randomly in the α-phase while their broad faces are aligned consistently in the $[56\bar{5}]_\beta$ direction forming the *lamellar structure* as shown in Fig. 1. In the scanning electron microscope (SEM) micrograph in the figure the bright area represents the β-phase and the dark area the α-phase; the concentration of the β-phase is 9%–10%. As is seen in Fig. 1, the layers do not have well-defined interfaces and the layer thickness changes from one to another. In contrast to the structural randomness, the orientation angles of the crystals are spatially invariant. The layers are oriented in an off-principal direction of both phases. The Burgers orientation relationships between the two phases in the colony microstructure are $[2\bar{1}\bar{1}0]_\alpha \| [1\bar{1}\bar{1}]_\beta$ and $(0001)_\alpha \| (101)_\beta$. It is assumed that the α- and β-phases satisfy the above ideal Burgers orientation relationships. Therefore, no common plane exists in which principal axes of both the α- and β-phases lie together. From the Burgers orientation relationships, the transformation of the crystallographic coordinate system of one of the phases (α-phase or β-phase) to another can be found. For example, to transform the α-phase elastic constants to a β-coordinate system, two successive rotations of the α-coordinate system, $\theta_z=-24.74°$ and $\theta_y=-45.0°$, are required (the angle of rotation is about the coordinate axis indicated as the subscript). We will discuss the coordinate rotations in more detail in Sec. II C.

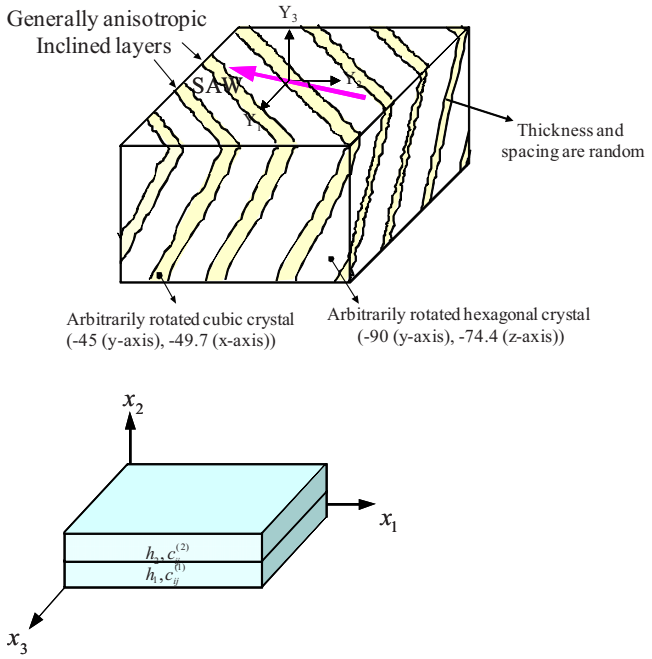


FIG. 2. (Color online) A schematic of the model Ti-6242 α/β single colony material microstructure; it shows SAW propagation direction, which varies during measurement by rotating of the line-focus transducer. The RVE of a periodic anisotropic multilayer is shown in the bottom.

B. Model anisotropic microstructure for SAW analysis of Ti-6242 α/β single colony

As described in Sec. II A and illustrated in Fig. 1, the Ti-6242 α/β single colony microstructure is formed by thin crystalline layers with different orientations of the main crystallographic systems. Therefore the sample surface, on which the SAW propagates, cannot be a plane of symmetry for at least one of the α- or β-phases, and for this phase the surface plane is a plane of general elastic anisotropy. Generally, both the α- and β-phases are anisotropic in the sample surface plane. We select the main coordinate system on the sample surface as shown in Fig. 2 with axis Y_3 perpendicular and (Y_1Y_2) plane in the sample surface. As shown in Fig. 2 a multilayered crystalline structure is formed by layers inclined to the sample surface, on which surface waves propagate and AM measurements are performed. The direction of wave propagation relative to the Y_2 sample axis is varied during measurements.

The Ti-6242 α/β single colony has a quasi-random lamellar microstructure exhibiting the following characteristics: (1) thin β-phase plates (white) are distributed randomly in the α-phase matrix, (2) the interfaces are semi-coherent,

and (3) the thickness of the β-phase plates varies from one to the other. These microstructural characteristics give rise to several difficulties in analyzing the SAW propagation in this material. Due to the randomness the wave propagation analysis for layered media with deterministic microstructures (aligned flat layers with constant thicknesses) cannot be applied. Furthermore, for analysis of SAW propagation on a vertically inclined anisotropic layer, one needs to employ numerical methods that will complicate property inversion. To account for the effects of the microstructure and to model the SAW propagation in the medium shown in Fig. 2 we will describe it as a quasi-static homogeneous medium with effective elastic constants (Sec. IV). In our measurements the SAW wavelength is much larger than the microstructural layer thickness and thus the velocity measured by the AM is also averaged over the SAW propagation distance (averaging microstructural elastic characteristics of α- and β-phases). In Sec. III, the ultrasonic measurement method is first described, next (Sec. IV) the microstructure homogenization and simulation of SAW propagation are discussed, and finally, in Sec. V, the inversion problem and determination of the α- and β-phase elastic constants are presented.

C. Coordinate transformations

To model SAW propagation and eventually use it for elastic constraints' determination, we need to transform the tensor of elastic constants of each material phase to the sample surface coordinate system (Y_1, Y_2, Y_3), as shown in Fig. 2. This is accomplished in two steps. First, the elastic constants of both α- and β-phases, which are described in their own crystallographic coordinate systems $(x,y,z)_\alpha$ and $(x,y,z)_\beta$, are transformed to the common X_1, X_2, X_3 coordinate system attached to the β-phase plate (as shown in Fig. 1). The X_2 axis of this system is perpendicular to the broad face of the plate $([56\bar{5}]_\beta)$. These transformations are performed in such a way that the z axis for each phase becomes aligned with the X_2 axis. The transformation is done by two successive rotations: $\theta_{y_\alpha} = -45.0°$ and $\theta_{x_\alpha} = -49.68°$ for α-phase, and $\theta_{z_\beta} = -74.4°$ and $\theta_{y_\beta} = -90°$ for β-phase. These angles are obtained from the Burgers relationships. Finally, the second transformation of elastic constant tensors is performed from the X_1, X_2, X_3 coordinate system to the physical (sample) coordinate system Y_1, Y_2, Y_3 (Fig. 1). This is done by two rotations, $\theta_{X_1} = 14.0°$ and $\theta_{X_2} = -46.6°$. The angle of rotation is about the coordinate axis indicated as the subscript.

This two step transformation is necessary since the effective elastic constants obtained are obtained in the X_1, X_2, X_3 coordinate (Sec. IV A). However, the prediction of SAW velocities (Sec. IV C) has to be performed in the sample surface coordinate system (Y_1, Y_2, Y_3).

III. EXPERIMENT: MEASUREMENT OF ANGLE DEPENDENCE OF SAW VELOCITY BY TIME-RESOLVED LINE-FOCUS AM

A. Experimental setup

The experimental setup consists of an ultrasonic line-focus transducer, mechanical alignment devices, a four-axis motion control system, a pulser/receiver, an oscilloscope, and a computer for controlling the system and waveform acquisition. The sample is placed in the immersion tank and the transducer is rotated by a precision motor. At each rotation angle, the transducer, which is aligned parallel with the sample surface, acquires the ultrasonic signals at different lift-off distances from the sample surface. Figure 3 shows a schematic of the experimental system.

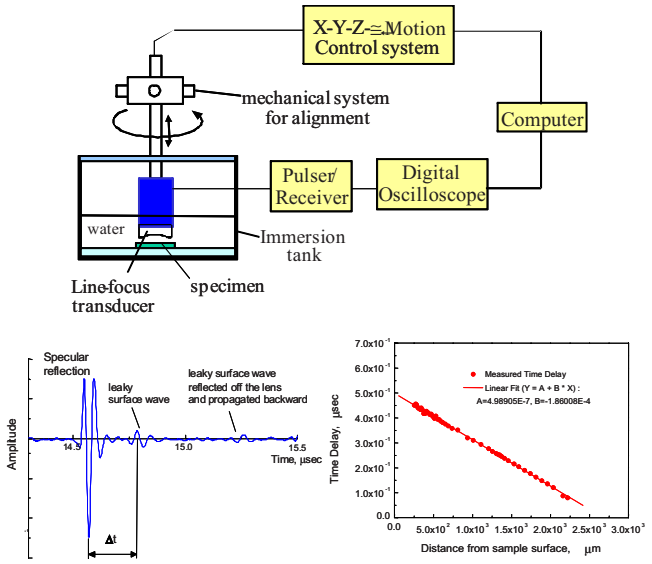


FIG. 3. (Color online) A schematic of experimental system with a typical ultrasonic signal and an example of measured time delay versus lift-off distance shown in the bottom.

The 75 MHz broad band ultrasonic transducer with a line-focus cylindrical lens was used in this study. The cylindrical lens has a radius of 6.35 mm, a focal length of 5.08 mm, and an aperture half angle of 38°. The transducer emits a short acoustic pulse with duration about 0.1 μs. The center frequency of the received SAW signal shifts down to 35 MHz due to the leakage of the high frequencies to water. The transducer is first focused on the sample and the lift-off distance is decreased until the SAW is generated. The specular reflection and SAW signals are resolved in the time domain and the time delay between these is measured as shown in the bottom of the figure. This time delay increases linearly as the transducer approaches the sample surface. At each propagation angle, the time delay is measured as a function of the transducer lift-off distance as shown in the figure. The proportionality constant m between the time delay and the lift-off distance is found by linear regression. Finally, the SAW velocity is determined[13] by $V_R = V_w[(V_w/m) - (V_w/2m)^2]^{-1/2}$, where V_w is the speed of sound in water.

B. Samples

Three different materials were used for measurements: the Ti-6242 α/β single colony sample, the α-phase Ti single crystal, and an X-cut quartz crystal.

(1) The Ti-6242 α/β single colony sample used for the AM measurements was a rectangular parallelepiped with dimensions $7.13 \times 3.17 \times 2.90$ mm^3. The density of the material is 4.42 g/cm^3. The wide face and the direction

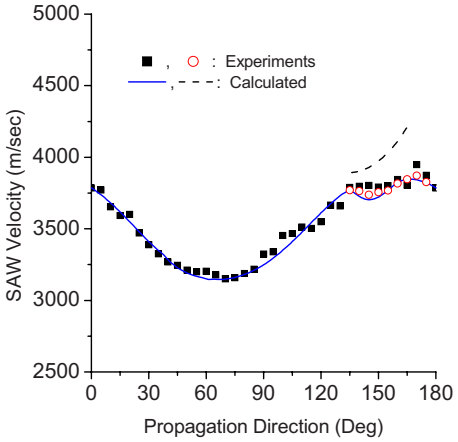


FIG. 4. (Color online) Measured (solid squares and open circles) and calculated SAW velocities in *X*-cut quartz sample. The dashed curve in 135°–165° is the theoretically calculated pseudo-surface waves.

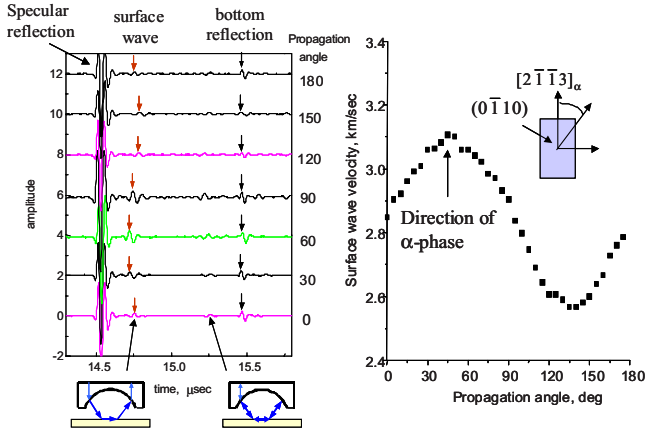


FIG. 5. (Color online) Time-domain signals and SAW velocity measured as a function of propagation angle.

of the long axis of the sample are $(01\bar{1}0)_\alpha$ and $[2\bar{1}\bar{1}3]_\alpha$, respectively. As shown schematically in Fig. 2, a multilayered crystalline microstructure is formed by layers inclined to the sample surface (Fig. 1), on which the AM measurements of SAW velocities are performed.

(2) The α-phase Ti single crystal sample had two flat surfaces: (0001) and $(01\bar{1}0)$ where the SAW velocities were measured. The sample was in a half cylindrical shape. The basal plane (0001) corresponds to the semi-circular top surface with radius about 5 mm while the surface $(01\bar{1}0)$ is a 12×8 mm^2 rectangle.

(3) An *X*-cut quartz crystal is a 1 mm thick disk with a radius of 3.5 cm; it was used to check our measurements since quartz has well-known properties.

C. SAW velocity measurements

To validate our AM measurements SAW velocities on the *X*-cut quartz sample were determined.[14] Quartz has trigonal symmetry with well-known elastic moduli:[15] $C_{11}=86.8$, $C_{12}=7.04$, $C_{13}=11.91$, $C_{14}=-18.04$, $C_{33}=105.75$, $C_{44}=58.2$, and $C_{66}=39.88$ (GPa). Figure 4 shows the measured and calculated (using above constants) SAW velocities in the *X*-cut plane at different orientation angles; they are in good agreement. At angles from 135° to 165° pseudo-surface waves are observed; this is shown as a broken curve with higher wave speeds.[16] The existence of the pseudo-SAWs in *X*-cut quartz in this limited range of angle has been theoretically[16] and experimentally[17] reported. Since the velocity of the pseudo-SAWs is quite close to that of the pure SAWs especially at angles around 135° (ratio 1.03), their time-domain signals are not fully separated at short propagation distances. Therefore, measurements of the pure SAW velocities in the range of angle where the pseudo-SAWs appear required a larger defocus to make sure that they are sufficiently resolved from each other in the time domain. Open circles and squares at angles above 135° are measurements of the pure SAW velocities at different transducer lens defocusing.

Several time-domain signals at different rotation angles in the α/β single colony sample and the SAW velocities determined from measured time delays are shown in Fig. 5. SAW signals appear after the specular reflection signals with a time delay changing with the propagation direction due to anisotropy. The reflection of the longitudinal wave from the sample bottom is identified. Leaky surface waves, which are reflected twice between the transducer face and the sample surface (as depicted in Fig. 5 by the transducer lens picture), appear between the SAW and bottom reflection signals. It should be noted that the SAW velocity has its maximum and minimum at propagation angles at about 45° and 135° from $[2\bar{1}\bar{1}3]_\alpha$, which is close to the normal direction to the β-phase (57.8° from $[2\bar{1}\bar{1}3]_\alpha$).

Measurements for the α-phase single crystal are shown in Fig. 6. Since it has the hexagonal structure, the SAW velocity in (0001) plane is constant. It is interesting to note that the SAW velocity in the $(01\bar{1}0)$ plane exhibits a maximum in the direction [0001] and coincides with the SAW velocity in the basal plane.

IV. SAW PROPAGATION AND QUASI-STATIC HOMOGENIZATION OF GENERALLY ANISOTROPIC LAMELLAR STRUCTURE

A. Homogenization: Quasi-static effective elastic constants of generally anisotropic multilayers

To simulate the SAW propagation and to account for the effects of the microstructure we replace the actual model microstructure of the medium shown in Fig. 2 with a homog-

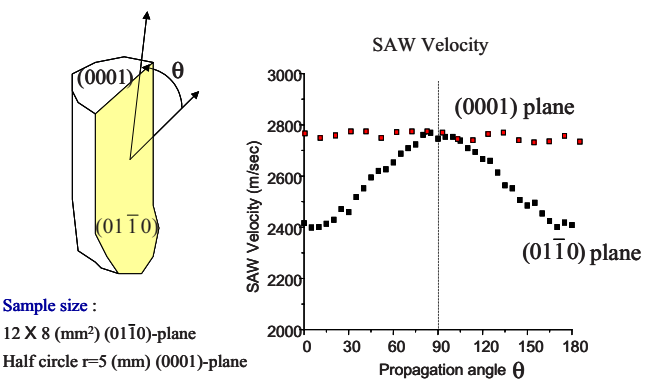


FIG. 6. (Color online) SAW propagation in Ti-α single crystal.

enized (averaged) medium with effective elastic constants. Let us consider a multilayer periodically stacked with a repetitive unit. It is assumed that each layer is generally anisotropic and its principal axes are arbitrarily oriented. A representative volume element (RVE) consisting of layers taken from the infinitely multilayered periodic medium is shown in Fig. 2. It is well known that the periodic medium model provides good approximate solutions for the effective quasi-static elastic constants of a quasi-random medium.[18] A perfectly welded contact at boundaries between layers is assumed. Layers have thicknesses $h_1, h_2, \ldots, h_n$ in the x_2 direction. The constitutive equation for the ith layer can be written in the contracted notation as

$$\begin{Bmatrix} \sigma_1^{(i)} \\ \sigma_3^{(i)} \\ \sigma_{13}^{(i)} \\ \sigma_2^{(i)} \\ \sigma_{23}^{(i)} \\ \sigma_{12}^{(i)} \end{Bmatrix} = \begin{bmatrix} C_{11}^{(i)} & C_{13}^{(i)} & C_{15}^{(i)} & C_{12}^{(i)} & C_{14}^{(i)} & C_{16}^{(i)} \\ C_{13}^{(i)} & C_{33}^{(i)} & C_{35}^{(i)} & C_{32}^{(i)} & C_{34}^{(i)} & C_{36}^{(i)} \\ C_{15}^{(i)} & C_{35}^{(i)} & C_{55}^{(i)} & C_{52}^{(i)} & C_{54}^{(i)} & C_{65}^{(i)} \\ C_{12}^{(i)} & C_{32}^{(i)} & C_{52}^{(i)} & C_{22}^{(i)} & C_{24}^{(i)} & C_{26}^{(i)} \\ C_{14}^{(i)} & C_{34}^{(i)} & C_{54}^{(i)} & C_{24}^{(i)} & C_{44}^{(i)} & C_{46}^{(i)} \\ C_{16}^{(i)} & C_{36}^{(i)} & C_{56}^{(i)} & C_{26}^{(i)} & C_{46}^{(i)} & C_{66}^{(i)} \end{bmatrix} \begin{Bmatrix} \varepsilon_1^{(i)} \\ \varepsilon_3^{(i)} \\ \gamma_{13}^{(i)} \\ \varepsilon_2^{(i)} \\ \gamma_{23}^{(i)} \\ \gamma_{12}^{(i)} \end{Bmatrix}, \tag{1}$$

where σ_{ij} and ε_{ij} are components of stress and strain, and $\gamma_{ij}=2\varepsilon_{ij}(i\neq j)$ is the engineering shear strain (indices 11, 22, and 33 are written as 1, 2, and 3). Note that the constitutive equation is represented such that in-plane and out-of-plane stress and strain components are partitioned. Moreover, the elastic constant matrix $C_{ij}^{(i)}$ is expressed in full matrix form to consider arbitrary anisotropic layers. In matrix form, Eq. (1) is

$$\begin{Bmatrix} \sigma_{\parallel}^{(i)} \\ \sigma_{\perp}^{(i)} \end{Bmatrix} = \begin{bmatrix} C_{\parallel}^{(i)} & C_{\times}^{(i)T} \\ C_{\times}^{(i)} & C_{\perp}^{(i)} \end{bmatrix} \begin{Bmatrix} \varepsilon_{\parallel}^{(i)} \\ \varepsilon_{\perp}^{(i)} \end{Bmatrix}, \tag{2}$$

where the matrices $C_{\parallel}^{(i)}$, $C_{\times}^{(i)}$, and $C_{\perp}^{(i)}$ are defined as those relating the in-plane stress $(\sigma_{\parallel}^{(i)})$ to the in-plane strain $(\varepsilon_{\parallel}^{(i)})$, the out-of-plane stress $(\sigma_{\perp}^{(i)})$ to the in-plane strain, and the out-of-plane stress to the out-of-plane strain $(\varepsilon_{\perp}^{(i)})$, respectively.

With this simple geometry and homogeneous remote loads the exact effective elastic constants of a periodic generally anisotropic cell can be derived. In this work the procedure is the generalization to general anisotropy of the method described in Ref. 19. Due to the importance of this problem for homogenization of composite materials somewhat equivalent averaging procedures have been employed in the past for different arrangements of applied loads. Chou *et al.*[20] analyzed the effective elastic constants of a multilayered composite with monoclinic layers. Pagano[21] derived the effective moduli of laminates in the plate bending theory. Norris[22] proposed the most general homogenization formulation for a three dimensional inclusion and then derived it to the multilayer case in the framework of the coordinate invariance. Nemat-Nasser and Hori[18] applied the general homogenization formulation to the one dimensional case, and in-plane motions are not considered. To simplify application of the inversion model described later we present below and in the Appendix a simple and straightforward method to obtain the effective moduli of our model microstructure, which was derived directly using the corresponding moduli of the each phase.

Considering continuity of displacement at the layer interface and uniform in-plane loading, compatibility of in-plane strains leads to equality between the effective strain and the strain in every layer of the system,

$$\bar{\varepsilon}_{\parallel} = \varepsilon_{\parallel}^{(1)} = \varepsilon_{\parallel}^{(2)} = \cdots = \varepsilon_{\parallel}^{(n)}, \tag{3}$$

where the overbar is used to denote effective quantities. From the stress averaging in the loading direction: $\bar{\sigma} = \int_V \sigma(\mathbf{x}) dV(\mathbf{x})$ over the volume V of the RVE, the effective stress is

$$\bar{\sigma}_{\parallel} = \sigma_{\parallel}^{(1)} v_1 + \sigma_{\parallel}^{(2)} v_2 + \cdots + \sigma_{\parallel}^{(n)} v_n. \tag{4}$$

Here, $v_i = h_i/h$ is the volume fraction of the ith layer where $h=\Sigma_i h_i$ and $\sigma_{\parallel}^{(i)}$ is the averaged stress in the layer i. Note that the stress in different layers is different to satisfy strain compatibility [Eq. (3)], and its value is obtained from Hooke's law for each individual layer.

The effective out-of-plane strain in the representative volume under uniform out-of-plane loading is given as the volume average of the strain in layers,

$$\bar{\varepsilon}_{\perp} = \varepsilon_{\perp}^{(1)} v_1 + \varepsilon_{\perp}^{(2)} v_2 + \cdots + \varepsilon_{\perp}^{(n)} v_n. \tag{5}$$

The stress compatibility is

$$\bar{\sigma}_{\perp} = \sigma_{\perp}^{(1)} = \sigma_{\perp}^{(2)} = \cdots = \sigma_{\perp}^{(n)}. \tag{6}$$

As shown in Appendix, the expressions for the effective elastic constants are derived in terms of the elastic constants of the individual layers using the compatibility conditions (3)–(6). For a bi-layer medium, they are expressed as

$$\bar{C}_{\parallel} = [v_1 C_{\parallel}^{(1)} + v_2 C_{\parallel}^{(2)}] - v_1 v_2 [C_{\times}^{(1)} - C_{\times}^{(2)}][v_1 C_{\parallel}^{(2)} + v_2 C_{\parallel}^{(1)}]^{-1} [C_{\times}^{(1)} - C_{\times}^{(2)}]^T, \tag{7}$$

$$\bar{C}_{\perp} = C_{\perp}^{(1)} [v_1 C_{\parallel}^{(2)} + v_2 C_{\parallel}^{(1)}]^{-1} C_{\perp}^{(2)}, \tag{8}$$

$$\bar{C}_{\times} = v_1 C_{\times}^{(1)} [v_1 C_{\perp}^{(2)} + v_2 C_{\perp}^{(1)}]^{-1} C_{\perp}^{(2)} + v_2 C_{\times}^{(2)} [v_1 C_{\perp}^{(2)} + v_2 C_{\perp}^{(1)}]^{-1} C_{\perp}^{(1)}. \tag{9}$$

Now the effective elastic constants of the Ti-6242 α/β single colony can be determined. Elastic constants reconstructed from SAW measurements in Table I are used for calculations (for the α-crystal with 1–3 plane as the basal plane of the hexagonal symmetry and the β-crystal). Using the orientation relationships between the α- and β-crystal layers and the rotation angles for the coordinate transformation presented in this section, the stiffness matrices for the α- and β-crystals are transformed to a reference coordinate system that is attached to the wide face of the β-plate with the x_2 axis aligned with the surface normal. Then, the effective stiffness matrix in the reference coordinate system is obtained using Eqs. (7)–(9) as follows:

TABLE I. List of reconstructed elastic constants of α- and β-phase Ti.

C_{11}	C_{22}	C_{12}	C_{13}	C_{44}	Material
			α-phase (hexagonal) elastic constants (GPa)		
141	163	76.9	57.9	48.7	Reconstruction from α/β single colony Ti-6%Al-2%Sn-4%Zr-2%Mo (present work)
136	163	78.0	68.5	40.6	Reconstruction from α single crystal Ti-6%Al (present work)
162	180	92.0	69.0	46.7	Pure α single crystal[a]
			β-phase (cubic) elastic constants (GPa)		
135		113		54.9	Reconstruction from α/β single colony Ti-6%Al-2%Sn-4%Zr-2%Mo (present work)
140		118		44.2	β-Ti-15%Cr alloy brine quenched[b]

[a]Reference 10.
[b]Reference 11.

$$\bar{C}_{ij} = \begin{bmatrix} 143.3 & 59.1 & 78.6 & 3.33 & 0 & 0 \\ & 164.7 & 60.0 & -0.016 & 0 & 0 \\ & & 142.1 & -3.72 & 0 & 0 \\ & \text{SYM} & & 46.8 & 0 & 0 \\ & & & & 32.4 & 1.54 \\ & & & & & 44.9 \end{bmatrix} \text{(GPa)}. \tag{10}$$

Note that the effective elastic constants of the α/β single colony show a monoclinic anisotropy in the reference coordinate system (X_1, X_2, X_3) in Fig. 1. Since SAW velocity measurements will be done on a non-principal plane of the effective medium, the elastic constants are transformed from the reference coordinate system to the coordinate system (Y_1, Y_2, Y_3) attached to the measurement plane, Fig. 2. Therefore, a calculation of SAW velocity in a generally anisotropic (triclinic) medium is required.

B. Effect of frequency

The above homogenization is static and is exact at zero frequency. Recently, Wang and Rokhlin[23] and Rokhlin and Wang[24] developed a homogenization scheme for predicting frequency-dependent effective elastic constants of periodic multilayered anisotropic media based on the use of Floquet wave analysis. This theory gives a set of frequency-dependent elastic constants of the effective medium in the frequency range below the first cutoff frequency of the Floquet wave. Here, the static and dynamic effective elastic constants are compared to ensure the assumption of weak dispersion of the SAW in the α/β single colony.

The calculations of effective elastic constants are performed for a multilayer model consisting of α- and β-phase Ti crystal layers as shown in Fig. 7. The thicknesses of the α- and β-layers are 45 and 5 μm, respectively. The α-layer is rotated by 90° with respect to an axis in the basal plane so that the effective medium will exhibit orthotropic anisotropy. In Fig. 7, the static effective constants are presented as dots at zero frequency. Perfect agreement of the static effective constants with the dynamic ones at $\omega=0$ is obtained. Small dispersion is observed at frequencies above 20 MHz for this perfectly periodic structure. This dispersion effect will be much reduced in the Ti-6242 α/β single colony due to the layer thickness randomness and imperfectness of periodicity (Fig. 2), and it is believed that the static approximation is valid in the higher frequency range. For this reason the static effective elastic constants are used in the calculation of SAW velocities.

C. Simulation of SAW velocity

The SAW velocity in the Ti-6242 α/β single colony is calculated after its homogenization. The computation algorithm employs a recently developed, stable, and efficient method to compute wave propagations in generally anisotropic multilayered piezoelectric media.[7,24,25] The governing equation for the wave propagation is written in the state equation form

$$\frac{d\boldsymbol{\xi}(z)}{dz} = i\mathbf{A}\boldsymbol{\xi}, \tag{11}$$

where $\boldsymbol{\xi}$ is the state vector for the anisotropic medium (including piezoelectric). $\boldsymbol{\xi}$ is composed of the generalized displacement $\mathbf{u}=[u_1, u_2, u_3]$ and stress $\boldsymbol{\sigma}=[\sigma_{13}, \sigma_{23}, \sigma_{33}]$ com-

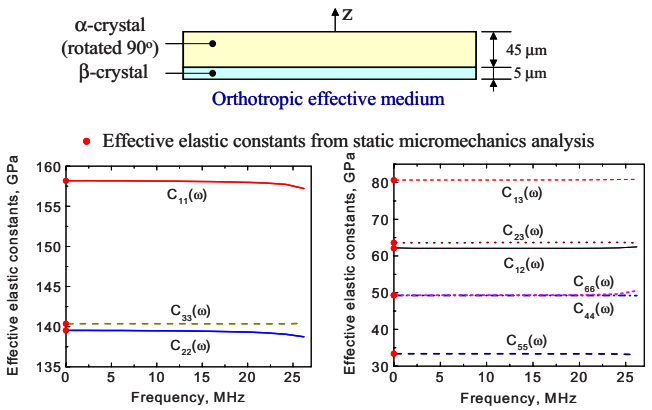


FIG. 7. (Color online) Comparison of static and frequency-dependent effective elastic constants for orthotropic layered medium composed of α- and β-phase Ti crystals.

ponents as defined in Ref. 25, and **A** is the fundamental acoustic tensor depending on the elastic constants, the density, and the frequency[7,25] [z (or 3) direction is the normal to the surface in which SAWs propagate].

The solution of the differential equation [Eq. (11)] is expressed using the eigenvalues k_z and eigenvector matrix **W** of **A**. There are six eigenvalues and eigenvectors representing six plane harmonic waves (for piezoelectric medium, quartz, the number of waves is eight[25]). Three propagate (or exponentially decay) in the $+z$ direction, and we denote the eigenvalues as $\boldsymbol{\beta}_z^+ = \mathbf{I}\ \mathrm{diag}(k_z^{+1}, k_z^{+2}, k_z^{+3})$ [$\mathbf{I}(3\times 3)$ is the unit matrix] and the corresponding eigenvector matrix as

$$\mathbf{W}^+ = \begin{bmatrix} \mathbf{P}^+ \\ \mathbf{D}^+ \end{bmatrix},$$

where $\mathbf{P}^+$ corresponds to the general displacement portion and $\mathbf{D}^+$ to the general stress portion in the eigenvector matrix $\mathbf{W}^+$ and the other three are waves propagating (or exponentially decaying) in the $-z$ direction whose eigenvalues we denote as $\boldsymbol{\beta}_z^- = \mathbf{I}\ \mathrm{diag}(k_z^{-1}, k_z^{-2}, k_z^{-3})$ and the eigenvector matrix as

$$\mathbf{W}^- = \begin{bmatrix} \mathbf{P}^- \\ \mathbf{D}^- \end{bmatrix}.$$

The canonical (diagonalized) form of the matrix **A** is represented as

$$\mathbf{A} = \mathbf{W}\boldsymbol{\beta}_z\mathbf{W}^{-1} = \begin{bmatrix} \mathbf{P}^- & \mathbf{P}^+ \\ \mathbf{D}^- & \mathbf{D}^+ \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_z^- & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_z^+ \end{bmatrix} \begin{bmatrix} \mathbf{P}^- & \mathbf{P}^+ \\ \mathbf{D}^- & \mathbf{D}^+ \end{bmatrix}^{-1},$$

where

$$\boldsymbol{\beta}_z = \begin{bmatrix} \boldsymbol{\beta}_z^- & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_z^+ \end{bmatrix}.$$

The elastic wave solution can be obtained in the form of a compliance matrix,[25] which relates displacements in z and $(z+h)$ planes to the stress

$$\begin{bmatrix} \mathbf{U}(z+h) \\ \mathbf{U}(z) \end{bmatrix} = \mathbf{S} \begin{bmatrix} \mathbf{T}(z+h) \\ \mathbf{T}(z) \end{bmatrix},$$

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+ \\ \mathbf{P}^-\mathbf{H}^- & \mathbf{P}^+ \end{bmatrix} \begin{bmatrix} \mathbf{D}^- & \mathbf{D}^+\mathbf{H}^+ \\ \mathbf{D}^-\mathbf{H}^- & \mathbf{D}^+ \end{bmatrix}^{-1}, \tag{12}$$

where matrix $\mathbf{H}^{\pm}(3\times 3) = \mathbf{I}\ \mathrm{diag}[e^{\pm ik_z^{\pm 1}h}, e^{\pm ik_z^{\pm 2}h}, e^{\pm ik_z^{\pm 3}h}]$.

The characteristic equation for the leaky surface wave on anisotropic semispace is given by

$$S_{ss}^{33} + \Lambda = 0, \tag{13}$$

where S_{ss}^{33} is the (3,3) element in the 3×3 compliance matrix for a semispace[24,25] and the second term $\Lambda = -\cos\theta/(i\omega\rho_f V_f)$ represents the fluid loading effect, ρ_f is the density and V_f is the velocity of the fluid, and θ is the SAW leakage angle. The leaky surface wave velocity is obtained by finding the roots of Eq. (13). The elements of the compliance matrix $\boldsymbol{S}$ can be obtained by direct canonical decomposition as discussed above of the fundamental acoustic tensor **A** or using the second-order[26] or high-order asymptotic expansions[7] by subdividing the substrate into very thin layers and applying the recursive algorithm for the layered system[7,24,25] to obtain the compliance matrix. The recursive asymptotic method significantly simplifies the analysis while preserving the same level of accuracy as in the exact matrix methods.[24,25] To avoid the large number of subdivisions by considering a semispace, the substrate is subdivided into two layers; the top layer is the material in which the SAW velocity is analyzed and the bottom layer is an artificial layer that has perfectly matched acoustic properties[7] (called perfect matching layer) and absorbs any incident wave energy. The SAW velocity in the Ti-6242 α/β single colony is calculated by solving the characteristic equation obtained from the total compliance matrix for the layers.

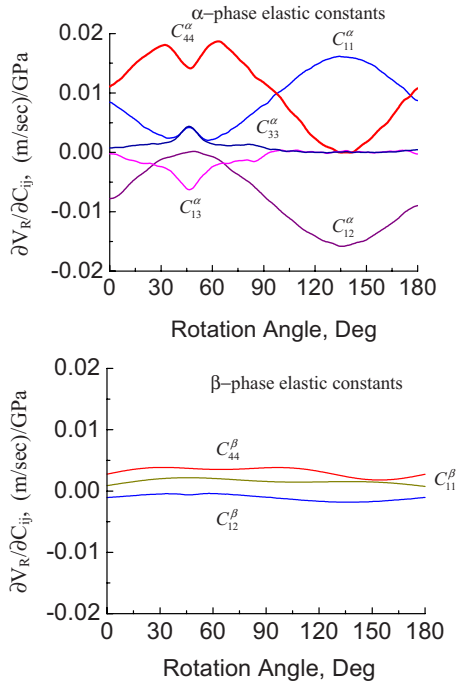


FIG. 8. (Color online) Sensitivity of surface wave velocity to elastic constants as a function of propagation angle in Ti α/β single colony sample: α-phase elastic constants (top) and β-phase elastic constants (bottom).

V. RECONSTRUCTION OF ELASTIC CONSTANTS

A. Sensitivity analysis

Figure 8 shows simulated sensitivities of the SAW velocity V to the elastic constants $(\partial V/\partial C_{ij})$ as a function of rotation angle in the Ti-6242 α/β single colony in the plane of measurement. The sensitivity depends strongly on propagation direction and in this specific plane the sensitivity for the α-phase elastic constants C_{11}^{α}, C_{44}^{α}, and C_{12}^{α} is relatively high and thus these constants can be determined more reliably from the SAW velocity measurements. Since the volume fraction of the β-phase is only 9%–10%, β-phase elastic constants have relatively smaller sensitivity to reconstruction. This is a limitation determined by the material microstructure and it cannot be overcome by choosing a different measurement plane. However, in the numerical reconstruction algorithm described below, these three elastic constants are treated with higher weight factors. Figure 9 shows the

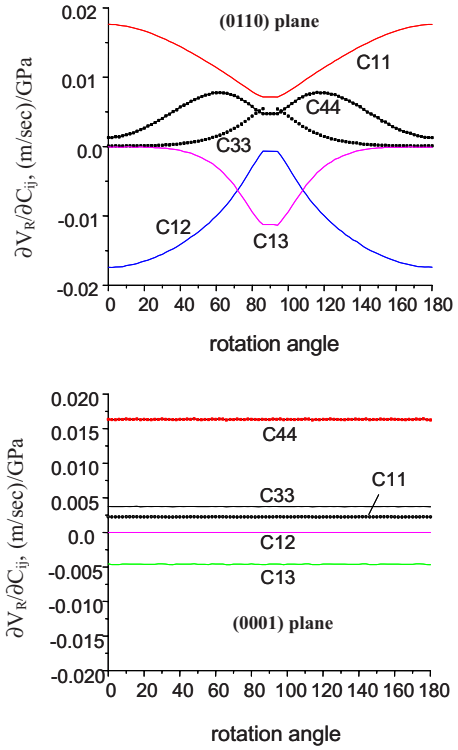


FIG. 9. (Color online) Sensitivity of surface wave velocity to elastic constants as a function of propagation angle in Ti α single crystal sample in (0110) plane (top) and in (0001) plane (bottom).

sensitivity of SAW velocity to the elastic constants of the Ti-6α single crystal sample as a function of rotation angle. Strong angle dependence similar to that in the single colony sample is observed in the $(01\bar{1}0)$ plane whereas no angle dependence is observed in the (0001) plane that is the plane of isotropy of this hexagonal crystal. Moreover, the elastic constant C_{12} has no sensitivity in this plane. For this reason, the elastic constants are reconstructed using the SAW velocities in the $(01\bar{1}0)$ plane.

B. Reconstruction algorithm and results

The algorithm for determining the elastic constants of the single colony sample is shown in Fig. 10. The reconstruction is based on the SAW velocities measured from the experiment and predicted by the theoretical model. The inverse solution for a set of elastic constants is implemented as a

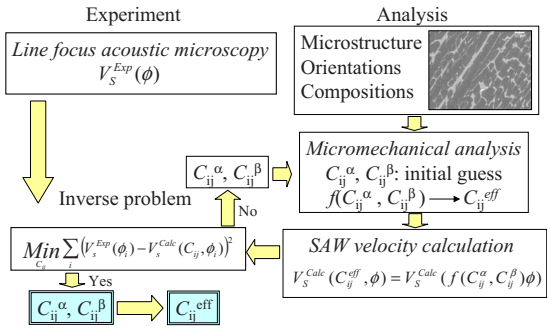


FIG. 10. (Color online) Block diagram illustrating the procedure to determine elastic constants from the measured and predicted SAW velocities.

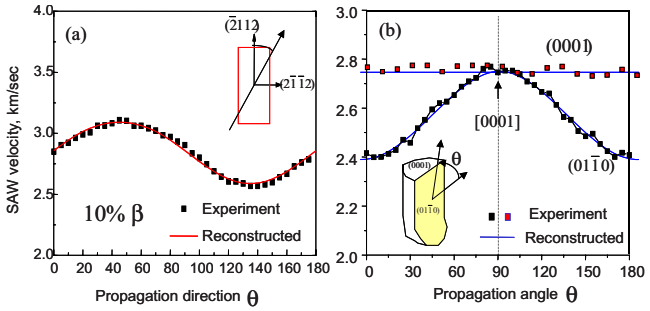


FIG. 11. (Color online) Comparisons of measured SAW velocities with reconstructed ones for a α/β single colony [(a), left] and α single crystal [(b), right].

nonlinear least squares optimization scheme[27] that minimizes the square error between the measured and calculated SAW velocities

$$\min_{C_{ij}} \sum_i (V_R^{\exp}(\varphi_i) - V_R^{\text{calc}}(C_{ij}, \varphi_i))^2. \tag{14}$$

SAW velocities ($V_R^{\exp}$) measured at 36 different propagation angles, shown in Fig. 5, are used in the reconstruction of elastic constants for the Ti-6242 α/β single colony sample. For the Ti-6α sample, 36 SAW velocity measurements at different angles for the $(01\bar{1}0)$ plane and the additional 36 for the (0001) planes, shown in Fig. 6, are used simultaneously in the reconstruction. However, since the SAW velocity in the (0001) plane is constant, the additional data in this plane work as a single constraint. The measured and reconstructed SAW velocities are compared in Fig. 11. It is seen that 36 data points are sufficient to describe the angle dependence of the SAW velocity. Since the total number of elastic constants of the constituent phases (which is 8:5 for α-phase in hcp and 3 for β-phase in cubic) is less than that of the effective elastic constants (13 for monoclinic material), it is advantageous to perform the error minimization in our least squares inversion process on the set of the constituent elastic constants rather than the effective elastic constants. However, the SAW velocity is calculated with the effective elastic constants, which are in turn calculated from the constituent elastic constants that are being searched for. Once the constituent elastic constants are found, the effective elastic constants are calculated using Eqs. (7)–(9). Reconstruction of the α-phase elastic constants from the α single crystal, which is more straightforward, is also performed in the same nonlinear optimization algorithm.

The present inversion problem defined by Eq. (14) has been shown to be sensitive to the initial guesses as is the case for many numerical inversion problems. This is because no inversion algorithm guarantees unconditional convergence. To ensure convergence to the correct result, literature values for the α- and β-Ti elastic constants shown in Table I are used as the initial guesses; that gives reasonable results for reconstruction. Note that when the initial guesses are arbitrarily chosen, the algorithm sometimes does not converge. To evaluate the reconstruction accuracy, the inversion problem is solved for different initial guesses around values available in literature. The reconstructed results fall in a

maximum deviation range of ±3%. Therefore, the accuracy of the reconstruction should be within this range.

In Fig. 11, the measured SAW velocities are compared with those calculated using the reconstructed α- and β-phase elastic constants. The quality of the fit indicates the quality of minimization and minimization error achieved. The reconstructed elastic constants are presented in Table I and compared with the literature data. The effective elastic constants are shown in matrix (10). It is seen in Table I that the α-phase elastic constants reconstructed from Ti-6242 and Ti-6 samples are quite close to each other; this indicates the consistency in the measurement and the reconstruction procedure. We assert that deviation of the elastic constants obtained in our study from those in literature is that these materials have different compositions and different processing histories. It is known that such differences may significantly influence the elastic properties of metal alloys.

VI. CONCLUSION

The α- and β-phase elastic constants of Ti-6242 α/β single colony and Ti-6α-phase crystals are determined by means of time-resolved line-focus AM. SAW propagation in a Ti-6242 α/β single colony is analyzed using the derived exact quasi-static effective elastic constants with the microstructural characteristics of the material taken into account. AM experiments are performed to measure SAW velocities in these materials. Elastic constants are inversely determined from the measured and predicted SAW velocities applying the nonlinear least squares minimization method. To calculate SAW velocity during iterations in the inversion algorithm the effective elastic constants are first determined using guessed values for the elastic constants of each phase. Since the total number of unknown elastic constants of two phases is less than that of the effective medium it is advantageous to perform search in the multidimensional space and thus the reconstruction of the constituent elastic constants for each phase rather than that for the effective elastic constants.

ACKNOWLEDGMENTS

This work was performed as a part of the project "Evaluation and Microstructure-based Modeling of Cold Dwell Fatigue in Ti-6242" through the Airworthiness Assurance Center of Excellence sponsored by the Federal Aviation Administration (FAA) under Contract No. 97-C-001. The authors are thankful to Professor M. J. Mills for providing samples and information on the single colony microstructure.

APPENDIX: EFFECTIVE QUASI-STATIC ELASTIC CONSTANTS

The effective elastic constants are defined from Hooke's law between average stresses $\bar{\sigma}$ and strains $\bar{\varepsilon}$ under remotely applied homogeneous load. The expressions for the effective elastic constants are derived in terms of the elastic constants of the individual layers using the compatibility conditions (3)–(6). From Eqs. (2) and (6), the following expression can be obtained:

$$[C_{\times}^{(i)} - C_{\times}^{(j)}]\bar{\varepsilon}_{\parallel} = -C_{\perp}^{(i)}\varepsilon_{\perp}^{(i)} + C_{\perp}^{(j)}\varepsilon_{\perp}^{(j)}. \tag{A1}$$

By solving the system of linear equations comprising Eqs. (4) and (A1), the out-of-plane strain in the ith layer is obtained as follows:

$$\varepsilon_{\perp}^{(i)} = D_{\perp}^{(i)}\bar{\varepsilon}_{\perp} - D_{\perp}^{(i)}\sum_{j\neq i} v_j C_{\perp}^{(j)^{-1}}[C_{\times}^{(i)} - C_{\times}^{(j)}]\bar{\varepsilon}_{\parallel}, \tag{A2}$$

where

$$D_{\perp}^{(i)} = \left[v_i I + \sum_{j\neq i} v_j C_{\perp}^{(j)^{-1}} C_{\perp}^{(i)}\right]^{-1}, \tag{A3}$$

and I denotes the identity matrix. From Eqs. (2) and (3), the effective out-of-plane stress is expressed as

$$\bar{\sigma}_{\perp} = C_{\times}^{(i)}\bar{\varepsilon}_{\parallel} + C_{\perp}^{(i)}\varepsilon_{\perp}^{(i)}. \tag{A4}$$

Substituting Eq. (A2) into Eq. (A4), one obtains a sub-matrix of the effective elastic constant

$$\bar{C}_{\perp} = C_{\perp}^{(i)} D_{\perp}^{(i)}. \tag{A5}$$

Further simplification can be made using standard matrix manipulations

$$\bar{C}_{\perp} = \left(\sum_i v_i C_{\perp}^{(i)^{-1}}\right)^{-1}. \tag{A6}$$

Similarly from Eqs. (2), (3), and (5), the effective in-plane stress is expressed as

$$\bar{\sigma}_{\parallel} = \sum_i v_i C_{\parallel}^{(i)}\bar{\varepsilon}_{\parallel} + \sum_i v_i C_{\times}^{(i)^T}\varepsilon_{\perp}^{(i)}. \tag{A7}$$

Now substituting Eq. (A2) into Eq. (A7) one can get the remaining sub-matrices of effective elastic constants

$$\bar{C}_{\parallel} = \sum_i v_i \left\{ C_{\parallel}^{(i)} - C_{\times}^{(i)^T} D_{\perp}^{(i)} \sum_{j\neq i} v_j C_{\perp}^{(i)^{-1}}[C_{\times}^{(i)} - C_{\times}^{(j)}] \right\} \tag{A8}$$

and

$$\bar{C}_{\times} = \sum_i v_i D_{\perp}^{(i)^T} C_{\times}^{(i)}. \tag{A9}$$

Equations (A6), (A8), and (A9) represent effective elastic constants of a medium with generally anisotropic multilayers.

[1] A. Briggs, *Acoustic Microscopy* (Oxford University Press, Oxford, 1992).
[2] J. O. Kim, J. D. Achenbach, M. Shinn, and S. A. Barnett, "Effective elastic constants and acoustic properties of single-crystal TiN/NbN superlattices," J. Mater. Res. **7**, 2248–2256 (1992).
[3] J. D. Achenbach, J. O. Kim, and Y. C. Lee, "Measuring thin-film elastic constants by line-focus acoustic microscopy," in *Advances in Acoustic Microscopy*, edited by A. Briggs (Plenum, New York, 1995), Vol. **1**, pp. 153–208.
[4] L. Wang, S. I. Rokhlin, and N. N. Hsu, "Time resolved focus acoustic microscopy of Composites," in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1999), Vol. **18B**, pp. 1321–1328.
[5] L. Wang and S. I. Rokhlin, "Time resolved line focus acoustic microscopy of layered anisotropic media: Application to composites," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1231–1244 (2002).
[6] D. E. Chimenti, "Rayleigh wave dispersion on vertically laminated composite surfaces," J. Acoust. Soc. Am. **92**, 492–498 (1992).
[7] L. Wang and S. I. Rokhlin, "Modeling of wave propagation in layered piezoelectric media by a recursive asymptotic method," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1060–1071 (2004).

[8]S. Ankem and H. Margolin, "Alpha-beta interface sliding in Ti-Mn alloys," Metall. Trans. A **14A**, 500–503 (1983).
[9]S. Suri, G. B. Viswanathan, T. Neeraj, D. H. Hou, and M. J. Mills, "Room temperature deformation and mechanisms of slip transmission in oriented single-colony crystals of an alpha/beta titanium alloy," Acta Mater. **47**, 1019–1034 (1999).
[10]E. S. Fisher and C. J. Renken, "Single-crystal elastic moduli and the hcp→bcc transformation in Ti, Zr, and Hf," Phys. Rev. **135**, A482–A494 (1964).
[11]E. S. Fisher and D. Dever, in *The Science, Technology and Application of Titanium*, edited by R. I. Jaffee and N. E. Promisel (Pergamon, New York, 1970), pp. 373–381.
[12]M. F. Savage, "Microstructural and mechanistic study of low temperature creep and dwell fatigue in single colony α/β titanium alloys," Ph.D. thesis, Ohio State University, Columbus, Ohio (2000).
[13]D. Xiang, N. Hsu, and G. V. Blessing, "Material characterization by a time-resolved and polarization-sensitive ultrasonic technique," in *Review of Progress in QNDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1996), Vol. **15B**, pp. 1431–1438.
[14]R. T. Webster, "X-cut quartz for improved surface acoustic wave temperature stability," J. Appl. Phys. **56**, 1540–1542 (1984).
[15]H. J. McSkimin, P. Andreatch, Jr., and R. N. Thurston, "Elastic moduli of quartz versus hydrostatic pressure at 25° and −195.8°," J. Appl. Phys. **36**, 1624–1632 (1965).
[16]G. W. Farnell, "Properties of elastic surface waves," in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1970), Vol. **VI**, pp. 109–166.
[17]F. R. Rollin, "Ultrasonic examination of liquid-solid boundaries using a right-angle reflector technique," J. Acoust. Soc. Am. **44**, 431–434 (1968).
[18]S. Nemat-Nasser and M. Hori, *Micromechanics: Overall Properties of Heterogeneous Solids* (Elsevier Science, New York, 1999).
[19]A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (Elsevier, Amsterdam, Netherlands, 1995).
[20]P. C. Chou, J. Carleone, and C. M. Hsu, "Elastic constants of layered media," J. Compos. Mater. **6**, 80–93 (1972).
[21]N. J. Pagano, "Exact moduli of anisotropic laminates," in *Composite Materials*, edited by G. P. Sendeckyj (Academic, New York, 1974), Vol. **2**, pp. 23–45.
[22]A. N. Norris, "The effective moduli of layered media—A new look at an old problem," in *Micromechanics and Inhomogeneity: The Toshio Mura Anniversary Volume*, edited G. J. Weng, M. Taya, and H. Abe (Springer-Verlag, Berlin, 1990), pp. 321–339.
[23]L. Wang and S. I. Rokhlin, "Floquet wave homogenization in periodic anisotropic media," J. Acoust. Soc. Am. **112**, 38–45 (2002).
[24]S. I. Rokhlin and L. Wang, "An efficient stable recursive algorithm for elastic wave propagation in layered anisotropic media," J. Acoust. Soc. Am. **112**, 822–834 (2002).
[25]L. Wang and S. I. Rokhlin, "A compliance/stiffness matrix formulation of general Green's function and effective permittivity for piezoelectric multilayers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 453–463 (2004).
[26]S. I. Rokhlin and W. Huang, "Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids: II. Second order asymptotic boundary condition," J. Acoust. Soc. Am. **94**, 3405–3420 (1993).
[27]Y. C. Chu and S. I. Rokhlin, "Comparative analysis of through transmission ultrasonic bulk wave methods for phase velocity measurements in anisotropic materials," J. Acoust. Soc. Am. **95**, 3204–3212 (1994).

Sound propagation in and low frequency noise absorption by helium-filled porous material

Y. S. Choy[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Lixi Huang and Chunqi Wang
Department of Mechanical Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong

(Received 20 January 2009; revised 17 September 2009; accepted 20 September 2009)

Low-frequency noise is difficult to deal with by traditional porous material due to its inherent high acoustic impedance. This study seeks to extend the effective range of sound absorption to lower frequencies by filling a low density gas, such as helium, in the porous material. Compared with conventional air-filled absorption material, the helium-filled porous material has a much reduced characteristic impedance; hence, a good impedance matching with pure air becomes more feasible at low frequencies. The acoustic properties of a series of helium-filled porous materials are investigated with a specially designed test rig. The characteristic of the sound propagation in a helium-filled porous material is established and validated experimentally. Based on the measured acoustic properties, the sound absorption performance of a helium-filled absorber (HA) of finite thickness is studied numerically as well as experimentally. For a random incidence field, the HA is found to perform much better than the air-filled absorber at low frequencies. The main advantage of HA lies in the middle range of oblique incidence angles where wave refraction in the absorber enhances sound absorption. The advantage of HA as duct lining is demonstrated both numerically and experimentally. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3257182]

PACS number(s): 43.55.Ev, 43.20.Jr [KVH] Pages: 3008–3019

I. INTRODUCTION

The porous sound absorption material has been the backbone of the noise control industry. But there are still technical challenges for the porous materials to deal with the low frequency noise effectively. The main technical challenge is associated with its high magnitude of the characteristic impedance at low frequencies. As the sound wave propagates through the porous material, the porous fibers exert a drag force on the oscillating air particles. At the same time, the irregular changes of air flow direction by the fibers cause a virtual mass. The drag force and the virtual mass can be grouped together as an inertia term to form the complex density of the porous material (Zwikker and Kosten, 1949). The imaginary part of the complex density derives from the conversion of the friction force, which is proportional to the air particle velocity, to an inertia term associated with the acceleration. This imaginary part is inversely proportional to the frequency and thus becomes very large at low frequencies. Therefore, the characteristic impedance of the commonly used porous material is very large at low frequencies. Because of the high impedance mismatch between the porous material and the air, it is hard for the incident sound wave to enter the porous material and most of the incident wave will be reflected, hence the poor absorption performance at low frequencies. The large characteristic impedance might be reduced by increasing the porosity of the sound absorption material so that more sound waves can penetrate the absorber easily, but it requires a very thick absorption layer to dissipate sufficient amount of the low-frequency noise. Roughly speaking, a quarter wavelength is a desirable thickness for such absorber. For instance, the desirable thickness is about 8.5×10^{-1} m at the frequency of 100 Hz, which implies that the absorber required would be very bulky for the low-frequency noise control.

Aiming for a new sound absorbing device which works efficiently in the low frequency range, it is proposed to fill the pores of the porous sound absorption material with helium gas in this study, hence a helium-filled absorber (HA). The potential advantages of HA over the conventional air-filled absorber (AA) can be explained in two aspects. First, the low density of the helium gas helps to reduce the large characteristic impedance of the absorber, which is the major reason that causes the impedance mismatch at low frequency. Other light gases, such as hydrogen, can also be adopted for this purpose, but helium is chemically more stable than hydrogen. So, the inert gas of helium is preferred. The second potential benefit of HA arises from the higher speed of sound in helium than in air, the ratio being about 3. According to the rule of refraction, when sound wave is incident on HA at an oblique direction, which is by far the more usual situation encountered in practice than the perfectly normal incidence, the penetrated wave is bent toward the air-absorber interface. In contrast, the penetrated wave is bent toward the normal direction of the air-absorber interface for AA. Hence, the penetrated wave travels a longer path in HA before it is reflected back to the incident medium, which in turn implies more energy dissipation within a thin absorption layer.

[a)]Author to whom correspondence should be addressed. Electronic mail: mmyschoy@polyu.edu.hk

The investigation of the absorption performance of the helium-filled porous material requires the knowledge of its bulk acoustics properties such as the effective density and speed of sound, which does not seem to be available in existing literature. For air-filled porous material, empirical formulas have been well established based on the extensive measurements (Delany and Bazley, 1970; Mechel, 1976; Kirby and Cumming, 1999). These empirical formulas and the so-called semiempirical models (e.g., Allard and Champoux, 1992) are simple in form and convenient for use, but they are only intended for the air-filled porous material. On the other hand, efforts have also been made to characterize these properties analytically with various physical parameters of the porous media as well as the filling gas, for instance, the theoretical models developed by Johnson *et al.* (1987), Attenborough (1982, 1983, 1987), and Allard *et al.* (1986, 1989). The analytical models mentioned above apply to porous materials with any kind of filling gases, but the implementation of these models requires the knowledge of a number of parameters, some of which are difficult to measure in a common acoustic laboratory. In the current study, a dimensional analysis is carried out to determine the relationship between the bulk acoustic properties of the porous materials and various relevant physical parameters. Instead of seeking to establish a specific functionality between the acoustic properties and the relevant parameters, the dimensional analysis identifies dimensionless groups which can help predict the bulk acoustic properties of helium-filled porous materials based on the available data of air-filled porous medium. With proper assumptions made, the same functional relationship is deduced for absorber filled with air and that filled with helium as far as the porosity and geometrical shape of the fibers are the same. A series of experiments is carried out to determine the effective density and speed of sound of the helium-filled materials with different fiber diameters so that the acoustic properties of the helium-filled porous materials can be characterized. The experimental studies also validate the functional relations established in the dimensional analysis. Empirical regression curves which describe the dependency of the effective density and speed of sound on controlling parameters such as viscosity, size of the fiber, and density of the filling gas are deduced. With the acoustic properties obtained, the absorption behavior of a helium-filled finite-layer absorber is studied and compared with the commonly used AA for both random incidence and the grazing incidence in duct lining.

II. DIMENSIONAL ANALYSIS

Dimensional analysis is adopted to determine the relationships qualitatively between the bulk acoustic properties of the porous sound absorption materials (i.e., the complex characteristic impedance and propagation constant) and various other physical parameters. These physical variables can be classified into two main groups: geometrically related and fluid related. The geometrically related variables refer to those associated with the arrangement of the porous skeleton, e.g., the fiber diameter, the porosity derived from the ratio of the fiber diameter and the interfiber distance, and the wavelength. The relevant fluid-related variables include the fluid density, the speed of sound, which characterizes the fluid compressibility, and fluid viscosity. As the fluid passes through the pores, viscous damping occurs due to the friction between the fluid and the skeleton of the porous material. In order to investigate the variations in the characteristic impedance of the porous material filled with different gases, the two constitutional components of the characteristic impedance, i.e., the complex speed of sound c and the complex density ρ inside the porous material, are analyzed separately,

$$c = F(\mu, f, d, \rho_{\text{gas}}, c_{\text{gas}}, d_L) \quad \text{and} \quad \rho = G(\mu, f, d, \rho_{\text{gas}}, c_{\text{gas}}, d_L), \tag{1}$$

where μ is the dynamic viscosity, f is the frequency, d is the fiber diameter, ρ_{gas} is the fluid density, c_{gas} is the sound speed of the fluid in free field, and d_L is the mean distance between fibers. The effect of heat transfer can also be included in a manner similar to that for the viscous effect, but no extra control parameter arises when the Prandtl number of the fluid media is close to unity. Based on Buckingham theorem (Buckingham, 1914) in dimensional analysis, both the dependent variables (c and ρ) and the independent variables (μ, f, d, d_L, ρ_{gas}, and c_{gas}) are expressed in terms of the three fundamental physical quantities such as mass, length, and time. As a result, the normalized complex speed of sound c/c_{gas} and the normalized complex density ρ/ρ_{gas} depend on three dimensionless groups as follows:

$$\frac{c}{c_{\text{gas}}} = F\left(H, \frac{fd}{c_{\text{gas}}}, \frac{f\rho_{\text{gas}}d^2}{\mu}\right) \quad \text{and} \quad \frac{\rho}{\rho_{\text{gas}}} = G\left(H, \frac{fd}{c_{\text{gas}}}, \frac{f\rho_{\text{gas}}d^2}{\mu}\right). \tag{2}$$

The first dimensionless variable H in Eq. (2) is the porosity of the porous material, which is a crucial parameter for controlling the flow resistance. The second independent dimensionless variable is expressed as the fiber diameter normalized by the wavelength in the pure fluid, denoted by λ. The third independent dimensionless variable, $f\rho_{\text{gas}}d^2/\mu$, is proportional to frequency, the fluid density, and the square of the pore size but inversely proportional to the fluid viscosity. This term can be interpreted as the ratio of the fiber diameter to the boundary layer thickness δ with the power of 2, where $\delta = 5\sqrt{\mu/\rho\omega}$ (Lighthill, 1978).

The effects of the second and third dimensionless groups on the acoustical properties of porous material can be analyzed and compared as follows. Except for ultrasonic frequency ranges, the wavelength λ is normally much larger than the fiber diameter. Since the main geometric effect is represented by porosity and the main dynamic (force) effect derives from the fluid viscosity, the ratio d/λ may be interpreted as an indication of the effect of the fiber "individuality," which is not considered when the absorption material is treated as an equivalent fluid. On the other hand, the ratio of the fiber diameter to the acoustic boundary layer thickness d/δ signifies the main dynamic effect that the fibers have on the oscillating fluid particles, and it is typically much larger than d/λ. Therefore, the effect of the second independent variable fd/c_{gas} (i.e., d/λ) can be neglected, much like the way the weak influence of the Reynolds number has on lift

and drag coefficients that is excluded when the Reynolds number is very large (Lighthill, 1978). The number of independent variables is thus reduced from 3 to 2. Before the above assumption is adopted, the following numerical comparison between the three length scales is given for one specific configuration for reference. For a rectangular duct of height $h=0.1$ m, in room air condition, the boundary layer thickness δ and the wavelength λ are found to be $\delta=2.4\times10^{-4}$ m and $\lambda=3.43\times10^{-1}$ m at frequency $f=1000$ Hz. If the fiber diameter is $d=3\times10^{-5}$ m, then $d/\lambda=8.7\times10^{-5}$ and $d/\delta=0.13$. The corresponding ratios in helium are $d/\lambda=2.9\times10^{-5}$ and $d/\delta=0.046$, respectively. On neglecting the effect of d/λ, the normalized parameters are expressed as the new functions of porosity H and $f\rho_{\text{gas}}d^2/\mu$, denoted below by a new symbol E,

$$\frac{c}{c_{\text{gas}}}=F(H,E) \quad \text{and} \quad \frac{\rho}{\rho_{\text{gas}}}=G(H,E) \quad \text{where} \quad E=\frac{f\rho_{\text{gas}}d^2}{\mu}. \tag{3}$$

A series of experiments is carried out to establish the nature of these two functions with respect to E for a few given porosity H in Sec. III. The experimental evidence will also validate that the same functions of $F(H,E)$ and $G(H,E)$ do apply to both helium and air-filled absorption materials of the same geometrical shape. As already mentioned in the Introduction, specific functionalities have been established between the bulk acoustic properties and a series of physical parameters in previous theoretical models (Johnson *et al.* 1987; Attenborough, 1982, 1983, 1987; Allard *et al.*, 1986, 1989). According to theses models, for given geometrical configuration, the effective density normalized by that of the pure filling gas is dependent on the single variable $E_0=\rho_{\text{gas}}f/\mu$. For example, the normalized effective density in Johnson's model (Johnson *et al.*, 1987) can be expressed as

$$\frac{\rho}{\rho_{\text{gas}}}=\alpha_\infty\left(1+i\frac{H}{2\pi\alpha_\infty\kappa E_0}\sqrt{1-i\frac{8\pi\alpha_\infty^2\kappa^2}{H^2\Lambda^2}E_0}\right),$$

where the tortuosity α_∞, the Darcy constant κ, the porosity H, and the characteristic length Λ only depend on the geometrical configuration of the porous media. Thus, Johnson's model indicates that ρ/ρ_{gas} is solely determined by $E_0=\rho_{\text{gas}}f/\mu$ for given geometrical parameters. The effective speed of sound can be deduced from either Attenborough's (1982, 1983, 1987) or Allard *et al.*'s (1986, 1989) model. For given geometrical parameters, the theoretical models suggest that the ratio c/c_{gas} is a function of $E_0=\rho_{\text{gas}}f/\mu$, the specific heat ratio γ, and the Prandtl number B. For the helium gas and air considered in this study, it can be shown that the ratio c/c_{gas} is mainly determined by E_0 and the effect of γ and B can be neglected. In other words, for given geometrical parameters, the ratio c/c_{gas} can be approximately expressed as a function of E_0 in the theoretical models developed by Attenborough (1982, 1983, 1987) and Allard *et al.* (1986, 1989). Note that the fiber diameter d in the dimensionless parameter $E=f\rho_{\text{gas}}d^2/\mu$ in Eq. (3) is constant for given porous material. It can be seen that functional relations in Eq. (3) is consistent with the three above-mentioned theoretical models.

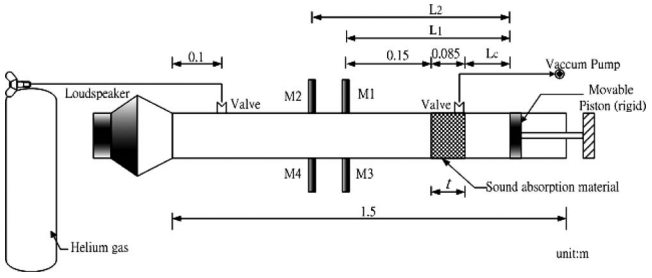


FIG. 1. Experimental set-up for the measurement of acoustical properties of porous materials filled with air or helium (dimensions: m).

III. EXPERIMENTAL STUDY AND REGRESSION ANALYSES

Three kinds of materials with different fiber diameters are tested with the experimental set-up shown in Fig. 1. The tube is specially designed for filling different kinds of gases under air-tight condition, namely, no leakage of gas. To ascertain this condition, the tube is calibrated by measuring and checking the speed of sound when helium gas is filled.

A. Calibration of the helium tube

The speed of sound of the helium gas in the test rig is measured with the test specimen (sound absorption material) absent in the experimental set-up. As shown in Fig. 1, when an incident sound wave hits the rigid piston, it will be reflected totally, and standing wave occurs in the tube. Below the first cut-on frequency, the sound pressures measured at M1 and M2 satisfy

$$p_1\cos(k_{\text{gas}}L_2)=p_2\cos(k_{\text{gas}}L_1), \tag{4}$$

where L_1 and L_2 are the distances between the surface of the movable piston and the microphones M1 and M2, respectively. To solve Eq. (4) for wavenumber k_{gas}, it is first expanded as

$$k_{\text{gas}}=k_0+\Delta k \quad \text{with} \quad k_0=\omega/c_{0,\text{gas}}, \tag{5}$$

where $c_{0,\text{gas}}$ is the nominal speed of sound in the filling gas (helium or air) indicated by the handbook (Zuckerwar, 2002), and Δk is the deviation of the measured wavenumber from the nominal value k_0 due to gaseous impurities and variation in environmental conditions. Assuming a small deviation Δk, the following approximation

$$e^{ik_{\text{gas}}L}\approx e^{ik_0L}(1+\Delta kL) \tag{6}$$

may be applied. By substituting Eq. (6) into Eq. (4), it gives

$$\Delta k=\frac{p_2\cos(k_0L_1)-p_1\cos(k_0L_2)}{p_2L_1\sin(k_0L_1)-p_1L_2\sin(k_0L_2)}. \tag{7}$$

Expanding Δk into the real and imaginary parts such that $\Delta k=\Delta k_r+i\Delta k_i$, the actual speed of sound in the gas is found as

$$c_{\text{gas}}=\omega/(k_0+\Delta k_r). \tag{8}$$

Figure 2 shows the measured (open circle) and the nominal (solid line) speed of sound of helium gas in room condition, together with the measured (triangle) and nominal (dashed line) speed of sound of air inside the same tube. The nominal

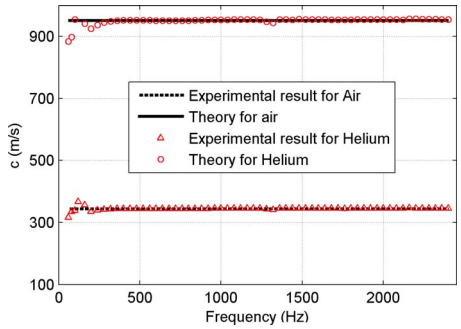


FIG. 2. (Color online) The speed of sound of helium- and air-filled inside the experimental testing tube.

speed of sound in air is 343 m/s for the particular temperature condition, while the experimental result is 340 m/s on average. For the helium gas, the nominal speed of sound is 970 m/s, while the measured value is about 975 m/s. On the whole, the measured speed of sound is in good agreement with the nominal value, especially in the relatively high frequency range (say, above 250 Hz). The measured speed of sound deviates from the nominal value gradually as the frequency approaches the very low frequency range (below 250 Hz). The largest deviation is observed at the lowest frequency, with the largest relative error of about 7% at the lowest frequency tested (120 Hz). The increased deviation may be attributed to the usual data uncertainty at extremely low frequencies due to the effect of the rig vibration. Nevertheless, it is considered to be acceptable for measuring the acoustical parameters of gas-filled porous material by using this test rig.

B. Methodology

The characteristic impedance and propagation constant are the prime parameters to be determined in order to find out the effective density and the speed of sound inside the porous material. Decomposition of these two parameters from the measurement of sound wave requires two distinct impedance conditions. Smith and Parrot (1983) achieved the distinct impedances by using two different thicknesses of material backed by a rigid cavity and this is called the two-thickness method. This method is not convenient for the current study because of the need to change the specimen inside the tube. On the other hand, Utsuno *et al.* (1989) introduced the improved two-cavity method in which two distinct impedances can be offered by changing the acoustic impedance behind the material. This can be achieved by adjusting the depth of the space behind the material. The thickness of the material remains the same in this method, so it is adopted for the current investigation of the acoustic properties of helium-filled material.

The experimental set-up for testing different kinds of absorption materials filled with different gases is shown in Fig. 1. The test tube, which has an internal diameter of 6×10^{-2} m and a total length of 1.5 m, is terminated by a movable rigid piston. The other end of the tube is connected to a loudspeaker. In order to avoid leakage of the filling gas, the surface of the cone of the loudspeaker is covered by a thin plastic membrane. The first cut-on frequency of the test tube is about 2850 Hz when air is filled and 7900 Hz when helium is filled. Since the density of helium gas is far less than that of air, the measurement frequency range was extended above the first cut-on frequency but still below the second cut-on frequency, so that the enough overlapping range of $E=f\rho_{\text{gas}}d^2/\mu$ between helium and air can be observed. For instance, the range of E for helium is extended from $E=2.8$ (below the first cut-on) to $E=5$ in Figs. 4(1a)–4(1d). It is known that, below the second cut-on frequency, both plane-wave mode and the first acoustic mode may occur in the test tube. As shown in Fig. 1, four microphones are used to resolve the plane-wave components from the standing wave pattern in the tube. The four microphones were calibrated before the measurement according to ISO 10534-2 (1998). By averaging the pressures at M1 (M2) and M3 (M4), the first acoustic mode is eliminated and only the plane-wave mode is kept, see Eqs. (9) and (10). Two microphone spacings were used for different frequency ranges during the measurement. In the low frequency range ($f \leq 1900$ Hz for air-filled tube and $f \leq 5300$ Hz for helium-filled tube), the separation distance between the two pairs of microphones was 8×10^{-2} m. In the high frequency range (1900 Hz $< f \leq 5000$ Hz for air-filled tube and 5300 Hz $< f \leq 11000$ Hz for helium-filled tube), a smaller separation distance of 3×10^{-2} m was used. There are two valves placed at 6×10^{-2} and 7.8×10^{-1} m away from the left end of the tube for air suction and gas injection, respectively. The movable piston is first used to place the testing samples in the appropriate position. Once the position of the specimen is fixed, the rigid piston is moved backward to create the space of depth L_c, which is measured by the scale attached onto the tube. The distance between the front surface of the testing material and the microphone M1 is adjusted to be 1.5×10^{-1} m. An air pump shown in Fig. 1 is used to discharge the air from the tube via a valve at one end while the helium gas is injected simultaneously into the tube from the other end. The pressure in the tube is monitored via a pressure gauge during the entire injecting and discharging procedure that lasts for at least 5 min. This is to ensure that the tube is completely filled with helium gas with minimum trace of air remaining in the tube. After the injection of helium gas, a signal with pre-set frequency range is generated from the loudspeaker, and the transfer function between the two microphones, M1 and M2, can be acquired. The incident noise is simulated by this loudspeaker. Two pairs of 0.5 in. microphones are used together with a conditioning amplifier to acquire the reasonably high signal-to-noise ratio. By using four microphones, the frequency range can be extended to 5000 Hz when it is air-filled and 11 000 Hz when it is helium-filled.

Formulations for the calculations of the characteristic impedance and propagation constant are briefly presented below. With the sound pressures measured by microphones M1–M4, the plane-wave components of the incident wave I and the reflected wave R can be found as follows:

$$I = \frac{(p_1 + p_3)e^{-ik_{\text{gas}}L_2} - (p_2 + p_4)e^{-ik_{\text{gas}}L_1}}{-i4\,\sin[k_{\text{gas}}(L_2 - L_1)]}, \tag{9}$$

$$R = \frac{(p_2 + p_4)e^{ik_{\text{gas}}L_1} - (p_1 + p_3)e^{ik_{\text{gas}}L_2}}{-i4\,\sin[k_{\text{gas}}(L_2 - L_1)]}. \tag{10}$$

As a result, the acoustic impedance Z at the surface of the porous material is

$$Z = \frac{I + R}{I - R}\rho_{\text{gas}}c_{\text{gas}}, \tag{11}$$

and it can be expressed in terms of the characteristic impedance Z_c, the propagation constant k of the porous material, as well as the thickness of the porous material t under test (Utsuno *et al.*, 1989),

$$Z = Z_c \cdot \frac{-i\rho_{\text{gas}}c_{\text{gas}}\,\cot(k_{\text{gas}}L_c)\cosh(kt) + Z_c\,\sinh(kt)}{-i\rho_{\text{gas}}c_{\text{gas}}\,\cot(k_{\text{gas}}L_c)\sinh(kt) + Z_c\,\cosh(kt)}, \tag{12}$$

where ρ_{gas}, c_{gas}, and k_{gas} are, respectively, the density, the speed of sound, and the wavenumber of the pure gas in free space. In order to get the second surface impedance, the cavity depth is changed from L_c to L_c' by adjusting the movable piston. The resulting surface impedance of the testing sample is denoted as Z'. During the process of cavity depth change, the thickness of the testing sample remains unchanged. With the two surface acoustic impedances Z and Z' measured, the characteristic impedance Z_c and the propagation constant k inside the porous material can be found as

$$Z_c = \pm\sqrt{\frac{ZZ'(-i\rho_{\text{gas}}c_{\text{gas}})[\cot(k_{\text{gas}}L_c) - \cot(k_{\text{gas}}L_c')] + \rho_{\text{gas}}^2c_{\text{gas}}^2\,\cot(k_{\text{gas}}L_c)\cot(k_{\text{gas}}L_c')(Z - Z')}{-i\rho_{\text{gas}}c_{\text{gas}}[\cot(k_{\text{gas}}L_c) - \cot(k_{\text{gas}}L_c')] - (Z - Z')}} \tag{13}$$

and

$$k = \frac{1}{i2t}\ln\left[\frac{(Z + Z_c)}{Z - Z_c}\cdot\frac{-i\rho_{\text{gas}}c_{\text{gas}}\,\cot(k_{\text{gas}}L_c) + Z_c}{-i\rho_{\text{gas}}c_{\text{gas}}\,\cot(k_{\text{gas}}L_c) - Z_c}\right]. \tag{14}$$

The complex speed of sound and density of the porous material are then calculated by $c = \omega/k$ and $\rho = Z_c/c$, respectively. Note that the cavity depth L_c and L_c' should be carefully chosen so that two distinct sets of data can be measured. In this study, the cavity depths are $L_c = 4 \times 10^{-2}$ m and $L_c' = 1.1 \times 10^{-1}$ m for AA. For HA, since the wavelength of the helium gas is longer, a set of larger cavity depth, $L_c = 5 \times 10^{-2}$ m and $L_c' = 2.45 \times 10^{-1}$ m, is chosen.

C. Results

Three different porous materials are tested. They are as follows: (1) polyethylene foam with diameters of 2×10^{-4}, 3×10^{-4}, and 5×10^{-4} m; (2) polyester fiber material with a fiber diameter of 3×10^{-5} m; and (3) metal bead with a pore size of 4×10^{-3} m. All the testing samples are prepared as a circular cylinder of 6×10^{-2} m in diameter and 8.5×10^{-2} m in length. Following the apparent density method (Palacio, *et al.*, 1999), the porosity of the polyethylene foam, polyester fibers, and metal bead are found to be $H = 0.97$, 0.97, and 0.2, respectively. In the following presentation, the measured complex speed of sound c and effective density ρ of the porous materials are normalized as c/c_{gas} and ρ/ρ_{gas}, where c_{gas} and ρ_{gas} are the speed of sound and density of the corresponding filling gas in free space, respectively.

The results for the air-filled polyethylene foam are presented in Fig. 3. Figures 3(a) and 3(b) show the real and imaginary components of the normalized speed of sound as a function of $E = f\rho_{\text{gas}}d^2/\mu$. The physical meaning of the dimensionless parameter E has been explained in Sec. II. Figures 3(c) and 3(d) show the real and imaginary components of the normalized effective density as a function of E. The open circles, triangles, and stars shown on the graphs represent the measured results with fiber diameters of 2×10^{-4}, 3×10^{-4}, and 5×10^{-4} m, respectively. It can be observed that the results for different diameters show similar dependency on E. Although deviations may be observed at certain points, for instance, the imaginary part of c/c_{gas} and the imaginary part of ρ/ρ_{gas} when E is between 0.8 and 1.2, the three sets of data overlap with each other on the whole. This overlapping trend appears to confirm that, for given porosity H, the normalized complex speed of sound c and effective density ρ of the porous materials are decided by the dimen-

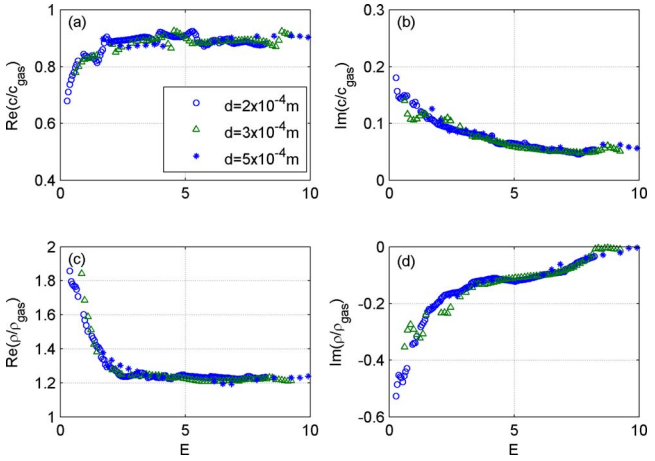


FIG. 3. (Color online) Characteristic of the polyethylene foam with different diameters and filled with air: (a) is the real part and (b) is the imaginary part of sound speed of the fluid in the porous material relative to that of air, respectively. (c) is the real part and (d) is imaginary part of effective density of porous material relative to that of air, respectively.

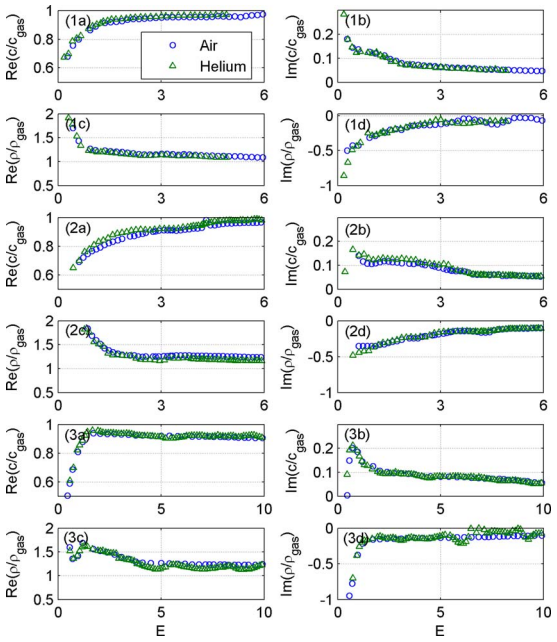


FIG. 4. (Color online) Characteristic of the polyethylene foam filled with helium (triangles) and air (circles). (1)–(3) represent the sponges with fiber diameters of 2×10^{-4}, 3×10^{-4}, and 5×10^{-4} m, respectively. For each set of subfigures, (a) is the real part and (b) is the imaginary part of sound speed of the fluid in the porous material relative to that of pure gas, respectively. (c) is the real part and (d) is the imaginary part of effective density of porous material relative to that of pure gas, respectively.

sionless parameter $E=f\rho_{\text{gas}}d^2/\mu$. This property can be employed to predict the acoustic properties of the air-filled porous material of other fiber diameters with the same porosity.

Figure 4 compares the measured acoustic properties for the helium-filled (triangles) and the air-filled (open circles) polyethylene foam as a function of $E=f\rho_{\text{gas}}d^2/\mu$. Three sets of subfigures, (1)–(3), are plotted for three different fiber diameters, 2×10^{-4}, 3×10^{-4}, and 5×10^{-4} m, respectively. In general, the normalized acoustic properties of the air-filled polyethylene foam coincide with those of the helium-filled polyethylene foam with little deviations. In the case of fiber diameter $d=2\times10^{-4}$ m, the maximum deviation of 0.06 is observed for the imaginary part of the normalized effective density, as shown in Fig. 4(1d). In the cases of fiber diameters $d=3\times10^{-4}$ and 5×10^{-4} m, the maximum deviation is found to be 0.05 in the low frequency range, as shown in Figs. 4(2d) and 4(3d). The deviation of 0.05 in the imaginary part of the effective density will cause a deviation of the normal absorption coefficient as large as 0.06, which can be neglected in many engineering applications. In other words, although the filling gases differ, the normalized acoustic properties of the porous materials show almost the same dependency on the dimensionless parameter, $E=f\rho_{\text{gas}}d^2/\mu$. Additionally, the smooth variation in the experimental results suggests that generalized regression curves can be made to characterize the normalized parameters of c/c_{gas} and ρ/ρ_{gas} for porous materials filled with different gases.

Glass fibers and polyester fibers are very commonly used in industry such as duct lining and acoustical treatment in room. Polyester fibers with fiber diameter $d=3\times10^{-5}$ m are tested here. Figure 5 shows the measured normalized

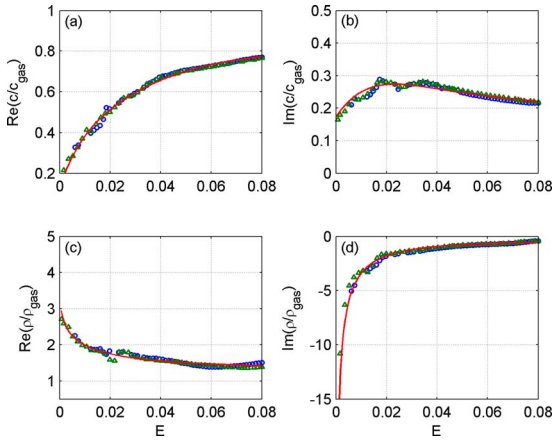


FIG. 5. (Color online) Characteristic of the polyester fiber with a fiber diameter of 3×10^{-5} m and filled with air (circles) and helium (triangles): (a) is the real part and (b) is the imaginary part of sound speed. (c) is the real part and (d) is the imaginary part of effective density, respectively. The solid lines show the regression curves based on the measured data of the air-filled polyester fiber.

acoustical parameters of air-filled (circles) and helium-filled (triangles) polyester fibers. Again, good agreement can be observed between the air-filled and helium-filled polyester fibers except at $E=0.02$ in Fig. 5(c), where a deviation of about 0.25 is observed for the real part of the normalized effective density. The same measurements are repeated for the air-filled and helium-filled metal beads with the diameter of 4×10^{-3} m, and similar observations can be made, as shown in Fig. 6.

Regression curves are deduced based on the measured data for air-filled polyester fiber in the forms of

$$\frac{c}{c_{\text{gas}}}=\sum_{m=0}^{4}a_mE^m+i\sum_{n=0}^{6}b_nE^n, \tag{15}$$

$$\frac{\rho}{\rho_{\text{gas}}}=\sum_{m=0}^{2}r_m(\log E)^m+i\sum_{n=0}^{4}s_n(\log E), \tag{16}$$

where a_m, b_n, s_n, and r_m are the regression parameters. The orders of the polynomial curves and the unknown regression

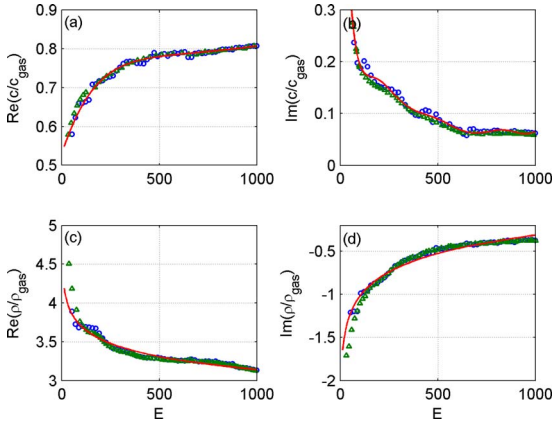


FIG. 6. (Color online) Characteristic of metal beads with diameter of 4×10^{-3} m and filled with air (circles) and helium (triangles): (a) is the real part and (b) is imaginary part of sound speed, (c) is the real part and (d) is the imaginary part of effective density, respectively. The solid lines show the regression curves based on the measured data of the air-filled metal beads.

TABLE I. The regression parameters for the empirical formulas for polyester fiber.

	$E \geq 0.12$	$E < 0.12$
a_0	0.792	0.147
a_1	0.511	27.89
a_2	−0.613	−553.62
a_3	0.284	5 081.88
a_4	−0.0467	−16 739.9
b_0	0.254	0.173
b_1	−0.742	11.65
b_2	1.32	−479.74
b_3	−1.311	8 926.6
b_4	0.717	−87 247
b_5	−0.202	430 398.8
b_6	0.023	−843 642.1
r_0	1.27	1.27
r_1	0.0154	0.0154
r_2	0.0291	0.0291
s_0	−0.06	−0.06
s_1	0.173	0.173
s_2	0.016	0.016
s_3	−0.017	−0.017
s_4	−0.01	−0.01

parameters are determined by minimizing the mean deviations between the measured results and the polynomial curves. The standard least-squares procedure is followed in the minimization of the mean deviations. Table I lists the regression parameters for the air-filled polyester fibers. The corresponding regression curves are plotted in Fig. 5. Although the regression curves are deduced using the measured results of the air-filled polyester fiber, they also match the measured results of the helium-filled materials.

As a summary, the complex speed of sound and the effective density of a variety of helium-filled porous materials are measured with an especially designed test rig. Since the helium gas changes the virtual mass and the flow resistance of the porous material greatly, the measured acoustic properties differ from those of the commonly used air-filled porous materials. However, it is found that the normalized parameters of c/c_{gas} and ρ/ρ_{gas} follow the same functional relationships of $c/c_{\text{gas}}=F(H,E)$ and $\rho/\rho_{\text{gas}}=G(H,E)$, where H is the porosity and $E=f\rho_{\text{gas}}d^2/\mu$. That is to say, following the normalization scheme described in Eq. (3), the acoustic properties of the helium-filled porous materials can be predicted from the existing data of the air-filled porous materials. This observation facilitates the analytical studies on the acoustic behaviors of porous materials saturated by gases other than air.

IV. ABSORPTION OF A HELIUM-FILLED FINITE-LAYER ABSORBER

To demonstrate the potential advantages of the helium gas filling in the porous absorption material, the performance of a HA of finite thickness is studied for both free field and confined space (duct). The polyester fiber is selected as the sound absorption material. With the empirical formulas established in Eqs. (15) and (16), the acoustical properties of the polyester fiber with different fiber diameters can be predicted.

A. Absorption in free field

In free field, the noise sources can be plane waves at normal, oblique, or diffuse incidence. For a finite absorption layer of thickness t in front of a rigid wall, the absorption coefficient at oblique incidence θ_0 is (Attenborough and Vér, 2006)

$$\alpha(\theta_0)=\frac{4\,\text{Re}(Z)\cdot Z_0\cos\theta_0}{[\text{Re}(Z\cos\theta_0)+Z_0]^2+\text{Im}(Z\cos\theta_0)^2}$$

with

$$Z=\frac{(Z_c/\rho_0 c_0)}{\cos\theta}\coth(ikt\cos\theta), \tag{17}$$

where the incidence angle θ_0 is measured from the direction normal to the absorber surface, θ is the complex refraction angle in the absorber layer, and $Z_0=1$ is the specific acoustic impedance of air. According to the rule of refraction (Snell's law), the complex refraction angle satisfies the relationship

$$\cos^2\theta=1-\frac{k_0^2}{k^2}\sin^2\theta_0, \tag{18}$$

where $k_0=\omega/c_0$ is the free field wavenumber in air. In most situations, the sound will be incident on an absorptive material from a multitude of incident angles at once. Therefore, to account for a wide range of incidence angles, the statistic absorption coefficient α_{ran} is usually determined by summing the absorption coefficients at all angles of incidence, as defined below:

$$\alpha_{\text{ran}}=\int_0^{\pi/2}\alpha(\theta_0)\sin(2\theta_0)d\theta_0. \tag{19}$$

The absorption performance of the HA is calculated and compared with that of the AA, as shown in Fig. 7. Both absorbers have a thickness of $t=5\times10^{-2}$ m, while the porosity of the polyester fiber is $H=0.97$. Figures 7(a) and 7(b) plot the contours of the absorption coefficient under random incidence, α_{ran}, as a function of the fiber diameter d and the normalized frequency

$$f_d=ft/c_0, \tag{20}$$

where t is the thickness of the finite absorption layer, and c_0 is the free field speed of sound in air. Figure 7(a) shows that AA can achieve at least $\alpha_{\text{ran}}=0.5$ for normalized frequency $f_d>0.04$. For HA shown in Fig. 7(b), $\alpha_{\text{ran}}\geq0.5$ can be achieved for $f_d>0.014$. For higher absorption coefficient, e.g., $\alpha_{\text{ran}}=0.7$, the lower frequency limits of AA and HA are $f_d=0.062$ and 0.034, respectively. The above comparison suggests that HA has a great potential to enhance the sound absorption performance in the low frequency range. Meanwhile, the comparison between the two contour plots also shows that the most effective fiber diameters of the two absorbers differ from each other. Aiming to maximize the absorption capability of the finite-layer absorber, an optimiza-

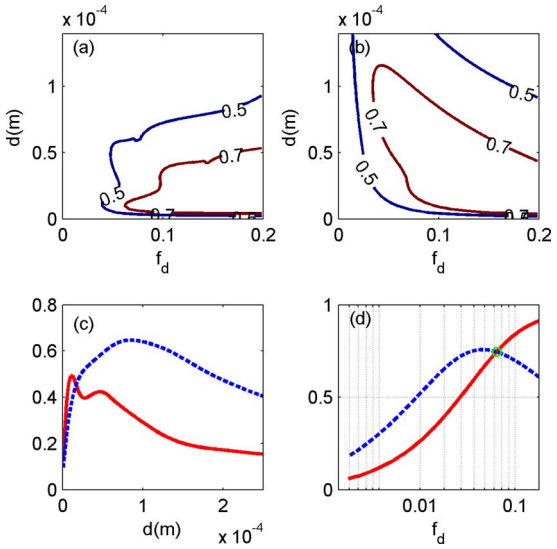


FIG. 7. (Color online) Performance optimization of the finite-layer absorbers (thickness $t=5\times10^{-2}$ m; porosity $H=0.97$). The top row shows the contour plots of the statistic absorption coefficient α_{ran} as a function of normalized frequency f_d and fiber diameter d for (a) the air-filled absorber and (b) the helium-filled absorber, respectively. (c) The variation in the overall (frequency integrated) absorption coefficient $\bar{\alpha}_{\text{ran}}$ as a function of the fiber diameter d (solid: air-filled; dashed: helium-filled). The two open circles mark the optimal diameters for the two absorbers. (d) Spectral comparison of the optimal absorption curves between the air-filled (solid) and the helium-filled (dashed) absorbers.

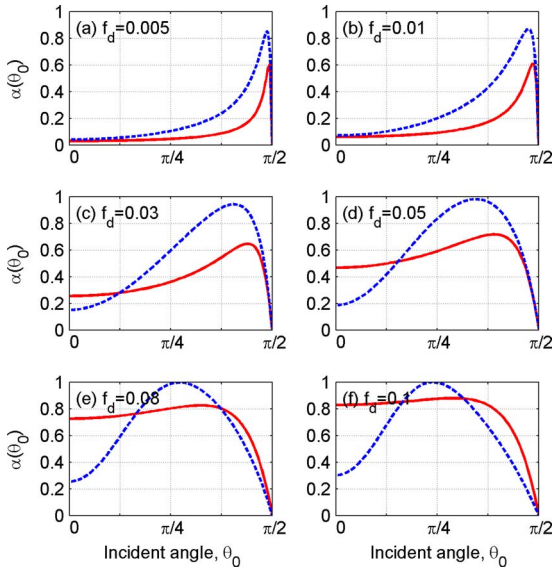


FIG. 8. (Color online) Absorption coefficients of the air-filled (solid lines) and helium-filled (dashed line) absorbers as a function of the angle of incidence θ_0. The porosity of the polyester fiber is $H=0.97$, the fiber diameter is $d=1\times10^{-5}$ m for the air-filled absorber, and $d=9\times10^{-5}$ m for the helium-filled absorber. The thickness for both absorbers is $t=5\times10^{-2}$ m.

tion procedure is necessary to find the most effective fiber diameters.

For given porosity $H=0.97$ and thickness $t=5\times10^{-2}$ m, the absorber performance is optimized by searching for the optimal fiber diameter d_{opt} so that the overall absorption coefficient is the highest within the frequency range of interest, $f_d=[f_1,f_2]$. Here, the overall absorption coefficient $\bar{\alpha}_{\text{ran}}$ is simply defined as

$$\bar{\alpha}_{\text{ran}}=\frac{\int_{f_1}^{f_2}\alpha_{\text{ran}}df_d}{f_2-f_1}. \tag{21}$$

Because special emphasis is put on the low frequency noise, the normalized frequency range of [0.00625,0.1] is investigated. The corresponding dimensional frequency range can be calculated as [42.5 Hz, 680 Hz], which falls into the low-to-medium frequency range in the RC II curves (Blazier, 1997). Figure 7(c) shows the variation in the overall absorption coefficient $\bar{\alpha}_{\text{ran}}$ in the specified frequency range as a function of fiber diameter d. The optimal fiber diameters for HA (dashed line) and AA (solid line) are marked by open circles. Figure 7(d) shows the statistic absorption coefficients predicted with the optimal fiber diameters. At very low frequencies (e.g., $f_d<0.04$), HA (dashed line) outperforms the AA (solid line) significantly. The overall noise absorption level of HA almost doubles that of the AA in this frequency region. The advantage of HA extends to the frequency range around $f_d=0.08$, beyond which the performance of HA becomes worse than that of AA. The reasons for the performance degradation of HA at high frequency are discussed in Sec. IV B.

Figure 8 compares the absorption coefficient $\alpha(\theta_0)$ of HA (dashed lines) and AA (solid lines) as a function of the angle of incidence. Figures 8(a)–8(f) depict the variation in the absorption coefficients at different frequencies. Two observations can be made here. First, at normal incidence angle ($\theta_0=0$), HA shows no advantages over AA. As shown in Figs. 8(a) and 8(b), both absorbers have little absorption at very low frequencies if the incident sound is perpendicular to the air-absorber interface. As the frequency increases, AA performs much better than HA, as illustrated in Figs. 8(c)–8(f). Second, the advantage of HA lies in its strong absorption capability at oblique incidence, especially in the middle range of the angle of incidence. As shown in Figs. 8(d)–8(f), almost total absorption occurs at $\theta_0=\pi/3$ ($f_d=0.05$) and at $\theta_0=\pi/4$ ($f_d=0.08,0.1$). As shown by the weighting function $\sin(2\theta_0)$ in Eq. (19), the absorption at angles around $\theta_0=\pi/4$ dominates the total noise absorption for random incidence. Therefore, in comparison with AA, HA demonstrates much better performance at low to middle frequencies. However, as the frequency increases further, this advantage of HA is offset by its relatively low absorption efficiency at small angles of incidence, as shown in Fig. 8(f).

B. Impedance analysis

The distinct absorption behavior of HA is analyzed by its unique impedance matching condition. As illustrated in Eq. (17), the surface impedance of the finite-layer absorber can be decomposed into two parts. The first part is the specific characteristic impedance of the sound absorbing material, Z_c/ρ_0c_0, where ρ_0c_0 is the characteristic impedance of air in free field. The second part is $\coth(ikt\cos\theta)/\cos\theta$, which characterizes the multiple reflected sound waves in the thin layer. At normal incidence, this part reduces to $\coth(ikt)$ as $\cos\theta=1$. In the following, the above two parts are simply called the characteristic impedance and the cavity impedance

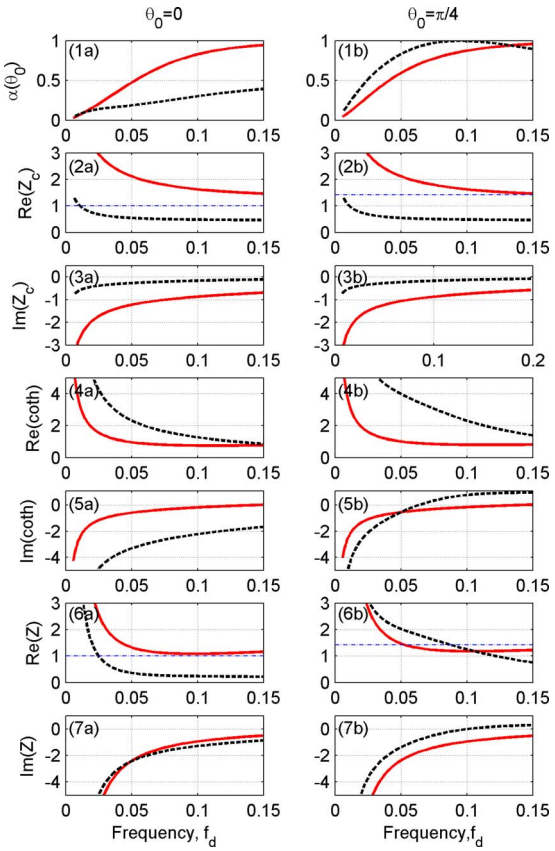


FIG. 9. (Color online) Impedance analysis of the air-filled (solid lines) and helium-filled (dashed lines) absorbers with the values of perfect match shown in dash-dot lines in some subfigures. The left column is for the normal incidence of $\theta_0=0$. The right column is for the oblique incidence with $\theta_0=\pi/4$. In (4a) and 5(a), "coth" means the cavity impedance $\coth(ikt\cos\theta)/\cos\theta$.

for convenience. Figure 9 shows the absorption coefficients and the corresponding acoustic impedances of the two absorbers at $\theta_0=0$ and $\pi/4$ (dashed lines: helium-filled; solid lines: air-filled). The physical properties of the sound absorption materials are the same as in Fig. 8.

The left column in Fig. 9 shows the results at normal incidence, $\theta_0=0$. At normal incidence, Figs. 9(2a) and 9(3a) just present the characteristic impedances of the helium-filled and air-filled polyester fibers normalized by that of air by setting $Z_0=1$. The characteristic impedance of the air-filled polyester fiber is too high, especially in the very low frequency range. The impedance mismatch between the air-filled polyester fiber and the incident wave reduces the chance of incident sound waves entering the absorber and being dissipated. On the contrary, the characteristic impedance of the helium-filled polyester fiber decreases from high to just under the value of unity as the frequency increases. Hence, an approximate impedance matching between the helium-filled polyester fiber and the incident field $(Z_0=1)$ can be observed in the frequency range $f_d \in [0.01, 0.03]$. However, in the case of normal incidence, HA does not perform better than AA, as indicated in Fig. 9(1a). The performance degradation of HA can be ascribed to its higher speed of sound and correspondingly longer acoustic wavelength. The longer acoustic wavelength makes the helium-filled absorber layer acoustically thinner than that of the air-filled

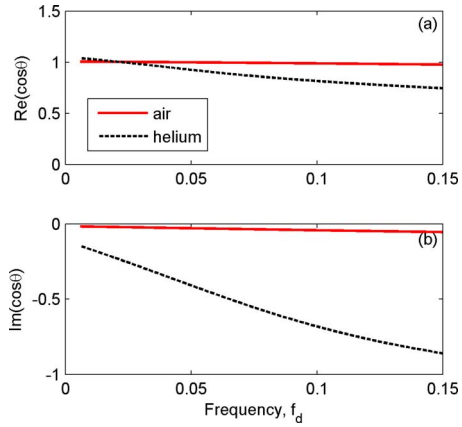


FIG. 10. (Color online) The cosine function of the complex refraction angle in the air-filled (solid lines) and the helium-filled (dashed lines) polyester fibers. (a) Real component and (b) imaginary component.

layer. Mathematically, it turns out to be a larger magnitude of the cavity impedance, $\coth(ikt)$, as shown in Figs. 9(4a) and 9(5a). Therefore, the total surface impedance of the helium-filled absorber, $Z_c \coth(ikt)/\rho_0 c_0$, still mismatches the impedance of the incident sound field $(Z_0=1)$ at low frequencies, as shown in Figs. 9(6a) and 9(7a). In other words, although the reduced characteristic impedance allows the incident noise to penetrate easily the helium-filled absorber, the thickness t is so thin compared with the acoustic wavelength that the incident noise has little chance of being dissipated before it is reflected by the rigid wall.

The right column in Fig. 9 shows the results when the angle of incidence is $\theta_0=\pi/4$. It can be seen that both absorbers have a better absorption performance than that at the normal incidence, but HA (dashed line) obviously outperforms AA (solid line) in this case. In order to find out the reasons for such performance difference, the effect of the oblique incidence angle on the acoustical behavior of AA is first investigated. Since the sound speed of the air-filled porous material is much slower than that in free space, the penetrated sound wave is bent toward the normal direction in the absorption layer. As shown in Fig. 10, for AA, the $\cos\theta$ is approximately equal to 1, which means that the penetrated sound wave is almost perpendicular to the air-absorber interface just as in the case of normal incidence. Hence, the acoustic impedance of AA is almost the same as that at normal incidence, see Eq. (17). The slight performance improvement of AA mainly derives from the changed impedance matching conditions relative to the situation of normal incidence. At oblique incidence, the absorption coefficient $\alpha(\theta_0)$ for the finite-layer absorber depends on how well the surface impedance Z matches the field impedance $Z_0/\cos\theta_0$ instead of Z_0. Note that the lack of impedance match at low frequencies is due to the overmatched surface impedance Z in many cases. Since the impedance mismatch is reduced by the increased field impedance $Z_0/\cos\theta_0=\sqrt{2}$ instead of $Z_0=1$, the absorption performance of AA is slightly improved in the low frequency range.

The effect of the oblique incidence angle on HA is more complicated but it can be shown to be beneficial. For HA, the characteristic impedances shown in Figs. 9(2b) and 9(3b)

remain the same as those in Figs. 9(2a) and 9(3a) (dashed line). However, as the incident field impedance becomes $Z_0/\cos\theta_0=\sqrt{2}$, the characteristic impedance of the helium-filled polyester fiber gets undermatched even at low frequencies. The cavity impedance, $\coth(ikt\cos\theta)/\cos\theta$, changes greatly due to the complex refraction angle θ compared with the results for the normal incidence, see Figs. 9(4a) and 9(5a), the magnitude of the cavity impedance decreases significantly, although it is still much larger than that of the characteristic impedance, as shown in Figs. 9(4b) and 9(5b). The final surface impedances shown in Figs. 9(6b) and 9(7b) represent a compromise between the two competing factors, Z_c and $\coth(ikt\cos\theta)/\cos\theta$. For HA, since the large magnitude of $\coth(ikt\cos\theta)/\cos\theta$ is compensated by the small characteristic impedance Z_c, almost perfect impedance match is obtained around the frequency $f_d=0.08$. Accordingly, nearly total absorption is observed in Fig. 9(1a) at this frequency. Alternatively, the impedance matching at oblique incidence can be explained partially by the change in wave propagation direction in the helium-filled porous material. Note that the sound speed in helium gas is about three times as large as that of the air. According to the rule of refraction, when sound wave is incident on the helium-filled absorber at an oblique direction, the penetrated wave is bent toward the air-absorber interface. This results in a longer path of sound propagation and hence more energy dissipation than in the case of air-filled absorber in which the wave is bent toward the normal direction.

As the frequency further increases, the effect of the cavity impedance $\coth(ikt\cos\theta)/\cos\theta$ is reduced and the fact that the characteristic impedance of helium is below that of air takes over. The mismatch of impedance at the air-absorber interface emerges for HA, which leads to a reduction in absorption. This explains the appearance of the peaks in Figs. 8(d) and 9(1b) in the absorption curves for HA.

C. Effect of the protective membrane and duct lining

In practice, an impervious membrane should be installed to separate the helium gas from the ambient air at the air/helium interface. There are two major concerns regarding the use of the protective membrane. The first concern is the gradual loss of helium gas through membrane permeation and the other is the acoustic effect of the membrane on sound penetration. The two concerns are, in fact, related as a thin membrane is required to minimize the negative acoustic impact of the membrane, but an ultrathin membrane will cause serious helium permeation leading to significant loss of helium content and pressure. The much slower infiltration of air into the absorber is negligible in comparison. The question of engineering concern is how long the helium absorber can sustain its nominal performance before the permeation of helium gets so pronounced that its absorption characteristics approach that of the common air-filled sound absorption material. Suppose an aluminum foil of the thickness 2×10^{-5} m is used. The permeation of helium in the porous material can be estimated as follows. The measured permeability coefficient of helium through the aluminum is about $\sigma=1.4\times10^{-21}$ $\mathrm{m^2/Pa\ s}$ (Gerlach *et al.*, 2001). The actual mass permeation rate is proportional to the difference of helium partial pressures at the two sides of the barrier, i.e., membrane, and is inversely proportional to the barrier thickness. The volume flux per unit membrane surface area has the dimension of speed, m/s, which can be understood as the rate of reduction in the thickness of the helium content if the rest is held at the same pressure. For the standard pressure difference of 10^5 Pa, and a membrane thickness of 2×10^{-5} m, the permeation flux is calculated as

$$\dot{q}=\frac{\sigma\Delta p}{\text{thickness}}=\frac{1.4\times10^{-21}\ \mathrm{m^2\ Pa\ s}\times10^5\ \mathrm{Pa}}{0.02\times10^{-3}\ \mathrm{m}}$$
$$=7\times10^{-12}\ \mathrm{m/s}=2.2\times10^{-4}\ \mathrm{m/year}.$$

If the helium-filled absorption layer has a thickness of 5×10^{-2} m, the rate of 2.2×10^{-4} m/year represents a loss of $2.2/500=0.44\%$ of helium. Since the molecule size of oxygen and many other air components is much bigger than that of helium, their permeation rate into the helium absorption layer is much lower than the helium gas. Based on the above analysis, it is reasonable to expect that the helium-filled absorption layer can work for at least 10 years before noticeable changes in acoustic performance may occur.

Another concern for the protective membrane is related to the structural impedance introduced by the membrane. The main effect of a protective membrane layer is its inertia when it is not stretched (Ingard, 1994, p. 1–6; Mechel, 2002, Chap. H.18, Eq. (5); Attenborough and Vér, 2006, p. 272). This is appreciated by looking at the vibration dynamics of the unstretched membrane. If the membrane is located at $y=0$, and sound is incident from $y<0$ at a certain oblique angle, one has (Allard, 1993; Mechel, 2002)

$$m_s i\omega v=p_{\mathrm{air}}-p_{\mathrm{por}}, \tag{22}$$

where m_s, v, p_{air}, and p_{por} are the mass per unit area of the membrane, vibration velocity of the membrane into the porous material, the sound pressures in the air, and porous material sides of the membrane, respectively. The impedance presented to the incident wave, $p_{\mathrm{air}}/v=p_{\mathrm{por}}/v+m_s i\omega$, is the sum of the membrane inertia contribution plus the impedance of the porous material without protection layer, p_{por}/v, which, for the typical case of a layer of finite thickness of t is $p_{\mathrm{por}}/v=(\rho c/\cos\theta)\coth(ikt\cos\theta)$ (Attenborough and Vér 2006, p. 253), where θ is the angle of the sound wave propagating in the sound absorption material. Note that the refracted angle θ is governed by Snell's law, namely, the wavenumber along the interface in the two sides of the air-absorber interface, which is not changed by the presence of the membrane. For the low frequency range considered in this study, the extra structural impedance $m_s i\omega$ reduces the reactance of the total impedance of the absorber. According to Figs. 9(7a) and 9(7b), such changes would lead to lower resonance frequency or lower frequency where absorption is good.

As a measure to evaluate the membrane effect on the absorption performance of the porous material, the transmission loss (TL) of a porous duct lining with and without a protective membrane is measured and compared. The test rig

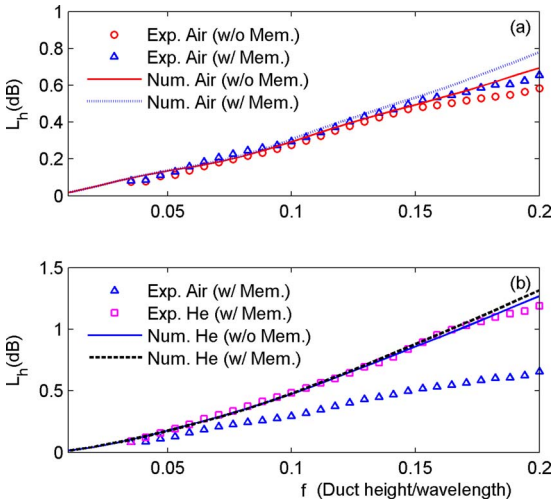


FIG. 11. (Color online) Transmission loss of duct lining. (a) Experimental results for air-filled duct lining covered with membrane (triangle) and without membrane (circle) and theoretical predictions for air-filled duct lining without membrane (solid line) and with membrane (dashed line). (b) Comparison between the experimental results for air-filled (triangle) and helium-filled (square) duct linings covered with membrane, together with comparison between the theoretical predictions for helium-filled duct lining without membrane (solid line) and that with membrane (dashed line).

consists of a rectangular duct with two side-branch cavities flush-mounted on opposite duct walls (Choy and Huang, 2003). The cross section of the duct is 0.2×0.2 m^2, and the two cavities are 5×10^{-2} m in depth and 1 m in length. The polyester fibers with fiber diameter $d=3\times10^{-5}$ m (as investigated in Fig. 5) are filled into the two side-branch cavities to form the so-called duct lining. In order to evaluate the effect of the protective membrane (material: aluminum; thickness: 2×10^{-5} m; aerial density 0.054 kg/m^2), the air-saturated porous material is first tested. Figure 11(a) compares the measured attenuation rate with (open triangles) and without the protective membrane (open circles). The dimensionless frequency f is normalized by the duct height and the speed of sound in air. On the whole, the membrane can improve the low frequency performance, but only slightly. The predicted attenuation rate, L_h, with a finite element model matches the experimental results, especially at low frequencies, where L_h is defined as the reduction in sound dissipation per unit axial distance equal to half of the height of the open part of the lined duct. When helium gas is filled in the side-branch cavities, the protective membrane must be used to separate the helium-filled porous material and the air in the main duct. The open squares in Fig. 11(b) show the measured attenuation rate when helium gas is filled in the side-branch cavities. Again, the measured attenuation rate is in good agreement with the numerical results (dashed line). Compared with the air-filled duct lining (open triangles), the helium-filled duct lining almost doubles the attenuation rate over the calculated frequency range. However, it does not necessarily mean that the helium-filled duct lining outperforms the air-filled duct lining under all design conditions. The particular example shown in Fig. 11 is only to demonstrate the effect of the membrane in a real application. Actually, numerical investigations show that the helium-filled duct lining usually performs better in the low frequency regime, but the air-filled duct linings outperforms in the relatively higher frequency range.

V. CONCLUSIONS

This study is motivated by the desire to extend the effective range of sound absorption by porous material down to lower frequencies, where high acoustic impedance in traditional absorbers prevents sound from ever penetrating into the absorber. The replacement of air by helium gas inside an absorber is shown to have the potential to provide a good low frequency absorber. As a first step in that direction, this study establishes the characteristic of sound propagation in a helium-filled absorption material by experimentation with various fiber configurations. The results are used to predict that a helium-filled absorber of finite thickness would indeed perform well in the low frequency regime. Equivalently, the thickness of such absorber can be reduced for a given absorption coefficient. The conclusion is also confirmed by experimental results in the setting of duct lining. Specific conclusions for this study are as follows.

(1) The speed of sound and complex density, when normalized by those in the pure fluid filling the absorption material, can be expressed as a function of the ratio of fiber dimension, e.g., diameter, to the boundary layer thickness, while the ratio of fiber diameter to the wavelength is less important. The functional relation is the same for absorber filled with air and that filled with helium as far as the porosity and geometrical shape of the fibers are the same. This is experimentally validated.
(2) Helium filled absorption material has a much reduced dimensional characteristic impedance, and so a good impedance match with the impedance of pure air becomes more feasible. This factor allows better sound penetration through the air-absorber interface toward low frequencies.
(3) The performance of an absorber of finite thickness depends also on the sound reflection by the rigid backing and the effective acoustic impedance seen by the incident wave is further increased when the thickness to wavelength ratio is very low. On this account, the performance of a helium-filled absorber is no better, or even worse, than that of ordinary air-filled absorber due to the fact that the wavelength of sound in helium is longer.
(4) When sound is incident at an oblique angle, however, sound is refracted and the comparison between the two absorbers changes in favor of helium, which has a much higher speed of sound. For air-filled absorber, sound penetrating the absorption layer is bent toward the direction perpendicular to the air-absorber interface, reducing the effective distance over which sound can travel in the absorber and get dissipated. For helium-filled absorber, however, the refraction bends the wave toward the air-absorber interface and prolongs the sound travel distance. Such desirable refraction effectively reduces the impedance amplification by the rigid wall backing, and

makes it possible to establish a near-perfect impedance match between the obliquely incident sound in air and the interface impedance of the absorber.

(5) The most important oblique incidence angle is 45° for random incidence. When parameters are chosen suitably, a helium-filled absorber may establish such a near-perfect absorption at certain low frequencies while an air-filled absorber has a monotonous decline of performance with decreasing frequency. The appearance of a spectral peak of absorption coefficient in the spectrum is related to the fact that the characteristic impedance of the helium-filled absorption material can be even lower than the acoustic impedance to be matched for the oblique incidence.

ACKNOWLEDGMENTS

The first author thanks the Hong Kong Polytechnic University for its funding support (Grant No. A-SA43). Likewise, support from a seed fund for applied research at the University of Hong Kong is also acknowledged.

Allard, J. F. (**1993**). *Propagation of Sound in Porous Media Modeling Sound Absorbing Materials* (Elsevier Applied Science, New York).

Allard, J. F., Aknine, A., and Depollier, C. (**1986**). "Acoustical properties of partially reticulated foams with high and medium flow resistance," J. Acoust. Soc. Am. **79**, 1734–1740.

Allard, J. F., and Champoux, Y. (**1992**). "New empirical equations for sound propagation in rigid frame fibrous material," J. Acoust. Soc. Am. **91**, 3346–3353.

Allard, J. F., Depollier, C., Rebillard, P., Lauriks, W., and Corps, A. (**1989**). "Inhomogeneous Biot waves in layered media," J. Appl. Phys. **66**, 2278–2284.

Attenborough, K. (**1982**). "Acoustical characteristics of porous materials," Phys. Rep. **82**, 179–227.

Attenborough, K. (**1983**). "Acoustical characteristics of rigid fibrous absorbents and granular materials," J. Acoust. Soc. Am. **73**, 785–799.

Attenborough, K. (**1987**). "On the acoustic slow wave in air-filled granular materials," J. Acoust. Soc. Am. **81**, 93–102.

Attenborough, K., and Vér, I. L. (**2006**). "Sound-absorbing materials and sound absorbers," in *Noise and Vibration Control Engineering: Principles and Applications*, edited by L. L. Beranek and I. L. Vér (Wiley, New York), Chap. 8.

Blazier, W. E., Jr. (**1997**). "RC Mark II: A refined procedure for rating the noise of heating, ventilating, and air conditioning (HVAC) systems in buildings," Noise Control Eng. J. **45**, 243–250.

Buckingham, E. (**1914**). "On physically similar systems; Illustrations of the use of dimensional equations," Phys. Rev. **4**, 345–376.

Choy, Y. S., and Huang, L. (**2003**). "Drum silencer with shallow cavity filled with helium," J. Acoust. Soc. Am. **114**, 1477–1486.

Delany, M. E., and Bazley, E. N. (**1970**). "Acoustical properties of fibrous materials," Appl. Acoust. **3**, 105–116.

Gerlach, A., Keller, W., Schulz, J., and Schumacher, K. (**2001**). "Gas permeability of adhesives and their application for hermetic packaging of microcomponents," Microsyst. Technol. **7**, 17–22.

Ingard, K. U. (**1994**). *Notes on Sound Absorption Technology* (Noise Control Foundation, New York).

ISO 10534-2, (**1998**). "Determination of sound absorption coefficient and impedance in impedance tubes," International Organization for Standardization.

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

Kirby, R., and Cumming, A. (**1999**). "Prediction of the bulk acoustic properties of fibrous materials at low frequencies," Appl. Acoust. **56**, 101–125.

Lighthill, J. (**1978**). *Waves in Fluids* (Cambridge University Press, Cambridge).

Mechel, F. P. (**1976**). "Model theory for fibrous absorber. Part I: Regular Fibre arrangement; Part II: Model consisting of elementary cells and numerical results," Acustica **36**, 53–89.

Mechel, F. P. (**2002**). *Formulas of Acoustics* (Springer, New York).

Palacio, L., Prádanos, P., Calvo, J. I., and Hernández, A. (**1999**). "Porosity measurements by a gas penetration method and other techniques applied to membrane characterization," Thin Solid Films **348**, 22–29.

Smith, C. D., and Parrot, T. L. (**1983**). "Comparison of three methods for measuring acoustic properties of bulk materials," J. Acoust. Soc. Am. **74**, 1577–1582.

Utsuno, H., Tanaka, T., Fujikawa, T., and Seybert, A. F. (**1989**). "Transfer function method for measuring characteristic impedance and propagation constant of porous materials," J. Acoust. Soc. Am. **86**, 637–643.

Zuckerwar, A. J. (**2002**). *Handbook of the Speed of Sound in Real Gases* (Academic, San Diego, CA).

Zwikker, C. and Kosten, C. W. (**1949**). *Sound Absorbing Materials* (Elsevier, New York).

Comparison of three measurement techniques for the normal absorption coefficient of sound absorbing materials in the free field

Kunikazu Hirosawa,[a] Kazuhiro Takashima, Hiroshi Nakagawa, Makoto Kon, and Aki Yamamoto
System Division, Nittobo Acoustic Engineering Co., Ltd., 1-21-10 Midori, Sumida-ku, Tokyo 130-0021, Japan

Walter Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, BE-3001 Heverlee, Belgium

(Received 23 April 2009; revised 3 September 2009; accepted 7 September 2009)

Three different techniques for evaluating the absorption coefficient of sound absorbing materials in free field conditions are discussed. One technique measures the acoustic impedance at one point nearby a specimen, the other two techniques evaluate the impedance from the transfer function of two sound pressures and two particle velocities at two points. These are called "PU-method," "PP-method," and "UU-method," respectively. An iterative algorithm to estimate the acoustic impedance of the locally reactive specimen in the spherical wave field is also applied. First, the effect of receiver positions, specimen areas, and source heights to the measured normal absorption coefficient is investigated by the boundary element method. According to these investigations, the PU-method is most stable against the effect of specimen area, and the UU-method is easily affected by that effect. Closer source to the specimen distance is advantageous for the signal to noise ratio of these measurement techniques, but correction for the effect of the spherical wave field has to be applied. As a finding, the iterative algorithm works for all of three techniques. Finally, the PU-method is applied experimentally with a pressure-velocity sensor and a loudspeaker in a hemi-anechoic room. As a result, the calculated results have been verified.
 [DOI: 10.1121/1.3242355]

PACS number(s): 43.55.Ev, 43.58.Bh, 43.20.El, 43.20.Rz [NX] Pages: 3020–3027

I. INTRODUCTION

There are several methods to measure the absorption coefficient or the acoustic impedance of acoustical materials. The method with impedance tubes[1] may generally be used for the normal incident absorption coefficient and the acoustic impedance. The method has advantages that a specimen can be small and, especially for the transfer-function method, the measuring time is short. In this method, however, the dimensions of the impedance tubes limit the frequency range, and a specimen needs to be cut carefully from an original absorbing material. Moreover, the measured results of the large material block itself and of the small cut specimen tend to differ when the specimen is a poroelastic material.[2]

Measurement methods in the free field[3–8] have a possibility to overcome the mentioned problem. Since a specimen does not need to be cut out from an original material in the free field methods, it can be used to measure the essential properties of a material regardless of the specimen size. A typical method is called the transfer function method using the sound pressures at two microphones.[5–7] Recently a new PU-probe was developed,[9] it can measure the sound pressure and the particle velocity simultaneously at the same point. Measurement methods with PU-probe have been developed;[10,11] the acoustic impedance at one point nearby the specimen is measured by this method. Some of the free field methods use a plane-wave hypothesis that the field around the receiver positions is assumed to be a plane-wave field. In actual situation, the field around the receivers is better approximated as a spherical wave field, especially at low frequencies. The reflection factor should be calculated by the method taking the spherical wave field into account.[12] However, it cannot be estimated by the direct measurement or calculation because the spherical wave field solution needs the acoustic impedance of the specimen or the reflection coefficient for the plane wave. They are the values to be identified, nevertheless, to estimate the values; the solutions of the spherical wave field model are required. This contradiction can be resolved by an iterative determination algorithm[7,11] which estimates the acoustic impedance of the specimen iteratively. Therefore, the reflection coefficient, the acoustic impedance, or the absorption coefficient of the specimen is estimated by the iterative determination.

The edge effect of a specimen or unwanted reflected waves from wedges of an anechoic room or reflection objects can affect the values measured by these free field methods. To avoid that, it can be expected that a measurement using the particle velocity has robustness against these effects. Be-

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

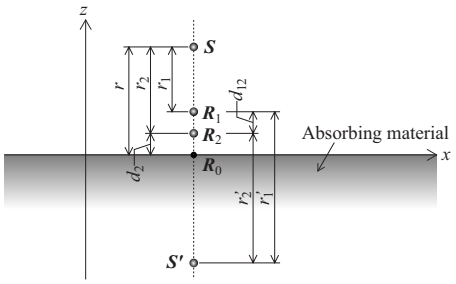


FIG. 1. Illustration of the setup for three measurement techniques. Here, $\boldsymbol{S}$ is the real point source, $\boldsymbol{S}'$ is the imaginary point source, $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are receiver points, and $\boldsymbol{R}_0$ is the intersection of the line $\boldsymbol{R}_1\boldsymbol{R}_2$ with the surface of the absorbing material.

cause the particle velocity is a vector and the normal component for a specimen can be considered to be robust against the effect from the other directions. Therefore, the new measurement technique can be considered to use the particle velocities at two points. In this paper, three measurement techniques in the free field, which are measurements of two sound pressures by two microphones, a sound pressure and a particle velocity by PU-probe, and two particle velocities by two velocity sensors, are compared with regard to *in-situ* measurement technique of the normal absorption coefficient. Here, the measurement using two sound pressures is called "PP-method," the measurement using a sound pressure and a particle velocity is called "PU-method," and the measurement using two particle velocities is called "UU-method." First, the edge effect caused by various sizes of the specimen, the sound source heights, and the receiver positions to the normal absorption coefficient are investigated by the boundary element method (BEM) instead of physical experiments. Then, the PU-method is investigated by some experiments. The iterative determination algorithm based on the method proposed by Allard and Champoux[7] is applied to all results of PP-, PU-, and UU-methods on these investigations.

In this paper, the angular frequency is expressed as ω, time is t, the imaginary unit is $j=\sqrt{-1}$, and all time-varying quantities should obey the time dependence $\exp(-j\omega t)$.

II. MEASUREMENT TECHNIQUES

Figure 1 shows the setup for examining three measurement techniques. r_1 is the distance between $\boldsymbol{S}$ and $\boldsymbol{R}_1$, r_2 is $\boldsymbol{S}$ and $\boldsymbol{R}_2$, r_1' is $\boldsymbol{S}'$ and $\boldsymbol{R}_1$, r_2' is $\boldsymbol{S}'$ and $\boldsymbol{R}_2$, and r is $\boldsymbol{S}$ and $\boldsymbol{R}_0$, respectively. d_{12} is the distance between $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, and d_2 is $\boldsymbol{R}_2$ and $\boldsymbol{R}_0$. The three measurement techniques are assumed that the absorbing material has a locally reacting surface.

A. PU-method

In Fig. 1, $\boldsymbol{R}_2$ is the receiver position for this method. Since the specific acoustic impedance $Z_n(\boldsymbol{R}_2)$ is the ratio of the sound pressure $p(\boldsymbol{R}_2)$ and the $-z$ directional particle velocity $u_n(\boldsymbol{R}_2)$, the plane-wave reflection coefficient R_p is given by[10]

$$R_p = \frac{\dfrac{e^{jkr_2}}{r_2}}{\dfrac{e^{jkr_2'}}{r_2'}} \cdot \frac{Z_n(\boldsymbol{R}_2)\dfrac{1-jkr_2}{jkr_2} - \rho c}{Z_n(\boldsymbol{R}_2)\dfrac{1-jkr_2'}{jkr_2'} + \rho c}, \tag{1}$$

where ρ is the density of air, c is the sound speed in air, and $k=\omega/c$ is the wave number.

B. PP-method

The plane-wave reflection coefficient R_p is given by[7]

$$R_p = \frac{\dfrac{e^{jkr_2}}{r_2} - H(\omega)\dfrac{e^{jkr_1}}{r_1}}{H(\omega)\dfrac{e^{jkr_1'}}{r_1'} - \dfrac{e^{jkr_2'}}{r_2'}}, \tag{2}$$

where $H(\omega)=p(\boldsymbol{R}_2)/p(\boldsymbol{R}_1)$ is the transfer function between the sound pressures at receivers $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$.

C. UU-method

The plane-wave reflection coefficient R_p is given by

$$R_p = \frac{\dfrac{1-jkr_2}{r_2}\dfrac{e^{jkr_2}}{r_2} - H'(\omega)\dfrac{1-jkr_1}{r_1}\dfrac{e^{jkr_1}}{r_1}}{\dfrac{1-jkr_2'}{r_2'}\dfrac{e^{jkr_2'}}{r_2'} - H'(\omega)\dfrac{1-jkr_1'}{r_1'}\dfrac{e^{jkr_1'}}{r_1'}}, \tag{3}$$

where $H'(\omega)=u_n(\boldsymbol{R}_2)/u_n(\boldsymbol{R}_1)$ is the transfer function between the $-z$ directional particle velocities at receivers $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$.

D. Iterative determination algorithm

The plane-wave reflection coefficient R_p in Eqs. (1)–(3) is derived from the assumptions that a distance between a source and a receiver is very long and a field around a receiver is regarded as the plane-wave field approximately. However, these assumptions are not correct especially in the low frequencies, and the reflection coefficient R_p for the plane wave estimated by Eqs. (1)–(3) is not valid. R_p should be estimated for the spherical wave field; nevertheless R_p cannot be estimated by direct measurement or calculation using Eqs. (1)–(3). Because estimating the reflection coefficient for the spherical wave needs the reflection coefficient for the plane wave which is unknown and is the value to be identified. In this paper, this discrepancy is resolved by the iterative determination algorithm based on the method proposed by Allard and Champoux.[7] The method can iteratively obtain the acoustic impedance Z_n of a specimen from the reflection coefficient R_p directly estimated by Eqs. (1)–(3), and then the reflection and absorption coefficients are calculated from Z_n. Ingard's[13] exact solution is applied to the spherical field solution in this algorithm.

1. Ingard's exact solution

Figure 2 shows the geometry of the sound propagation above the infinite plane with the finite specific acoustic impedance Z_n. r_1 is the distance between $\boldsymbol{S}$ and $\boldsymbol{R}_1$, r_1' is the distance between $\boldsymbol{S}'$ and $\boldsymbol{R}_1$, and the angle Θ is defined as the

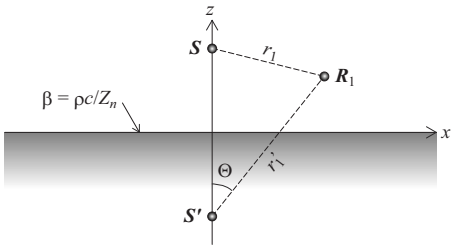


FIG. 2. Geometry of the sound propagation above an infinite plane with a finite acoustic impedance. Here, $\boldsymbol{S}$ is the real point source, $\boldsymbol{S}'$ is the imaginary point source, and $\boldsymbol{R}_1$ is receiver point.

line $\boldsymbol{S}$,$\boldsymbol{R}_1$ and the z axis. $\beta=\rho c/Z_n$ is the normalized specific acoustic admittance of the locally reacting infinite plane. According to Ingard's[13] exact solution, the velocity potential $\varphi(\boldsymbol{R}_1)$ is given by

$$\varphi(\boldsymbol{R}_1)=\frac{e^{jkr_1}}{r_1}+\{R_p+(1-R_p)F\}\frac{e^{jkr_1'}}{r_1'},$$

$$F=1-(\beta+\gamma_0)(kr_1')\times\int_0^\infty\frac{e^{-kr_1's}ds}{\sqrt{(1+\beta\gamma_0+js)^2-(1-\gamma_0^2)(1-\beta^2)}},$$

$$\gamma_0=\cos\Theta,\quad R_p=(\gamma_0-\beta)/(\gamma_0+\beta), \tag{4}$$

where R_p is the plane-wave reflection coefficient of the infinite plane.

The sound pressure $p(\boldsymbol{R}_1)$ and the particle velocity $u(\boldsymbol{R}_1)$ are given by

$$p(\boldsymbol{R}_1)=\rho\frac{\partial\varphi(\boldsymbol{R}_1)}{\partial t}, \tag{5}$$

$$u(\boldsymbol{R}_1)=-\operatorname{grad}\varphi(\boldsymbol{R}_1). \tag{6}$$

Since the normal incidence is considered in this paper, Eq. (4) essentially means the same as the solution for perpendicular incidence according to Nobile and Hayek,[14] and only expressions for both solutions are different when $\gamma_0=1$ in Eq. (4). Champoux and L'Espérance[5] discussed the gradient of the velocity potential developed by Nobile and Hayek that describes the particle velocity in the spherical wave field above the locally reacting impedance plane. The gradient of F in Eq. (4) can be derived from the same way as Champoux and L'Espérance.[5] In this paper, the integrals in Eq. (4) and the gradient of F are computed by the Gauss–Legendre quadrature with discretized integral intervals. Even though these integrations are from zero to infinity, these integrands $f(s)$ can converge at zero with $s\to\infty$. Therefore, the truncated upper bound U of these integrals can be introduced, and it can be estimated with the following condition:

$$\left|\int_0^U f(s)ds-\int_0^{U+\Delta U}f(s)ds\right|\Big/\left|\int_0^U f(s)ds\right|\le C, \tag{7}$$

where ΔU is the discretized integral interval and $C=10^{-13}$ in this paper.

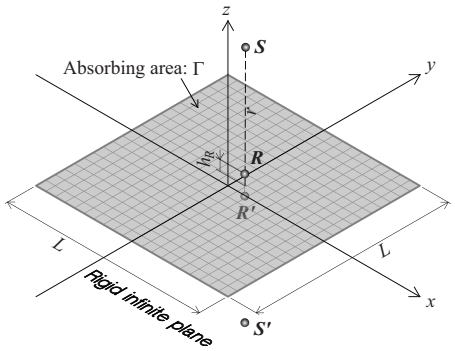


FIG. 3. Illustration of the boundary element analysis model. Here, $\boldsymbol{S}$ is the real point source, $\boldsymbol{S}'$ is the imaginary point source, $\boldsymbol{R}$ is the receiver point, and h_R is the receiver height.

2. Iterative procedure

The iterative determination algorithm is as follows

(1) The normal acoustic impedance $Z_n^{(0)}$ $(i=0)$ at $\boldsymbol{R}_0$ in Fig. 1 is estimated by the following equation substituting R_p given by Eqs. (1)–(3):

$$Z_n^{(i)}=\frac{\rho c(1+R_p)}{(1-R_p)(1-1/jkr)}. \tag{8}$$

Since $Z_n^{(0)}$ is obtained from the measured values, it is used as the reference value in this algorithm. i is the iterative index.

(2) $Z_n^{(0)}$ is set as the initial value of the desired normal specific acoustic impedance Z_n of the absorbing material surface, viz., $Z_n=Z_n^{(0)}$.

(3) Z_n is substituted for Eq. (4), and the sound pressures and the particle velocities at $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are calculated through Eqs. (5) and (6).

(4) In order to estimate R_p, $Z_n(\boldsymbol{R}_2)=p(\boldsymbol{R}_2)/u(\boldsymbol{R}_2)$ in PU-method is substituted for Eq. (1), $H(\omega)=p(\boldsymbol{R}_2)/p(\boldsymbol{R}_1)$ in PP-method is substituted for Eq. (2), or $H'(\omega)=u_n(\boldsymbol{R}_2)/u_n(\boldsymbol{R}_1)$ in UU-method is substituted for Eq. (3).

(5) $Z_n^{(i)}$ is calculated by Eq. (8) substituting R_p.

(6) It is judged that whether $Z_n^{(i)}$ is equal to $Z_n^{(0)}$ or not. If $Z_n^{(i)}\neq Z_n^{(0)}$, go to (7). Otherwise, go to (8).

(7) Z_n is corrected by the following equation:

$$Z_n=Z_n+Z_n^{(i)}-Z_n^{(0)}. \tag{9}$$

Returning to (3), and the updated Z_n is substituted for Eq. (4) again. The above procedure is repeated until $Z_n^{(i+1)}$ is equal to $Z_n^{(0)}$.

(8) The plane-wave reflection coefficient and the absorption coefficient of the absorbing material are estimated from Z_n determined by this algorithm.

In this paper, the convergence condition of (6) in the iterative procedure is that the relative error between $Z_n^{(i)}$ and $Z_n^{(0)}$ is less than 10^{-6}.

III. NUMERICAL INVESTIGATION WITH BEM

A. Analysis model

Figure 3 shows three-dimensional hemi-free field; x-y plane is the rigid infinite plane. There is the absorbing area Γ

which has L (m) $\times L$ (m) area, and the center of the absorbing area is set to the origin. The distance between $\boldsymbol{S}$ and Γ is r (m), and the height of $\boldsymbol{R}$ from Γ is h_R (m).

The mirror image method is applied to the analysis model, and Γ is modeled as a locally reacting plane with the normal specific acoustic impedance Z_n. The Helmholtz–Huygens integral of the sound pressure $p(\boldsymbol{R})$, which describes the sound field, can be written as follows:[15,16]

$$p(\boldsymbol{R}) = G(\boldsymbol{S},\boldsymbol{R}) + G(\boldsymbol{S}',\boldsymbol{R}) + \frac{2jk\rho c}{Z_n}\int\int_{\Gamma} p(\boldsymbol{r}_\Gamma)G(\boldsymbol{r}_\Gamma,\boldsymbol{R})d\Gamma, \quad (10)$$

where G is the Green's function in the three-dimensional free field, and $\boldsymbol{r}_\Gamma$ is an arbitrary point on Γ. When $\boldsymbol{R}$ is on Γ, Eq. (10) becomes the Fredholm integral equation of the second kind which has unknown values $p(\boldsymbol{r}_\Gamma)$. The solutions are obtained by solving the linear simultaneous equations after discretization of Γ into boundary elements. The boundary element is quadrilateral constant element and the maximum element length is less than 1/8 wavelength at the frequency.

In this paper, the absorbing area Γ is assumed to be a glass wool with thickness of 25 mm (flow resistivity of 55 000 N s/m^4), and Z_n of Γ is estimated by the empirical equation proposed by Allard and Champoux.[17]

The whole edge effect should be considered with this formulation of the BEM in frequency domain.

B. Effect of receiver positions

To investigate the normal absorption coefficient affected by the effect of receiver positions, the sound pressures and the particle velocities at four receiver points $\boldsymbol{R}_2^{(1)}=(0,0,0.01)$, $\boldsymbol{R}_2^{(2)}=(0.1,0,0.01)$, $\boldsymbol{R}_2^{(3)}=(0.1,0.1,0.01)$, and $\boldsymbol{R}_2^{(4)}=(0.1,0.05,0.01)$ with $d_{12}=0.01$ m are calculated by the BEM. The distance r is fixed at 1.0 m, and the absorbing area Γ is 1.0×1.0 m^2.

Figure 4(a) shows the normal absorption coefficient estimated by the PU-method, Fig. 4(b) shows that by the PP-method, and Fig. 4(c) shows that by the UU-method, respectively.

The legend "TMM" in these figures describes the normal absorption coefficient calculated by the transfer matrix method[18] (TMM) using the material characteristics. The propagation constant and the characteristic impedance in the TMM are also estimated by the empirical equation of Allard and Champoux.[17] The situation supposed in the TMM is that a plane wave propagates into an absorbing material with infinite area, and the edge effect is not considered. Therefore, the normal absorption coefficient calculated by the TMM is the reference value on the numerical investigation in this paper.

From these results, differences of the normal absorption coefficient caused by different receiver positions are found. The differences can be caused by the diffracted waves from the edges of the absorbing area Γ, and especially $\boldsymbol{R}^{(1)}$ is the center of Γ, and the diffracted waves equally affect the sound pressure and the particle velocity at $\boldsymbol{R}^{(1)}$ in phase. In contrast, the effect of the diffracted waves at $\boldsymbol{R}^{(4)}$ is dispersed as a result of the different distances between the edges and $\boldsymbol{R}^{(4)}$, and these figures also indicate that the values at $\boldsymbol{R}^{(4)}$ are most stable. Therefore, the receiver positions of these measurement techniques should not be set to the center. In what follows, the normal absorption coefficient at $\boldsymbol{R}^{(4)}$ is investigated.

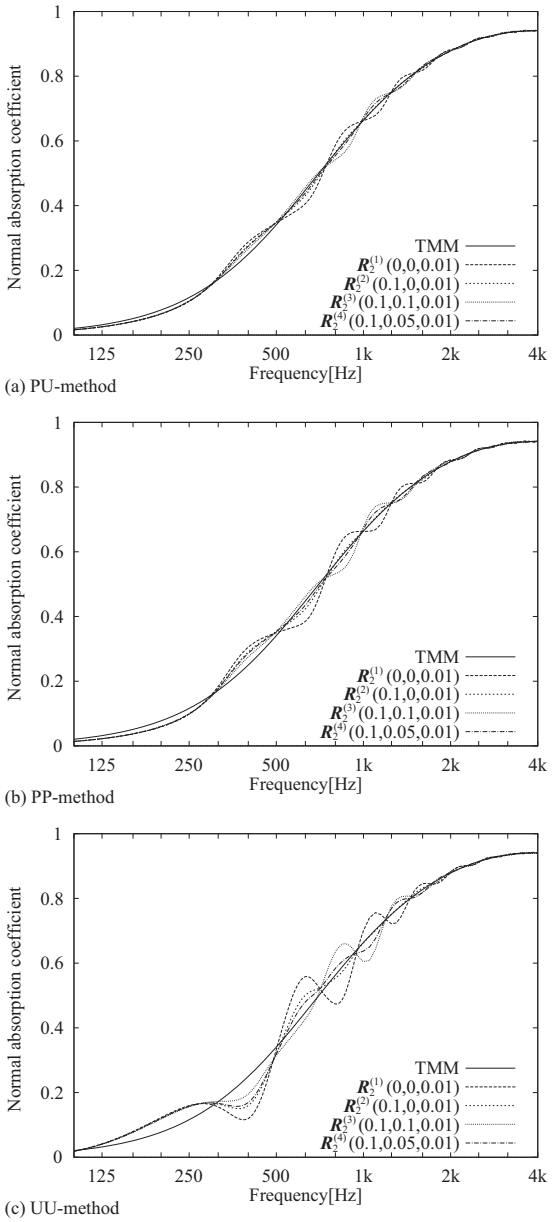
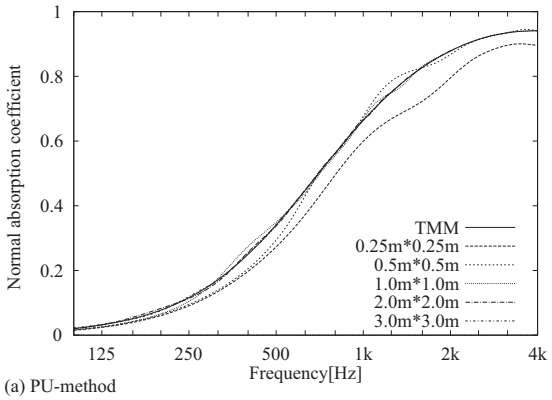


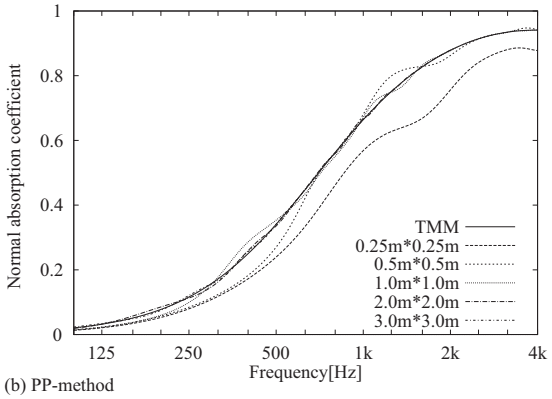
FIG. 4. Comparison of the normal absorption coefficients calculated by the BEM: r is 1.0 m and Γ is 1.0×1.0 m^2. Effect of receiver positions: (a) is estimated by the PU-method, (b) by the PP-method, and (c) by the UU-method.

C. Effect of absorbing areas

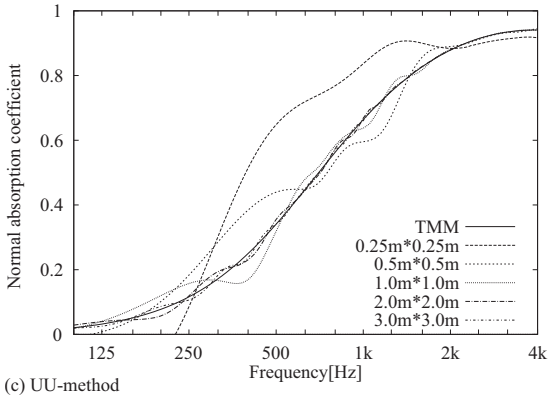
The normal absorption coefficient estimated by the PU-method is shown in Fig. 5(a), that by the PP-method in Fig. 5(b), and that by the UU-method in Fig. 5(c), respectively. Compared absorbing areas are 0.25×0.25, 0.5×0.5, 1.0×1.0, 2.0×2.0, and 3.0×3.0 m^2. Positions are $\boldsymbol{S}=(0.1,0.05,1.0)$, $\boldsymbol{R}_1=(0.1,0.05,0.02)$, and $\boldsymbol{R}_2=(0.1,\ 0.05,\ 0.01)$, and the distance r is 1.0 m.



(a) PU-method

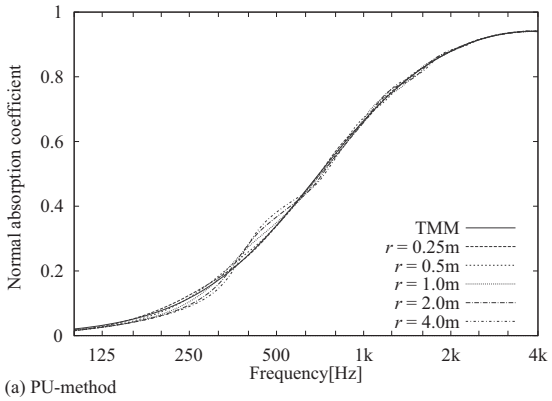


(b) PP-method

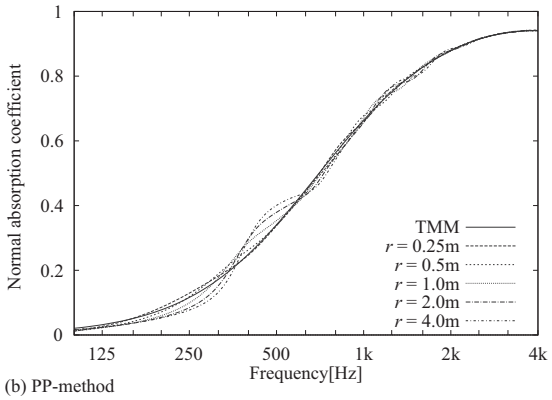


(c) UU-method

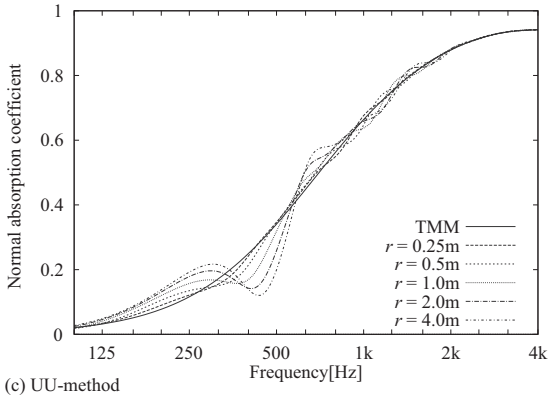
FIG. 5. Comparison of the normal absorption coefficients calculated by the BEM: r is 1.0 m. Effect of absorbing areas: (a) is estimated by the PU-method, (b) by the PP-method, and (c) by the UU-method.



(a) PU-method



(b) PP-method



(c) UU-method

FIG. 6. Comparison of the normal absorption coefficients calculated by the BEM: Γ is 1.0×1.0 m^2. Effect of source heights: (a) is estimated by the PU-method, (b) by the PP-method, and (c) by the UU-method.

The frequency characteristic of the normal absorption coefficient of the PU-method is very similar to that of the PP-method; however, that of the PP-method is slightly more deviated than that of the PU-method. On the other hand, that of the UU-method differs from those of the PU- and the PP-methods, and it is more deviated than the others. The normal absorption coefficient of the three techniques tends to converge at that of the TMM when the area becomes larger. Therefore, it is found that the PU-method is most robust against the effect of absorbing areas in the three measurement techniques. In addition, the area has to be larger than 1.0×1.0 m^2 in the case of the PU-method.

D. Effect of source heights

The normal absorption coefficient estimated by the methods of the PU, the PP, and the UU is compared on the different distances of the source height r=0.25, 0.5, 1.0, 2.0, and 4.0 m. Figure 6(a) shows the normal absorption coefficient estimated by the PU-method, Fig. 6(b) shows that by the PP-method, and Fig. 6(c) shows that by the UU-method. The positions are $\boldsymbol{S}=(0.1,0.05,r)$, $\boldsymbol{R}_1=(0.1,0.05,0.02)$, and $\boldsymbol{R}_2=(0.1,0.05,0.01)$, and the absorbing area is 1.0×1.0 m^2.

In all three measurement techniques, if $\boldsymbol{S}$ is closer to the sample surface Γ, the normal absorption coefficient of the finite absorbing area converges to that of infinite absorbing area described by the TMM. The reason for this phenomenon is the diffraction effect on the edges of the absorbing area. This effect is relatively larger when the source is higher, and it indicates that a closer source to the specimen is advantageous. Comparing three measurement techniques, the PU-method is also most stable against the effect of the absorbing area.

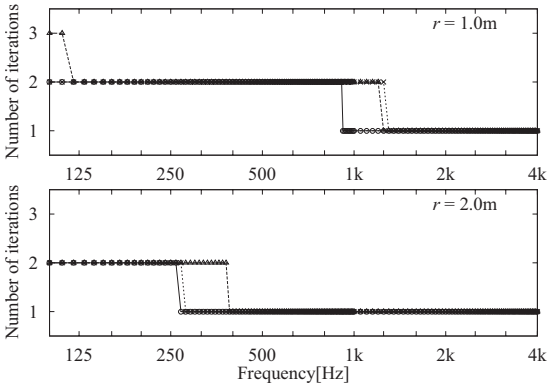


FIG. 7. Comparison on the number of iterations for the iterative determination algorithm calculated by the BEM: Γ is 1.0×1.0 m^2. —○—: the PU-method, ---△---: the PP-method, and ···×···: the UU-method.

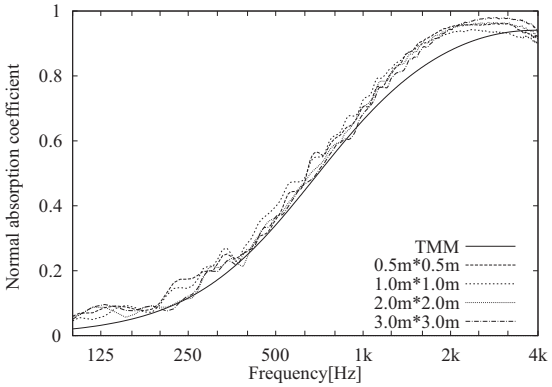


FIG. 9. Comparison of the normal absorption coefficients measured by the PU-method for different absorbing areas: $r=1$ m.

E. Number of iterations

The number of iterations is shown in Fig. 7 with respect to comparison of three measurement techniques. The upper side of the figure shows the number under the condition of $r=1.0$ m, the lower side of the figure shows the number under $r=2.0$ m, and the absorbing area is 1.0×1.0 m^2.

Since the longer distance r, the sound field around the receivers is closer to the plane-wave field, the number of iterations on $r=2.0$ m is less than those of $r=1.0$ m. The number for the PU-method is less than the others. On the other hand, the number for the PP-method in low frequency is more than the others in the case of $r=1.0$ m. Therefore, the convergence of the PU-method is faster than the others.

IV. EXPERIMENTAL INVESTIGATION OF THE PU-METHOD

A. Measurement procedure

The measurements described in the following were performed in a hemi-anechoic room. The dimensions of this hemi-anechoic room are the width of 4.8 m, the depth of 6.2 m, and the height of 3.6 m, and the cutoff frequency of the absorbing wedge is around 125 Hz. The sound was radiated by the full range loudspeaker with the diameter 13 cm (DIATONE DS-5). For improvement of signal to noise (S/N) ratio in low frequency, the sound source signal was used as the logarithmic swept sine signal,[19] and the impulse responses of the sound pressure and the particle velocity at PU-probe were measured. The raised cosine time window was applied to the impulse responses for elimination of unwanted reflected waves. Generating the sound signal, measuring the impulse responses, transforming into the frequency domain by fast Fourier transform, and postprocessing were processed by the personal computer.

The calibration of PU-probe was made by the free field calibration method.[10] It is shown that the acoustic impedance measured by PU-probe is corrected by the theoretical acoustic impedance in the free field. The measured normal absorption coefficient was smoothed by a moving average.

The measured porous material was a glass wool board with the thickness of 25 mm, the density of 98.5 kg/m^3, and the flow resistivity of 55 124 N s/m^4. This flow resistivity was measured by the method according to ISO 4638.[20]

B. Measurement results

Figure 8 shows comparison of the normal absorption coefficient calculated by the TMM, measured by the impedance tube with the diameter of 40 mm, and measured by the PU-method with $r=1$ m. Figure 10 shows the same comparison, however, for the PU-method with $r=2$ m. The infinite absorbing area for the PU-method was realized by spreading the material all over the floor in the hemi-anechoic room. The measured result by the PU-method corresponds to the measured value by the impedance tube and the calculated

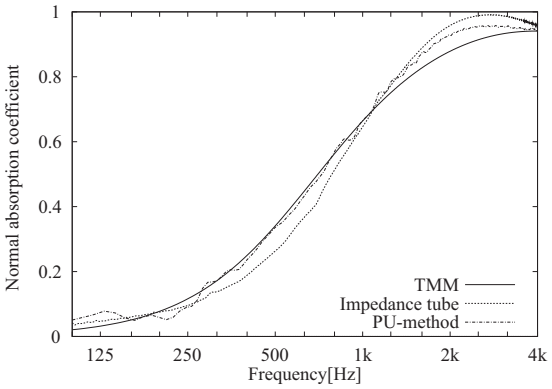


FIG. 8. Comparison of the normal absorption coefficients. The condition of "PU-method" is that $r=1$ m and the absorbing area is infinite.

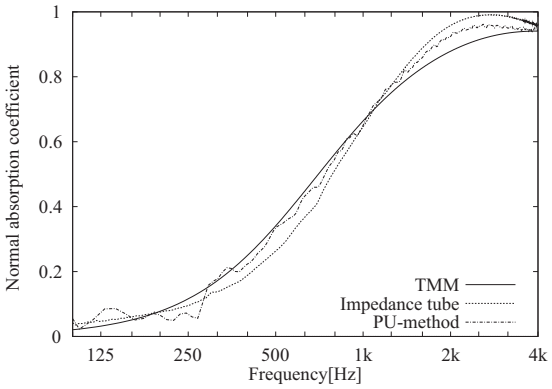


FIG. 10. Comparison of the normal absorption coefficients. The condition of "PU-method" is that $r=2$ m and the absorbing area is infinite.

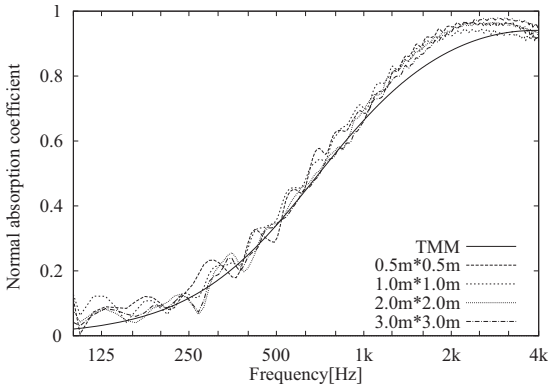


FIG. 11. Comparison of the normal absorption coefficients measured by the PU-method: $r=2$ m for different absorbing areas.

value by the TMM approximately in both cases of $r=1$ and 2 m. In other words, the TMM agrees with the measured values, and this shows that the numerical calculation has been verified by these results.

Figure 9 shows the normal absorption coefficient with different absorbing areas when $r=1$ m, and Fig. 11 shows that when $r=2$ m. As the BEM analysis of Fig. 5 points out, these figures indicate that the normal absorption coefficient measured by the PU-method is hardly changed by the different absorbing areas. The deviation of the measured normal absorption coefficient at frequencies lower than 300 Hz can be found, and the same deviation can also be found in the values calculated by the BEM. This deviation can be caused by the effects of absorbing areas and unwanted reflected waves. Since the absorption coefficient of a fibrous material at low frequencies is a small value, the normal component of the particle velocity close to the material becomes very small at those frequencies. Therefore, the effect of the diffracted waves from the material edges and the unwanted reflected waves to the receiver can be relatively bigger than that of the direct sound radiated from the sound source. Especially on this experiment, since the cutoff frequency of the absorbing wedges in this hemi-anechoic room is around 125 Hz, the reflected waves from the wedges may affect the measured values at low frequencies. Nevertheless, a larger absorbing area gives more stability to the normal absorption coefficient, and the value becomes closer to the TMM, because the effect of absorbing areas becomes small.

The deviation of the normal absorption coefficient with $r=1$ m is smaller than that with $r=2$ m, and this indicates that more stable measured values can be obtained by having the sound source closer to the specimen. In this study, the sound source, the loudspeaker with the diameter of 13 cm has been used, and it is not a real point source, even though the iterative determination algorithm premises the point source. If the loudspeaker is set close to the specimen, the measurement may lose the validity of the assumption. Therefore, it is important that the sound source regarded as the point source is used and the radiated sound is big enough to keep sufficient S/N ratio in low frequency.

V. CONCLUSIONS

In this study, three measurement techniques in the free field, which are the PU-method, the PP-method, and the UU-method, have been compared and discussed with the numerical analysis by the BEM and the experiment. The specimen used was a 25 mm glass wool and a locally reacting surface has been assumed, which is a reasonable assumption, given the high flow resistivity of the sample. As a result, following outcomes have been found.

- The PU-method is most stable against the effect of the absorbing area, and the UU-method is easily affected by that effect contrary to the original expectation.
- The major cause for variation in the normal absorption coefficient is the source height rather than the absorbing area. Because the source is higher, the direct source to receiver signal is relatively weaker compared to the effect from the diffraction of the edges. Therefore, a source closer to the specimen is advantageous for these measurement methods.
- The iterative determination algorithm works for the PU- and the UU-methods as well as the PP-method.
- The number of iteration of the PU-method is smaller than that of the PP- and the UU-methods.

In this study, since the loudspeaker with the diameter of 13 cm has been used for radiation of the sound in low frequency around 100 Hz, the loudspeaker could not be set very close to the specimen. If the loudspeaker realizing the ideal point source can be used, especially on the PU-method, more accurate measurement can be expected.

[1]ISO 10534, "Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 1 and 2," (1998).
[2]T. E. Vigran, L. Kelders, W. Lauriks, P. Leclaire, and T. F. Johansen, "Prediction and measurements of the influence of boundary conditions in a standing wave tube," Acta. Acust. Acust. **83**, 419–423 (1997).
[3]U. Ingard and R. H. Bolt, "A free field method of measuring the absorption coefficient of acoustic materials," J. Acoust. Soc. Am. **23**, 509–516 (1951).
[4]T. M. Barry, "Measurement of the absorption spectrum using correlation/spectral techniques," J. Acoust. Soc. Am. **55**, 1349–1351 (1974).
[5]Y. Champoux and A. L'Espérance, "Numerical evaluation of errors associated with the measurement of acoustic impedance in a free field using two microphones and a spectrum analyzer," J. Acoust. Soc. Am. **84**, 30–38 (1988).
[6]J. F. Allard, Y. Champoux, and J. Nicolas, "Pressure variation above a layer of absorbing material and impedance measurement at oblique incident and low frequencies," J. Acoust. Soc. Am. **86**, 766–770 (1989).
[7]J. F. Allard and Y. Champoux, "In situ two-microphone technique for the measurement of the acoustic surface impedance of materials," Noise Control Eng. J. **32**, 15–23 (1989).
[8]J.-F. Li and M. Hodgson, "Use of pseudo-random sequences and a single microphone to measure surface impedance at oblique incidence," J. Acoust. Soc. Am. **102**, 2200–2210 (1997).
[9]H. de Bree, "An overview of microflown technologies," Acta. Acust. Acust. **89**, 163–172 (2003).
[10]R. Lanoye, G. Vermeir, W. Lauriks, R. Kruse, and V. Mellert, "Measuring the free field acoustic impedance and absorption coefficient of sound absorbing materials with a combined particle velocity-pressure sensor," J. Acoust. Soc. Am. **119**, 2826–2831 (2006).
[11]J. D. Alvarez and F. Jacobsen, "An iterative method for determining the surface impedance of acoustic materials in situ," in Proceedings of the Inter-Noise 2008 (2008).
[12]K. Hirosawa, H. Nakagawa, M. Kon, and A. Yamamoto, "Investigation of absorption coefficient measurement of acoustical materials by boundary element method using particle velocity and sound pressure sensor in free field," Acoust. Sci. & Tech. **30**(6), 442–445 (2009).

[13]U. Ingard, "On the reflection of a spherical sound wave from an infinite plane," J. Acoust. Soc. Am. **23**, 329–335 (1951).
[14]M. A. Nobile and S. I. Hayek, "Acoustic propagation over an impedance plane," J. Acoust. Soc. Am. **78**, 1325–1336 (1985).
[15]T. Terai, "On calculation of sound fields around three dimensional objects by integral equation methods," J. Sound Vib. **69**, 71–100 (1980).
[16]Y. Kawai and H. Meotoiwa, "Estimation of the area effect of sound absorbent surfaces by using a boundary integral equation," Acoust. Sci. & Tech. **26**, 123–127 (2005).
[17]J. F. Allard and Y. Champoux, "New empirical equations for sound propagation in rigid frame fibrous materials," J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[18]B. Brouard, D. Lafarge, and J. F. Allard, "A general method of modelling sound propagation in layered media," J. Sound Vib. **183**, 129–142 (1995).
[19]T. Fujimoto, "A study of tsp signal getting higher sn ratio at low frequency bands," in Proceedings of the 1999 Autumn Meeting of Acoustical Society of Japan, pp. 433–434, in Japanese.
[20]ISO 4638, "Polymeric materials, cellular flexible—Determination of air flow permeability," (1984).

Transmission characteristics of a tee-junction in a rectangular duct at higher-order modes

Siu-Kit Lau[a)]
Architectural Engineering Program, University of Nebraska-Lincoln, 1110 South 67th Street, Omaha, Nebraska 68182-0681

Kwan-Hao Leung[b)]
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, People's Republic of China

(Received 20 February 2009; revised 1 September 2009; accepted 29 September 2009)

Numerical and experimental studies were undertaken to characterize the noise transmission and scattering properties in higher-order modes across the tee-junction of a rectangular duct used in ventilation and air-conditioning systems. To measure these properties, a formulation of a transmission matrix based on the transfer function and a two-microphone method was devised. The measurement of modal sound transmission and scattering coefficients is demonstrated for a duct element in a rectangular duct. The results of numerical simulations were verified by experiments. The results show that sound transmissions of fundamental mode and higher-order modes across the main duct are high at the eigen-frequencies of the main duct and sidebranch. Weak modal coupling of the branch-modes and the traveling wave in the main duct is observed at or very close to the eigen-frequencies of the sidebranch, which shifts excitation of the higher-order branch-modes at higher frequencies. A decrease in sound transmission and increase in sound scattering into higher-order modes occur with excitation of the axial branch-mode. Excitation of the longitudinal branch-mode due to branch-end reflection also results in lower sound transmission of higher-order modes across the junction along the main duct.
 [DOI: 10.1121/1.3257209]

PACS number(s): 43.55.Rg, 43.20.Mv [SFW] Pages: 3028–3039

I. INTRODUCTION

Sound propagation in duct systems is well known as an important problem in building, automotive, and aeronautic noise control. Duct systems consist of a network of various elements (e.g., bends and constrictions) in addition to straight sections. Such elements of duct systems (other than straight ducts) alter the acoustic impedance and thus causes the reflection of sound in the system.

In recent decades, the sound-transmission characteristics of some duct elements have been investigated at low frequency, including bends,[1] sidebranches,[2,3] constrictions,[4] and expansions;[5] however, textbooks on engineering design (e.g., Refs. 6 and 7) typically contain few predictions of sound-transmission loss at very low frequency for these conventional passive duct elements. Inaccurate estimates of sound-transmission loss arise when the frequency is close to and higher than the first cut-on frequency of the duct. The effective control of such noise in ducts would require a comprehensive characterization and specific measurement procedure for sound transmission and scattering across duct elements in higher-order modes.

Sidebranches are one of the most common elements in duct networks, commonly used for flow separation when coupled with straight ducts and other elements. Tang and Li[2] undertook theoretical and experimental studies of the sound-transmission loss of double sidebranches using plane-wave theory and the two-microphone method, respectively. The authors found that predictions based on plane-wave theory are only accurate when the sidebranch is distant from other duct elements or when the original sound-transmission loss of other duct elements is weak. Moreover, the sound excitation of axial branch-modes yields lower sound transmission across a tee-junction located along the main duct.[3] However, the studies are limited to the sound-transmission fundamental mode below the first eigen-frequency of the main duct.

For measurement of the sound reflection and transmission coefficients of a duct element, Parrondo *et al.*[8] extended the two-microphone method for measurement of the transmission and reflection matrix of a sidebranch, using three loudspeakers and six sound-pressure microphones. The investigations of the sound transmission of a sidebranch are again limited to analyses of propagation of fundamental mode along the main duct. In addition, use of the two-microphone method in experimental studies is restricted to low frequencies (i.e., below the first eigen-frequency of the duct) because of the fundamental-mode (plane wave) assumption; however, sound propagation at higher-order modes is also important for noise control, especially in the case of large ducts. For example, the typical dimension (i.e., width and/or height) of air-conditioning ducts is up to about 3 m,[9] which corresponds to a low first cut-on frequency of

[a)]Author to whom correspondence should be addressed. Electronic mail: slau3@unl.edu
[b)]Present address: Ove Arup & Partners Hong Kong Ltd., L5 Festival Walk, 80 Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong.

the duct (57 Hz). In such cases, sound propagation in higher-order modes becomes of greater practical importance in terms of noise control.

A theoretical solution for a tee-junction has been formulated using the Fourier transform and a mode matching technique;[10,11] however, there exist few case studies of such problems. The interaction of acoustic modes and the transmission and scattering of sound across a tee-junction in the duct has yet to be clearly addressed. It is therefore worthwhile to study the sound-transmission characteristics of a tee-junction and to conduct a more in-depth analysis of the problem for higher-order modes.

In the measurement of sound propagation in higher-order duct modes, the assumptions involved in the two-microphone method are invalid and the complexity of the measurement is increased compared with that for fundamental mode. A modal decomposition method has been suggested to overcome these problems.[12] Abom[13] proposed a direct method based on the measurement of transfer functions between the microphone pairs in order to separate the modes in a duct section into incident and reflected parts. To perform a modal decomposition of N number of modes requires the incident and reflected wave of each mode to be determined, as well as measurement of the sound pressures at $2\times N$ independent spatial locations with N independent source conditions. For circular ducts, this method can extend the measurement to frequencies below the first cut-on frequency of the radial mode when ten measurement positions are used (i.e., separation of the first five modes). A reduction in the number of measurement locations can be offset by measurements undertaken with different load conditions[14] at sites located downstream of the duct element. Previous works have generally considered circular ducts. Schultz *et al.*[15] determined the reflection and modal scattering coefficients of acoustic liners using the direct method in a square duct; however, few studies have applied the modal decomposition method to rectangular ducts and their elements.[16]

The aim of the present paper is to investigate and characterize sound propagation in higher-order modes across a tee-junction within a rectangular duct, using numerical (finite element method) and experimental (direct modal decomposition method) approaches. The resultant transmission matrix of a tee-junction is experimentally verified using the modal decomposition technique. We demonstrate a measurement procedure for sound transmission and the scattering coefficients of rectangular duct elements, which can be applied under the propagation conditions of higher-order modes. Also discussed are practical difficulties involved in measuring the transmission matrix of duct elements. The present study extends the work of Tang[3] to higher-order modes. It is hoped that the results will provide a more complete picture of the acoustical properties of tee-junctions (particularly for the air duct used in the ventilation and air-conditioning), thereby providing improved duct noise control and a useful experimental approach for investigating sound transmission and scattering within duct elements at higher-order modes.

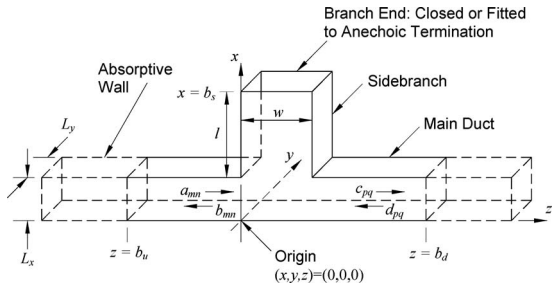


FIG. 1. Schematic diagram of sound transmission within a duct across a tee-junction, showing the employed coordinate system.

II. COMPUTATIONAL MODEL AND EXPERIMENTAL METHOD

Numerical and experimental studies were undertaken to characterize the sound transmission and scattering properties in higher-order modes across the tee-junction of a rectangular duct commonly used in ventilation and air-conditioning system. The methodologies are discussed in Secs. II A and II B, respectively.

A. Computational domain

Consider a tee-junction in a rectangular duct consisting of acoustically hard surfaces (see Fig. 1). All dimensions are normalized by the duct width L_x. The origin of the coordinate system is set at the left-side entry to the tee-junction. The width w and length l of the sidebranch can be varied relative to L_x. For simplicity, a two-dimensional system is considered in the numerical experiment. Axial modes in the axis along the sidebranch are considered. Some effects of spanwise modes along the y-axis of a three-dimensional system will be discussed later in the experimental investigation. A finite element computation is used to solve the inhomogeneous wave equation, given as

$$\nabla^2\hat{p} + k^2\hat{p} = \hat{q}, \tag{1}$$

where $\hat{p}$ and k are the acoustic pressure and the wavenumber of the sound, respectively; $\hat{q}$ denotes the source strength density of the excitation; and ∇^2 is the Laplacian operator over a specified domain with Neumann boundary conditions. The partial differential equation (PDE) solver within the software MATLAB was used to solve Eq. (1) in the present numerical experiments. The "assempde" of MATLAB function is used. The side walls of the duct and the sidebranch are acoustically rigid; therefore, the normal pressure gradient vanishes and the boundary condition is

$$\left.\frac{\partial\hat{p}}{\partial\hat{n}}\right|_{\text{wall}} = 0, \tag{2}$$

where $\hat{n}$ is the outward unit normal to the boundary. The main duct consists of two anechoic terminations at $z<-10L_x$ and $z>(10L_x+w)$. The higher-order modes are present in the main duct and sidebranch at frequencies above the first eigen-frequencies of the main duct and sidebranch, respectively. To damp down the higher-order modes before they reach the end boundaries, we adopt the impedance-matched anechoic termination of Tang and Lau,[4] whose

walls are acoustically absorptive. The boundary conditions of the absorptive side walls are given as

$$\frac{\partial \hat{p}}{\partial \hat{n}} \pm jk\gamma\hat{p} = 0, \tag{3}$$

where $j=\sqrt{-1}$. γ denotes an artificial absorption coefficient that is set at $0.01(z+b_u)^2$ and $0.01(z-b_d)^2$ for the upstream and downstream absorptive endings of the main duct, respectively, as in Ref. 3. $b_u=-10L_x$ and $b_d=(10L_x+w)$. A termination length greater than $10L_x$ is used in the analysis. The boundary condition at the exit of an anechoic termination is set as described in Eq. (3), with $\gamma=1$ for an outgoing wave. Two types of sidebranch endings are considered: infinitely long and closed sidebranches. They are tested for two situations of weak and strong branch-end reflections. For an infinitely long sidebranch, the anechoic termination is set at $x>b_s$ with $\gamma=0.01(x-b_s)^2$, where $b_s=L_x+l$. Although the acoustic properties are independent of sidebranch length in the case of an infinitely long sidebranch, the length is set to $l=5L_x$. For a closed sidebranch, the boundary condition described in Eq. (2) is set at $x=b_s$, $\forall z\in[0,w]$.

Upstream of the duct ($z<0$), an incident field a_{mn} propagates toward the tee-junction, as shown in Fig. 1. In the present study, the excitation of this incident field is applied at $z=b_u$. Sound is investigated at frequencies below the first symmetric duct mode ($m=2$ and $p=2$) of the x-axis in the main duct (i.e., $kL_x=2\pi$). For investigations of the higher-order-mode propagation, two types of excitation are considered:

$$\hat{q} = 2jk\delta(z-b_u) \quad \text{Excitation Type 1: } \forall\, x\in[0,L_x], \tag{4a}$$

$$\hat{q} = 2jk\delta(z-b_u) \quad \text{Excitation Type 2: } \forall\, x\in[0,0.5L_x], \tag{4b}$$

where δ denotes a delta function. The former (Excitation Type 1) refers to fundamental-mode (unit plane wave) excitation, while the latter (Excitation Type 2) is asymmetric wave excitation and the source is bounded at $x\in[0,0.5L_x]$. The modal amplitudes arising from Type 2 excitations can be found using the wave equation.[16] Both the fundamental mode and first mode are excited by asymmetric wave excitation in Eq. (4b) at frequencies of $\pi\leq kL_x<2\pi$. The resultant sound-pressure field of the fundamental mode and higher-order modes at either side of the junction and without the junction can be decomposed by the method of modal decomposition.[17] The resultant sound-pressure fields due to excitation of the first mode only can be found by subtracting the contribution of the unit plane-wave excitation with the scattering coefficients, as calculated using Eq. (4a).

B. Experimental methodology

At any location (x,y,z) within the duct upstream of the junction ($b_u<z<0$) in Fig. 1, the general solution of acoustic pressure to the inhomogeneous Helmholtz equation described in Eq. (1) in a three-dimensional system at driving frequency ω and time t can be written as

$$\hat{p}(x,y,z)e^{j\omega t} = \sum_{m,n=0}^{\infty} \psi_{m,n}(x,y)[a_{mn}e^{j(\omega t-k_{z,mn}z)} + b_{mn}e^{j(\omega t+k_{z,mn}z)}], \tag{5}$$

where $\psi_{m,n}(x,y)$ are the eigen-functions of the acoustic modal pressure distribution inside the duct. The integers m and n are the acoustic duct mode numbers of the x-axis and y-axis, respectively, upstream of the junction in the main duct, which $(m,n)=(0,0)$ denotes fundamental mode and others denote higher-order modes. The duct has a rectangular cross-section of dimension L_x by L_y. $k_{z,mn}$ is the wavenumber of the duct mode in the axial direction,[18] where

$$k_{z,mn}^2 = k^2 - k_{mn}^2, \tag{6}$$

and

$$k_{mn}^2 = \left(\frac{m\pi}{L_x}\right)^2 + \left(\frac{n\pi}{L_y}\right)^2. \tag{7}$$

a_{mn} and b_{mn} are the complex acoustic pressure modal amplitudes of the traveling waves of (m,n) mode toward the junction (incident wave) and in the upstream part (reflected wave) of the main duct, respectively, as shown in Fig. 1. To decompose the incident and reflected sound-pressure fields (a_{mn} and b_{mn}, respectively), the microphones should be located with some transverse and axial separation in the area upstream of the junction.

A simple approach is to divide the microphones into two groups separated by an axial distance, u. For the decomposition of N number of modes, at least N microphones at each group are placed on the same axial location. Therefore, $2\times N$ microphones are used on each side of the tee-junction along the main duct. In the matrix form of Eq. (5), a vector of sound pressures at the same axial location z and η number of independent transverse locations inside the duct can be written simply as the product of the acoustic mode shape matrix at points $\mathbf{\Phi}$, the propagating mode matrix $\mathbf{Z}$, and the complex acoustic pressure modal amplitude vectors of incident and reflected waves ($\mathbf{a}$ and $\mathbf{b}$, respectively) due to the following sources:

$$\hat{\mathbf{p}} = \begin{bmatrix} \hat{p}(x_1,y_1,z) \\ \hat{p}(x_2,y_2,z) \\ \vdots \\ \hat{p}(x_\eta,y_\eta,z) \end{bmatrix} = \mathbf{\Phi}^T\mathbf{Z}\mathbf{a} + \mathbf{\Phi}^T\mathbf{Z}^{-1}\mathbf{b}, \tag{8}$$

where

$$\mathbf{\Phi} = \begin{bmatrix} \psi_{0,0}(x_1,y_1) & \psi_{0,0}(x_2,y_2) & \cdots & \psi_{0,0}(x_\eta,y_\eta) \\ \psi_{1,0}(x_1,y_1) & \psi_{1,0}(x_2,y_2) & \cdots & \psi_{1,0}(x_\eta,y_\eta) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{m,n}(x_1,y_1) & \psi_{m,n}(x_2,y_2) & \cdots & \psi_{m,n}(x_\eta,y_\eta) \end{bmatrix},$$

and $\mathbf{Z}=\mathrm{diag}(e^{-jk_{z,00}z},\ e^{-jk_{z,10}z},\cdots e^{-jk_{z,mn}z})$ is a diagonal matrix. The superscript T denotes matrix transpose. To solve the incident and reflected waves, additional vectors of sound pressures are required with σ number of linear independent source conditions to form a matrix such that

$$\hat{\mathbf{p}} = \mathbf{\Phi}^T \mathbf{Z} \mathbf{A} + \mathbf{\Phi}^T \mathbf{Z}^{-1} \mathbf{B}, \tag{9}$$

where **A** and **B** are $[\{\mathbf{a}\}_1\{\mathbf{a}\}_2 \ldots \{\mathbf{a}\}_\sigma]$ and $[\{\mathbf{b}\}_1\{\mathbf{b}\}_2 \ldots \{\mathbf{b}\}_\sigma]$, respectively. $\{\mathbf{a}\}_i$ and $\{\mathbf{b}\}_i$ are the complex acoustic pressure modal amplitude vectors of incident and reflected waves, respectively, due to ith source condition. $\hat{\mathbf{p}} = [\{\hat{\mathbf{p}}\}_1\{\hat{\mathbf{p}}\}_2 \cdots \{\hat{\mathbf{p}}\}_\sigma]$. The transfer matrix, $\mathbf{H_{ab}}$, of the two groups of microphones upstream of the junction at $z = z_a$ and z_b, respectively, is defined as

$$\mathbf{H_{ab}} \hat{\mathbf{p}}_\mathbf{a} = \hat{\mathbf{p}}_\mathbf{b}, \tag{10}$$

where the subscripts **a** and **b** denote the quantities associated with the axial location at z_a and z_b, respectively, in the upstream with an axial separation of $z_b - z_a = u$. Substituting Eq. (9) into Eq. (10), we find that the reflection matrix of the junction (or other duct elements) can be estimated from experimental data as

$$\mathbf{B} \mathbf{A}^{-1} = \mathbf{Z_a} \mathbf{R_{ab}} \mathbf{Z_a}, \tag{11}$$

where

$$\mathbf{R_{ab}} = (\mathbf{\Phi}_\mathbf{b}^T \mathbf{Z}_\mathbf{u}^{-1} - \mathbf{H_{ab}} \mathbf{\Phi}_\mathbf{a}^\mathbf{T})^{-1} (\mathbf{H_{ab}} \mathbf{\Phi}_\mathbf{a}^T - \mathbf{\Phi}_\mathbf{b}^T \mathbf{Z_u}), \tag{12}$$

and

$$\mathbf{Z_u} = \mathrm{diag}\left(e^{-jk_{z,00}(z_b - z_a)}, \quad e^{-jk_{z,10}(z_b - z_a)}, \quad \ldots, \quad e^{-jk_{z,mn}(z_b - z_a)}\right). \tag{13}$$

The amplitude of the diagonal elements in the diagonal matrix $\mathbf{Z_a}$ is unity; therefore, the modal sound-power reflection coefficients upstream of the tee-junction can be found from the values of the elements of $\mathbf{R_{ab}}$ in Eq. (12). Substituting Eq. (11) into Eq. (9), the sound pressures at $z = z_a$ can be rewritten as

$$\hat{\mathbf{p}}_\mathbf{a} = \mathbf{\Phi}_\mathbf{a}^T (\mathbf{I} + \mathbf{R_{ab}}) \mathbf{Z_a} \mathbf{A}. \tag{14}$$

To solve the incident and reflected sound-pressure fields downstream of the main duct, suppose there are two other groups of microphones downstream of the tee-junction at axial locations of z_c and z_d of the main duct, respectively. Equations (8)–(13) can then be applied. The sound pressure at $z = z_c$ can be found as

$$\hat{\mathbf{p}}_\mathbf{c} = \mathbf{\Phi}_\mathbf{c}^T (\mathbf{I} + \mathbf{R_{cd}}) \mathbf{Z_c} \mathbf{C}, \tag{15}$$

I is the unit matrix. **C** denotes the modal amplitude matrix of complex acoustic pressure of wave transmitted downstream with σ number of linear independent source conditions, which consists of a complex acoustic pressure amplitude c_{pq} of (p,q) mode. The integers p and q are the acoustic duct mode numbers of the x-axis and y-axis, respectively, downstream of the junction in the main duct. $\mathbf{R_{cd}}$ is the modal amplitude matrix of the reflection coefficients of the ending downstream (right-hand end) of the duct, which can be found by Eq. (12) with subscripts **c** and **d** instead of **a** and **b**, respectively. By Eq. (14) and (15), the transmission matrix of the junction can be written as

$$\mathbf{C} \mathbf{A}^{-1} = \mathbf{Z_c}^{-1} \mathbf{T} \mathbf{Z_a}, \tag{16}$$

TABLE I. Eigen-frequencies, k_{mn}, of the duct normalized by the first eigen-frequency, $k_{mn=(1,0)}$.

Acoustic modes (m,n)	$k_{mn}/k_{mn=(1,0)}$
(1,0)	1.00
(0,1)	1.15
(1,1)	1.53
(2,0)	2.00

$$\mathbf{T} = [\mathbf{\Phi}_\mathbf{c}^T (\mathbf{I} + \mathbf{R_{cd}})]^{-1} \mathbf{H_{ac}} [\mathbf{\Phi}_\mathbf{a}^T (\mathbf{I} + \mathbf{R_{ab}})], \tag{17}$$

where the subscripts **c** and **d** denote the quantities associated with the axial locations at z_c and z_d, respectively, of the main duct and the area downstream of the junction. $\mathbf{H_{ac}} = \hat{\mathbf{p}}_\mathbf{c} \hat{\mathbf{p}}_\mathbf{a}^{-1}$ is the transfer matrix of sound-pressure groups located at z_a and z_c.

Equation (16) can be further simplified by removing a microphone layer when an anechoic termination is installed at the downstream ending where $\mathbf{R_{cd}} = 0$ or when a termination with specified modal reflection is installed. For a modal decomposition of N number of modes, measurements should be carried out with at least $N = \eta$ and $N = \sigma$. The solution of the over-determined case can be found with $N < \eta$. **T** is the modal complex amplitude matrix of the transmission coefficient. This matrix is independent of upstream conditions, which consist of N^2 coefficients:

$$T_{mn,pq}, \quad m,n,p,q = 1,2, \ldots, N,$$

representing the sound transmission or scattering coefficient amplitude from the (m,n) mode upstream of the junction in the main duct to the (p,q) mode downstream of the junction. The forms of Eqs. (12) and (17) are similar to those of the well-known two-microphone method for fundamental mode.

An experiment was performed to verify and extend the results of the numerical investigation. A test rig was constructed in which the main duct used in the measurements in Fig. 1 was made of 20-mm-thick Perspex with a rectangular cross-section of dimensions $L_x = 173$ mm and $L_y = 150$ mm. This set-up yielded a first eigen-frequency of 991 Hz in the main duct. Table I shows the normalized eigen-frequencies of the first four modes of the main duct. The sidebranch has the same cross-sectional dimensions as the main duct, with a sidebranch length l of 1 m. The end of the sidebranch can be closed or fitted with an anechoic termination. A loudspeaker was mounted at one end of the main duct; the other end was terminated anechoically using a length of highly absorbent duct end, as described in Ref. 2, to eliminate the reflection of sound produced at the other end of the main duct. The sound-power reflection coefficient was less than 0.6% at the frequency of interest, indicating that an anechoic condition was establish.

To resolve sound transmission and scattering of the first four modes, four microphones are required at each of four axial locations (i.e., 16 microphones in total). The microphones were flush mounted in the sides of the duct. The number and locations of the microphones were chosen to observe the desired acoustic modes, which are the (0,0), (1,0), (0,1), and (1,1) modes. The locations of microphones are listed in Table II. The measurement microphones were

TABLE II. Locations of microphones within the duct (mm).

Microphone No.	Transverse location (x,y)	Axial location, z
1	(153, 0)	−670
2	(0, 20)	
3	(173, 130)	
4	(20, 150)	
5	(20, 0)	−620
6	(0, 130)	
7	(173, 20)	
8	(153, 150)	
9	(153, 0)	600
10	(0, 20)	
11	(173, 130)	
12	(20, 150)	
13	(20, 0)	650
14	(0, 130)	
15	(173, 20)	
16	(153, 150)	

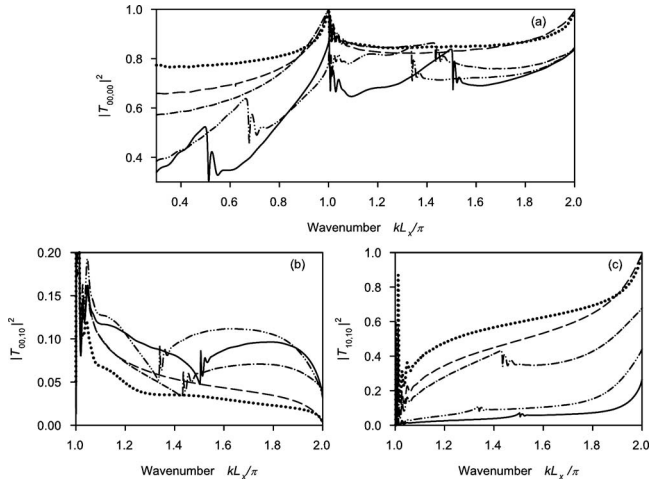


FIG. 2. Sound-power transmission and scattering across an infinitely long sidebranch along the main duct.(a) $|T_{00,00}|^2$, (b) $|T_{00,10}|^2$, and (c) $|T_{10,10}|^2$. (·····) $w/L_x=0.3$; (– – – –) $w/L_x=0.5$; (–·–·–) $w/L_x=0.7$; (–··–··–) $w/L_x=1.5$; (—) $w/L_x=2.0$.

Bruel & Kjaer (B&K) Type 4935 microphones (7 mm diameter). For calibration of the amplitude and phase of all channels, each microphone was mounted in an intensity coupler (B&K Type 3541) with a reference microphone. The same acoustic pressure was exposed to the microphones, as only the fundamental mode could propagate inside the intensity coupler. The transfer functions of microphones were measured using a white noise generator (B&K Type ZI 0055) and then used to correct the response of each microphone. The phase mis-match among the microphones was less than $\pm 2.7°$ at frequencies below 2 kHz. For the source conditions employed in measuring the higher-order modes, four independent source conditions were generated with four different restrictor plates, as described in Schultz *et al.*[15]

III. RESULTS AND DISCUSSION

A. Numerical investigation

As mentioned above, we investigated sound-wave propagation up to the eigen-frequency of the second x-axis mode ($m=2$ and $p=2$) at $kL_x=2\pi$ in the main duct. A two-dimensional system is considered in this section ($n=0$ and $q=0$). The sound transmission and scattering coefficients are estimated from the computed data at $[(z-w)/L_x]=8$ in order to eliminate the influence of evanescent waves. The numerical experiment focuses on the sound transmission and scattering of the fundamental mode ($m=0$) and first asymmetric mode ($m=1$) across a tee-junction along the main duct.

1. Infinitely long sidebranch

a. Sound-power transmission of the fundamental mode (0,0) at higher frequency, $|T_{00,00}|^2$. For an infinitely long sidebranch, the tee-junction is known to be a high-pass filter at frequencies below the first eigen-frequency of the main duct, $k_{mn=(1,0)}=\pi/L_x$.[3] Figure 2(a) shows the effects of w/L_x on the sound-power transmission of a fundamental mode $|T_{00,00}|^2$ across the junction along the main duct at higher frequencies below the eigen-frequencies of the (2,0) modes $[k \leq k_{mn=(2,0)}]$. $|T_{00,00}|^2$ denotes the sound-power transmission-coefficient from fundament mode upstream of the junction $(m,n)=(0,0)$ to fundamental mode downstream of the junction $(p,q)=(0,0)$ in the main duct. A slight fluctuation in $|T_{00,00}|^2$ can be observed at frequencies just above $k_{mn=(1,0)}$, which can be caused by the strong evanescent waves close to the eigen-frequencies. At frequencies below the first eigen-frequency of the sidebranch, $k<\pi/w$, only a fundamental mode can propagate in the sidebranch. The peaks in $|T_{00,00}|^2$ occur at the eigen-frequencies of the x-axis modes in the main duct, $k_{mn=(m,0)}$, as shown in Fig. 2(a) with $w/L_x=0.3$ and 0.5, where it approaches unity at $kL_x=\pi$ and 2π. This indicates strong sound transmission of fundamental mode at the eigen-frequencies of axial mode in x-axis at the main duct. $|T_{00,00}|^2$ declines gradually and reaches a minimum at frequencies between two eigen-frequencies of x-axis modes in the main duct. The present study was extended to higher frequencies, and the same observations were found at other higher-order modes (data not shown).

A greater amount of acoustic power is transmitted into the sidebranch with increasing w/L_x by branch sound-power division,[19] resulting in a rapid decline in $|T_{00,00}|^2$ as the frequency increases from $k=k_{mn=(1,0)}=\pi/L_x$ and a lowering in the minimum point between $k_{mn=(m,0)}$, as shown in Fig. 2(a) with $w/L_x=0.5$ (cf. $w/L_x=0.3$). For large w/L_x, the nonplanar branch-mode, which is the higher-order axial mode of the side branch in the axis along the main duct z-axis, can be excited at a frequency of $k>\pi/w$; however, a local maximum of $|T_{00,00}|^2$ can be found at or very close to the eigen-frequencies of the sidebranch, $k=\alpha\pi/w$, where α is the axial branch-mode number for the z-axis.

An example with $w/L_x=0.7$ is shown in Fig. 2(a), where a local maximum in $|T_{00,00}|^2$ can be observed at the frequency of the first asymmetric branch-mode, $kL_x=1.429\pi$. For frequencies just above the eigen-frequencies of the sidebranch, we find a sharp fall in $|T_{00,00}|^2$ followed by fluctuating values. Subsequently, the trend resumes, increasing as the frequency approaches another higher eigen-frequency of an x-axis mode in the main duct; however, the peak of

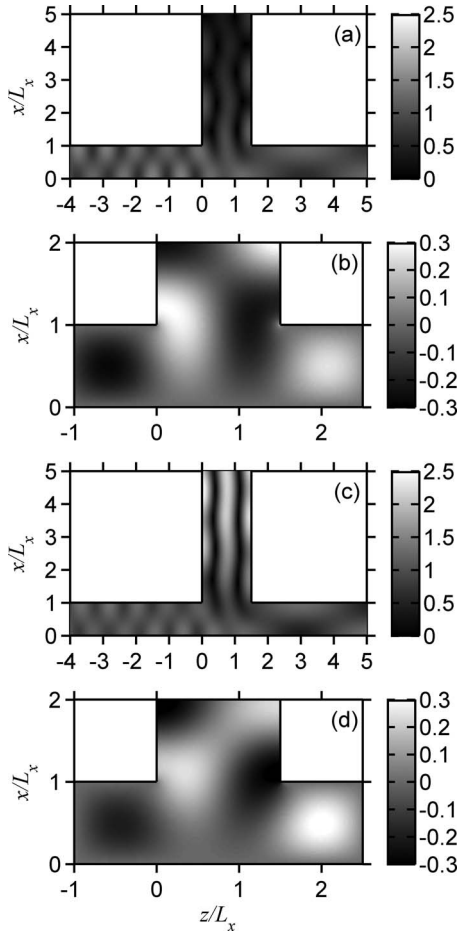


FIG. 3. Sound field around the tee-junction between the main duct and an infinitely long sidebranch with $w/L_x=1.5$ due to fundamental-mode excitation. (a) $kL_x=1.333\pi$, pressure magnitude (Pa); (b) $kL_x=1.333\pi$, particle-velocity profile (toward the sidebranch), $\rho c v_x$(kg m^{-1} s^{-2}); (c) $kL_x=1.340\pi$, pressure magnitude (Pa); (d) $kL_x=1.340\pi$, particle-velocity profile (toward the sidebranch), $\rho c v_x$(kg m^{-1} s^{-2}). ρ, c, and v_x are the density of the media (kg/m^3), speed of sound in the media (m/s), and acoustic particle velocity (m/s), respectively.

$|T_{00,00}|^2$ at the eigen-frequencies of the main duct declines when the non-planar branch-mode is excited, as shown in Fig. 2(a) with $w/L_x=0.7$, 1.5, and 2.0 (cf. $w/L_x=0.3$ and 0.5). Similar patterns are found for higher branch-modes, e.g., Fig. 2(a) with $w/L_x=1.5$ and 2.0. The local maxima in $|T_{00,00}|^2$ are also found at or close to the eigen-frequency of the second branch-mode, $k=2L_x/w$. These maxima and subsequent fall at higher-order mode will now be discussed with the visualization of the sound pressure and particle-velocity fields in the rest of this sub-section.

Figure 3(a) shows the sound-pressure field at the eigen-frequency of the second branch-mode, $k=2L_x/w$, with $w/L_x=1.5$, due to fundamental-mode excitation. Higher sound pressure can be observed inside the junction at $x=0$. The distribution of acoustic velocity in x-axis toward the sidebranch, v_x (m/s), is asymmetric [Fig. 3(b)]; consequently, the symmetric branch-mode is suppressed at $kL_x=1.333\pi$. This mismatch results in increased $|T_{00,00}|^2$ in the main duct at and very close to the eigen-frequencies of the sidebranch [see Fig. 2(a)]. The situation is markedly different at frequencies slightly above the eigen-frequencies of the sidebranch, as shown in Fig. 3(c). This change occurs because the wavelength is slightly smaller than the width of the sidebranch at the frequencies immediately above the eigen-frequency of the sidebranch. This leads to an extension in the distribution of acoustic velocity in axis, v_x (m/s), toward the sidebranch at $x=L_x$ [Fig. 3(d)] and thus more effective excitation of the symmetric branch-mode. Strong resonance of the higher branch-mode is evident in Fig. 3(c). As a result of the large amount of acoustic power taken away by the strong resonance of the branch-mode, a slump occurs in the sound transmission $|T_{00,00}|^2$ across the junction along the main duct at frequencies slightly above the eigen-frequency of the sidebranch [see Fig. 2(a)].

The results also confirm that higher-order branch-mode excitation is shifted at higher frequency by the relatively uniform particle-velocity field toward the sidebranch, as evident in the results of Tang,[3] although this previous study did not explicitly specify the effects of the higher-order branch-mode. The present results indicate weak coupling of the traveling wave in the main duct and branch-modes at or very close to the eigen-frequencies of the sidebranch. This weak coupling arises because of the mismatch between the particle-velocity field and branch-mode at the entry to the sidebranch, resulting in a shift in excitation of the higher-order branch-mode at higher frequency and thus an increase in the sound transmission of the fundamental mode across the junction along the main duct at the eigen-frequencies of the sidebranch. As the frequency increases, the non-planar branch-mode is excited and thus acoustic power is transmitted into the sidebranch. This causes the decrease in $|T_{00,00}|^2$.

b. Sound-power scattering of the fundamental mode (0,0) into the first mode (1,0), $|T_{00,10}|^2$. Because of acoustic scattering and diffraction at the junction in the main duct, all modes are able to propagate downstream of the junction due to excitation by the fundamental mode and other modes at the entry to the junction at $z=0$. Figure 2(b) shows that the sidebranch acts as a low-pass filter for sound scattering of the fundamental mode into the first duct mode of the x-axis $|T_{00,10}|^2$ at frequencies of $kL_x=\pi$ to 2π. The first mode of the x-axis is able to propagate through the duct at frequencies above $k_{mn=(1,0)}=\pi/L_x$. Generally, $|T_{00,10}|^2$ shows a gradual decrease with increasing frequency from $k_{mn=(1,0)}$. Again, a slight fluctuation in $|T_{00,10}|^2$ at frequencies just above $k_{mn=(1,0)}$ can be caused by the strong evanescent waves at and close to the eigen-frequencies. For $k<L_x/w$, an increase in $|T_{00,10}|^2$ is observed with increasing w/L_x, as shown in Fig. 2(b) with $w/L_x=0.5$ (cf. $w/L_x=0.3$). $|T_{00,10}|^2$ vanishes at the eigen-frequency of the (2,0) duct mode, $k=2\pi/L_x$.

Figure 4 shows the sound-pressure field due to fundamental-mode excitation with small $w/L_x(=0.5)$. A high sound pressure occurs inside the junction at $x=0$ at frequencies above and close to $k_{mn=(1,0)}$ [Fig. 4(a)]. With increasing frequency, the region of high sound pressure is extended to the exit of the junction, and the strength is reduced [Fig. 4(b)]. Sound pressure is again high inside the junction at the frequency close to $k_{mn=(2,0)}$; however, this area is located close to the center of the junction, as shown in Fig. 4(c). This positioning results in inefficient excitation of the first asym-

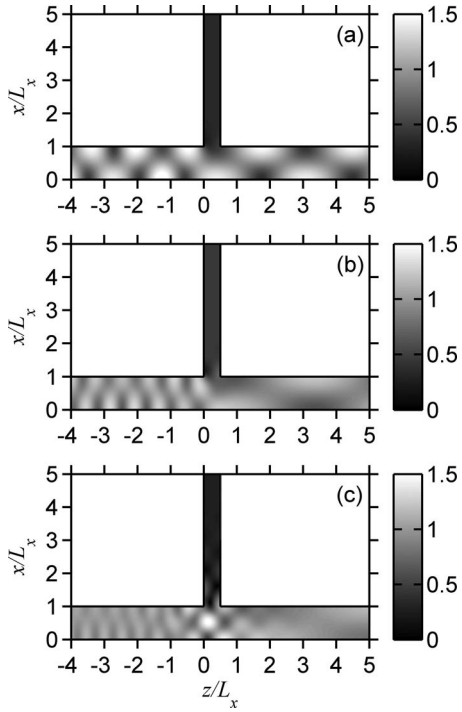


FIG. 4. Sound-pressure field (Pa) around the tee-junction between the main duct and an infinitely long sidebranch with $w/L_x=0.5$ due to fundamental-mode excitation. (a) $kL_x=1.050\pi$, (b) $kL_x=1.500\pi$, and (c) $kL_x=1.950\pi$.

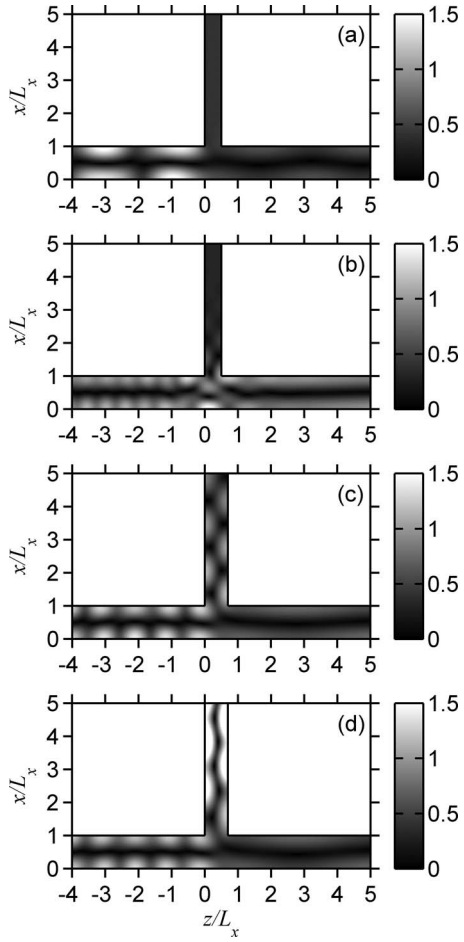


FIG. 5. Sound-pressure field (Pa) around the tee-junction between the main duct and an infinitely long sidebranch due to unit (1,0) mode excitation. (a) $kL_x=1.100\pi$, $w/L_x=0.5$; (b) $kL_x=1.950\pi$, $w/L_x=0.5$; (c) $kL_x=1.429\pi$, $w/L_x=0.7$; (d) $kL_x=1.436\pi$, $w/L_x=0.7$.

metric mode and the vanishing of sound scattered into the $(p,q)=(1,0)$ mode by the fundamental mode downstream of the junction along the main duct at a frequency close to $k_{mn=(2,0)}$. For $k>L_x/w$, a marked drop in $|T_{00,10}|^2$ occurs as frequency increases from $k_{mn=(1,0)}$ to the eigen-frequencies of the sidebranch. $|T_{00,10}|^2$ reaches a local minimum at or close to the eigen-frequencies of the sidebranch ($k=\alpha L_x/w$). Figures 2(a) and 2(b) shows weak modal coupling of the branch-mode and traveling wave in the junction at or close to the eigen-frequencies of the sidebranch, as a consequence of high $|T_{00,00}|^2$ and low $|T_{00,10}|^2$. The sudden fall in $|T_{00,00}|^2$ [Fig. 2(a)] at the frequency slightly above the eigen-frequencies of the sidebranch results in a small jump in $|T_{00,10}|^2$ [Fig. 2(b)] due to the non-uniform distribution of pressure at the exit of the tee-junction along the main duct; thus, some acoustic power is scattered into the (1,0) mode [Fig. 3(c)], followed by minor fluctuations with increasing frequency. For large w/L_x, a more effective scattering of the fundamental mode into the (1,0) mode can be observed at frequencies above $k=L_x/w$ [Fig. 2(b)], with excitation of the non-planar branch-mode and thus non-uniform distribution of pressure at $z=w$.

c. Sound-power transmission of the first mode (1,0), $|T_{10,10}|^2$. Similar to the situation for $|T_{00,00}|^2$ at frequencies below $k_{mn=(1,0)}$,[3] the tee-junction acts as a high-pass filter for the sound-power transmission of the first duct mode $|T_{10,10}|^2$ at frequencies between $k_{mn=(1,0)}=\pi/L_x$ and $k_{mn=(2,0)}=2\pi/L_x$ of the x-axis modes, as shown in Fig. 2(c); however, the vanishing of $|T_{10,10}|^2$ can be observed at or close to $k_{mn=(1,0)}$. Again, a fluctuation in $|T_{10,10}|^2$ at frequencies just above the eigen-frequency of the main duct $k_{mn=(1,0)}$ occurs due to the strong evanescent waves. Sound transmission of the (1,0) mode, $|T_{10,10}|^2$, becomes increasingly effective with increasing frequency, approaching $k_{mn=(2,0)}$. A rapid increase in $|T_{10,10}|^2$ occurs with increasing frequency from $k_{mn=(1,0)}$, and sound transmission of the (1,0) mode becomes more effective with decreasing w/L_x because less acoustic power is taken away by the sidebranch. For frequencies below $k=L_x/w$, $|T_{10,10}|^2$ approaches unity at $k_{mn=(2,0)}$, which is an eigen-frequency of the higher-order x-axis mode. Again, a sharp fall in $|T_{10,10}|^2$ is followed by fluctuating values at frequencies just above the eigen-frequencies of the sidebranch [e.g., Fig. 2(c) with $w/L_x=0.7$, 1.5, and 2.0].

Figure 5 shows the sound-pressure field around the tee-junction arising from excitation of the (1,0) mode only in the area upstream of the duct. A strong standing wave of the first mode (1,0) is observed upstream of the duct [Fig. 5(a)], indicating the occurrence of high modal reflection at frequencies close to that of the (1,0) mode. With increasing frequency, high sound pressure occurs inside the junction at $x=0$; thus, effective transmission of the (1,0) mode is observed close to $k_{mn=(2,0)}$ [Fig. 5(b)]. Again, the non-planar branch-mode is suppressed at or close to the eigen-frequencies of the sidebranch [Fig. 5(c)] because of weak coupling of the branch-mode and traveling wave in the junction. With increasing frequency, a sudden drop in $|T_{10,10}|^2$

occurs in response to the improved distribution of acoustic velocity at the entry to the sidebranch (data not shown). Strong resonance of the branch-mode is observed at frequencies slightly higher than the eigen-frequencies of the sidebranch [Fig. 5(d)]; however, the effects of the non-planar branch-mode on the sound transmission of higher-order modes are reduced compared with the effects on fundamental-mode transmission, especially for higher-order branch-modes [Fig. 2(c)]. The scattering of other modes is minor, and is not considered here.

d. Discussion of infinitely long sidebranch. In general, sound transmissions of fundamental mode and higher-order modes are high at the eigen-frequencies of the main duct and branch. These are due to resonance of the main duct and weak modal coupling between the branch-mode and traveling wave in the junction, respectively. With the excitation of the non-planar branch-mode, a decrease in sound transmission can be observed. A fluctuation in sound transmission and scattering at frequencies just above the eigen-frequencies of the main duct occurs due to the strong evanescent waves.

2. Branch-end reflection

Two types of longitudinal branch resonance have been found to be critical for the sound transmission of fundamental mode at frequencies below $k_{mn=(1,0)}$ due to the end reflection of the sidebranch:[3] the "both-ends-closed" and "one-end-closed–one-end-open" types, which correspond to the creation of high and low pressures at the entry to the sidebranch, respectively. Moreover, $|T_{00,00}|^2$ approaches zero and unity for infinitely long and closed sidebranches, respectively, at very low frequency, $kd \rightarrow 0$. Here, we investigate the case of a closed sidebranch using numerical experiments, and the frequency is extended to the higher-order modes of the x-axis, $k_{mn=(2,0)}$. We consider the effects of two types of longitudinal branch resonance and non-planar branch-modes on sound transmission in higher-order modes.

Figures 6(a)–6(d) show the sound-power transmission of a fundamental mode $|T_{00,00}|^2$ and sound-power scattering of a fundamental mode into the (1,0) mode $|T_{00,10}|^2$, respectively, for a duct across a 1-m-long closed sidebranch along the main duct at various high frequencies between $kL_x=\pi$ and 2π. $|T_{00,00}|^2$ attains minima at certain frequencies in response to the two types of branch resonances. These two types of branch resonance are excitation of the longitudinal branch resonance in a low-pressure region created at the entry to the sidebranch at frequencies $1.4\pi < kL_x < 1.6\pi$ and excitation of the non-planar branch-mode.

Figure 7(a) shows the sound-pressure field at the minimum point in Fig. 6(a) with $w/L_x=0.3$, due to a one-end-closed–one-end-open type of longitudinal branch resonance at frequencies above $k_{mn=(1,0)}$. A low-pressure region is created inside the junction and at the entry to the sidebranch, while a high-pressure region is created in the wall opposite the sidebranch. This non-uniform distribution of sound pressure results in effective scattering of the fundamental mode into the asymmetric mode (1,0) downstream of the junction, along the main duct [Figs. 6(c) and 7(a)]. In addition to an increase in w/L_x, more inclined nodal and anti-nodal planes

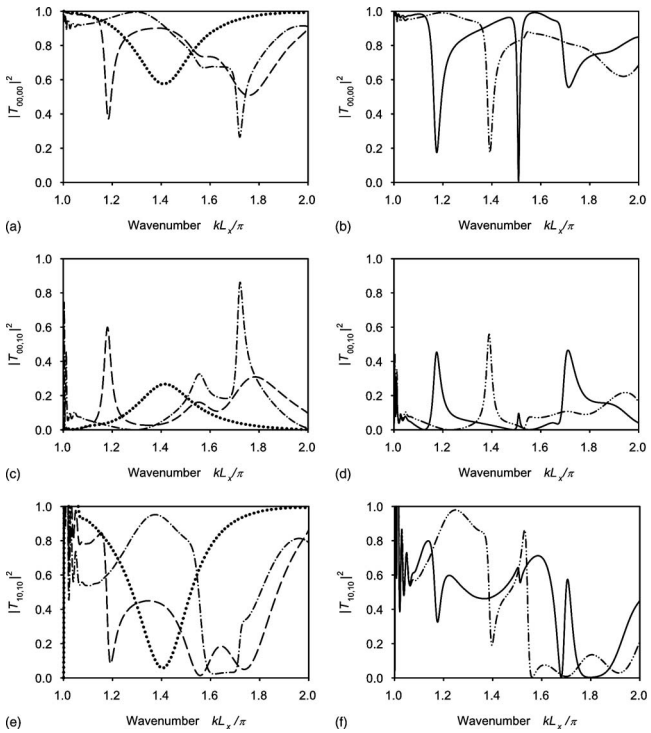


FIG. 6. Sound-power transmission and scattering across a closed sidebranch along the main duct. $l/L_x=1$. [(a) and (b)] $|T_{00,00}|^2$, [(c) and (d)] $|T_{00,10}|^2$, and [(e) and (f)] $|T_{10,10}|^2$. (·····) $w/L_x=0.3$; (– – – –) $w/L_x=0.9$; (–·–·–) $w/L_x=1.2$; (–··–··–) $w/L_x=1.5$; (—) $w/L_x=1.8$.

are observed [Fig. 7(b)]. The inclined nodal plane at the entry to the sidebranch extends to the center of the junction, and the high-pressure region in the wall opposite the sidebranch is reduced in size. These changes lead to an increase in the effectiveness of sound transmission of the fundamental mode $|T_{00,00}|^2$ and a reduction in the effectiveness of sound

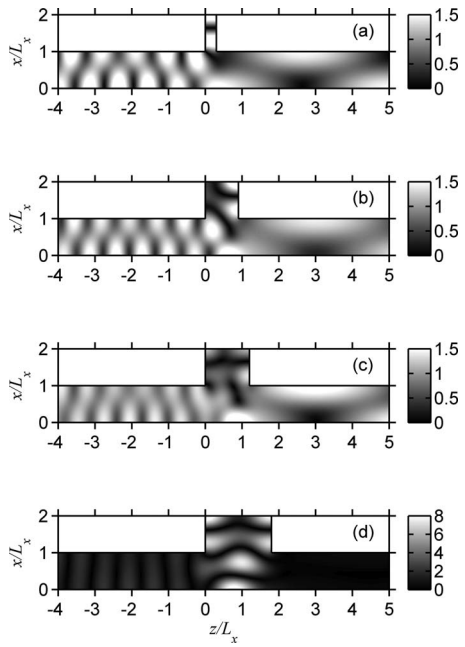


FIG. 7. Sound-pressure field (Pa) around the tee-junction between the main duct and a closed sidebranch due to fundamental-mode excitation, with $l/L_x=1$. (a) $kL_x=1.410\pi$; $w/L_x=0.3$; (b) $kL_x=1.552\pi$; $w/L_x=0.9$; (c) $kL_x=1.556\pi$, $w/L_x=1.2$; (d) $kL_x=1.510\pi$, $w/L_x=1.8$.

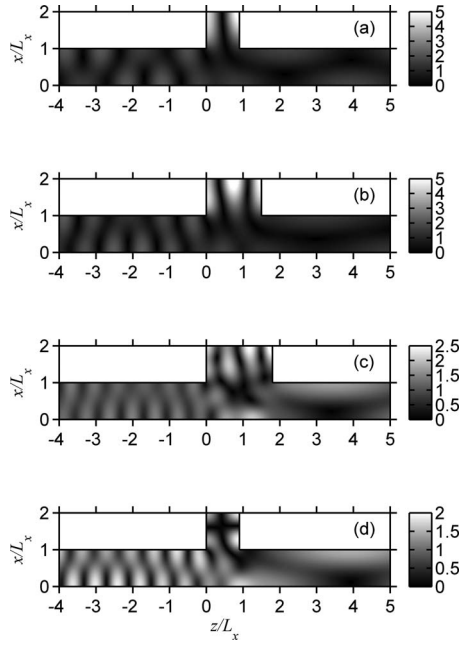


FIG. 8. Sound-pressure field (Pa) around the tee-junction between the main duct and a closed sidebranch due to fundamental-mode excitation, with $l/L_x=1$. (a) $kL_x=1.184\pi$, $w/L_x=0.9$; (b) $kL_x=1.392\pi$, $w/L_x=1.5$; (c) $kL_x=1.714\pi$, $w/L_x=1.8$; (d) $kL_x=1.756\pi$, $w/L_x=0.9$.

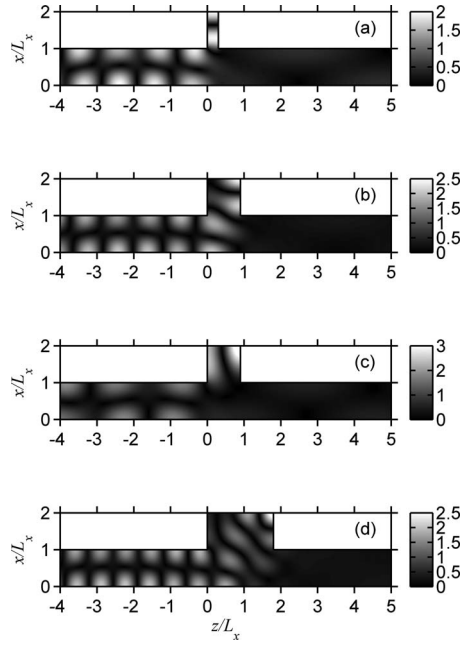


FIG. 9. Sound-pressure field (Pa) around the tee-junction between the main duct and a closed sidebranch due to unit (1,0) mode excitation, with $l/L_x=1$. (a) $kL_x=1.404\pi$, $w/L_x=0.3$; (b) $kL_x=1.558\pi$, $w/L_x=0.9$; (c) $kL_x=1.192\pi$, $w/L_x=0.9$; (d) $kL_x=1.796\pi$, $w/L_x=1.8$.

scattering into the first asymmetric mode $|T_{00,10}|^2$ downstream of the junction. The inclined nodal and anti-nodal planes also act to shift the frequency of the minimum points of $|T_{00,00}|^2$ toward the higher side with increasing w/L_x, e.g., the plot for $w/L_x=0.3$ and 0.9 in Figs. 6(a), 7(a), and 7(b). For large w/L_x, the co-excitation of the non-planar and longitudinal branch-modes results in a decrease in $|T_{00,00}|^2$ [Fig. 7(c) with $w/L_x=1.2$].

A more effective scattering of the fundamental mode into the first asymmetric duct mode $|T_{00,10}|^2$ is observed with excitation of the non-planar branch-mode, as shown in Figs. 6(c) and 7(c). Figures 7(a) and 7(b) show the inclined nodal plane (low-pressure region) at the entry to the sidebranch at higher frequencies but with the same l/L_x, for which the longitudinal branch resonance is subdued and $|T_{00,00}|^2$ does not vanish at frequencies higher than $k_{mn=(1,0)}$, in contrast to the vanishing observed at low frequencies (see Fig. 8 in Ref. 3). The excitation of strong resonance of longitudinal branch-modes results in vanishing $|T_{00,00}|^2$ and a reduction in the frequency of the minimum point [Fig. 7(d)]. The non-planar mode is suppressed because of the weak modal coupling of the non-planar branch-mode and the traveling wave at or close to the eigen-frequencies of the sidebranch. The development of a strong standing wave in the duct upstream of the junction and inside the sidebranch indicates the high reflection of sound and strong longitudinal branch resonance. There exists a uniform distribution of acoustic velocity (upward) at the entry to the sidebranch (data not shown), and $|T_{00,00}|^2$ and $|T_{00,10}|^2$ are low, as shown in Figs. 6(b) and 6(d). Scattering is less dependent on excitation of the longitudinal branch-mode [Figs. 6(c) and 6(d) at a frequency of $1.4\pi < kL_x < 1.6\pi$]. The minima in Figs. 6(a) and 6(b) outside the range of $1.4\pi < kL_x < 1.6\pi$ are caused by excitation of the non-planar branch-modes (e.g., Fig. 8); however, sound transmission does not completely vanish at higher frequencies [compare with the results reported by Tang[3] at frequencies below $k_{mn=(1,0)}$]. For excitation of the second branch-mode at frequencies of $kL_x=1.722\pi$, 1.392π, and 1.176π with $w/L_x=1.2$, 1.5, and 1.8, respectively, we observed that the eigen-frequencies of the sidebranch are shifted toward the high side, and $|T_{00,00}|^2$ decreases with increasing w/L_x, as shown in Figs. 6(a) and 6(b). Again, a strong scattering of the fundamental mode into the first asymmetric mode $|T_{00,10}|^2$ is observed with excitation of the non-planar branch-mode [Figs. 8(a)–8(c)]. In the case of co-excitation of the longitudinal branch-mode and non-planar branch-mode, we find a more broad band-stop action in $|T_{00,00}|^2$ [Figs. 6(a) and 8(d)].

Figures 6(e) and 6(f) show the effects of w/L_x on the sound-power transmission of the (1,0) mode $|T_{10,10}|^2$ across the junction and along the main duct at various frequencies between $k_{mn=(1,0)}$ and $k_{mn=(2,0)}$. Vanishing and minimum $|T_{10,10}|^2$ are found with excitation of the longitudinal branch-modes within the low-pressure region at the entry to the sidebranch and with excitation of non-planar branch-modes.

Figures 9(a) and 9(b) shows the sound-pressure fields at frequencies of vanishing $|T_{10,10}|^2$ due to the excitation of one-end-closed–one-end-open type of longitudinal branch resonance in the range $1.4\pi < kL_x < 1.6\pi$, which is similar to $|T_{00,00}|^2$ [cf. Figs. 7(a) and 7(b)]. Figures 9(c) and 9(d) show excitation of non-planar branch-modes at frequencies of a minimum and vanishing $|T_{10,10}|^2$, respectively, for the case of a closed sidebranch, which yields similar results to that in the case of $|T_{00,00}|^2$ [cf. Figs. 8(a) and 8(c)]. The both-ends-closed type of longitudinal branch resonance, which leads to a high-pressure region at the entry to the sidebranch, results in high $|T_{00,00}|^2$ and $|T_{10,10}|^2$ (data not shown).

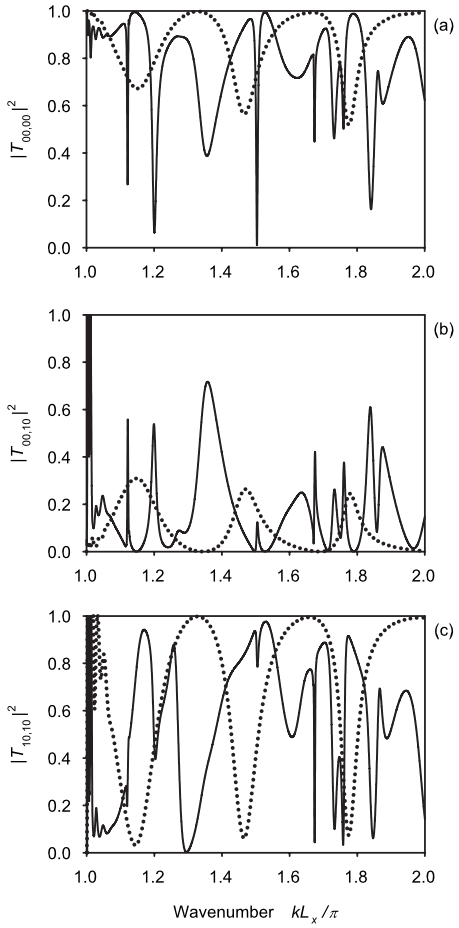


FIG. 10. Sound-power transmission and scattering across a closed sidebranch along the main duct, with $l/L_x=3$. (a) $|T_{00,00}|^2$, (b) $|T_{00,10}|^2$, and (c) $|T_{10,10}|^2$. (·····) $w/L_x=0.3$; (—) $w/L_x=1.8$.

With increasing sidebranch length, additional occurrences of branch resonance are observed; an example for $l/L_x=3$ with $w/L_x=0.3$ and 1.8 is shown in Fig. 10, involving the same mechanism as that described for $l/L_x=1$. The corresponding results are not discussed here.

To summarize this sub-section, the sound transmissions $|T_{00,00}|^2$ and $|T_{10,10}|^2$ are low with excitation of the non-planar branch-mode and excitation of the longitudinal branch resonance when a low-pressure region created at the entry to the sidebranch. However, an increase in $|T_{00,10}|^2$ occurs with excitation of the non-planar branch-mode.

B. Experimental results

Here, the formulations in Sec. II B are used for modal decomposition of the sound-pressure field upstream and downstream of the junction in the test rig within three-dimensional space. Figure 11 shows the sound-power transmission and scattering of fundamental mode with an anechoic sidebranch. Also plotted for comparison are the results of finite element modeling for $|T_{00,00}|^2$, $|T_{00,10}|^2$, and $|T_{10,10}|^2$. In terms of sound transmission and scattering, the results of the experiment show a reasonable match with the results of numerical modeling with anechoic termination at the end of the sidebranch. The sound transmission of fundamental mode, $|T_{00,00}|^2$, shows peaks close to the eigen-

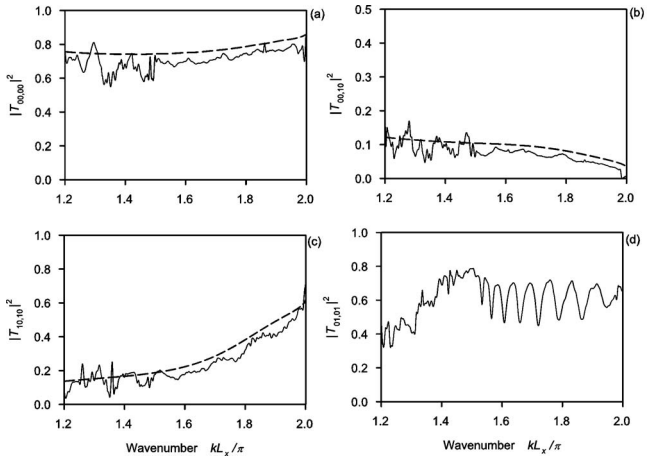


FIG. 11. Sound-power transmission and scattering across a tee-junction with anechoic termination along the main duct. (a) $|T_{00,00}|^2$, (b) $|T_{00,10}|^2$, (c) $|T_{10,10}|^2$, and (d) $|T_{01,01}|^2$. (—) measured data; (– – – – –) simulation results.

frequencies of the x-axis mode, $k_{mn=(m,0)}$ [Fig. 11(a)]. High sound scattering of the fundamental mode to the first mode $|T_{00,10}|^2$ is observed close to $k_{mn=(1,0)}$, although a gradual decrease is observed with increasing frequency [Fig. 11(b)]. As mentioned above, $|T_{00,10}|^2$ attains a minimum at the (2,0) mode, $k=2\pi/L_x$.

The experiment reveals a dramatic fluctuation in the results at frequencies close to the eigen-frequencies of the duct at $kL_x=\pi$ to 1.2π, reflecting the influence of the evanescent mode; therefore, those results are not shown here. For sound transmission of the first mode across the junction, $|T_{10,10}|^2$ shows a vanishing trend at $k_{mn=(1,0)}$. With increasing frequency, sound transmission of the first mode becomes increasingly effective, as shown in Fig. 11(c). Figure 11(d) shows the sound transmission of the first mode of the y-axis $|T_{01,01}|^2$, a scenario that cannot be evaluated using the two-dimensional numerical experiments described in Sec. III A. The upward trend in $|T_{01,01}|^2$ observed at frequencies from $k_{mn=(0,1)}$ to $k_{mn=(1,1)}$ [Fig. 11(d)] is similar to that of $|T_{00,00}|^2$ from $k_{mn=(0,0)}$ to $k_{mn=(1,0)}$ [i.e., from $k=0$ to π/L_x], as shown in Fig. 2 in Ref 3. This finding indicates the same mechanism of sound transmission for the non-planar mode of the y-axis as that for the fundamental mode. The fluctuation of $|T_{01,01}|^2$ at frequencies above the eigen-frequency of the (1,1) mode at $kL_x=1.53\pi$ is due to modal coupling of the spanwise branch-mode and the traveling wave at the junction.[16]

Figure 12 shows $|T_{00,00}|^2$, $|T_{00,10}|^2$, $|T_{10,10}|^2$, and $|T_{01,01}|^2$ with a 1-m-long closed sidebranch. The finite element model slightly overestimates the sound transmission and scattering compared with measured values, especially for cases of sound transmission for higher-order modes. This discrepancy possibly reflects the occurrence of modal damping within the duct, especially at high frequencies; however, the results are with acceptable engineering tolerance.

As mentioned above, the minima in $|T_{00,00}|^2$ and $|T_{10,10}|^2$ are found in response to excitation of the branch resonances, as shown in Figs. 12(a) and 12(c). The modal decomposition method can satisfactorily solve the modal transmission and scattering coefficients. Figure 12(d) indicates several minima in $|T_{01,01}|^2$ at some frequencies between $kL_x=1.5\pi$ and 2π.

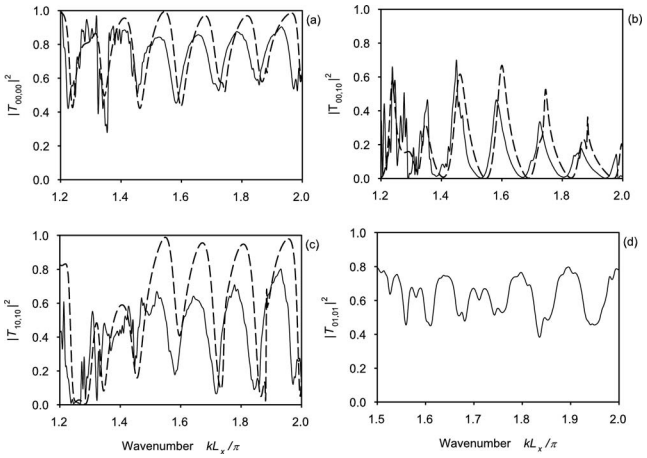


FIG. 12. Sound-power transmission and scattering across a 1-m-long closed tee-junction along the main duct. (a) $|T_{00,00}|^2$, (b) $|T_{00,10}|^2$, (c) $|T_{10,10}|^2$, and (d) $|T_{01,01}|^2$. (—) measured data; (– – – – –) simulation results.

The fluctuation in $|T_{01,01}|^2$ at higher frequencies is possibly due to branch resonances. High sound transmission of the higher-order modes is observed in Figs. 11(c), 11(d), 12(c), and 12(d), thereby confirming the importance of sound propagation at higher-order modes.

Figures 13(a) and 13(b) shows the sound scattering of the (0,1) mode into the (1,1) mode and sound transmission of the (1,1) mode across the junction with an anechoic termination, respectively, revealing downward and upward trends. Fluctuations are due to modal coupling of the spanwise branch-mode and the traveling wave in the duct. Again, Figs. 11(b) and 13(a) reveal high sound scattering into higher-order modes at the frequency of the transmitted higher-order modes. Figures 11(c) and 13(b) demonstrate that sound transmission of high-order modes and the fundamental mode (as shown in Ref. 3) is increased from its eigen-frequency to a higher eigen-frequency. Finally, Figs. 13(c) and 13(d) show

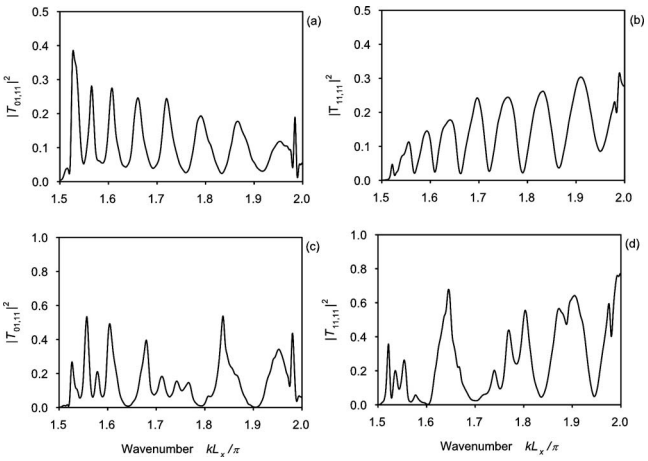


FIG. 13. Experimental results of sound-power transmission and scattering across a 1-m-long closed tee-junction along the main duct. [(a) and (b)] $|T_{01,11}|^2$ and $|T_{11,11}|^2$ across a tee-junction with anechoic termination along the main duct, respectively; [(c) and (d)] $|T_{01,11}|^2$ and $|T_{11,11}|^2$ across a 1-m-long closed tee-junction along the main duct, respectively. (—) measured data; (– – – – –) simulation results.

that high sound scattering and transmission (with end reflection of the sidebranch) can be observed at higher frequencies in higher-order modes.

The numerical results in Sec. III A are verified in this sub-section using direct modal decomposition method. The experimental method for measurement of sound transmission and scattering of rectangular duct in higher-order modes is demonstrated. There exists a difficulty in the measurement at frequencies close to the eigen-frequency of the main duct due to the strong evanescent waves. For the tests with branch-end reflection, the measured values are slightly less than that of the numerical simulations due to the occurrence of modal damping in the experiments.

IV. CONCLUSIONS

We investigated the sound-transmission characteristics of a tee-junction along an infinitely long rectangular duct using numerical simulations and model experiments. The analysis considered infinitely long sidebranches and closed sidebranches. Sound transmission and scattering of the first four acoustic modes across the tee-junction of the duct were decomposed and investigated experimentally.

For an infinitely long sidebranch, two acoustical multi-band filters in series can be observed in the sound transmission of fundamental mode. The tee-junction acts as a multi-band filter for the fundamental mode. The pass-bands of the filter are found at the eigen-frequencies of the axial mode in the axis along the sidebranch, along which the sound transmission of fundamental mode gradually increases from a very low frequency and approaches a maximum at the eigen-frequencies of the axial mode of the main duct. An increase in sidebranch width results in higher attenuation at the stop-bands of the filter as the sidebranch takes away larger amounts of acoustic power. Another relevant mechanism is the weak coupling of the branch-mode and traveling wave in the junction at the entry to the sidebranch, which provides multi-band filtering with pass-bands at or very close to the eigen-frequencies of the sidebranch, thereby shifting excitation of the higher-order non-planar branch-modes at higher frequency. Sound transmission decreased with excitation of the non-planar branch-mode; however, this effect is reduced for sound transmission of higher-order modes across the junction along the main duct because of weaker coupling of the branch-modes and traveling wave in the junction. Sound transmission of the axial mode along the sidebranch axis vanishes at its eigen-frequencies and increases as the frequency approaches the eigen-frequency of the immediately higher axial mode along the same axis of the sidebranch. For sound scattering into higher acoustic modes, the sidebranch acts as a low pass filter; for example, sound scattering from fundamental mode into the first asymmetric mode of the duct in the axis along the sidebranch becomes more effective as the frequency approaches the eigen-frequency of the first asymmetric mode; it also decreases with increasing frequency. Sound scattering into higher-order modes is increased with excitation of the non-planar branch-mode, and

an increase in sidebranch width results in an increase in sound scattering and the reduced transmission of higher-order modes.

In the case of a closed sidebranch, lower sound transmission of higher-order modes are observed with excitation of the longitudinal branch-mode in the low-pressure region created at the entry to the sidebranch and with excitation of the non-planar branch-mode. The inclined nodal plane of the longitudinal branch-mode means that sound transmission does not vanish at higher frequencies; however, vanishing of the sound transmission of fundamental mode and a reduction in the effectiveness of higher-order-mode scattering are observed for a strong standing wave of the longitudinal branch-mode and suppression of the non-planar branch-mode.

A procedure of measurement of sound transmission and scattering of a duct element in a rectangular duct is reported, using a modal decomposition approach based on the transfer functions of microphones located at different axial positions. Further simplification of the formulation is possible if the modal reflection condition is known downstream of the duct termination. The present study experimentally verified the sound-transmission coefficients of a tee-junction. The experiment results are reasonably consistent with the simulation results; however, the simulation of a tee-junction with branch-end reflection and a hard-wall overestimated the sound transmission and scattering, possibly due to modal damping at high frequencies. Fluctuations in sound transmission and scattering are observed due to coupling of the spanwise branch-mode and traveling wave.

The sound-transmission characteristics of a tee-junction at higher-order mode have been shown. Such problems frequently arise when the sound propagation in waveguide is at frequencies higher than first eigen-frequency of the duct. Ventilation duct is a representative example. A procedure of the measurement of sound transmission of scattering at higher-order mode is demonstrated in the present study.

ACKNOWLEDGMENTS

This work was fully supported by a grant from the Hong Kong Polytechnic University (Project A/C No. G-U362). We would like to thank Professor S. K. Tang (Department of Building Services Engineering, The Hong Kong Polytechnic University) for support with finite element computations and for valuable advice.

[1]A. Cummings, "Sound transmission in curved duct bends," J. Sound Vib. **35**, 451–477 (1974).
[2]S. K. Tang and F. Y. C. Li, "On low frequency sound transmission loss of double sidebranches: A comparison between theory and experiment," J. Acoust. Soc. Am. **113**, 3215–3225 (2003).
[3]S. K. Tang, "Sound transmission characteristics of tee-junctions and the associated length corrections," J. Acoust. Soc. Am. **115**, 218–227 (2004).
[4]S. K. Tang and C. K. Lau, "Sound transmission across a smooth nonuniform section in an infinitely long duct," J. Acoust. Soc. Am. **112**, 2602–2611 (2002).
[5]C. K. Lau and S. K. Tang, "Sound transmission across duct constrictions with and without tapered sections," J. Acoust. Soc. Am. **117**, 3679–3685 (2005).
[6]D. D. Reynolds and J. M. Bledsoe, *Algorithms for HVAC Acoustics* (American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 1991).
[7]*ASHRAE Handbook—Applications* (American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2007).
[8]J. L. Parrondo, J. Fernández, I. García, and E. Ruiz, "Noise transmission through duct divisions in air circuits, considered as three-port acoustic systems," J. Sound Vib. **296**, 183–194 (2006).
[9]*ASHRAE Handbook—Fundamentals* (American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2005).
[10]J. Y. Kwon, H. H. Park, and H. J. Eom, "Acoustic scattering from two junctions in a rectangular waveguide," J. Acoust. Soc. Am. **103**, 1209–1212 (1998).
[11]J. Y. Kwon and H. J. Eom, "Acoustic hybrid junction in a rectangular waveguide," J. Acoust. Soc. Am. **107**, 1868–1873 (2000).
[12]C. J. Moore, "Measurement of radial and circumferential modes in annular and circular fan ducts," J. Sound Vib. **62**, 235–256 (1979).
[13]M. Abom, "Modal decomposition in ducts based on transfer function measurements between microphone pairs," J. Sound Vib. **135**, 95–114 (1989).
[14]A. Sitel, J. M. Ville, and F. Foucart, "Multiload procedure to measure the acoustic scattering matrix of a duct discontinuity for higher order mode propagation conditions," J. Acoust. Soc. Am. **120**, 2478–2490 (2006).
[15]T. Schultz, L. N. Cattafesta III, and M. Sheplak, "Modal decomposition method for acoustic impedance testing in square ducts," J. Acoust. Soc. Am. **120**, 3750–3758 (2006).
[16]T. C. Redmore and K. A. Mulholland, "The application of mode coupling theory to the transmission of sound in the sidebranch of a rectangular duct system," J. Sound Vib. **85**, 323–331 (1982).
[17]V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. I. Theory," J. Acoust. Soc. Am. **100**, 2034–2048 (1996).
[18]M. L. Munjal, *Acoustics of Ducts and Mufflers with Application to Exhaust and Ventilation System Design* (Wiley, New York, 1987).
[19]*ASHRAE Handbook—HVAC Applications* (American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, 2007).

Transmission loss of double panels filled with porogranular materials

Jean-Daniel Chazot[a)]
Centre de Recherche de Royallieu, Université de Technologie de Compiègne, CNRS UMR 6253 Roberval, BP 20529, 60205 Compiègne Cedex, France

Jean-Louis Guyader
Laboratoire Vibrations Acoustique de l'INSA de Lyon, 25 bis, avenue Jean Capelle, 69621 Villeurbanne Cedex, France

(Received 10 April 2009; revised 2 September 2009; accepted 17 September 2009)

Sound transmission through hollow structures found its interest in several industrial domains such as building acoustics, automotive industry, and aeronautics. However, in practice, hollow structures are often filled with porous materials to improve acoustic properties without adding an excessive mass. Recently a lot of interest arises for granular materials of low density that can be an alternative to standard absorbing materials. This paper aims to predict vibro-acoustic behavior of double panels filled with porogranular materials by using the patch-mobility method recently published. Biot's theory is a basic tool for the description of porous material but is quite difficult to use in practice, mostly because of the solid phase characterization. The original simplified Biot's model (fluid-fluid model) for porogranular material permitting a considerable reduction in data necessary for calculation has been recently published. The aim of the present paper is to propose a model to predict sound transmission through a double panel filled with a porogranular material. The method is an extension of a previous paper to take into account the porogranular material through fluid-fluid Biot's model. After a global overview of the method, the case of a double panel filled with expanded polystyrene beads is studied and a comparison with measurements is realized.
 [DOI: 10.1121/1.3245033]

PACS number(s): 43.55.Rg, 43.55.Ev [NX] Pages: 3040–3048

I. INTRODUCTION

Double glazed panels are of great interest to increase sound insulation in several domains. Basically the air gap between the panels permits to decouple the motions of the receiving panel from the excited one. As shown by London in his work on infinite single[1] and double panels,[2] the gain of sound insulation is mainly obtained thanks to the decoupling of the two panels at high frequency. However, at the double panel resonance frequency, the mass-air-mass resonance phenomenon deteriorates the good sound insulation properties of the double panel. On the other hand, Brunskog[3] showed that the effect of the air layer viscosity could have some influence when the air layer thickness is small enough. This phenomenon is, however, not sufficient to improve significantly the sound insulation of a double panel.

Multilayered structures are also widely employed for sound absorption and vibration damping in several applications such as buildings, automobiles, aircraft, etc. These light weight structures are designed to improve sound insulation without adding an excessive mass that is critical in such applications. To predict the vibro-acoustic behavior of these multilayered structures, several methods are available. Analytical formulations have first been used by Guyader and Lesueur[4,5] for orthotropic multilayered plates and by Nilsson and Nilsson[6] for sandwich structures with honeycomb and foam cores. The transfer matrix method described by Munjal and Sastry[7,8] can also be used to predict the response of infinite multilayered poroelastic panels to incident oblique plane waves. However, these methods are not adapted to predict the vibro-acoustic behavior of finite skew plates. On the contrary, the finite element method used by Panneton and Atalla[9] to predict the sound transmission through finite multilayer systems with poroelastic materials is well adapted to model complex finite geometries. Nevertheless, the main drawback of finite element models comes from the significant computational time required. The patch-mobility method (PMM) used by Ouisse *et al.*[10] to couple linear acoustic problems and by Chazot and Guyader[11] to couple vibro-acoustic problems can also be used efficiently. Indeed, the use of a mobility technique permits to characterize each component of the vibro-acoustic problem separately, either analytically or numerically, and then to calculate the global response solving the interaction equation. If one element is modified, only its own characterization has to be calculated before solving interaction equations. In structural coupling, a similar mobility technique has already been used by Moorhouse[12] and by Naji.[13] An extension to energy mobility has also been developed by Orefice *et al.*[14]

In order to improve the sound insulation properties of hollow structures such as double panels, poroelastic materials are also widely employed. The gain in sound insulation obtained by these materials is due to the large viscous and

[a)]Author to whom correspondence should be addressed. Electronic mail: jean-daniel.chazot@utc.fr

thermal dissipations in the pores. In order to model the vibro-acoustic behavior of such materials, homogenized Biot's theory[15] is generally used. This theory enables to take into account the solid and the fluid phase and their elastic, inertial, and visco-thermal interactions. This model have been implemented with the transfer matrix method by Lauriks *et al.*,[16] and with the finite element method by Panneton and Atalla.[17] An enhanced mixed displacement-pressure formulation has also been developed by Atalla *et al.*[18] However, due to the high computation cost still required to implement this model, simplified numerical methods have been developed,[19,20] as well as simplified poroelastic models. Two classical simplifications of Biot's theory can hence be cited: the rigid frame formulation[21] and the limp formulation.[22,23] In both cases, the solid phase elasticity is not taken into account. These two simplifications lead therefore to an equivalent dissipative fluid formulation. However, in the limp formulation, inertial effects of the solid phase are taken into account. Finally, when the solid phase elasticity is taking part in the vibro-acoustic response of the system, it is possible to only neglect the solid phase shear stress. This simplification leads to a fluid-fluid model presented by Chazot and Guyader.[24] This model is well adapted to describe the behavior of materials with a low shear modulus.[25,26]

Recently the interest on granular material has increased and particular granular materials have been studied such as recycled rubber particles.[27] In vibration damping, such materials have proven to be efficient.[28,29] In sound absorption, the behavior of granular materials is similar to poroelastic materials. Classical Biot's theory can thus be employed directly[30] or with the simplified fluid-fluid model[31] for light and non-cohesive porogranular materials.

This paper presents a method to predict the sound transmission through a double panel filled with a porogranular material. It is an extension of the double panel transmission loss (TL) model with the patch-mobility method developed in Ref. 11, when the simplified fluid-fluid model presented in Ref. 24 is used to take into account the porogranular material. After a global overview of the method, the case of a double panel filled with expanded polystyrene beads is studied and a comparison with measurements is realized.

II. VIBRO-ACOUSTIC MODEL OF A DOUBLE PANEL FILLED WITH A POROGRANULAR MATERIAL

In this section, the PMM detailed in Ref. 11 is applied to calculate the vibro-acoustic response of a double panel containing a porogranular material. The double panel is coupled to a cavity and radiates in an open external space (see Fig. 1). Only the main lines of the method are reminded here.

In the PMM, the averaged velocity on patch i of the subsystem A after coupling, $\langle V_A\rangle_i$, is calculated by summing the averaged velocity before coupling $\langle \tilde{V}_A\rangle_i$ and the averaged velocity responses due to coupling surface forces with other subsystems $\langle F_{\tilde{A}/A}\rangle_j$ on patch j such as

$$\langle V_A\rangle_i = \langle \tilde{V}_A\rangle_i + \sum_{j=1}^{N_{AB}} \langle\langle Y_A\rangle\rangle_{ij}\langle F_{\tilde{A}/A}\rangle_j. \tag{1}$$

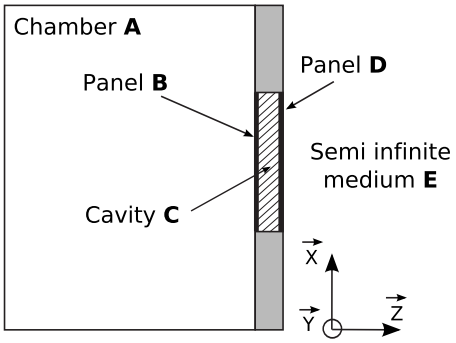


FIG. 1. Sketch of a double panel coupled with a source room and a semi-infinite medium.

The averaged velocity response of the subsystem A on patch i due to a coupling force on patch j is calculated with the patch mobility defined as the ratio of the averaged patch i velocity to the averaged patch j surface force:

$$\langle\langle Y_A\rangle\rangle_{ij} = \frac{\langle V\rangle_i}{\langle F\rangle_j}, \tag{2}$$

where

$$\langle V\rangle_i = \frac{\int_{Si} V(x,y)dxdy}{Si} \tag{3}$$

and

$$\langle F\rangle_j = \frac{\int_{Sj} F(x,y)dxdy}{Sj}. \tag{4}$$

External excitations are taken into account, thanks to the velocities before coupling, and continuity conditions over coupling surfaces are necessary to get rid of the unknown coupling surface forces.

In the case of the double panel containing a porogranular material, the system is first divided into five sub-systems as depicted in Fig. 1: the emission chamber A, the first panel B, the cavity C filled with a porogranular material, the second panel D, and the semi-infinite medium E. Existing coupling surfaces between subsystems i and j are defined: S_{ij}.

The complete set of equations given in Ref. 11 that describes all the physical interactions taking part in the vibro-acoustic response of the double panel can be slightly simplified because the pressure radiated by the panel into the source room is negligible compared to the excitation provided by the source. The coupling of the first panel with the source room is therefore omitted, and the double panel excitation is introduced with a velocity before coupling of the first panel.

For each subsystem, simplified governing equations writes the following.

For the first panel,

$$\langle V_B\rangle_i = \langle \tilde{V}_B\rangle_i + \sum_{j=1}^{N_{BC}} \langle\langle Y_B\rangle\rangle_{ij}\langle F_{C/B}\rangle_j. \tag{5}$$

For the cavity face BC ($\forall i \in S_{BC}$),

$$\langle V_{C1}\rangle_i = \sum_{j=1}^{N_{BC}} \langle\langle Y_{C1}\rangle\rangle_{ij}\langle F_{B/C}\rangle_j + \sum_{k=1}^{N_{CD}} \langle\langle Y_{C1}\rangle\rangle_{ik}\langle F_{D/C}\rangle_k. \quad (6)$$

For the cavity face CD ($\forall i \in S_{CD}$),

$$\langle V_{C2}\rangle_i = \sum_{k=1}^{N_{CD}} \langle\langle Y_{C2}\rangle\rangle_{ik}\langle F_{D/C}\rangle_k + \sum_{j=1}^{N_{BC}} \langle\langle Y_{C2}\rangle\rangle_{ij}\langle F_{B/C}\rangle_j. \quad (7)$$

For the second panel,

$$\langle V_D\rangle_i = \sum_{k=1}^{N_{CD}} \langle\langle Y_D\rangle\rangle_{ik}\langle F_{C/D}\rangle_k. \quad (8)$$

Excitation appears in Eq. (5) as a patch velocity vector before coupling $\langle \tilde{V}_B\rangle_i$. It is calculated from plate B patch transfer mobility (PTM) $\langle\langle Y_B\rangle\rangle_{ij}$, and from boundary pressure in the source room, as

$$\langle \tilde{V}_B\rangle_i = \sum_j \langle\langle Y_B\rangle\rangle_{ij}\langle P_{(\text{blocked})}\rangle_j S_j, \quad (9)$$

where blocked patch pressures $\langle P_{(\text{blocked})}\rangle_j$ are calculated by integrating over patches the pressure, generated by a source in the emission chamber, over the panel assumed to be rigid (blocked).

Finally, velocity continuity conditions over coupling surfaces S_{BC} and S_{CD} are written: $\forall i \in S_{BC}$: $\langle V_B\rangle_i = \langle V_C\rangle_i$ and $\forall i \in S_{CD}$: $\langle V_C\rangle_i = \langle V_D\rangle_i$. It yields to

$$\begin{Bmatrix} \langle \tilde{V}_1\rangle_i \\ 0 \end{Bmatrix} = [\mathbf{Y}] \begin{Bmatrix} \langle F_{B/C}\rangle_j \\ \langle F_{D/C}\rangle_k \end{Bmatrix} \quad (10)$$

with

$$[\mathbf{Y}] = \begin{bmatrix} \langle\langle Y_{C1}\rangle\rangle_{ij} + \langle\langle Y_B\rangle_i\rangle_j & \langle\langle Y_{C1}\rangle\rangle_{ik} \\ \langle\langle Y_{C2}\rangle\rangle_{ij} & \langle\langle Y_{C2}\rangle\rangle_{ik} + \langle\langle Y_D\rangle\rangle_{ik} \end{bmatrix}.$$

Solving the linear system (10) allows to get coupling forces $\langle F_{B/C}\rangle_j$ and $\langle F_{D/C}\rangle_k$ and then to calculate patch velocities after coupling by using Eqs. (5)–(8).

Sections III–V present the panel patch mobilities, porogranular patch mobilities, and radiation patch mobilities that are used to calculate the response of the double panel filled with a porogranular material. In all the following calculations, harmonic motions of angular frequency ω are assumed, and for sake of simplicity, time dependence $e^{+j\omega t}$ is omitted.

III. PANEL PATCH MOBILITIES

The panel patch mobility can be calculated from the Love–Kirchhoff equation of motion (flexural vibration of thin plates) using a modal expansion of the plate transverse displacement with simply supported boundary conditions. Details of these calculations are given in Ref. 11. The panel patch mobility hence obtained writes

$$\langle\langle YP\rangle\rangle_{ij} = \frac{j\omega}{S_i S_j}\sum_{pq} \frac{\int_{S_i}\Phi_{pq}(x,y)dS \cdot \int_{S_j}\Phi_{pq}(x,y)dS}{M_{pq}[\omega_{pq}^{*2} - \omega^2]}, \quad (11)$$

where

$$K_{pq}^* = \int_S D^*\Phi_{pq}\nabla^4(\Phi_{pq})dS, \quad M_{pq} = \int_S \rho h \Phi_{pq}^2 dS,$$

$$\omega_{pq}^* = \sqrt{\frac{K_{pq}^*}{M_{pq}}}, \quad D^* = \frac{E^* h^3}{12(1-\nu^2)}.$$

E^* is complex Young modulus which takes into account structural damping such as $E^* = E(1+j\eta_s)$, and η_s is the damping loss factor. Mode shapes of simply supported plates are used: $\Phi_{pq}(x,y) = \sin((p\pi/L_x)x)\sin((q\pi/L_y)y)$, where L_x and L_y are the plate dimensions. In some cases, and, in particular, for a diffuse field excitation with large panel dimensions, the transmission loss of the panel does not depend on the boundary conditions. However, as mentioned in Refs. 32 and 33, real boundary conditions have to be taken into account at low frequency and for small panel dimensions. A modal expansion of clamped panel given in Ref. 32 can then be used.

IV. POROGRANULAR PATCH MOBILITIES

Different simplifications can be brought to full Biot's theory depending on the studied material elastic characteristics. A simplification was hence proposed in Ref. 24 to model light and non-cohesive porogranular materials. The basic idea of this model is to modify the elastic law of the solid phase to suppress shear stresses. The granular material phase is also assumed to behave like a fluid with a hydrostatic stress tensor such as $\sigma^s = -p_s I$, with I the identity matrix and p_s the equivalent solid pressure that represents the hydrostatic stress of the granular material phase. The resulting fluid-fluid model is presented in Eq. (12).

$$\begin{aligned} &[A\Delta p_s + p_s] + B\Delta p = 0 \\ &[C\Delta p + p] + D\Delta p_s = 0 \end{aligned} \quad \text{on } \Omega$$

$$\mathbf{V_s}\cdot\mathbf{n} = \mathbf{V_f}\cdot\mathbf{n} = \begin{cases} \langle V_n\rangle_j & \text{on excitation patch } j \\ 0 & \text{elsewhere} \end{cases} \quad \text{on } \partial\Omega \quad (12)$$

p_s and $\mathbf{V_s}$ stand for the equivalent pressure and velocity in the solid phase, p and $\mathbf{V_f}$ are the fluid pressure and velocity, $\mathbf{n}$ denotes here the outward normal unit vector to the poroelastic domain, and A, B, C, and D are elastic coefficients given in Appendix B. These coefficients are related to elastic and inertial coupling coefficients.

Boundary conditions given on the surface $\partial\Omega$ are chosen to obtain the transfer impedance between two patches i and j. A uniform normal velocity $\langle V_n\rangle_j$ is therefore applied on excitation patch j, and the resulting pressure p_t is averaged over observation patch i. These boundary conditions are then applied over the porogranular material using coupling conditions by considering an elastic-poroelastic interface as defined in Ref. 34. The resulting coupling conditions are given

by Eq. (13). These equations are based on the flow, the fluid velocity, and the total normal stress continuity conditions. On excitation patch, the solid and fluid normal velocities are thus equal to the normal excitation velocity, while the total pressure over the observation patch is given by the average pressure over the two phases.

$$\mathbf{V_s}\cdot\mathbf{n}\begin{cases}\langle V_n\rangle_j & \text{on excitation patch } j\\ 0 & \text{elsewhere,}\end{cases}$$

$$\mathbf{V_f}\cdot\mathbf{n}\begin{cases}\langle V_n\rangle_j & \text{on excitation patch } j\\ 0 & \text{elsewhere,}\end{cases} \tag{13}$$

$$\langle p_t\rangle_i = (1-\phi)\langle p_s\rangle_i + \phi\langle p\rangle_i,$$

where ϕ is the porosity and the normal averaged velocity $\langle V_n\rangle_j$ is considered constant on patch j. The solid and fluid velocities, $\mathbf{V_s}$ and $\mathbf{V_f}$, used in boundary conditions are also related to the solid and fluid pressure gradients through Eq. (14):

$$\begin{aligned}(1-\phi)\nabla p_s &= -(\rho_{11}\mathbf{V_s}+\rho_{12}\mathbf{V_f})j\omega,\\ \phi\nabla p &= -(\rho_{12}\mathbf{V_s}+\rho_{22}\mathbf{V_f})j\omega.\end{aligned} \tag{14}$$

In order to obtain the weak formulation associated with the fluid-fluid model, Eq. (12) are multiplied with a trial function and integrated over the domain Ω. Green's identities enable then to introduce the boundary conditions leading to the weak formulation given in Eq. (15):

$$\begin{aligned}&\int_\Omega (A\Delta\Psi_s+\Psi_s)p_s d\Omega - \int_{\partial\Omega} A(\nabla\Psi_s p_s)\mathbf{n}dS\\ &\quad - j\omega A\frac{\rho_{11}+\rho_{12}}{1-\phi}\langle\Psi_s\rangle_j\langle V_n\rangle_j S_j + \int_\Omega B(p\Delta\Psi_s)d\Omega\\ &\quad - \int_{\partial\Omega} B(\nabla\Psi_s p)\mathbf{n}dS - j\omega B\frac{\rho_{12}+\rho_{22}}{\phi}\langle\Psi_s\rangle_j\langle V_n\rangle_j S_j = 0,\\ &\int_\Omega (C\Delta\Psi+\Psi)p d\Omega - \int_{\partial\Omega} C(\nabla\Psi p)\mathbf{n}dS\\ &\quad - j\omega C\frac{\rho_{12}+\rho_{22}}{\phi}\langle\Psi\rangle_j\langle V_n\rangle_j S_j + \int_\Omega D(p_s\Delta\Psi)d\Omega\\ &\quad - \int_{\partial\Omega} D(\nabla\Psi p_s)\mathbf{n}dS - j\omega D\frac{\rho_{11}+\rho_{12}}{1-\phi}\langle\Psi\rangle_j\langle V_n\rangle_j S_j = 0,\end{aligned} \tag{15}$$

where Ψ_s and Ψ are the solid and fluid test functions. Averaged test functions write $\langle\Psi_s\rangle_j=\int_{s_j}\Psi_s dS/S_j$ and $\langle\Psi\rangle_j=\int_{s_j}\Psi dS/S_j$.

The Rayleigh–Ritz method is then applied to the weak formulation of fluid-fluid Biot's model, and modal expansions of the solid and fluid pressure fields are introduced:

$$p_s(M) = \sum_{pqr} a_{pqr}\cdot\psi_{pqr}(M), \tag{16}$$

$$p(M) = \sum_{pqr} b_{pqr}\cdot\psi_{pqr}(M). \tag{17}$$

The fluid pressure is hence decomposed over the basis of rigid-walled cavity mode shapes $\psi_{pqr}(M)$ such as $\psi_{pqr}(M)=\cos(p\pi x/L_x)\cos(g\pi y/L_y)\cos(r\pi z/L_z)$, with L_x, L_y, and L_z the cavity dimensions. The solid phase being also assumed to behave like a fluid, the equivalent solid pressure p_s is decomposed over the same basis. Notice also that velocity boundary conditions can be employed in the weak formulation (15) with a basis of rigid-walled cavity mode shapes. The resulting velocity field will then be incorrect locally on the excitation patch, but the weak formulation ensures, however, that the pressure response is correct. Thus, it is possible to calculate the impedance matrix [Eq. (21)] and finally to obtain the mobility matrix by inversion [Eq. (22)].

Solid and fluid test functions are also replaced by a mode shape $\psi_{pqr}(M)$:

$$\begin{aligned}\Psi_s &= \psi_{pqr}(M),\\ \Psi &= \psi_{pqr}(M).\end{aligned} \tag{18}$$

Modal orthogonality and null normal derivatives of test functions on the boundary surfaces lead the Rayleigh–Ritz method associated with the weak formulation (15) to Eq. (19).

$$\begin{aligned}a_{pqr} &= \frac{Bk_{pqr}^2 N_{pqr}}{I_{pqr}}b_{pqr} + \left(\frac{A}{I_{pqr}}\cdot\frac{\rho_{11}+\rho_{12}}{1-\phi} + \frac{B}{I_{pqr}}\cdot\frac{\rho_{22}+\rho_{12}}{\phi}\right)\\ &\quad\times j\omega\langle V_n\rangle_j\langle\psi_{pqr}\rangle_j S_j,\\ b_{pqr} &= \left(\frac{D}{J_{pqr}}\cdot\frac{\rho_{11}+\rho_{12}}{1-\phi} + \frac{(BD-AC)k_{pqr}^2+C}{J_{pqr}}\cdot\frac{\rho_{22}+\rho_{12}}{\phi}\right)\\ &\quad\times j\omega\langle V_n\rangle_j\langle\psi_{pqr}\rangle_j S_j,\end{aligned} \tag{19}$$

with $k_{pqr}^2=(p\pi/L_x)^2+(q\pi/L_y)^2+(r\pi/L_z)^2$, $I_{pqr}=N_{pqr}(1-Ak_{pqr}^2)$, $J_{pqr}=N_{pqr}(1-(A+C)k_{pqr}^2+(AC-BD)k_{pqr}^4)$, and $N_{pqr}=\int_\Omega\psi_{pqr}^2 d\Omega$.

The total averaged pressure on patch i is therefore given by

$$\langle p_t\rangle_i = \sum_{pqr}((1-\phi)a_{pqr}+\phi b_{pqr})\langle\psi_{pqr}\rangle_i, \tag{20}$$

and Patch transfer impedance writes

$$\langle\langle Z\rangle\rangle_{ij} = \frac{\langle p_t\rangle_i S_i}{\langle V_n\rangle_j}. \tag{21}$$

Porogranular patch mobilities can finally be deduced by inverting the patch transfer impedance matrix:

$$[Y] = [Z]^{-1}. \tag{22}$$

V. RADIATED POWER INTO THE SEMI-INFINITE MEDIUM

Radiated power is calculated from patch velocities and radiated patch pressures, and can be written using radiation patch impedances (23):

$$\Pi_{\text{rad}} = \frac{1}{2}\sum_i \text{Re}\{\langle V\rangle_i^*\langle P_{\text{rad}}\rangle_i S_i\},$$

$$= \frac{1}{2}(\mathrm{Re}[Z_{\mathrm{rad}}])\{V\}'\{V_i S_i\}^*. \quad (23)$$

The expressions of the radiation patch impedances are given in Appendix A.

VI. TRANSMISSION LOSS CALCULATIONS

A standard room response modal expansion is used to calculate the mean quadratic room pressure P_r^2 and the blocked patch pressures $\langle P\rangle_i$ over the incident panel. The resulting expressions are, respectively, given in Eqs. (24) and (25). Mode shapes ψ_{pqr} of a parallelepipedic-shaped room with rigid walls are used, and detailed calculations of modal amplitudes A_{pqr} are given in Ref. 11.

$$P_r^2 = \sum_{pqr} \frac{|A_{pqr}^2|\varepsilon_p\varepsilon_q\varepsilon_r}{8}, \quad (24)$$

with

$$\varepsilon_p = \begin{cases} 1 & \text{if } p = 0 \\ 2 & \text{if } p \neq 0, \end{cases}$$

$$\langle P\rangle_i = \int_{Si} \left(\sum_{p,q,r} A_{pqr}\psi_{pqr}(x,y,z) \right) dS. \quad (25)$$

The blocked patch pressures are used in Eq. (9) to calculate panel patch velocities before coupling, and enable to solve the set of Eqs. (5)–(8). As stated in Ref. 11, the great advantage of using blocked patch pressures instead of summing incident plane waves is to avoid the integration of plane waves over incidence angles to model a diffuse field leading to tremendous calculation time compared to the present method. Moreover, the blocked patch pressure calculation depends only on the source room configuration and is totally separated from the transmission loss calculations of a panel. Hence, once the source room is characterized with blocked patch pressures, transmission loss calculation of any panel is possible.

Finally, using the mean quadratic room pressure and radiated patch pressures, the panel transmission loss can be calculated. TL is indeed defined as the ratio of transmitted power Π_t to incident power Π_i: $\mathrm{TL} = 10\log(\Pi_i/\Pi_t)$. Transmitted power can be calculated, thanks to the mobility method with Eq. (23). On the other hand, according to the procedure of standard measurement of transmission loss, incident power is deduced from mean quadratic room pressure using a classical result of Sabine's theory: $\Pi_i = SP_r^2/4\rho_0 c$. Of course the assumption of Sabine's theory is not exactly verified for standard rectangular rooms especially at low frequency, but is nevertheless of general use for sake of simplicity.

VII. RESULTS

A. Calculated porogranular patch mobilities

The porogranular material we used was made of expanded polystyrene beads non-cohesive and of low density (19 kg/m^3). Its properties are given in Table I. These data come from Ref. 24 where the measurement technique is also reported. It is based on impedance tube measurements when no preload is applied to the granular material. When the material is maintained in a limited space with an initial load due to the gravity or to the experimental mounting, the bulk modulus can increase like in a fluid submitted to a pressure load. The chosen value of the bulk modulus is discussed in Sec. VII B.

TABLE I. Parameters of the porogranular material (after Ref. 24).

Parameter	Value
Porosity ϕ	0.41
Tortuosity α_∞	1.2855
Permeability k_0	9.29×10^{-9} m^2
Viscous characteristics length Λ	0.2 mm
Thermal characteristic length Λ'	0.34 mm
Bulk modulus	$0.35(1+0.18i)$ MPa.

The comparison of patch input mobilities of porogranular material and air cavity is presented in Fig. 2. The magnitude of the patch input mobility of the porogranular material appears to be lower than the air cavity. This is due to higher stiffness of the porogranular material. On the other hand, the phase of the air cavity input mobility is constant around 90 while the phase of the porogranular input mobility decreases with frequency. This phase illustrates the total dissipations, viscous and thermal, that take place in the porogranular material. Indeed, the media does not dissipate energy when the phase is equal to 90, and dissipations appear when the phase is different from 90. The porogranular material is therefore more dissipative at high frequency than at low frequency.

Surprisingly also, patch mobilities do not exhibit a modal behavior in the studied frequency band. In Ref. 24, this non-modal behavior is explained by the fact that direct patch-mobility calculation derives from a Dirichlet problem with a null boundary pressure except on excited patch. Cavity resonances with Dirichlet boundary conditions are obviously different from that of a cavity with Neumann boundary conditions. The first resonance of the rectangular cavity with Dirichlet boundary conditions is equal to $c/2\pi \times\sqrt{(pi/L_x)^2+(pi/L_y)^2+(pi/L_z)^2}$ that is to say 17 000 Hz. That explains why no resonance appears in the studied frequency band, even if our calculation is indirect and use inverse impedance matrix calculated with rigid wall cavity basis.

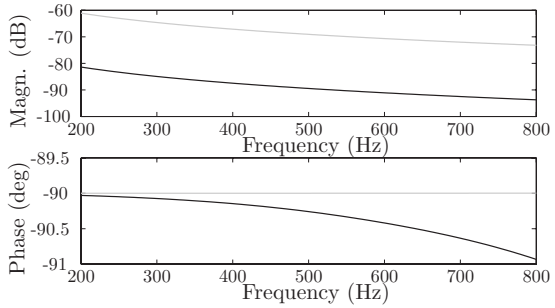


FIG. 2. Patch input mobility—black line: porogranular mobility; gray line: air cavity mobility (after Ref. 11)—patch 42—cavity: L_x=1.5 m, L_y=0.96 m, L_z=0.01 m—patch size: Δx=7.9 cm, Δy=7.4 cm—Patch 42: X=[0.24 m;0.32 m], Y=[0.15 m;0.22 m].

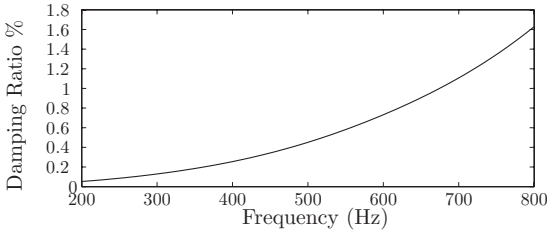


FIG. 3. Damping ratio evaluated from the patch input mobility—patch 42—cavity: L_x=1.5 m, L_y=0.96 m, L_z=0.01 m—patch size: Δx=7.9 cm, Δy=7.4 cm—patch 42: X=[0.24 m;0.32 m], Y=[0.15 m;0.22 m].

To quantify the damping associated with the porogranular material, the patch input impedance is identified as a complex stiffness K^* with an imaginary part characterized by a damping loss factor η such as $K^*=K(1+j\eta)$. The corresponding mobility has the following form:

$$Y = -\frac{j\omega S}{K(1+j\eta)}. \tag{26}$$

One can deduce that $\eta=\mathrm{Re}(Y)/\mathrm{Im}(Y)$. This damping loss factor η presented in Fig. 3 shows the expected ability of the porogranular material to dissipate energy. The damping loss factor is increasing with frequency and reach 0.016 at 800 Hz. Of course this analysis only characterizes the dissipation observed at a given patch, and cannot be seen as a complete characterization of the damping properties of the porogranular material that is not under study in this paper. However, the dissipation is of major importance to limit the transmission at the mass-air mass frequency.

In Fig. 4, the patch transfer mobility (PTM) of a porogranular material and air cavity are compared. The magnitude of the PTM is lower for the porogranular material due to a higher stiffness. However, the phase of the porogranular transfer mobility varies with frequency. These variations are related to the propagation time of solid and fluid waves between excitation and observation points and also to viscothermal dissipations inside the material.

B. Comparison of measured and calculated TL

Figure 5 presents the experimental setup. Measurements of transmitted power were done using intensity technique with a 50 mm intensity probe located at 20 cm from the

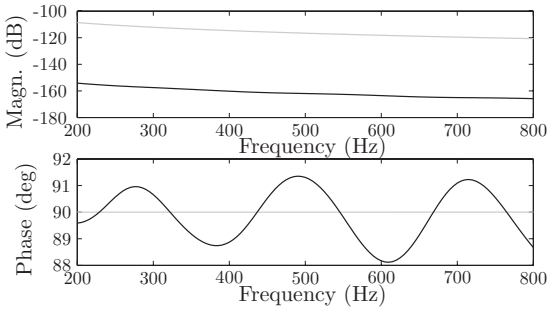


FIG. 4. Patch transfer mobility—black line: porogranular mobility; gray line: air cavity mobility (after Ref. 11)—patches 42 and 72—cavity: L_x=1.5 m, L_y=0.96 m, L_z=0.01 m—patch size: Δx=7.9 cm, Δy=7.4 cm—patch 42: X=[0.24 m;0.32 m], Y=[0.15 m;0.22 m], Z=0—patch 72: X=[1.11 m;1.18 m], Y=[0.22 m;0.30 m], Z=0.01 m.

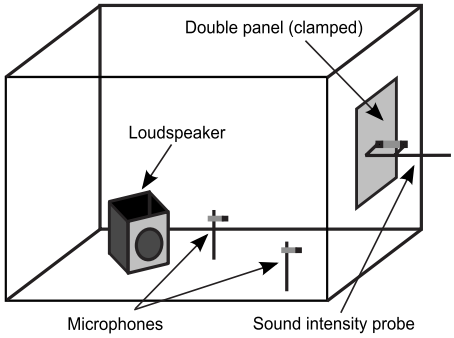


FIG. 5. Experimental setup.

radiating panel. 100 points have been used to measure radiated intensity. Source room is a large reverberant room (lx=11.5 m, ly=8.69 m, lz=4.03 m, cutt off frequency =187 Hz) where walls are not parallel and with diffusers in the corners. Five microphones in the reverberant chamber enabled to measure reverberant pressure. Receiving room is a large anechoic space treated to be isolated from exterior noise.

Transmission loss of a double aluminum panel filled with a porogranular material was measured. The double panel dimensions were as follows: width 0.96 m, length 1.5 m, thicknesses 2 and 1.5 mm, intermediate cavity thickness 1 cm. Panel critical frequencies are 7961 Hz (respectively, 5971 Hz) for the aluminum panel of 1.5 mm (respectively, 2 mm) thickness. The porogranular material tested was made of expanded polystyren beads of very low density (19 kg/m^3) and non-cohesive. Acoustic parameters of this material are given in Table I.

The experiment presents the case of a cavity not fully filled in order to allow a relative motion between beads. It is indeed important to not couple the two panels in order to keep the good insulation properties of the double panel. A small gap of air is hence kept at the top of the cavity. However, due to gravity a small preload on the porogranular material exists.

Figure 6 presents comparative transmission loss of a double panel measured and calculated with the patch-mobility method. The model gives good comparative results. The bulk modulus of the porogranular material solid phase given in Ref. 24 has, however, been modified to get these results. As mentioned previously, the material being maintained in a limited space with an initial load, the bulk modulus is also increased compared to data in Ref. 24, obtained in standing wave tube, on a layer of porogranular material totally free of compression. A good comparison of transmission loss was observed for a bulk modulus increased to 0.35 MPa.

Moreover, the bulk modulus depends also on the behavior of the granular material. Basically, three behaviors can be distinguished:[35] quasistatic, rapid granular flow, and intermediate behaviors. In the present case of a cavity not fully filled with a light porogranular material, the intermediate behavior is the more likely. Indeed, the beads can slightly move in the cavity but a rapid granular flow is not possible due to the small cavity thickness and to the small acoustic displace-

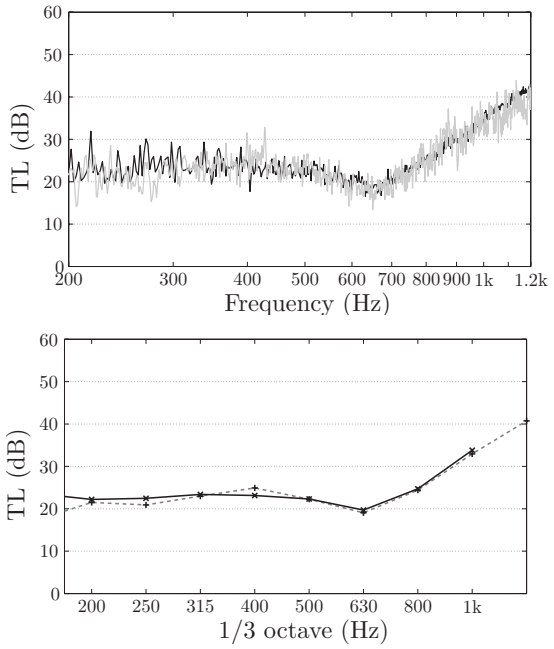


FIG. 6. Comparison with experiment—black line: model; gray line: experiment—double aluminum panel (L_x=0.96 m, L_z=1.5 m), 2 and 1.5 mm thickness, separated by a cavity of 1 cm filled with expanded polystyrene beads, centered on point X=6 m, Y=0 m, Z=1.75 m—room dimensions: lx=11.5 m, ly=8.69 m, lz=4.03 m—patch size: ΔX=0.08 m, ΔZ=0.074 m—source at Xs=2 m, Ys=4 m, Zs=1 m, amplitude: S_0=2—cut off frequency: 187 Hz—structural damping: 0.01.

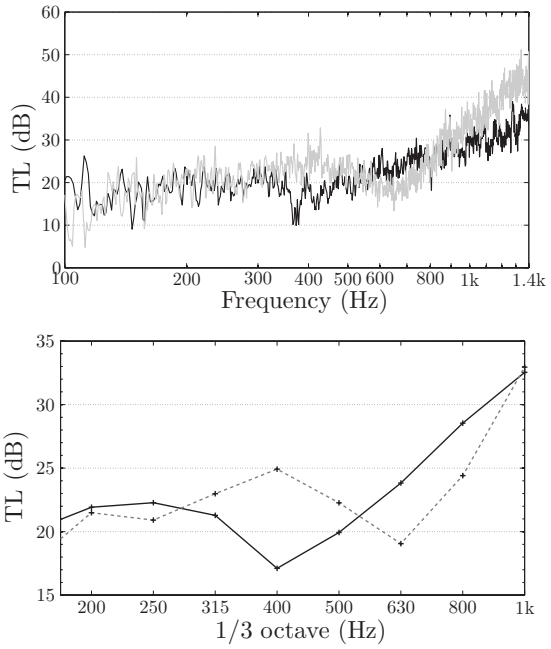


FIG. 7. Experimental comparisons—black solid line: air cavity; gray dotted line: cavity filled with expanded polystyrene beads—double aluminum panel (L_x=0.96 m, L_z=1.5 m), 2 and 1.5 mm thickness, separated by a cavity of 1 cm, centered on point X=6 m, Y=0 m, Z=1.75 m—room dimensions: lx=11.5 m, ly=8.69 m, lz=4.03 m—source at Xs=2 m, Ys=4 m, Zs=1 m—cut off frequency: 187 Hz.

ments. However, if the filling of the cavity is increased, the preload on the material can increase due to a compacting of the beads and the quasistatic behavior can become predominant. In that case, it will be necessary to change the bulk modulus in the model in order to take into account this change of behavior. However, a linear increase in the bulk modulus with the preload is unlikely because of the point contact between the beads. This non-linear behavior is, for example, well known for fibrous materials where the apparent stiffness depends on the strain level at which the sample is tested.[36]

Figure 7 presents a comparison of experimental results obtained for double panels filled with a porogranular material and with an air cavity. In the case of a cavity filled with a porogranular material, the dip at mass-air-mass resonance frequency is visible at 630 Hz instead of 400 Hz. This slight increase is due to the stiffness added by the porogranular material. At higher frequency, the increase of TL is still present and confirms that higher stiffness does not suppress the double plate behavior. Moreover, the transmission loss gradient after the mass-air-mass resonance frequency is increased of 40% with the porogranular material. This study on a standard granular material shows therefore that the TL of a double panel can be improved. However, this particular case is certainly not the optimal combination of material and cavity filling. Indeed, the frequency resonance can still be modified and adapted through the material properties and the cavity filling. Practical applications in building and transportation industries of hollow panels filled with porogranular material are therefore promising.

VIII. CONCLUSIONS

In this paper, the patch-mobility method was used to predict the sound transmission loss through finite double panels filled with a porogranular material. Porogranular material was taken into account with Biot's fluid-fluid model based on a simplification of the standard Biot's poroelastic model. This simplification was an alternative to other simplifications like the limp model or the rigid frame model. It permitted to neglect the shear stress and to take into account the solid frame stiffness in order to describe the double panel behavior. Numerical results obtained with this model were compared to transmission loss measurements made on a double aluminum panel including expanded polystyrene beads. It appeared hence that the TL trends were correctly predicted by this model. However, a modified bulk modulus was necessary to take into account the static compression of beads induced by the mounting of the double panel. Finally, very light porogranular materials placed in double plate cavity appeared to be an interesting way to improve sound insulation through double panels.

APPENDIX A: RADIATION PATCH IMPEDANCE

The radiation patch impedance $\langle\langle Z\rangle\rangle_{ij}$ is defined as the ratio of averaged patch i radiated pressure to averaged patch j velocity: $\langle\langle Z\rangle\rangle_{ij}=\langle P_{\mathrm{rad}}\rangle_i/\langle V\rangle_j$.

Calculations are based on Rayleigh's integral and are detailed in Ref. 11. When emitting and receiving patches are different (i.e., $i\neq j$), Eq. (A1) is used to calculate radiation impedance where d_{ij} is the distance between two patches central points. A particular case is defined when emitting and receiving patches are the same (i.e., $i=j$). In this case, radia-

tion impedance expression is given by Eq. (A2) where a_i is the radius of a circular patch of surface S_i. The approximation employed to get these expressions is valid for small enough patch surface, and in general a discretization based on the standard $\lambda/6$ criterion is sufficient.

$$\langle\langle Z\rangle\rangle_{ij} = \frac{\langle P_{\text{rad}}\rangle_i}{\langle V\rangle_j} = \frac{1}{2\pi}\rho_0 j \frac{\exp^{-jkd_{ij}}}{d_{ij}} S_j, \tag{A1}$$

$$\langle\langle Z\rangle\rangle_{ii} = \frac{\langle P_{\text{rad}}\rangle_i}{\langle V\rangle_i} = \rho_0 c[1 - \exp^{-jka_i}]. \tag{A2}$$

APPENDIX B: FLUID-FLUID MODEL

The presented fluid-fluid model is detailed in Ref. 24. Elastic coefficients used in the fluid-fluid model write

$$A = F(P\rho_{22} - Q\rho_{12}), \quad B = F(-P\rho_{12} + Q\rho_{11})\phi/(1-\phi),$$

$$C = F(-Q\rho_{12} + R\rho_{11}), \quad D = F(Q\rho_{22} - R\rho_{12})(1-\phi)/\phi, \tag{B1}$$

$$F = [(\rho_{22}\rho_{11} - \rho_{12}^2)\omega^2]^{-1}.$$

These coefficients are related to the porosity ϕ, elastic coefficients P, Q, and R, and equivalent densities ρ_{ij}. Elastic coefficients and equivalent densities are standard coefficients used in Biot–Allard's formulation with time dependence $e^{+j\omega t}$ and are, respectively, recalled in Eqs. (B2) and (B3).

$$P = \frac{4}{3}N + K_b + \frac{(1-\phi)^2}{\phi}K_f,$$

$$Q = \frac{R(1-\phi)}{\phi}, \tag{B2}$$

$$R = \phi K_f,$$

$$\rho_{11} = \rho_s + \phi(\rho_f - \rho_0),$$

$$\rho_{12} = \phi(\rho_0 - \rho_f), \tag{B3}$$

$$\rho_{22} = \phi\rho_f.$$

N, K_b, and ρ_s are, respectively, the shear modulus, the bulk modulus, and the density of the solid frame. Viscous and thermal dissipations are, respectively, taken into account by frequency dependent expressions of fluid density ρ_f and dynamic fluid compressibility K_f. Johnson–Allard's expressions[37] given in Eq. (B4) are used to describe the microgeometry structure with five parameters: porosity ϕ, tortuosity α_∞, viscous permeability k_0 (related to air flow resistivity by $\sigma = \eta/k_0$ with k_0 the dynamic fluid viscosity), and viscous and thermal characteristic lengths (respectively, Λ and Λ'). Pr is the Prandtl number.

$$\rho_f = \rho_0\alpha_\infty\left(1 + \frac{\eta\phi}{j\rho_0\alpha_\infty k_0\omega}\sqrt{1 + 4\frac{k_0^2\alpha_\infty^2 j\omega\rho_0}{\Lambda^2\phi^2\eta}}\right), \tag{B4}$$

$$K_f = \frac{\gamma P_0}{\gamma - (\gamma-1)\left(1 + \frac{8\eta}{j\Lambda'^2\,\text{Pr}\,\omega\rho_0}\sqrt{1 + j\rho_0\frac{\text{Pr}\,\Lambda'^2\omega}{16\eta}}\right)^{-1}}.$$

However, two different definitions can be employed for the elastic coefficient P. If one considers a light and noncohesive porogranular material without shear stress, the elastic coefficient P stands for the skeleton bulk modulus with a coupling part due to the fluid phase: $P = K_b + ((1-\phi)^2/\phi)K_f$. Compared to the classical coefficient P used in Biot's model, the shear term $\frac{4}{3}N$ has been removed by simply considering a fluid behavior in the stress-strain relations (i.e., hydrostatic stress tensor). On the contrary, when the aim is to neglect the shear stress in a material that does not behave like a fluid, it is also possible to keep the same coefficient P than the one used in Biot's formulation: $P = (4/3)N + K_b + ((1-\phi)^2/\phi)K_f$. This enables to keep the same dispersion equation in both formulations. Of course, when N is very small compared to $K_b + ((1-\phi)^2/\phi)K_f$, both definitions of P are similar. In the following, the first definition is employed.

[1] A. London, "Transmission of reverberant sound through single walls," J. Res. Natl. Bur. Stand. **42**, 605–615 (1949).
[2] A. London, "Transmission of reverberant sound through double walls," J. Acoust. Soc. Am. **22**, 270–279 (1950).
[3] J. Brunskog, "The influence of finite cavities on the sound insulation of double-plate structures," J. Acoust. Soc. Am. **117**, 3727–3739 (2005).
[4] J. Guyader and C. Lesueur, "Acoustic transmission through orthotropic multilayered plates, part ii: Transmission loss," J. Sound Vib. **58**, 69–86 (1978).
[5] J. Guyader and C. Lesueur, "Transmission of reverberant sound through orthotropic viscoelastic multilayered plates," J. Sound Vib. **70**, 319–332 (1980).
[6] E. Nilsson and A. Nilsson, "Prediction and measurement of some dynamic properties of sandwich structures with honeycomb and foam cores," J. Sound Vib. **251**, 409–430 (2002).
[7] M. Munjal, "Response of a multi-layered infinite plate to an incident oblique plane wave by means of transfer matrices," J. Sound Vib. **162**, 333–343 (1993).
[8] J. Sastry and M. Munjal, "A transfer matrix approach for evaluation of the response of a multi-layer infinite plate to a two-dimensional pressure excitation," J. Sound Vib. **182**, 109–128 (1995).
[9] R. Panneton and N. Atalla, "Numerical prediction of sound transmission through finite multilayer systems with poroelastic materials," J. Acoust. Soc. Am. **100**, 346–354 (1996).
[10] M. Ouisse, L. Maxit, C. Cacciolati, and J. Guyader, "Patch transfer functions as a tool to couple linear acoustics problems," ASME J. Vibr. Acoust. **127**, 458–466 (2005).
[11] J. Chazot and J. Guyader, "Prediction of transmission loss of double panels with a patch-mobility method," J. Acoust. Soc. Am. **121**, 267–278 (2007).
[12] A. T. Moorhouse, "A dimensionless mobility formulation for evaluation of force and moment excitation of structures," J. Acoust. Soc. Am. **112**, 972–980 (2002).
[13] S. Naji, "Etude des transmissions vibratoires par une méthode de mobilité mixte dans les assemblages par surface (Study on vibration transmissions with a mobility method)," Ph.D. thesis, Universit Claude Bernard Lyon I (1993).
[14] G. Orefice, C. Cacciolati, and J. Guyader, "The energy mobility," J. Sound Vib. **254**, 269–295 (2002).
[15] M. Biot, "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**, 168–178 (1962).
[16] W. Lauriks, A. Cops, J. Allard, C. Depollier, and P. Rebillard, "Modelization at oblique incidence of layered porous materials with impervious screens," J. Acoust. Soc. Am. **87**, 1200–1206 (1990).
[17] R. Panneton and N. Atalla, "An efficient finite element scheme for solving the three dimensional poroelasticity problem in acoustics," J. Acoust. Soc. Am. **101**, 3287–3298 (1997).
[18] N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure

formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1452 (1998).
[19]L. Jaouen, B. Brouard, N. Atalla, and C. Langlois, "A simplified numerical model for a plate backed by a thin foam layer in the low frequency range," J. Acoust. Soc. Am. **280**, 681–698 (2005).
[20]O. Dazel, C. Sgard, F. Lamarque, and N. Atalla, "An extension of complex modes for the resolution of finite-element poroelastic problems," J. Sound Vib. **253**, 421–445 (2002).
[21]C. Zwikker and C. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949).
[22]L. Beranek, "Acoustical properties of homogeneous isotropic rigid tiles and flexible blankets," J. Acoust. Soc. Am. **19**, 556–568 (1947).
[23]O. Doutres, N. Dauchez, J. Gnevaux, and O. Dazel, "Validity of the limp model for porous materials: A criterion based on the Biot theory," J. Acoust. Soc. Am. **122**, 2038–2048 (2007).
[24]J. Chazot and J. Guyader, "Acoustic modeling of light and non-cohesive poro-granular materials with a fluid/fluid model," Acta Mech. **195**, 227–247 (2008).
[25]B. Nennig, J. Chazot, E. Perrey-Debain, and M. Ben Tahar, "Influence of solid phase elasticity in poroelastic liners submitted to grazing flows," in Proceedings of Euronoise, Paris, France (2008).
[26]B. Nennig, J. Chazot, E. Perrey-Debain, and M. Ben Tahar, "Evaluation of simplified poroelastic models," in Proceedings of SAPEM, Bradford, UK (2008).
[27]Z. Hong, L. Bo, H. Guangsu, and H. Jia, "A novel composite sound absorber with recycled rubber particles," J. Sound Vib. **304**, 400–406 (2007).
[28]M. Saeki, "Impact damping with granular materials in a horizontally vibrating system," J. Sound Vib. **251**, 153–161 (2002).
[29]Z. Xu, M. Wang, and T. Chen, "An experimental study of particle damping for beams and plates," ASME J. Vibr. Acoust. **126**, 141–148 (2004).
[30]J. Allard, M. Henry, J. Tizianel, L. Kelders, and W. Lauriks, "Sound propagation in air-saturated random packings of beads," J. Acoust. Soc. Am. **104**, 2004–2007 (1998).
[31]J. Chazot and J. Guyader, "Prediction of sound transmission through double panels filled with granular materials," in Proceedings of Internoise, Honolulu, HI (2006), pp. 227–247.
[32]F. Xin, T. Lu, and C. Chen, "Vibroacoustic behavior of clamp mounted double-panel partition with enclosure air cavity," J. Acoust. Soc. Am. **124**, 3604–3612 (2008).
[33]F. Xin and T. Lu, "Analytical and experimental investigation on transmission loss of clamped double panels: Implication of boundary effects," J. Acoust. Soc. Am. **125**, 1506–1517 (2009).
[34]P. Debergue, R. Panneton, and N. Atalla, "Boundary conditions for the weak formulation of the mixed (u,p) poroelasticity problem," J. Acoust. Soc. Am. **106**, 2383–2390 (1999).
[35]R. Daniel, A. Poloski, and A. Sez, "A continuum constitutive model for cohesionless granular flows," Chem. Eng. Sci. **62**, 1343–1350 (2007).
[36]B. Song, J. Bolton, and Y. Kang, "Effect of circumferential edge constraint on the acoustical properties of glass fiber materials," J. Acoust. Soc. Am. **110**, 2902–2916 (2001).
[37]J. Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Chapman and Hall, London, 1993).

Measurement of characteristic impedance and wave number of porous material using pulse-tube and transfer-matrix methods

Liang Sun, Hong Hou,[a] Li-ying Dong, and Fang-rong Wan
College of Marine, Northwestern Polytechnical University, 710072 Xi'an, People's Republic of China

(Received 31 October 2008; revised 3 June 2009; accepted 7 September 2009)

The time domain implementation of the transfer-matrix method developed by Song and Bolton [J. Acoust. Soc. Am. **107**, 1131–1154 (2000)] for measuring the characteristic impedance and wave number of porous materials is described in this paper. The so called Butterworth impulse is generated in a standing wave tube with a flat frequency response over a wide frequency range. With only two microphone measurements, the transfer matrix of porous layers can easily be determined through the calculation of complex amplitudes of incident, reflected, and transmitted pulses. The procedure has been used to measure the acoustical properties of a fiber material, and good agreement was found between measured acoustical properties and predicted results by Delany and Bazley [Appl. Acoust. **3**, 105–116 (1971)] semiempirical formulas. Although the error associated with the sample-edge constraint still remains, the new method has a better frequency response as a result of the system calibration process, and the optimal inter-microphone distance is no longer required compared to the frequency domain implementation.
 [DOI: 10.1121/1.3242354]

PACS number(s): 43.58.Bh [RR] Pages: 3049–3056

I. INTRODUCTION

As an effective means of controlling environmental noise, absorptive material and its performance are always of interests to acoustic engineers and scientists. In understanding and evaluating the effects of a material in particular applications, detailed acoustic analyses normally require material properties, such as surface impedance, characteristic impedance, and wave number, or even material macroscopic properties, as the inputs.[1–4] Nowadays determination of these properties still relies on various experimental techniques.

Traditional measurement approaches, and also today's widely adopted approaches for characteristics impedance and wave number determination, include two-thickness,[5] two-cavity methods,[6–9] both of which are carried out in an impedance tube, also known as the standing wave tube. Basically it requires two separate measurements on the surface impedance of the material sample either with different thicknesses or different backing spaces. Many papers have been published with regard to the limitation, error analysis, improved data acquisition, and post-processing techniques based on the two methods.[7] Song and Bolton[10] gave a thorough and comprehensive review on the development history of the technology in their paper.

In the same paper, however, Song and Bolton[10] presented a new measurement approach that was developed based on the transfer-matrix theory. By utilizing the reciprocal principle naturally imposed to the material sample, the characteristic impedance and the wave number are related to the amplitudes of the four approaching waves on both sides of the sample, which successfully eliminated possible variations associated with mounting two samples in the whole process, possible errors induced by insufficient differences in the surface impedance between two thicknesses due to high flow resistivities, and errors for those high dissipative materials of which the surface impedance is not sensitive to backing conditions. Although it was made clear that the method does not apply to elastic porous materials, such as foam, because of the formula taken in transfer-matrix elements, testing result shows promising correlation with Delany and Bazley's empirical predictions for fiber materials that can be considered to be either limp or rigid.[10]

It is worth noting, however, that in these methods, data acquisition and processing were mostly done in the frequency domain, which inevitably leads to either conducting separate measurements on different sample configurations, or setting up a multi-channel data acquisition/processing system. The transfer-matrix method described in Ref. 11 needs four microphone measurements to retrieve the amplitudes of four approaching waves in the standing wave tube. In addition, an optimal inter-microphone spacing should be honored for a given frequency range of interest.

Recently, Jing and Fung[12] developed an impulsive sound source, which can generate repeatable sharp pulses with durations as short as 0.5 ms. Hou and co-workers[13,14] used the echo-pulse tube to measure the absorption coefficient of porous materials through the generation of a kind of pulse referred to as Butterworth pulse. The measurement was conducted in the time domain and in a way that the incident and the reflected pulses could be separated from each other such that the amplitude of each wave can be determined.[15] The technique was claimed to be an effective approach for *in situ* sound absorption coefficient measurements.

[a]Author to whom correspondence should be addressed. Electronic mail: houhong@nwpu.edu.cn

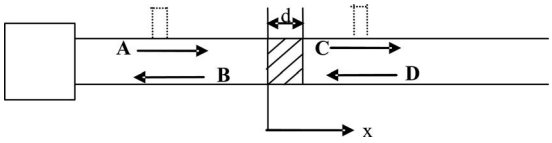


FIG. 1. Schematic of the echo-pulse tube.

The work presented here basically lies within the framework established by Song and Bolton.[10] The transfer-matrix representation of a porous material layer is still limited to be either limp or rigid. The echo-pulse technique is employed to directly measure the amplitudes of four approaching waves, which only requires two microphone measurements. For a chosen fiber material, the complex characteristic impedance and the complex wave number are extracted and compared to those predicted using the Delany and Bazley empirical model. It appears that the approach is able to deliver reliable measurements, and results are even more robust to the tube end conditions.

II. MEASUREMENT PRINCIPLE

Detailed transfer-matrix analysis can be found in Ref. 10. For a homogeneous, isotropic porous material layer that can be represented by what Allard described as the limp or rigid frame model, there is

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} \cos k_p d & i\rho_p c_p \sin k_p d \\ i \sin k_p d/\rho_p c_p & \cos k_p d \end{bmatrix}, \tag{1}$$

where T_{ij} is the transfer-matrix elements; and c_p, k_p, and ρ_p are complex sound speed, complex wave number, and the complex density of the material, respectively. In here, T_{ij} is related to the sound pressures P and normal acoustic particle velocities V on the two surfaces of material layer as

$$\begin{bmatrix} P \\ V \end{bmatrix}_{x=0} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} P \\ V \end{bmatrix}_{x=d}, \tag{2}$$

where $x=0$ and $x=d$ denote the two surfaces of the material sample, as shown in Fig. 1. Since P and V at $x=0$ and $x=d$ can also be expressed as the superposition of the four approaching waves in the tube, we have

$$P|_{x=0} = A + B, \quad V|_{x=0} = \frac{A - B}{\rho_0 c_0},$$

$$P|_{x=d} = Ce^{-ikd} + De^{ikd}, \quad V|_{x=d} = \frac{Ce^{-ikd} - De^{ikd}}{\rho_0 c_0}, \tag{3}$$

where A, B, C, and D are complex amplitudes of incident wave, reflected wave from the sample, transmitted wave, and reflected wave from the tube termination, respectively. ρ_0 is the ambient fluid density and c_0 is the ambient sound speed. With established four equations in Eq. (3), each transfer-matrix element T_{ij} can be expressed as a function of A, B, C, and D. Thus if the complex amplitudes of four approaching waves can be determined experimentally, the complex wave number and the characteristic impedance can be extracted through

$$\rho_p c_p = \sqrt{\frac{T_{12}}{T_{21}}}, \tag{4}$$

$$k_p = \frac{1}{d}\cos^{-1} T_{11} \quad \text{or} \quad k_p = \frac{1}{d}\sin^{-1}\sqrt{-T_{12}T_{21}}. \tag{5}$$

When a pulse is generated in a tube and propagates in the $x+$ direction toward the sample as the incident pulse (A), the first reflection will happen at the sample surface at $x=0$, and a reflected pulse (B) will be generated and propagates in the $x-$ direction. Meanwhile, the transmitted pulse (C) will be through the sample surface at $x=d$ and propagates along the $x+$ direction. This pulse will be reflected at the end of the tube as the second reflected pulse (D) unless an anechoic end condition is presented. Using M_1 and M_2 as two fixed reference locations in the upper and down streams of the sample, it can be expected that there would be a delay in the time domain between the incident pulse and the first reflected pulse passing through location M_1 and that between the transmitted pulse and the second reflected pulse through location M_2. If the pulse is short enough in width, the distance between location M_1 and the sample surface $x=0$, and the distance between M_2 and the tube end are long enough, the incident and the first reflected pulses can be well separated from each other, and so can the transmitted and the second reflected pulses be. As such, the amplitude of each wave can be determined without much difficulty given today's signal processing technology. It can be seen that this measurement approach only needs two microphone measurements at M_1 and M_2. As long as the pulses are well separated, the end condition of the tube would not impose any influence on the testing results. For all the testing results shown in this paper, the tube end was left open without any particular acoustical treatments.

III. EXPERIMENTAL PROCEDURE

Figure 2 shows the experimental setup for the measurements, which includes a standing wave tube, a speaker-pipe system, an NI DAQ6062E data acquisition board controlled by a computer, a BNC2029 adaptor, a power amplifier, and a pre-amplifier. The speaker system is a 5 in. Hi-Vi mid-frequency woofer positioned inside a steel cylinder of which the length is 15 cm. The sound wave from the speaker is guided through a 20 cm long conical reduction to a straight standing wave tube of 27 mm in diameter. The material sample is placed at 72 cm from the end of the tube, and 68 cm away from the speaker. Two measurement locations were chosen at 42 cm away from the sample for M_1, and at 25 cm away from the tube end for M_2. In the testing, only one microphone, a $\frac{1}{4}$ in. MPA416 condenser microphone, was used to record the data at two locations separately. The analogy input (AI) and analogy output (AO) channels of the NI DAQ 6062E board are programed to be synchronized and have the sampling rate of 100 kHz.

In Ref. 14, it has been shown that the so called Butterworth pulse is considered to be good for acoustic measurement because of its regular waveform, short duration, and

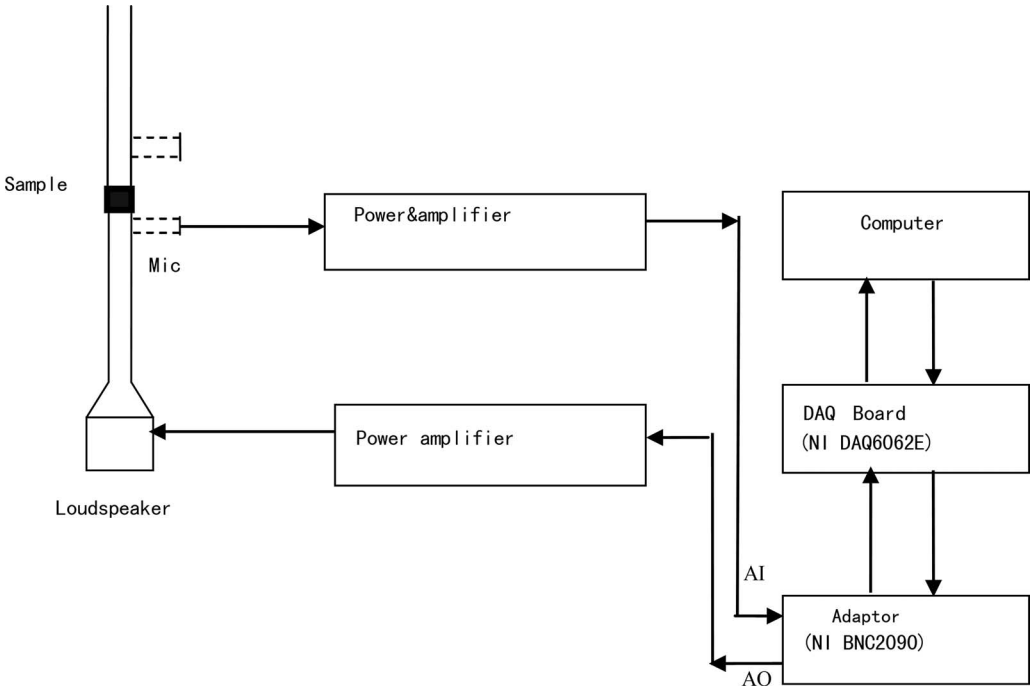


FIG. 2. Sound source configuration and test equipment.

wide frequency range. In this testing, this pulse was used. Figure 3 shows a typical Butterworth pulse waveform and its corresponding frequency spectrum.

In order to generate a desired acoustic pulse signal in the standing wave tube, the impulse response of the sound generator system, including the speaker, power amplifier, etc., should be calibrated so that the appropriate driving signal could be derived. For this purpose, an impulse of fifth-order Butterworth filter shape with 1 ms duration was fed into the system. The acoustic response signal was recorded by a microphone located in the straight standing wave tube. The frequency response of the sound generation system $H(\omega)$ thus can be obtained by

$$H(\omega) = \frac{H_a(\omega)}{H_e(\omega)}, \tag{6}$$

where $H_e(\omega)$ and $H_a(\omega)$ are the Fourier transforms of the excitation and response signals, respectively. For a given desired spectrum $H_b(\omega)$, which is the Butterworth pulse spectrum in here, the spectrum of the driving signal $H_d(\omega)$ can be computed by

$$H_d(\omega) = \frac{H_b(\omega)}{H(\omega)}. \tag{7}$$

Thus the driving signal in the time domain would simply be obtained through the inverse Fourier transform of Eq. (7).

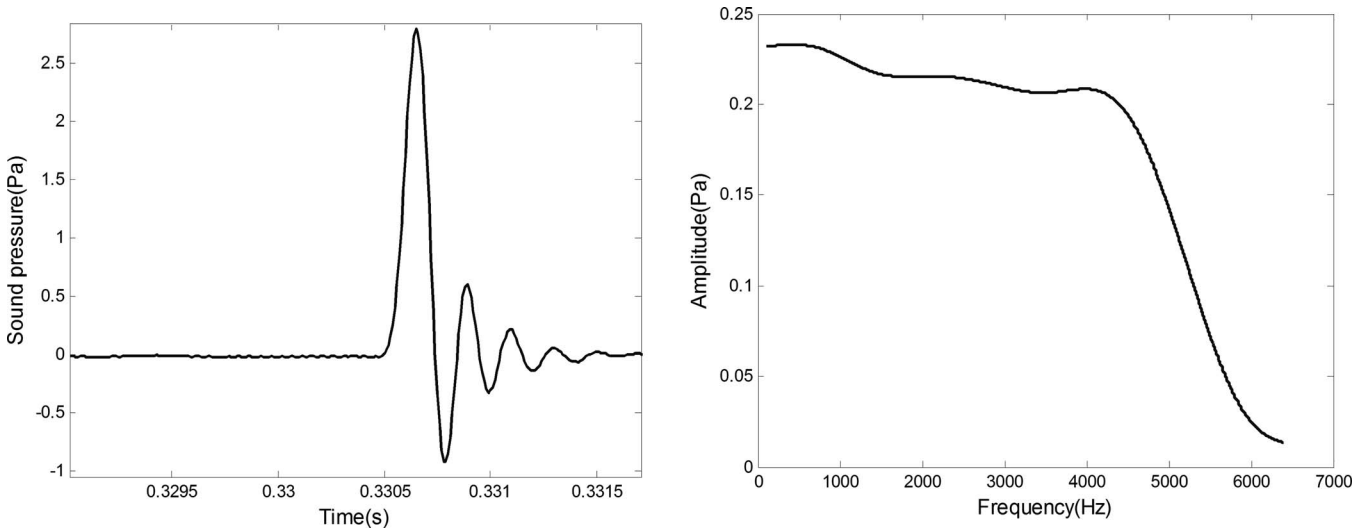


FIG. 3. The Butterworth pulse and its spectrum.

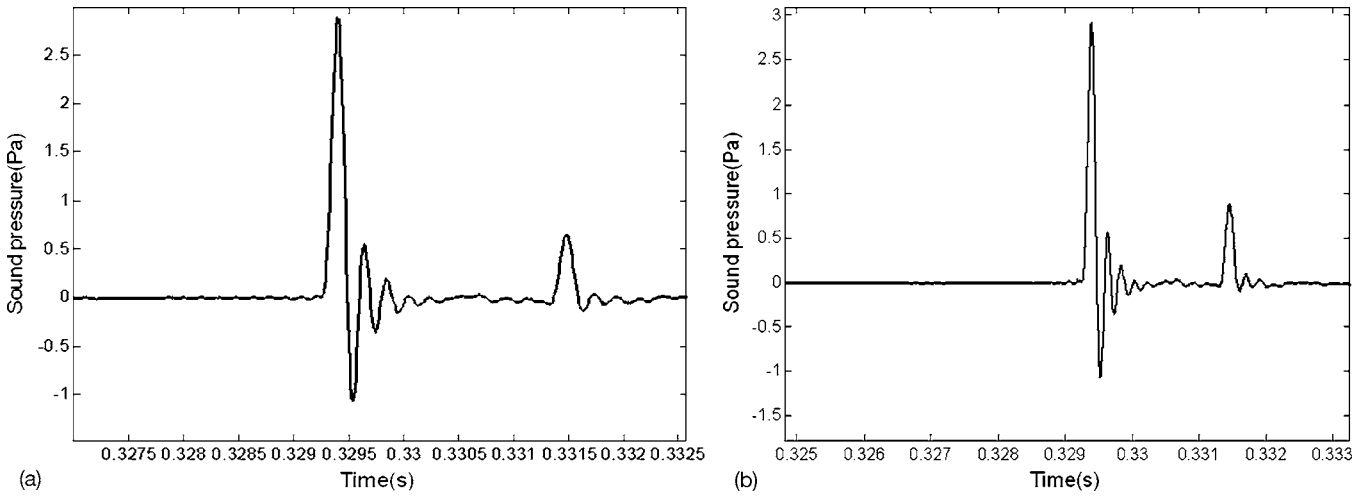


FIG. 4. Measured waveforms of incident and the first reflected pulses with 50 averages.

Compared to traditional methods or Song and Bolton's approach, this process seems to need extra effort in system setups. But for a particular measurement setup, this calibration only needs to be done once. Nevertheless, it should be noted that through this process, any deficiencies in the frequency response of the sound generation system will be compensated, which would improve the signal-to-noise ratio in the measurement.

IV. EXPERIMENTAL RESULTS

A glass fiber material was chosen for the validation testing. The material bulk density is 24 kg/m^3, and the measured flow resistivity is 13 100 Rayls/m. Two samples, 20 and 30 mm in thickness, were prepared. In testing, the pulse width is about 1 ms. For each microphone measurement, the synchronized time signal is averaged through 50 pulse generations to improve the signal-to-noise ratio. Measured waveforms of incident and the first reflected pulses (from the measured sample), transmitted, and the second reflected pulses (from the tube end) with respect to these two samples are shown in Figs. 4 and 5.

Rectangular window was adopted in this work to crop each pulse in the time domain. For each separated pulse, a 65 536 point fast fourier transform (FFT) was then performed to extract coefficients A, B, C, and D in the frequency domain. It should be noted that a successful separation of those pulses basically relies on the time delay between two adjacent pulses passing through the measurement point, which actually depends on the distance between the upstream measurement location and the sample, and that between the downstream measurement location to the tube end. For a pulse of 1 ms in duration, both distances should be greater than 17 cm (assume $c_0=340$ m/s) for two adjacent pulses be distinguished in the time domain. Previous study[14] has shown that as long as the pulses can be separated, very little impact to the final results was found with regard to measurement locations. This can be considered as another advantage over traditional multi-microphone method, which needs to identify the microphone and sample positions precisely.[15]

As a comparison, predictions made by utilizing the Delany and Bazley semiempirical formulas are also plotted.[16]

From Figs. 6–9, the measured complex wave number, complex material density, and the complex sound speed are shown. Good agreement can be seen between the method

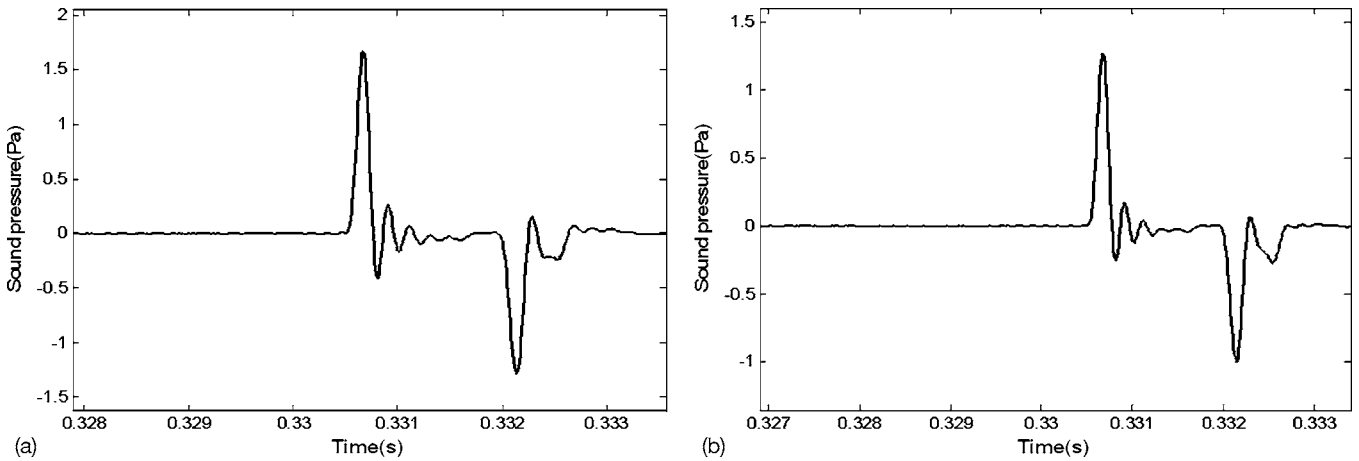


FIG. 5. Measured waveforms of transmitted and the second reflected pulses with 50 averages.

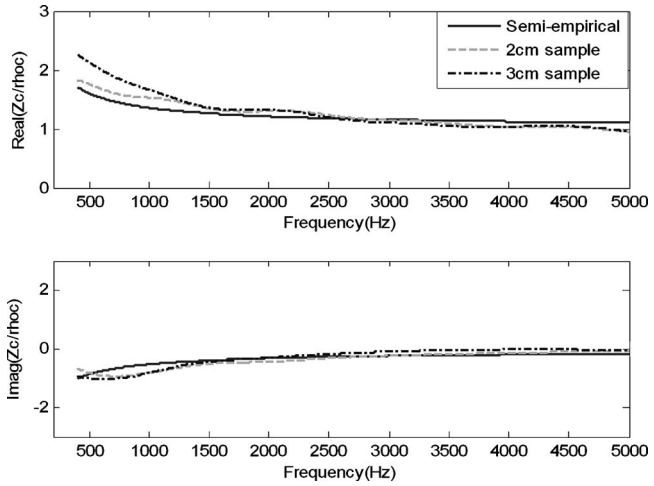


FIG. 6. Normalized characteristic impedance of fiber material determined by using transfer-matrix method and predicted using semiempirical formulas.

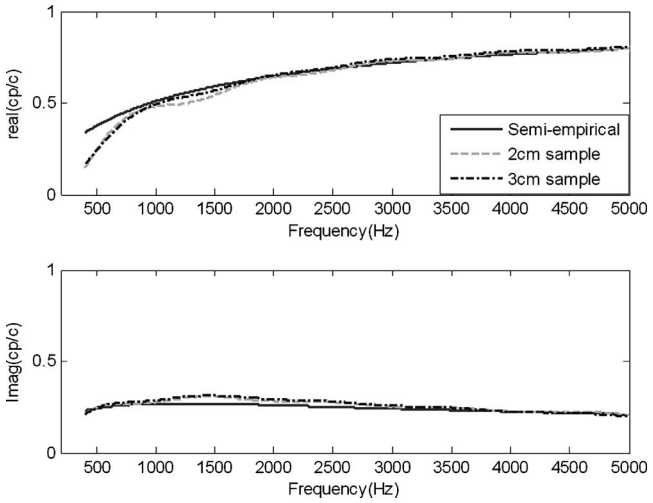


FIG. 8. Normalized complex sound speed of fiber material determined by using transfer-matrix method and predicted using semiempirical formulas.

presented in this paper and predicted results. Also, measured results are irrespective to the sample thickness as expected. Some disagreements occurred at lower frequencies might be related to the sample-edge constraints as discussed in Ref. 10.

V. SENSITIVITY AND ERROR ANALYSIS

To evaluate our new method correctly, uncertainty analysis of measurement system and computations is presented here followed by some disagreement analysis. From Eqs. (3) and (5), one can see that the calculated characteristic impedance and wave number values are mainly related with the amplitudes of four approaching waves A, B, C, and D, as well as the sample thickness d, sound speed c, and the characteristic impedance $\rho_0 c$ in air, which are closely related to environment temperature. Simulations are therefore made to investigate the impact of uncertainties of above contributors on the estimated characteristic impedance and wave number values. The following values are used in our simulations that are typical in most acoustic laboratories: Variations in amplitudes of all pulses are assumed to be $\pm 1\%$, and uncertainties in sample thickness are ± 1 mm. As sound speed and characteristic impedance of air are affected by ambient temperature, variation is chosen as ± 1 °C.

To analyze the impact of these uncertainties, the Monte Carlo method[17] is employed. For each variable, the Gaussian distribution is adopted with a standard deviation (σ) defined as above. A random perturbation, which is limited to $\pm 2\sigma$, is then added to the appropriate input variable to calculate the new Zc and kp values. This procedure is repeated until the mean values and standard deviations converge. In our simulations, convergence can be reached after 20 000 iterations. To ensure an accurate result, the analysis was run for 100 000 iterations.

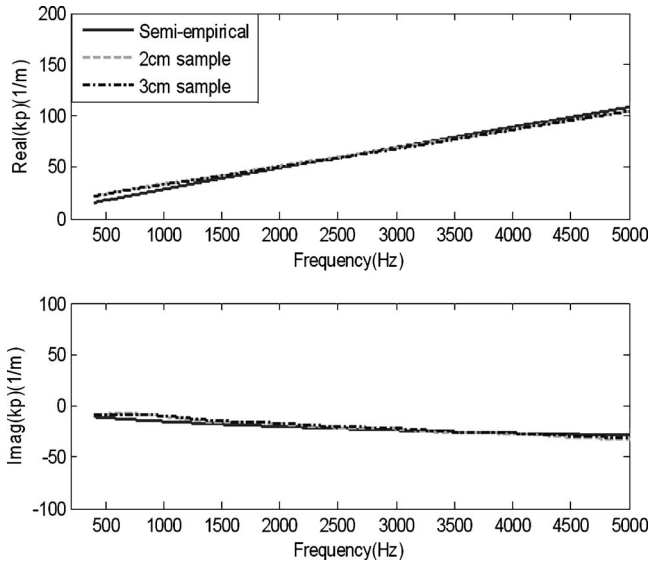


FIG. 7. Wave number within fiber material determined by using transfer-matrix method and predicted using semiempirical formulas.

A. Errors from measurement system itself

Measuring system can introduce errors in the process of pulse generation, signal sampling, and its conditioning.

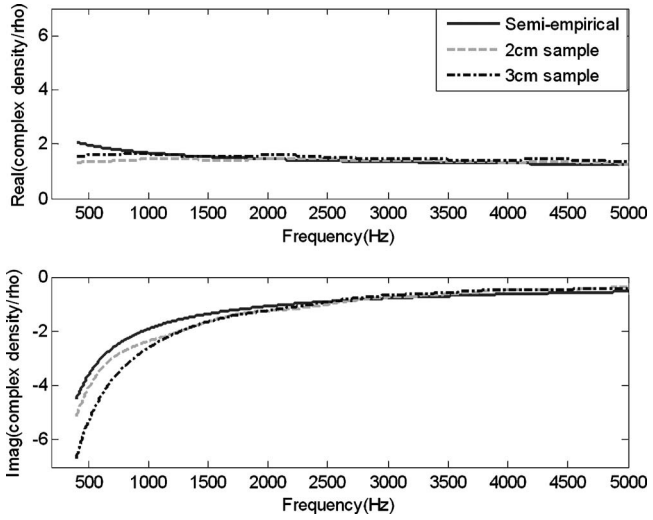


FIG. 9. Normalized complex density of fiber material determined by using transfer-matrix method and predicted using semiempirical formulas.

TABLE I. Effect of amplitude error of incident pulse on calculated results.

	Frequency (Hz)				
Error bounds	400	500	800	1000	2000
Re(Zc)	1.5%	1.63%	0.91%	0.47%	0.0429%
Im(Zc)	2.94%	2.17%	1.36%	1.1%	0.34%
Re(kp)	6.08%	6.06%	5.2%	4.26%	1.66%
Im(kp)	42.16%	41.5%	42.01%	42.59%	42.93%

TABLE II. Effect of amplitude error of reflected pulse on calculated results.

	Frequency (Hz)				
Error bounds	400	500	800	1000	2000
Re(Zc)	0.25%	0.19%	0.25%	0.32%	0.30%
Im(Zc)	0.72%	0.7%	0.62%	0.52%	0.23%
Re(kp)	1.72%	2.07%	2.42%	2.16%	1.16%
Im(kp)	1.78%	1.91%	2.48%	2.6%	2.14%

TABLE III. Effect of amplitude error of transmitted pulse on calculated results.

	Frequency (Hz)				
Error bounds	400	500	800	1000	2000
Re(Zc)	1.73%	1.81%	1.15%	0.78%	0.34%
Im(Zc)	2.22%	1.48%	0.75%	0.57%	0.10%
Re(kp)	4.35%	3.99%	2.81%	2.10%	0.49%
Im(kp)	40.53%	39.78%	39.61%	40.5%	40.6%

TABLE IV. Effect of measurement error in sample thickness on calculated results.

	Frequency (Hz)				
Error bounds	400	500	800	1000	2000
Re(Zc)	1.65%	1.38%	1.11%	1.06%	0.39%
Im(Zc)	1.28%	1.67%	1.71%	1.46%	1.29%
Re(kp)	4.33%	6.15%	8.08%	7.41%	6.31%
Im(kp)	22.6%	22.39%	30.55%	39.69%	47.23%

TABLE V. Effect of measurement error in temperature on calculated results.

	Frequency (Hz)				
Error bounds	400	500	800	1000	2000
Re(Zc)	0.11%	0.093%	0.075%	0.071%	0.026%
Im(Zc)	0.086%	0.11%	0.11%	0.098%	0.086%
Re(kp)	0.13%	0.03%	0.09%	0.05%	0.11%
Im(kp)	0.22%	0.25%	0.28%	0.26%	0.12%

TABLE VI. Effect of all factors on calculated results.

	Frequency (Hz)				
Error bounds	400	500	800	1000	2000
Re(Zc)	1.54%	1.29%	1.04%	0.99%	0.37%
Im(Zc)	1.19%	1.56%	1.60%	1.36%	1.20%
Re(kp)	6.34%	8.61%	12.08%	12.41%	16.4%
Im(kp)	22.53%	22.83%	30.83%	40.04%	47.29%

These errors will appear in the amplitudes of sampling pulses. In this part, amplitudes errors' effects on the final measuring results will be explored.

To simulate the experiment, we use the calculated results of characteristic impedance and wave number implemented by this new method as the baseline values. In the following tables, five frequencies 400, 500, 800, 1000, and 2000 Hz selected randomly are used to study the errors' effect on the final calculated values.

Table I details that error bound is more noticeable in wave number than that of characteristic impedance when amplitude of incident pulse slightly changes. In particular, it can reach an error bound about 42.93% for imaginary part of wave number. Another conclusion can be drawn that amplitude error can affect imaginary parts of Zc and kp more significantly than real parts of them.

It may be seen from Table II that error in reflected pulse amplitude can affect the calculated results slightly with a maximum error bound of 2.6%. Nonetheless, it follows the same rule as shown in Table I that imaginary part can be influenced more significantly than real part of them.

When seen from Table III, it also follows the same rules as indicated in Table I: Imaginary parts of both Zc and kp are significantly affected by the error in transmitted pulse amplitude, and big errors have also been introduced for it over whole frequency range for imaginary part of kp. So, it may be concluded that error in transmitted pulse amplitude is also a dominant contributor to the final whole errors.

B. Measurement error in sample thickness

It is of interest to investigate the impact of sample thickness error on the results of Zc and kp. This uncertainty is assumed to be ± 1 mm as 20 mm used as baseline thickness values for all calculations.

It can also be seen from Table IV that complex wave numbers are more sensitive to errors in sample thickness measurement than characteristic impedance. Particularly, the imaginary part of kp would be severely influenced by sample thickness error.

C. Error in temperature measurement

As our experiment is implemented at the temperature about 15 °C, the values of sound speed 340 m/s and characteristic impedance 415 N s/m^3 in air are adopted. At the same time, since they are dominated by temperature, its measurement error is given as ± 1 °C. The following values are utilized: sound speed range 339–341 m/s, corresponding with characteristic impedance varied from 414 to 416 N s/m^3. The simulation results are given in Table V.

Table V shows that temperature error impact on final calculated results is negligible in general.

D. Overall error

From Table VI, one can see that error bounds in characteristic impedance are much smaller than that of complex wave number. Besides this, the error bounds in imaginary part of complex wave number become bigger as frequency increases.

From above analysis, some conclusions can be drawn as follows: (1) The main error contributors result from errors in amplitudes of incident, transmitted pulses as well as in sample thickness; and (2) the effect of temperature measurement error on results can be negligible.

VI. CONCLUSIONS

The intention of the present work is to further improve the efficiency and accuracy of the transfer-matrix method for determining the fundamental acoustical properties of commonly used porous materials. By employing the echo-pulse technique in a standing wave tube, the microphone measurements can be cut to half compared to the four microphone measurements in the frequency domain implementation described by Song and Bolton.[10] In addition, with the help of impulse response calibration process, the signal-to-noise ratio in the measurement is improved due to a compensated frequency response of the sound generation system. The optimal inter-microphone distance to facilitate a better frequency response is no longer required in this approach as long as the incident, reflected, and the transmitted pulses can be separated. However, it should be noted that possible errors associated with the sample-edge constrain as illustrated in Ref. 10 still remain in the approach presented.

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China with Grant No. 10674111 and doctorate foundation of Northwestern Polytechnical University with Grant No. Cx200802. We are grateful to Dr. Chong Wang at General Motors Corporation for his helpful discussions. Our thanks also go to Professor Xiaolin Wang and Mrs. Jie Wang and Mr. Baojun Chang for their help in sample preparations and flow resistivity measurements.

[1]R. A. Scott, "The absorption of sound in a homogeneous porous medium," Proc. Phys. Soc. London **58**, 358–368 (1946).
[2]R. F. Lambert and J. S. Teaser, "Acoustic structure and propagation in highly porous, layered, fibrous materials," J. Acoust. Soc. Am. **76**, 1231–1237 (1984).
[3]L. L. Beranek, "Acoustical properties of homogenous isotropic rigid tiles and flexible blankets," J. Acoust. Soc. Am. **19**, 556–568 (1947).
[4]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[5]M. A. Ferrero and G. G. Sacerdote, "Parameters of sound propagation in granular absorption materials," Acustica **1**, 135–142 (1951).
[6]S. L. Yaniv, "Impedance tube measurement of propagation constant and characteristic impedance of porous acoustical material," J. Acoust. Soc. Am. **54**, 1138–1142 (1973).
[7]C. D. Smith and T. L. Parott, "Comparison of three methods for measuring acoustic properties of bulk materials," J. Acoust. Soc. Am. **74**, 1577–1582 (1983).
[8]H. Utsuno, T. Tanaka, T. Fujikawa, and A. F. Seybert, "Transfer function method for measuring characteristic impedance and propagation constant of porous materials," J. Acoust. Soc. Am. **86**, 637–643 (1989).
[9]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981).
[10]B. H. Song and J. S. Bolton, "A transfer function approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials," J. Acoust. Soc. Am. **107**, 1131–1154 (2000).
[11]R. T. Muehleisen, C. W. Beamer, and B. D. Tinianov, "Measurement and empirical model of the acoustic properties of reticulated vitreous carbon," J. Acoust. Soc. Am. **117**, 536–544 (2005).
[12]X. Jing and K. Y. Fung, "Generation of desired sound impulses," J. Sound Vib. **297**, 616–626 (2006).
[13]S. Xu and H. Hou, "A method to develop a dignally-controlled sound impulse generator," Journal of elementary electroacoustics **31**, 9–13 (2007).
[14]H. Hou, S. Liang, D. Liying, and W. Fangrong, "Sound absorption measurement in a circular pipe using echo-impulse method," Acta Metrologica Sinica (in press).
[15]S. H. Jang and J. G. Ih, "On the multiple microphone method for measuring induct acoustic properties in the presence of mean flow," J. Acoust. Soc. Am. **103**, 1520–1526 (1998).
[16]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[17]R. T. Muehleisen and C. W. Beamer IV, "Comparison of errors in the three- and four-microphone methods used in the measurement of the acoustic properties of porous material," J. Acoust. Soc. Am. **107**, 112–117 (2002).

High-rate synthetic aperture communications in shallow water

H. C. Song, W. S. Hodgkiss, and W. A. Kuperman
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

T. Akal
TUBITAK-MAN, Marmara Research Center, Earth and Marine Science Research Institute, Kocaeli, Turkey

M. Stevenson
Spawar Systems Center, San Diego, California 92152-5000

(Received 10 June 2009; revised 7 September 2009; accepted 24 September 2009)

Time reversal communication exploits spatial diversity to achieve spatial and temporal focusing in complex ocean environments. Spatial diversity can be provided easily by a vertical array in a waveguide. Alternatively, spatial diversity can be obtained from a virtual horizontal array generated by two elements, a transmitter and a receiver, due to relative motion between them, referred to as a synthetic aperture. This paper presents coherent synthetic aperture communication results from at-sea experiments conducted in two different frequency bands: (1) 2–4 kHz and (2) 8–20 kHz. Case (1) employs binary-phase shift-keying modulation, while case (2) involves up to eight-phase shift keying modulation with a data rate of 30 kbits/s divided by the number of transmissions (diversity) to be accumulated. The receiver utilizes time reversal diversity combining followed by a single channel equalizer, with frequent channel updates to accommodate the time-varying channel due to coupling of space and time in the presence of motion. Two to five consecutive transmissions from a source moving at 4 kts over 3–6 km range in shallow water are combined successfully after Doppler compensation, confirming the feasibility of coherent synthetic aperture communications using time reversal. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3257184]

PACS number(s): 43.60.Dh, 43.60.Gk, 43.60.Fg [NX] Pages: 3057–3061

I. INTRODUCTION

While our understanding of underwater acoustic communications has improved over the last 2 decades,[1] communications involving mobile assets (e.g., Gliders and AUVs) remains quite challenging. First, the slow propagation speed of acoustic waves (e.g., 1500 m/s) makes Doppler effects significant even for a relatively slow moving platform (e.g., 4 kts). Due to the broadband nature of underwater acoustic communication signals, Doppler due to source motion requires resampling of the received signal.[2,3] Moreover, the Doppler is time-varying and the temporal variability of underwater acoustic channels induces additional Doppler spreading. Second, underwater acoustic systems typically operate at low signal-to-noise ratios (SNRs) and thus require some form of spatial diversity (e.g., array) to enhance SNR and mitigate channel fading effects.

This paper is concerned with point-to-point coherent communication between a moving source and a single fixed receiver where spatial diversity is obtained from a virtual horizontal array generated by relative motion between them, referred to as synthetic aperture communications (SACs). It is motivated by our initial investigation into the feasibility of SAC in shallow water using the 2–4 kHz frequency band.[4] In that work, simple on/off keying was employed to minimize complexity and not require coherent demodulation. Furthermore, the motion was almost transverse resulting in a small Doppler shift. In that case, diversity appears to have come from the data being collected in an azimuthally inhomogeneous environment coupled with temporal channel variations between transmissions. Despite its practical significance, we are not aware of any previous work in the literature on SAC.

Based on the initial results reported in Ref. 4, this paper advances our understanding of SAC by analyzing the data from two experiments (FAF-04 and FAF-06) where transmissions were carried out in different frequency bands: (1) 2–4 kHz and (2) 8–20 kHz. Case (1) involved binary-phase shift-keying (2PSK or BPSK) coherent modulation, radial motion (significant Doppler), and transmission design suitable for SAC (interleaving). On the other hand, case (2) explores a higher frequency regime more typical for shallow water acoustic telemetry permitting the use of a large bandwidth with high-order constellations up to eight-phase shift keying (8PSK), achieving the potential of a high transmission rate. The receiver employed is time reversal diversity combining followed by a single channel decision-feedback equalizer (DFE) with frequent channel updates to accommodate the time-varying channel due to the coupling of space and time in the presence of motion.[5–8]

This paper is organized as follows. Section II describes the FAF-04 SAC experiment conducted in the 2–4 kHz frequency band. SAC results are presented involving a single receive element and multiple transmissions from a moving source with BPSK modulation. Section III describes the FAF-06 SAC experiment conducted in the 8–20 kHz frequency band and then demonstrates high data-rate SAC with up to 8PSK modulation.

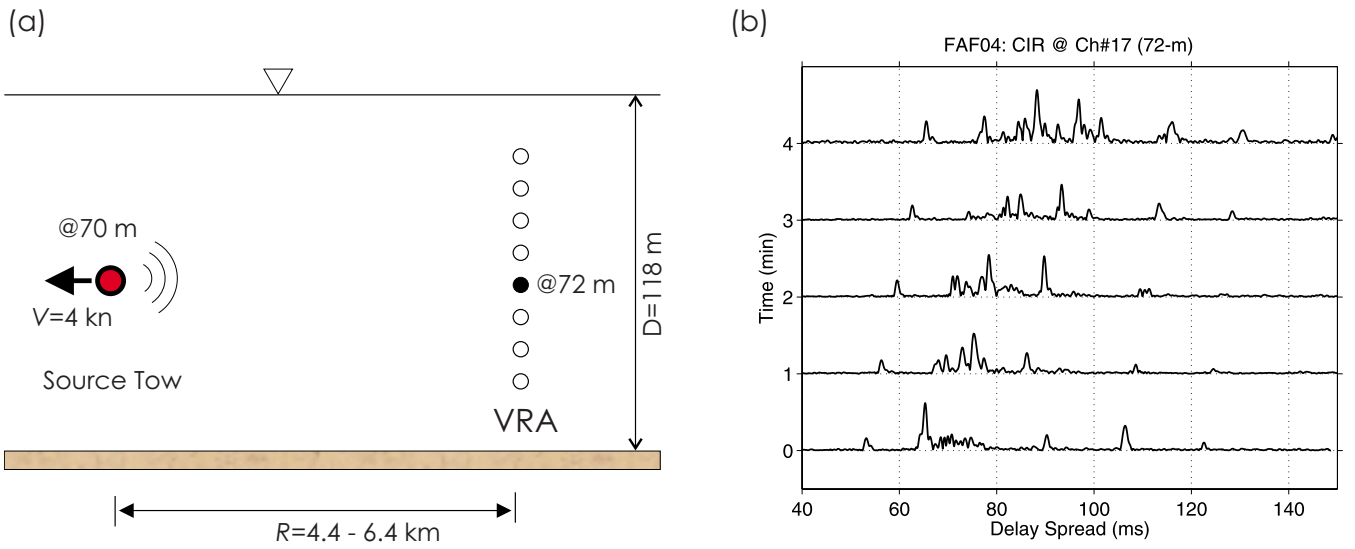


FIG. 1. (Color online) (a) Schematic of a synthetic aperture communication between a moving source towed at about 4 kts at 70-m depth and a fixed receiver at 72-m depth. (b) The first five channel responses (envelope) spaced 1-min (~120 m) apart.

II. FAF-04 EXPERIMENT

A. Experimental data

The FAF-04 experiment was conducted in July 2004 both north and south of Elba Island, off the west coast of Italy.[3] The SAC data discussed in this paper were collected in a flat region in 118-m deep water north of Elba on July 17 (JD199). A 32-element vertical receiver array (VRA) was deployed spanning the water column from 42 to 104 m with 2-m element spacing (~4 wavelengths). The VRA was moored for stable operation and a single element at 72-m depth (in the middle of the VRA) is chosen to investigate coherent SAC, as depicted in Fig. 1(a). The source was suspended from the R/V Alliance at 70-m depth. The ship was moving at a speed of 4 kts away from the VRA such that the virtual horizontal array (i.e., source track) is endfire (as viewed from above), resulting in a Doppler shift of −4.7 Hz at the 3 kHz carrier frequency. This is in contrast with the broadside aperture generated by the nearly transverse motion described in Ref. 4. On the other hand, the radial motion facilitated obtaining distinct Green's functions over a relatively short period of time. The source level was 194 dB re 1 μPa.

This paper analyzes the data collected during the 18-min-long source-tow run from 08:53 to 09:11 UTC on JD199 where the source range increased from 4.4 to 6.4 km. The detailed structure of the transmitted signal is illustrated in Fig. 2. Each 10-s data packet denoted by the letters $\{a,b,c,d\}$ consists of a 100-ms, 2–4 kHz linear frequency modulated (LFM) chirp as a channel probe, followed by a 9-s long communication sequence at a carrier frequency of 3 kHz. The symbol rate was $R=1$ kbits/s using BPSK modulation with a total of 9002 bits. Five data packets are concatenated together to form a single transmission with a duration of 50 s in the order $\{a,b,a,c,d\}$. The 50-s-long signal then was repeated every minute from the towed source during the entire 18-min run. The interleaving of communication sequences was suggested in Ref. 4 to keep the effective data rate as $R_e=R/M$ with M being the number of transmissions accumulated in generating the synthetic aperture. This is accomplished by filling in different data streams during the otherwise idle time between transmissions which are separated long enough to provide distinct (uncorrelated) Green's functions.

In this paper, each sequence will be treated either independently or combined coherently for synthetic aperture communications. During the 18-min-long source track, a total 90 transmissions were acquired over source ranges of 4.4–6.4 km. The first five channel responses spaced 1 min (~120 m) are shown in Fig. 1(b). It should be mentioned that a few transmissions (indices 43–45) are noisy and will be excluded in the processing.

B. Analysis

The application of time reversal processing to underwater acoustic communications has been studied extensively over the past several years. The receiver employed here incorporates time reversal diversity combining followed by a single channel decision-feedback equalizer with frequent channel updates to accommodate the time-varying

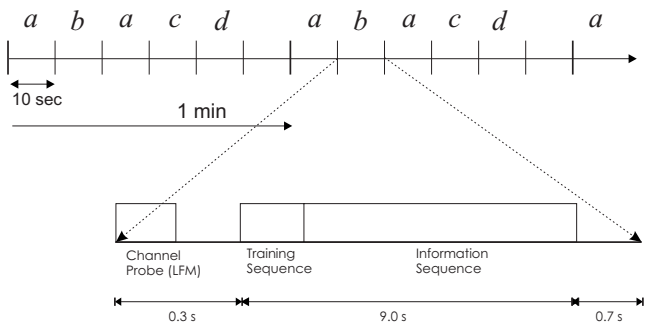


FIG. 2. Communication signal transmitted by the towed source during the synthetic aperture experiment (FAF-04). Each 10-s data packet denoted by letters $\{a,b,a,c,d\}$ consists of a channel probe followed by a 9-s communication sequence. Five data packets are concatenated together to form a single 50-s-long transmission with data packet a used twice. The 50-s-long signal was transmitted every minute during the entire 18-min-long run.

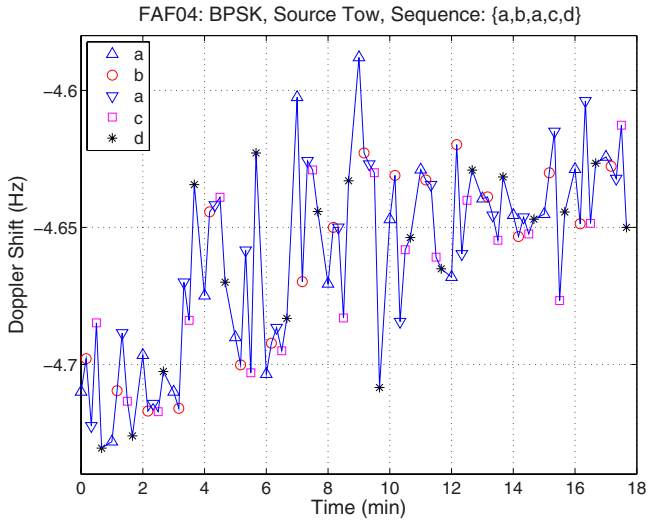


FIG. 3. (Color online) Estimate of the individual packet Doppler shifts due to source motion during the 18-min-long source-tow run. The sequence of data packet transmissions is consistent with Fig. 2.

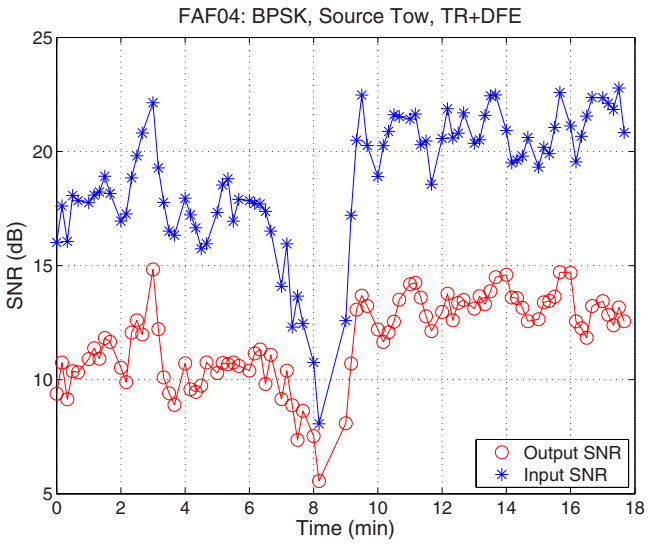


FIG. 4. (Color online) FAF-04: Performance of single reception ($M=1$) time reversal communications at receiver Ch 17 in terms of output SNR (◦) during the 18-min source motion (90 transmissions). The input SNR (∗) is displayed as a reference. Note that three noisy transmissions around 8–9 min are excluded.

channel.[6,8] Specifically, the receiver structure discussed in Ref. 8 is used for the analysis reported here with two additional components required by SAC to accommodate source/receiver motion. For each data packet, estimation of the Doppler shift at the carrier frequency is required, followed by resampling of the original broadband received signal.[3]

The resulting estimate of the Doppler shifts is shown in Fig. 3 for all 90 transmissions with four different data sequences (since a is used twice) alternating as in Fig. 2. The time-varying Doppler shift is around −4.65 Hz with a general decreasing trend in magnitude over time. A constant Doppler shift is assumed within each 10-s transmission and the residual Doppler shift is compensated for by a phase tracking algorithm in the demodulation process.

We first examine the performance of individual transmissions and then coherently combine multiple transmissions of the same sequence of data packets using time reversal to investigate the performance of synthetic aperture communications. Frequent channel updates using previously detected symbols will be conducted prior to time reversal combining, followed by a single channel adaptive DFE.[8]

The performance of each single transmission/reception ($M=1$) over the 18-min observation interval is shown in Fig. 4 (◦). Note that the output SNR closely follows the input SNR (∗). Recall that the three noisy receptions around 8–9 min are excluded in the processing. Due mostly to the high input SNR (i.e., 15–25 dB), however, even single receptions alone provide good performance (10–15 dB). To explore SAC, we add ambient noise recorded separately during the experiment to the individual receptions in order to lower the input SNR to approximately 3 dB. Thus, a single transmission/reception alone results in poor performance.

The results of coherent time reversal SAC are shown in Fig. 5 as a function of the number of receptions combined (M) for the five different data streams $\{a,b,a,c,d\}$ (see Figs. 2 and 3). These results corresponded to the initial reception at approximately 4.4 km. However, similar results were obtained through the entire range of transmissions (4.4–6.4 km). The overall performance is improved significantly, especially when two ($M=2$) or three ($M=3$) transmissions are combined together. Note that there is a minimal increase in output SNR for sequence b after $M=4$ (◦). Although not shown here, even two receptions of sequence a spaced only 20-s apart (∼40 m) provided a 5-dB output SNR, suggesting that the radial motion facilitates achieving spatial diversity over a relatively short period of time.

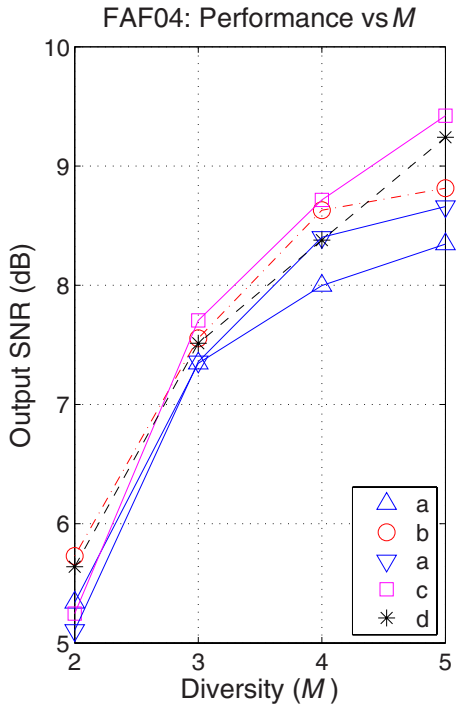


FIG. 5. (Color online) FAF-04: Performance of synthetic aperture time reversal communications versus the number of transmissions combined (M), with transmissions separated by 1 min. To each transmission, ambient noise recorded separately is added to lower the input SNR to about 3 dB such that each single transmission alone results in poor performance.

A fractionally sampled ($K=2$) adaptive DFE is employed. The number of feedforward and feedback equalizer taps are (80, 40) as compared to (120, 60) when $M=1$ (Fig. 4), due to the temporal compression achieved from time reversal diversity combining.

III. FAF-06 EXPERIMENT

A. Experimental data

The FAF-06 experiment was carried out June 21 to July 11, 2006, south of Elba Island. The SAC data discussed in this paper were collected on June 28 between a towed source and a fixed VRA. The source was an ITC-3013 transducer with an available bandwidth of 8–20 kHz. It was towed at about 4 kts over 3.4–4.2 km range at 60-m depth in 100-m deep water with the source level of 185 dB re 1 μPa. The VRA was deployed in 92-m water and consisted of 16 elements spanning a 56.25 m aperture with 3.75-m spacing.[9]

Two aspects of this experiment were different from our earlier FAF-04 SAC experiment. First, a higher frequency band was utilized which enabled our exploring a higher transmission rate with high-order modulations up to 8PSK (3 bits/symbol). Specifically, we investigated two different symbol rates: (1) 10 ksymbols/s using the full-band (8–20 kHz) centered at 15 kHz and (2) 5 ksymbols/s using half the bandwidth (8–14 kHz) centered at 11 kHz. The corresponding Doppler shifts were estimated to be about −19 and −15 Hz, at the respective center frequencies. Second, a raised-cosine filter is employed as the symbol shaping filter rather than the rectangular window used in FAF-04. For the raised-cosine filter, the signal bandwidth is controlled by a roll-off factor (i.e., $\beta=0.2$ for both symbol rates).[10]

Similar to FAF-04, six different data packets are concatenated together to form a single transmission with a duration of about 1 min. The order of the packets is {BPSK, QPSK, 8PSK} at half-band, followed by {BPSK, QPSK, 8PSK} at full-band. Each data pack is 8-s-long including a LFM channel probe which is useful for Doppler estimation via phase tracking and channel synchronization.[3] The 1-min-long signal then was repeated every minute from the towed source. A single element of the VRA at about 60-m depth (Ch 11) is chosen as the receiver with an input SNR of about 20 dB. An example of the channel response using the full bandwidth spaced 1-min apart (~120 m) is shown in Fig. 6, indicating less than 5-ms channel delay spread. In this case, the intersymbol interference spans about 50 symbols. A sample-spaced ($K=1$) adaptive DFE is employed after diversity combining and the number of taps for the feedforward and feedback filters is 30.

B. Analysis

The performance of high-frequency coherent SAC is displayed in Fig. 7 in terms of output SNR versus the number of receptions combined (M) for three different modulations (BPSK, QPSK, and 8PSK) using the full bandwidth available (denoted by F). In addition, the result of 8PSK modulation using only half the bandwidth (denoted by H) is included for comparison purposes. There are three important observations. First, the performance improves with coherent combination of multiple transmissions (M) as in Fig. 5 and remains consistent for a given M over different modulations, suggesting that a high-order constellation can be exploited to increase the data rate. In particular, 8PSK modulation can provide an effective data rate of $30/M$ kbits/s. Second, for a single reception ($M=1$) only BPSK modulation results in reasonable performance. On the other hand, the results in Fig. 7 confirm that diversity combining ($M>1$) is essential to mitigate channel fading. Finally, the performance of 8PSK

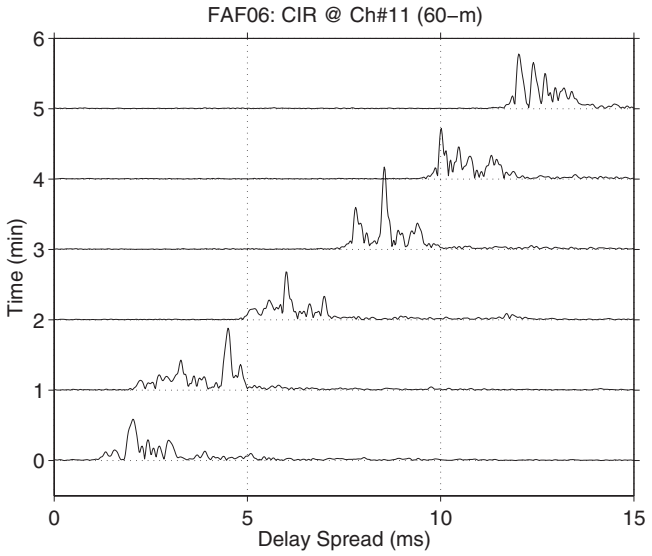


FIG. 6. FAF-06: The first six channel responses (envelope) spaced 1-min (~120 m) apart using the full bandwidth available (8–20 kHz). A source at 60-m depth was towed at about 4 kts over 3.4–4.2 km range away from a receiver chosen at 60-m depth (Ch 11).

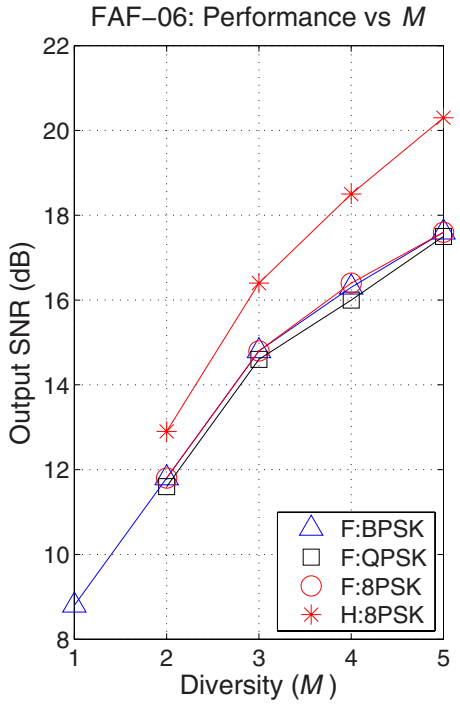


FIG. 7. (Color online) FAF-06: Performance of synthetic aperture communications versus the number of transmissions combined (M) for three different modulations using the full bandwidth available (8–20 kHz). The performance of 8PSK (*) using only half the bandwidth (8–14 kHz) also is displayed for comparison purposes.

using half the bandwidth available (∗) shows better performance (1–2 dB) over 8PSK using the full bandwidth available (∘) at the expense of lower transmission rate (5 ksymbols/s), indicating a trade-off between data rate and performance which is beyond the scope of this paper.

IV. SUMMARY AND CONCLUSIONS

Time reversal communications exploits spatial diversity to achieve spatial and temporal focusing in complex ocean environments. Spatial diversity easily can be provided by a vertical array in a waveguide. Alternatively, spatial diversity can be obtained from a virtual horizontal array generated by two elements, a transmitter and a receiver, due to relative motion between them, referred to as a synthetic aperture. This paper investigated the feasibility of coherent synthetic aperture communications through the analysis of data from two experiments conducted in different frequency bands, 2–4 and 8–20 kHz. Two to five consecutive transmissions from a source moving at 4 kts over 3–6 km range are combined successfully after Doppler compensation, confirming the feasibility of coherent synthetic aperture communications using time reversal. In particular, the large bandwidth (8–20 kHz) coupled with 8PSK modulation provided an effective data rate of $30/M$ kbits/s at about 3.5 km range in shallow water.

ACKNOWLEDGMENT

This work was supported by the Office of Navel Research under Grant No. N00014-06-1-0128.

[1] D. Kilfoyle and A. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).

[2] M. Johnson, L. Freitag, and M. Stojanovic, "Improved Doppler tracking and correction for underwater acoustic communications," in Proceedings of the IEEE Oceans'97 (1997), pp. 575–578.

[3] H. C. Song, W. S. Hodgkiss, W. A. Kuperman, W. Higley, K. Raghukumar, T. Akal, and M. Stevenson, "Spatial diversity in passive time reversal communications," J. Acoust. Soc. Am. **120**, 2067–2076 (2006).

[4] W. J. Higley, P. Roux, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Synthetic aperture time-reversal communications in shallow water: Experimental demonstration at sea," J. Acoust. Soc. Am. **118**, 2365–2372 (2005).

[5] J. Gomes, A. Silva, and S. Jesus, "Adaptive spatial combining for passive time-reversed communications," J. Acoust. Soc. Am. **124**, 1038–1053 (2008).

[6] A. Song, M. Badiey, H. C. Song, W. S. Hodgkiss, M. Porter, and KauaiEx, "Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx)," J. Acoust. Soc. Am. **123**, 856–865 (2008).

[7] J. Flynn, J. Ritcey, D. Rouseff, and W. Fox, "Multichannel equalization by decision-directed passive phase conjugation: Experimental results," IEEE J. Ocean. Eng. **29**, 824–836 (2004).

[8] H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "Basin-scale time reversal communications," J. Acoust. Soc. Am. **125**, 212–217 (2009).

[9] J. D. Skinner and W. S. Hodgkiss, "A networked/autonomous receiving array system," in Proceedings of OCEANS 2008 (2008).

[10] J. Proakis, *Digital Communications* (McGraw-Hill, New York, 2001).

A source separation approach to enhancing marine mammal vocalizations

M. Berke Gur[a)]
Department of Mechatronics Engineering, Bahcesehir University, Besiktas, Istanbul 34353, Turkey

Christopher Niezrecki
Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854

(Received 10 September 2009; accepted 2 October 2009)

A common problem in passive acoustic based marine mammal monitoring is the contamination of vocalizations by a noise source, such as a surface vessel. The conventional approach in improving the vocalization signal to noise ratio (SNR) is to suppress the unwanted noise sources by beamforming the measurements made using an array. In this paper, an alternative approach to multi-channel underwater signal enhancement is proposed. Specifically, a blind source separation algorithm that extracts the vocalization signal from two-channel noisy measurements is derived and implemented. The proposed algorithm uses a robust decorrelation criterion to separate the vocalization from background noise, and hence is suitable for low SNR measurements. To overcome the convergence limitations resulting from temporally correlated recordings, the supervised affine projection filter update rule is adapted to the unsupervised source separation framework. The proposed method is evaluated using real West Indian manatee (*Trichechus manatus latirostris*) vocalizations and watercraft emitted noise measurements made within a typical manatee habitat in Florida. The results suggest that the proposed algorithm can improve the detection range of a passive acoustic detector five times on average (for input SNR between −10 and 5 dB) using only two receivers. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3257549]

PACS number(s): 43.60.Fg, 43.30.Sf, 43.60.Mn [EJS] Pages: 3062–3070

I. INTRODUCTION

A frequently encountered problem in passive acoustic based marine mammal monitoring is the enhancement of vocalization signals in the presence of other interfering sources and ambient noise. The presence of dominant noise sources reduces the signal-to-noise ratio (SNR) of the measurements, can degrade detection and classification performance, or reduce the effective range of a passive acoustic monitoring system. A feasible implementation of a passive acoustic based monitoring system operating in noisy environments generally requires the enhancement of the vocalization signals. One such application related to marine mammals is the enhancement of manatee vocalizations for more effective passive acoustic based detection in the presence of recreational watercraft. In this paper, a two-channel second-order-statistics (SOS) based blind source separation (BSS) approach is developed and evaluated for enhancing manatee vocalizations.

The West Indian manatee (*Trichechus manatus latirostris*) was added to the endangered species list in 1967. In 1980, the U.S. Fish and Wildlife Service established a manatee protection plan. Within this plan, collisions with recreational boats and other watercraft were identified as the most significant cause of unnatural manatee mortalities. Accordingly, idle-speed and no-wake zones have been designated throughout the shallow Florida waterways where manatee-watercraft collisions are most likely to occur. However, a 2001 recovery review revealed that the rate of manatee-watercraft collision related mortalities continued to remain high despite measures taken.[1] Although the West Indian manatee was recently re-classified as a threatened species with a very high risk of extinction,[2] data collected by the Florida Fish and Wildlife Conservation Commission indicate that watercraft related mortalities remain steady at 25% among all manatee mortalities.[3] Several reasons for the ineffectiveness of the speed zones have been put forth in the recent literature. A factor that may contribute to the high rate of collisions is the lack of compliance of boaters to year-long and seasonal speed zones since these speed zones significantly increase travel times within the Florida channels. In a survey conducted at 15 sites in Florida, overall compliance rates to speed zones were reported as 58% and 63%.[4]

One possible solution to improve compliance rates to speed zones is an active boater warning system based on passive acoustic detection of manatee vocalizations and alerting nearby boaters of the presence of the animal. A typical manatee vocalization lasts between 0.1 and 0.5 s and may have several harmonics in the frequency band of 2–10 kHz [see Figs. 1(a) and 1(b)]. Detailed information on manatee vocalizations can be found in the works of Steel,[5] Nowacek *et al.*,[6] Phillips *et al.*,[7] and the references therein. It was shown by Niezrecki *et al.*[8] that a frequency domain energy detector is capable of satisfactorily detecting manatee vocalizations in moderate SNR measurements. However, as the noise levels increase relative to the vocalization source strength, a signal enhancement procedure becomes necessary prior to detection.[9]

In general, fluctuations in the underwater ambient noise statistics warrant adaptive signal enhancement algorithms. Several single channel, adaptive algorithms have been pro-

[a)]Author to whom correspondence should be addressed. Electronic mail: berke.gur@bahcesehir.edu.tr

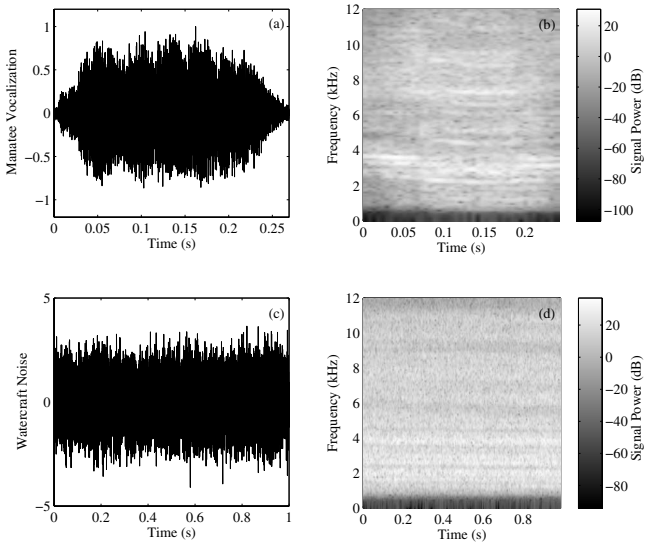


FIG. 1. [(a) and (c)] Time domain plots and [(b) and (d)] spectra of a manatee vocalization and watercraft emitted noise, respectively (after high-pass filtering).

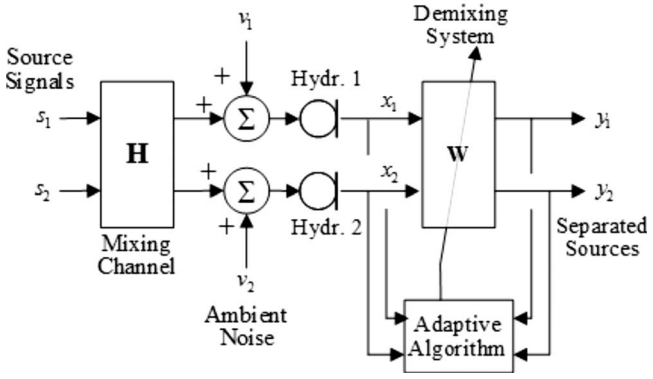


FIG. 2. The typical setup of a two-channel BSS problem.

posed for enhancing manatee vocalizations and other bioacoustic signals. Yan *et al.*[10,11] proposed an adaptive line enhancer for enhancing manatee vocalizations. An adaptive wavelet domain *ad-hoc* method for enhancing manatee vocalizations was developed by Gur and Niezrecki.[12] More recently, Ren *et al.*[13] proposed a wavelet domain non-linear adaptive filter for enhancing bioacoustic signals. The enhancement performances of conventional single channel adaptive filtering methods drastically decline as the SNR of the measurements decreases. However, if a significant portion of background noise is emitted from a point noise source such as a surface vessel, signal enhancement performance can substantially be improved through multi-channel adaptive algorithms.

The conventional multi-channel approach to signal enhancement is beamforming through which measurements from a uniform linear array (ULA) are constructively combined to enhance the signal incident from the target source location.[14] Adaptive beamforming algorithms can also be designed to suppress a dominant point noise source by placing a null in the corresponding incidence direction. Several passive acoustic based detection, localization, and classification systems that incorporate beamforming are described in the literature.[15,16] However, beamforming has some important drawbacks. It requires the precise knowledge of the target and noise source locations. In general, the location of the sources is not known a priori, and the direction of arrival (DOA) estimation for the vocalization signals must be performed. The estimation of DOA and the related problem of time difference of arrival (TDOA) estimation from noisy manatee vocalization recordings were investigated by Muanke and Niezrecki[17] in the context of source localization. The cited authors concluded that an input SNR of 8 dB or higher was necessary on all the input channels for accurately estimating the TDOA. Alternatively, the DOA can be estimated using blind sub-space algorithms such as the multiple signal classification algorithm or even by scanning each radial angle for high power incoming acoustic signals. However, the signal power of weak vocalization signals will generally be too low to determine the corresponding DOA using such methods. The ability of a beamformer to resolve the location of the target is inversely proportional to the length of the discrete aperture formed by the ULA. Combined with sensor spacing restrictions necessary to avoid spatial aliasing, enhancing marine mammal vocalizations requires an array that consists of many (on the order of 10 or more) hydrophones. Hence, beamforming is not a cost-effective solution to enhance manatee vocalizations in the numerous manatee idle-speed/no-wake zones within the Florida waterways.

As an alternative, BSS is a class of adaptive signal processing algorithms that serve for retrieving the original signals emitted from multiple point sources from multi-channel mixtures. These algorithms are referred to as "blind" because both the source signals and the mixing channels are assumed to be unknown. The signals emitted from multiple acoustic point sources are assumed to be statistically independent, which can be physically justified. Due to transmission through a multi-input, multi-output mixing channel, the acoustic signals measured at the receivers become statistically dependent. The original sources can be extracted from the measurements by solving for a separating solution that makes the multi-channel measurements statistically independent. The separated sources are retrieved from the measurements using unsupervised adaptive filtering (see Fig. 2 for a two-channel setup). BSS algorithms can be separated into several different groups based on the implementation of the statistical independence criterion. Most of these algorithms utilize the known or estimated probability density function (PDF) and/or higher-order-statistics (HOS) of the acoustic source signals to achieve source separation. Unfortunately, accurate estimates of the source PDF and HOS of weak vocalizations from noisy measurements are very difficult to obtain. However, a separate group of SOS based algorithms achieves source separation by making the outputs uncorrelated, which result in less complex and more robust algorithms for two-channel setups.

Source separation, particularly its applications in speech enhancement and communications, is a topic that has received high interest among researchers over the past decade (for a concise review, see Ref. 18). In the context of signal

enhancement, separation of the vocalization and noise signals suggests an improvement in the SNR of the vocalization signal. Therefore, BSS algorithms are versatile methods for enhancing marine mammal (e.g., manatee) vocalizations corrupted by an interfering point source such as watercraft navigating through the channel. Despite the extensive literature on speech and communications source separation, only a few studies implement source separation in the context of underwater acoustics. Gaeta *et al.*[19] suggested a HOS based blind separation of artificially mixed underwater acoustic signals. They numerically computed the channel impulse responses using ray propagation theory. Bonnifay *et al.*[20] incorporated prior knowledge of the channel impulse responses in a HOS based source separation algorithm and experimented with artificially mixed underwater communication signals. More recently, Mansour *et al.*[21] investigated blind separation of underwater acoustic signals (including artificial mixtures of ship noise and whale vocalizations) for passive acoustic tomography and reported that SOS based frequency domain iterative algorithms exploiting the non-stationarity of the source signals resulted in better separation performance.

In this study, a new SOS based BSS algorithm for enhancing marine mammal vocalizations is proposed and evaluated using real vocalization and watercraft emitted noise measurements. This paper is organized as follows. A novel SOS-BSS algorithm based on the affine projection (AP) filter update rule is proposed in Sec. II. The signal enhancement performance of the proposed SOS-BSS algorithm is evaluated using real measurements in Sec. III. The improvements in the detection range resulting from preprocessing the measurements with these algorithms are presented in Sec. IV. Lastly, the concluding remarks are provided in Sec. V.

II. THEORETICAL DEVELOPMENT

A. The underwater acoustic channel model

The mixing channel model is depicted in Fig. 3(a) where the two measurements $\mathbf{X}=[x_1\ x_2]^T$ are a result of a linear mixture of the two sources $\mathbf{S}=[s_1\ s_2]^T$ (assuming that the s_1 is the vocalization signal) and ambient noise $\mathbf{V}=[v_1\ v_2]^T$,

$$\mathbf{X}=\mathbf{H}*\mathbf{S}+\mathbf{V}, \tag{1}$$

where $*$ represents time domain convolution, and $\mathbf{H}$ is a matrix of finite impulse response (FIR) filters modeling multipath channel transmission. In the context of source separation, the source signals are assumed to be more dominant compared to ambient noise. The channel mixing process can be separated into two stages, as shown in Fig. 3(b); the propagation of the source signals through the convolutive channel is represented by the impulse responses h_{11} and h_{22}, and the cross-channel mixing of the sources is represented by the pseudochannel impulse responses $\tilde{h}_{12}=h_{12}/h_{11}$ and $\tilde{h}_{21}=h_{21}/h_{22}$. Simple decorrelation can be used to extract the source signals $\tilde{s}_1$ and $\tilde{s}_2$, but is not sufficient to retrieve the original sources s_1 and s_2. Additional known properties of the source signals such as non-stationarity[22] or temporal correlation[23] need to be exploited to perform simultaneous separation and deconvolution. Alternatively, deconvolution

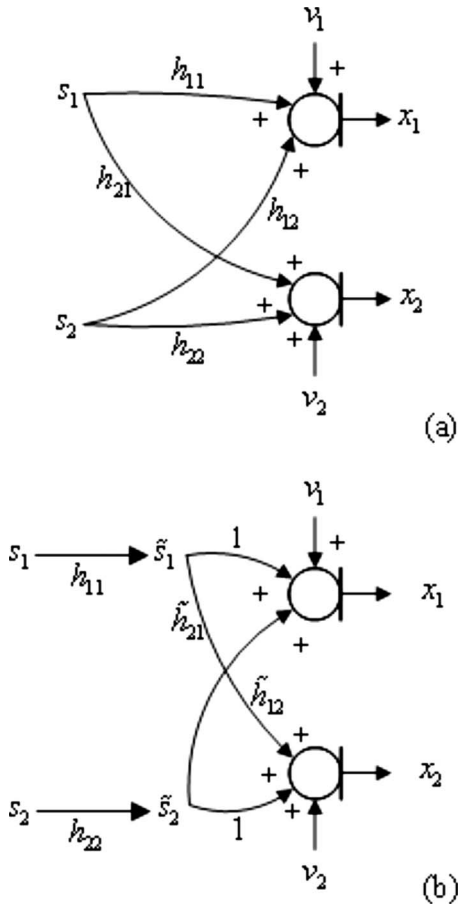


FIG. 3. (a) The two-channel transmission model and (b) the channel model with channel propagation and mixing stages decoupled.

can be performed postseparation.[24] Although $\tilde{s}_1$ is a filtered version of the original vocalization signal, it is sufficient to extract this filtered signal for detection purposes.

B. Second-order-statistics based blind source separation

Decorrelation based algorithms employ a SOS based cost function of the de-mixing system outputs for achieving source separation. Adaptively decorrelating the outputs

$$E[y_p(n)y_q(n-k)]=r_{y_py_q}(n,k)=0,\quad p\neq q=1,2, \tag{2}$$

where $E[\cdot]$ is the expectation operator, y_p is the output at the pth channel and $r_{y_py_q}(n,k)$ is the cross correlation coefficient between the outputs at time n and lag k, is anticipated to result in source separation. Assuming a two-channel mixing system, a cost function that uses the instantaneous estimate of the correlation between the two outputs at L lags combined with a stochastic gradient update rule results in the SOS based symmetric adaptive decorrelator (SAD) algorithm[25] with a filter update structure similar to the least mean squares (LMS) algorithm. The SAD algorithm is a simple and robust BSS algorithm suitable for source separation in the two-input, two-output configuration.

However, the SAD algorithm also inherits the limitations of the LMS algorithm. Due to the large eigenvalue spread of the input autocorrelation matrix, the convergence

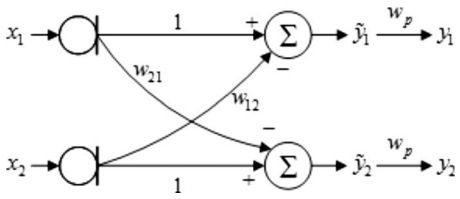


FIG. 4. The block diagram of the two-channel feedforward structured de-mixing system.

of the filter coefficients is rather slow in the presence of temporally correlated inputs (such as the harmonic manatee vocalizations that require small step sizes to prevent the separating filters from destabilizing).[26] The small step sizes required for stability reduce the ability of the de-mixing filters to adapt to changes in the noise signal statistics or to track the changes in the channel impulse responses due to moving watercraft. One possible solution to improve the convergence rate of the SAD algorithm is to update the adaptive filters using the recursive least squares (RLS) algorithm.[27] The resulting double-RLS (DRLS) algorithm eliminates the large eigenvalue spread of the filter inputs by pre-multiplying the inputs with a recursive estimate of the inverse of the cross-correlation matrix between the inputs and filtered outputs. However, the DRLS algorithm is prone to the same divergence phenomenon observed in supervised RLS algorithms caused by numerical instabilities.[28] To further aggravate the problem, the use of the input and output cross-covariance matrices in the DRLS algorithm (in contrast to the Toeplitz input autocorrelation matrix, in case of the supervised RLS algorithm) prevents fast and numerically more stable QR-decomposition based implementations.

C. The proposed method

The AP algorithm[29] was proposed in the context of supervised adaptive filtering as an intermediate algorithm between the LMS and RLS algorithms. Instead of the inverse of the full order $L \times L$ input autocorrelation matrix (where L is the filter order), the AP algorithm uses the inverse of a $K \times K$ lower order partial autocorrelation matrix (where $K < L$) to pre-whiten the input data. In this paper, an unsupervised BSS algorithm based on the supervised AP algorithm is derived and implemented for the two-channel configuration. The extension of the supervised AP algorithm to the unsupervised adaptive filtering framework was first proposed by Gabrea.[30] The cited author proposed a feedback structured block-update double affine projection (DAP) algorithm in the context of two-channel speech enhancement. However, the feedback structure may cause instability. In addition, the $K \times K$ partial autocorrelation matrix will always be incomplete due to the block-update restriction, leading to degraded separation performance. Hence, to circumvent these problems, a feedforward (FF) implementation of the DAP algorithm with sequential update (SU) of the filter coefficients is proposed. The block diagram of the feedforward setup is depicted in Fig. 4. The FF/SU DAP algorithm solves for the separating filters by computing the minimum squared norm filter update such that the updated filter decorrelates the past K lag cross-correlations between the intermediate outputs $\tilde{y}_1(n)$ and $\tilde{y}_2(n)$, where $K < L$ and L is the maximum order of the cross-channel filters. This criterion can be expressed as the cost function

$$\mathcal{J}_p(n) = (\|\mathbf{w}_{pq}(n+1) - \mathbf{w}_{pq}(n)\|_2)^2 + E[\tilde{\mathbf{y}}_{p,K}^T(n)\tilde{y}_q(n)]\boldsymbol{\lambda}, \quad p \neq q = 1,2, \tag{3}$$

where $\boldsymbol{\lambda}$ is the $K \times 1$ vector of Lagrange multipliers and the $K \times 1$ intermediate output vector $\tilde{\mathbf{y}}_{p,K}(n)$ is defined as

$$\tilde{\mathbf{y}}_{p,K}(n) = \mathbf{x}_{p,K}(n) - \mathbf{X}_q^H(n)\mathbf{w}_{pq}(n+1), \tag{4}$$

and where $\mathbf{X}_q(n) = [\mathbf{x}_{q,L}(n)\ \mathbf{x}_{q,L}(n-1)\ \cdots\ \mathbf{x}_{q,L}(n-K+1)]$ is an $L \times K$ matrix of the past filter outputs. Replacing the cross-correlations with their instantaneous estimates (i.e., $E[y_p y_q] = y_p y_q$), using a stochastic gradient descent optimization rule, and after some algebraic manipulation, the resulting filter update equations take the form

$$\mathbf{w}_{pq}(n+1) = \mathbf{w}_{pq}(n) + \mu\mathbf{X}_q(n)(\mathbf{X}_q^T(n)\mathbf{X}_q(n) + \delta\mathbf{I})^{-1}\tilde{\mathbf{y}}_{p,K}(n). \tag{5}$$

Here, μ is the step size and δ is a regularization term added diagonally to prevent numerical difficulties in inverting the possibly rank deficient matrix $\mathbf{X}_q^T\mathbf{X}_q$. More detailed derivations of the FF/SU DAP algorithm as well as some other related BSS algorithms can be found in Ref. 31.

D. Performance measures

Well defined, quantitative performance measures are necessary for objectively evaluating the signal enhancement performances of the proposed BSS algorithm. In this paper, two performance measures are utilized to evaluate the signal enhancement performance of the FF/SU DAP algorithm. The pre-denoising quality of a noisy vocalization signal is quantified in terms of the input SNR which is defined for each input channel as the ratio of the squared root-mean-square (rms) values of the vocalization and noise signals

$$\mathrm{SNR}_{\mathrm{in},p} = 10\log_{10}((s_{\mathrm{rms},p})^2/(v_{\mathrm{rms},p})^2), \quad p = 1,2, \tag{6}$$

where p is the channel index, and $s_p(n)$ and $v_p(n)$ are the vocalization and noise signals at channel p, respectively. The rms value of the length N signal $x(n)$ is defined as

$$x_{\mathrm{rms}} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(x(n))^2\right]^{1/2}. \tag{7}$$

The noise signal in the denominator of Eq. (6) represents all signals (including watercraft emitted noise) other than the vocalization signal. The rms value of the noise signal is computed over the duration of the vocalization signal. In general, the SNR at the input channels may vary. In contrast, a single output SNR is defined as

$$\mathrm{SNR}_{\mathrm{out}} = 10\log_{10}[((y_{\mathrm{rms},s})^2 - (y_{\mathrm{rms},v})^2)/(y_{\mathrm{rms},v})^2], \tag{8}$$

where $y_{\mathrm{rms},s}$ is the rms value of the enhanced estimate of the vocalization signal and $y_{\mathrm{rms},v}$ is the rms value of the output if no vocalization is present. Thus, $y_{\mathrm{rms},v}$ represents the noise residue that is not suppressed by the algorithm over the duration of the vocalization.

The signal-to-distortion ratio (SDR) is defined as the ratio of the vocalization signal and distortion power

$$\mathrm{SDR} = 10\log_{10}((s_{\mathrm{rms}})^2/(e_{\mathrm{rms}})^2), \tag{9}$$

where $s(n)$ is the vocalization signal, and the distortion is defined as the mean-squared-error

$$e_{\mathrm{rms}} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(e(n))^2\right]^{1/2} = \left[\frac{1}{N}\sum_{n=0}^{N-1}(s(n)-y_s(n))^2\right]^{1/2}, \tag{10}$$

and $y_s(n)$ is the enhanced estimate of the vocalization signal. The SDR is an indicator of how well the vocalization waveform is preserved, which strongly affects the detection rate of a matched filter detector or other similar correlation based detectors, and the accuracy of source localization algorithms based on TDOA estimates.

III. EXPERIMENTAL RESULTS

The underwater acoustic environment is very challenging in terms of propagation. The acoustic propagating channel has certain dynamics that cannot be modeled accurately with a FIR filter. Unfortunately, these undesired channel effects are more pronounced in shallow water channels, primarily due to the complex interaction of the acoustic waves with the waveguide boundaries, volumetric inhomogeneities in the water, and other uncertainties inherent to the underwater acoustic channel. The underwater acoustic channel is frequency selective and can significantly attenuate high frequency signal energy. Another factor that results in frequency selective attenuation of underwater acoustic signals is the Lloyd mirror effect. All of these factors reduce the coherence of the signals measured at different receivers, which may affect the convergence and signal enhancement performance of the BSS algorithms. Thus, it is essential to evaluate the proposed FF/SU DAP algorithm under real environments to be able to fully comprehend their signal enhancement performances.

To evaluate the performance of the FF/SU DAP algorithm under realistic conditions, real vocalization and noise data were recorded at Crystal River, FL. The test location where the background noise data were collected is at close proximity to known manatee habitats and a busy waterway, and thus represents a pilot site where a manatee vocalization detector may potentially be implemented. Therefore, the results presented here are expected to be good indicators of how these algorithms are expected to perform in-field.

In speech processing, speech signals are generally recorded in an anechoic chamber in order to obtain reverberant free source signals, and then convolved with an experimentally measured or numerically simulated reverberant impulse response. Thus, the source signals and the channel impulse responses are known a priori and can be used to evaluate the SNR and SDR performances of the algorithms as well as the convergence of the de-mixing filters to the optimum separating solution. Alternatively, to obtain more realistic mixtures, the speech signals are recorded separately (but in the same environment) and are numerically superposed to create the noisy measurements.[32] Both experimental setups used in speech processing are not feasible for marine mammal monitoring applications. Nevertheless, in an effort to generate as realistic noisy vocalization recordings as possible, the vocalizations in the manatee vocalization library[33] are convolved with actual measured underwater acoustic channel impulse responses and added to real background noise recordings. The channel impulse responses used to convolve manatee vocalizations are estimated from a series of broadband broadcast tests conducted at Crystal River, FL on the same day and location as the background noise recordings. A typical impulse response obtained from the broadcast tests is presented in Fig. 5 where the underwater speaker is placed 10.3 m away from the reference hydrophone.

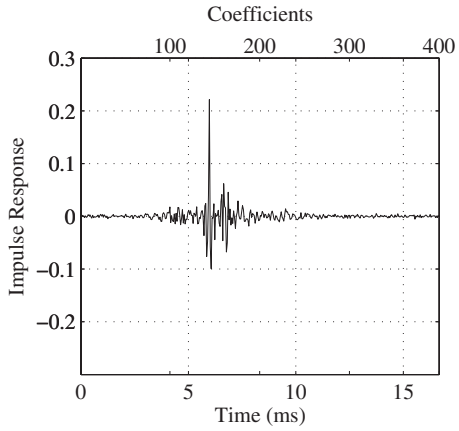


FIG. 5. A typical channel impulse response estimated from broadband broadcast tests conducted in Crystal River, FL.

The performance of the FF/SU DAP algorithm is evaluated in a two-channel setup which consists of a manatee and a watercraft as the only two active acoustic point sources. In general, the input SNR is a function of the vocalization and watercraft emitted noise source levels (SLs) as well as the channel attenuation. Channel attenuation is determined by the channel impulse responses, whereas both the vocalization and watercraft emitted noise SL may change.[7,9] To compensate for the variance in the SL of both the manatee vocalizations and watercraft emitted noise, the input SNR is controlled by scaling the power of the vocalization signal such that the input SNR at the reference channel of the vocalizations is equal to pre-specified values. Another important factor that reduces the SNR and the separation performance of the proposed BSS algorithm is the presence of extraneous noise. The ambient background noise levels measured at the Crystal River test site when no point source was active (i.e., no watercraft in the vicinity) were determined to be lower than −12 dB compared to typical watercraft emitted noise throughout the measurements. The diffuse ambient noise levels were used to estimate the lowest input SNR that the FF/SU DAP algorithm was expected to achieve acceptable signal enhancement performance. Since manatee vocalizations are generally 0.5 s or shorter in duration, several watercraft noise recordings, each of 1 s duration, were selected from the measurements [see Figs. 1(c) and 1(d)]. The approach direction and the relative speed of each watercraft were noted during the recordings. Both the vocalization and

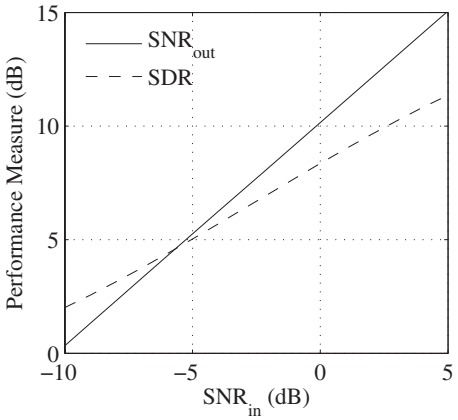


FIG. 6. The output SNR and SDR performance measures as a function of the input SNR for the FF/SU DAP algorithm averaged over ten manatee vocalizations and four different noise recordings.

noise signals were high-pass filtered using tenth order Butterworth filter with a cutoff frequency of 1 kHz. The filter order, step size, and cross correlation delay for the FF/SU DAP algorithm were set to $L=200$, $\mu=0.01$, and $K=10$, respectively.

The output SNR and SDR results averaged over ten manatee vocalizations and four different noise recordings are presented in Fig. 6. For these tests, it was assumed that the manatee was 1 m away from the reference hydrophone. The other hydrophone is located 9.3 m away. The input SNR at the vocalization reference channel is varied from −10 to 5 dB. The input SNR at the other channel is not manipulated, but rather is determined by the channel impulse responses, and is usually 5–10 dB lower than the input SNR. A typical time domain output of the FF/SU DAP algorithm is presented in Fig. 7 for −5 dB input SNR on channel 1.

As is discussed in Sec. II, the cross channel transfer function is more relevant for BSS algorithms. The results presented above are obtained for manatee vocalizations convolved with a channel impulse response between two hydrophones (separated by 9.3 m) where the manatee is assumed to be 1 m away from the reference hydrophone.

In the following tests, the effect of changing the distance between the manatee and the corresponding reference hydrophone on the performance of the FF/SU DAP algorithm is investigated. The separation between the hydrophones is fixed at 9.3 m, and the manatee is assumed to be at distances of 1, 4.7, 10.3, and 13.9 m away from the reference hydrophone. Although the distance between the hydrophones remains the same, the pseudochannel impulse response changes as the range of the manatee is increased. The output SNR and SDR results obtained with the manatee assumed to be at the four different locations are presented in Fig. 8. These results suggest that the signal enhancement performance of the FF/SU DAP algorithm is not significantly affected by changes in the channel transfer function between the manatee and the corresponding reference receiver, particularly at low input SNR values. Moreover, these results prove that the detection range resulting from processing the noisy vocalizations with the FF/SU DAP algorithm is only a function of the input SNR. This conclusion simplifies the detection range computations presented in Sec. IV.

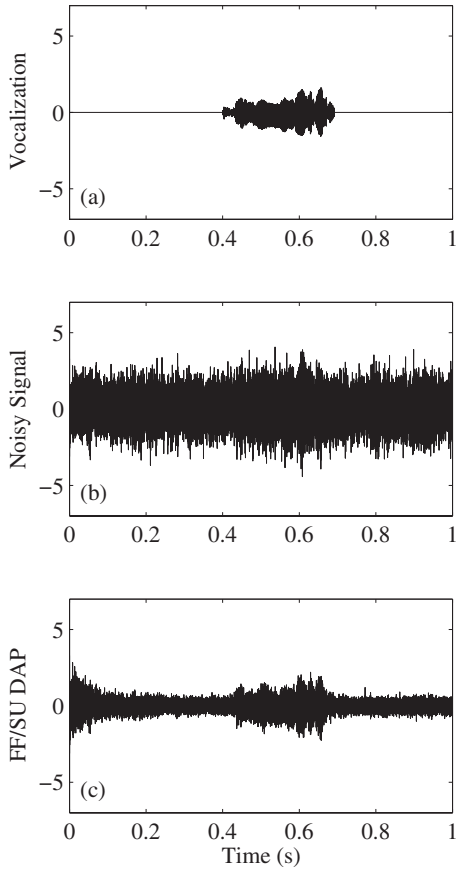


FIG. 7. (a) A clean manatee vocalization in the time domain, (b) the noisy measurement with input SNR equal to −5 dB on channel 1, and (c) the time domain output of the FF/SU DAP algorithm (plots are for the channel 1 measurements; the input SNR for channel 2 is −13.4 dB).

IV. DETECTION RANGE IMPROVEMENTS

The signal enhancement performance of the FF/SU DAP algorithm can be related to the improvements in the detection range through the passive sonar equation written in terms of the figure of merit (FOM)

$$\mathrm{FOM} = \mathrm{SL}_m - \max(\mathrm{NL}_a, \mathrm{SL}_w - \mathrm{TL}_w) + \mathrm{AG} - \mathrm{DT}, \quad (11)$$

where SL_m is the source level of the vocalizations, NL_a is the ambient noise levels, SL_w is the boat source level, TL_w is the transmission loss associated with the watercraft, AG is the array gain, and DT is the detection threshold of the vocalization detector.[9] The source levels are the sound pressure levels (referenced to 1 μPa) located 1 m from the source. The FOM represents the maximum allowable transmission loss, and hence the maximum range at which the animal can be detected by the passive acoustic system. Assuming that the transmission loss obeys a mixed spreading model,[9] the FOM can also be expressed as a function of the range of the manatee to the hydrophones

$$\mathrm{FOM} = 15 \log_{10}(r_m). \quad (12)$$

Equating Eqs. (11) and (12) and solving for the range of the manatee, one obtains

$$r_m = 10^{(\mathrm{SL}_m - \max(\mathrm{NL}_a, \mathrm{SL}_w - 15 \log_{10}(r_w)) + \mathrm{AG} - \mathrm{DT})/15}. \quad (13)$$

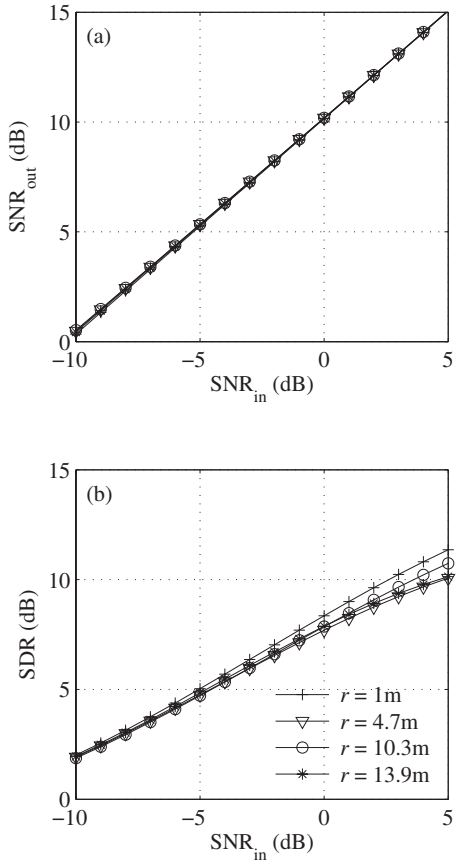


FIG. 8. (a) The output SNR and (b) the SDR performance measures for different manatee ranges as a function of the input SNR for the FF/SU DAP algorithm.

The ratio of the detection range resulting from the FF/SU DAP algorithm to high-pass filtering detection range can be computed using the relation

$$\frac{(r_m)_{\mathrm{DAP}}}{(r_m)_{\mathrm{HPF}}} = 10^{\mathrm{AG}/15}. \tag{14}$$

The AG term is defined as the improvement in the SNR with respect to the single receiver input SNR. Thus, for the FF/SU DAP algorithm, AG is defined as the improvement in the SNR,

$$\mathrm{AG} = \mathrm{SNR}_{\mathrm{out}} - \mathrm{SNR}_{\mathrm{in}}. \tag{15}$$

The average ratio of the detection ranges resulting from the FF/SU DAP algorithm over high-pass filtering for the Crystal River test cases is presented in Fig. 9. The FF/SU DAP algorithm achieves a relatively uniform output SNR performance over the input SNR values and can improve the detection range by a factor of 4.7 or higher for input SNR varying from −10 to 5 dB.

Next, the detection ranges resulting from processing the vocalizations with the FF/SU DAP algorithm are presented using an example. The watercraft noise and manatee vocalization SL are assumed to be 140 and 118 dB, respectively. Ambient noise is assumed to be 70 dB. For a given output SNR, it is necessary to compute the input SNR (and thus, the AG) for the FF/SU DAP algorithm. To obtain an analytic

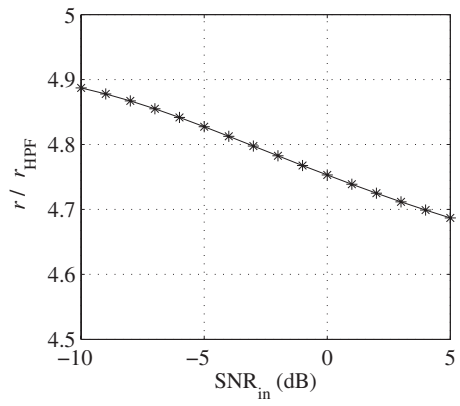


FIG. 9. The average ratio of the detection range resulting from the FF/SU DAP algorithm to high-pass filtering for the Crystal River tests.

relation between the input and the output SNRs, the output SNR for the Crystal River test cases is curve fitted with a first order polynomial

$$\mathrm{SNR}_{\mathrm{out}} = 0.9734 \cdot (\mathrm{SNR}_{\mathrm{in}}) + 10.2057 \tag{16}$$

in the least squared sense.[34] Equation (16) suggests that an average of 10.2 dB AG can be achieved with the FF/SU DAP algorithm. For the FF/SU DAP algorithm implemented together with a 3 dB DT passive acoustic detector, the AG is 10.4 dB and the minimum input SNR necessary for detection is computed as −7.4 dB. The ratio of the estimated detection range resulting from the FF/SU DAP algorithm and high-pass filtering is computed as 4.9 for a DT of 3 dB. The maximum detection range as a function of the detection threshold and range of the watercraft are depicted in Fig. 10. Although the detection range drops below 10 m when the watercraft is 100 m away from the receivers with the FF/SU DAP algorithm, the proposed method significantly increases the detection range (approximately five times) and the effective coverage area (approximately 25 times) of a passive acoustic based detector compared to a high-pass filter alone.

For the setup in this study, the manatee and watercraft are assumed to be closer to hydrophones 1 and 2, respec-

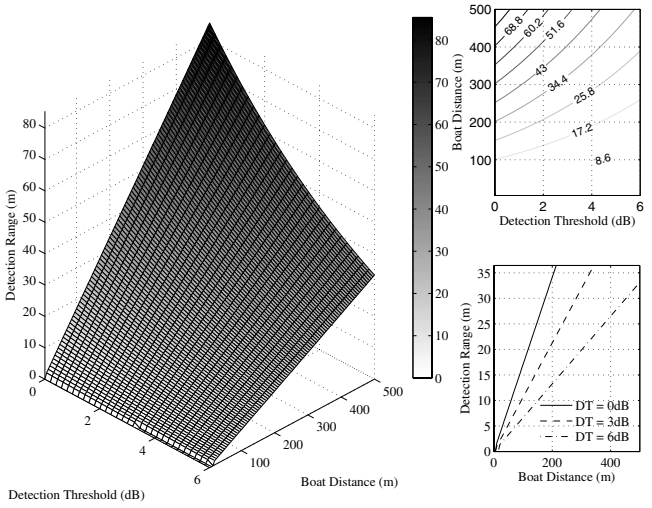


FIG. 10. Manatee vocalization detection ranges after processing with the FF/SU DAP algorithm in the presence of watercraft emitted noise.

tively. Accordingly, it can be shown that the separated vocalization and watercraft noise signals are retrieved from channels 1 and 2, respectively.[27] Since the location of the animal is not known in advance, both channels have to be monitored by a detector.

The separation performance of the FF/SU DAP algorithm depends on the coherence between the measurements. The separation and signal enhancement performance reduces as the cross-channel coherences drop. There are several factors that may affect inter-channel coherence. A greater separation between the hydrophones increases the coverage area of the passive detection system, but also reduces the coherence. Hence, there is a trade-off between coverage and performance. In addition, the acoustic signals recorded from a small recreational watercraft departing away from the hydrophones are relatively incoherent due to the scatter from cavitation of the propeller. Thus, the proposed system will enhance vocalization signals particularly in the presence of an approaching vessel. Increase in the ambient diffuse noise field (e.g., when precipitating) also results in reduced coherence. For cases in which the coherence drops below a certain level, the FF/SU DAP algorithm can be reconfigured to function as a single channel supervised adaptive noise canceller implemented with the affine projection update rule (by forcing one of the de-mixing filters to function as an all-pass unity filter).

V. CONCLUSIONS

The problem of enhancing marine mammal vocalizations in the presence of an interfering acoustic source is addressed in this paper. The conventional approach to underwater signal enhancement is to use expensive directional receivers such as vector sensors or to beamform the measurements from an array of many omni-directional hydrophones. As a low-cost alternative to these methods, a two-channel, unsupervised, adaptive source separation approach is proposed for enhancing noisy manatee vocalizations. The presented FF/SU DAP algorithm uses the affine projection update rule to be able to maintain a satisfactory convergence speed in the presence of dynamic underwater mixing channels (due to the motion of watercraft) and temporally correlated vocalizations. The FF/SU DAP algorithm is evaluated using real watercraft emitted noise and manatee vocalizations in which the input SNR is varied from −10 to 5 dB. The performance of the algorithm is evaluated in terms of the SNR and SDR performance measures. These experiments suggest that the proposed method extends the detection range on average five times compared to high-pass filtering alone. In contrast to conventional beamforming, the presented source separation approach does not require an array of many hydrophones and is not sensitive to uncertainties in the sensor locations. One limitation of the proposed FF/SU DAP algorithm is that the signal enhancement performance declines as the inter-channel coherence between the two channels is reduced (e.g., due to increases in diffuse ambient noise levels). For low coherence measurements, the implementation structure of the FF/SU DAP algorithm allows it to be reconfigured as the conventional supervised adaptive noise canceller which is a robust algorithm for enhancing signals in the presence of multiple uncorrelated noise measurements.

Although the proposed FF/SU DAP algorithm was implemented and evaluated using manatee vocalizations, the approach makes very generic assumptions about properties of the vocalization signals and the shallow underwater acoustic channel. Hence, the proposed method is suitable for extending to other marine mammal monitoring applications such as the detection and the classification of whale calls in deep oceans with minimal modifications to the algorithm parameters.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Sea Grant, Florida Fish and Wildlife Conservation Commission, University of Florida Marine Mammal Program, and the University of Massachusetts Lowell Provost's Office for providing financial support for this research.

[1]U.S. Fish and Wildlife Service, "Florida manatee recovery plan (*Trichechus manatus latirostris*)," 3rd revision, US Fish and Wildlife Service, 5430 Grosvenor Lane, Suite 110, Bethesda, MD 20814 (2001).
[2]Florida Fish and Wildlife Commission, "Florida manatee management plan. Florida Fish and Wildlife Commission," 620 South Meridian Str., Tallahassee, FL 32399-1600 (2007).
[3]Florida Fish and Wildlife Conservation Commission, "Manatee mortality statistics," http://research.myfwc.com/features/category_sub.asp?id=2241 (Last viewed 8/14/2009).
[4]J. F. Gorzelany, "Evaluation of boater compliance with manatee speed zones along the Gulf Coast of Florida," Coastal Manage. **32**, 215–226 (2004).
[5]C. Steel, "Vocalization patterns and corresponding behavior of the West Indian manatee (*Trichechus manatus*)," Ph.D. thesis, Florida Institute of Technology, Melbourne, FL (1982).
[6]D. Nowacek, B. M. Casper, R. S. Wells, S. M. Mowacek, and D. A. Mann, "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations," J. Acoust. Soc. Am. **114**, 66–69 (2003).
[7]R. Phillips, C. Niezrecki, and D. O. Beusse, "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428 (2004).
[8]C. Niezrecki, R. Phillips, M. Meyer, and D. O. Beusse, "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**, 1640–1647 (2003).
[9]R. Phillips, C. Niezrecki, and D. O. Beusse, "Theoretical detection ranges for acoustic based manatee avoidance technology," J. Acoust. Soc. Am. **120**, 153–163 (2006).
[10]Z. Yan, C. Niezrecki, L. N. Cattafesta III, and D. O. Beusse, "Background noise cancellation for improved acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **117**, 3566–3573 (2005).
[11]Z. Yan, C. Niezrecki, L. N. Cattafesta III, and D. O. Beusse, "Background noise cancellation of manatee vocalizations using an adaptive line enhancer," J. Acoust. Soc. Am. **120**, 145–152 (2006).
[12]B. M. Gur and C. Niezrecki, "Autocorrelation based denoising of manatee vocalizations using the undecimated discrete wavelet transform," J. Acoust. Soc. Am. **122**, 188–199 (2007).
[13]Y. Ren, M. T. Johnson, and J. Tao, "Perceptually motivated wavelet packet transform for bioacoustic signal enhancement," J. Acoust. Soc. Am. **124**, 316–327 (2008).
[14]B. Van Veen and K. M. Buckley, "Beamforming techniques for spatial filtering," in *The Digital Signal Processing Handbook*, edited by V. K. Madisetti and D. Williams (CRC, Boca Raton, FL, 1998), pp. 61/1–61/22.
[15]P. Miller and P. L. Tyack, "A small towed beamforming array to identify vocalization *resident* killer whales (*Orcinus orca*) concurrent with behavioral observation," Deep-Sea Res., Part II **45**, 1389–1405 (1998).
[16]F. Desharnais, G. R. Ebbeson, M.-N. R. Matthews, G. J. Heard, D. J. Thomson, and G. H. Brooke, "A generalized beamformer for localization

of marine mammals," Appl. Acoust. **67**, 1213–1225 (2006).
[17]P. B. Muanke and C. Niezrecki, "Manatee position estimation by passive acoustic localization," J. Acoust. Soc. Am. **121**, 2049–2059 (2007).
[18]M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "Convolutive blind source separation methods," in *Springer Handbook of Speech Processing*, edited by J. Benesty, S. M. Sondhi, and Y. Huang (Springer, Berlin, 2008), pp. 1065–1095.
[19]M. Gaeta, F. Briolle, and P. Esparcieux, "Blind separation of sources applied to convolutive mixtures in shallow water," in Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics (1997), pp. 340–343.
[20]S. Bonnifay, K. Yao, and C. Jutten, "Underwater acoustic signal separation based on prior estimation of the channel impulse response," in Proceedings of the ICASSP (2000), Vol. 5, pp. 3156–3159.
[21]A. Mansour, N. Benchekroun, and C. Gervaise, "Blind separation of underwater acoustic signals," in *Proceedings of the ICA, Lecture Series on Computer Science*, edited by J. Rosca, D. Erdogmus, J. C. Principe, and S. Haykin (Springer, Berlin, 2006), Vol. 3889, pp. 181–188.
[22]L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," IEEE Trans. Speech Audio Process. **8**, 320–327 (2000).
[23]H. Buchner, R. Aichner, and W. Kellerman, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," IEEE Trans. Speech Audio Process. **13**, 120–134 (2005).
[24]Z. H. Michalopoulou, "Underwater transient signal processing: Marine mammal identification, localization, and source signal deconvolution," in Proceedings of the ICASSP (1997), Vol. **1**, pp. 503–506.
[25]S. Van Gerven and D. Van Compernolle, "Signal separation by adaptive symmetric decorrelation: Stability, convergence, and uniqueness," IEEE Trans. Signal Process. **43**, 1602–1612 (1995).
[26]S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, Upper Saddle River, NJ, 2002), Chap. 5, pp. 231–320.
[27]E. Weinstein, M. Feder, and A. V. Oppenheim, "Multi-channel signal separation by decorrelation," IEEE Trans. Speech Audio Process. **1**, 405–413 (1993).
[28]G. E. Bottomley and S. T. Alexander, "A novel approach for stabilizing recursive least squares filters," IEEE Trans. Signal Process. **39**, 1770–1779 (1991).
[29]K. Ozeki and T. Umeda, "An adaptive filtering algorithm using orthogonal projection to an affine subspace and its properties," Electron. Commun. Jpn. **67**, 19–27 (1984).
[30]M. Gabrea, "Double affine projection algorithm-based speech enhancement algorithm," in Proceedings of the ICASSP (2003), Vol. **1**, pp. 904–907.
[31]B. M. Gur, "Adaptive enhancement of marine mammal vocalizations," Ph.D. thesis, University of Massachusetts, Lowell, MA (2008).
[32]D. Schobben, K. Torkkala, and P. Smaragdis, "Evaluation of blind signal separation methods," in Proceedings of the ICA (1999), pp. 261–266.
[33]T. O'Shea, "Manatee vocalization-catalog of sounds," produced by Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL (1981–1984).
[34]Z. Yan, "Background noise cancellation for acoustic detection of manatee vocalizations," Ph.D. thesis, University of Florida, Gainesville, FL (2006).

Passive cavitation imaging with ultrasound arrays

Vasant A. Salgaonkar, Saurabh Datta, Christy K. Holland, and T. Douglas Mast[a)]
Department of Biomedical Engineering, University of Cincinnati, 231 Albert Sabin Way, Cincinnati, Ohio 45267-0586

(Received 12 January 2009; revised 21 August 2009; accepted 4 September 2009)

A method is presented for passive imaging of cavitational acoustic emissions using an ultrasound array, with potential application in real-time monitoring of ultrasound ablation. To create such images, microbubble emissions were passively sensed by an imaging array and dynamically focused at multiple depths. In this paper, an analytic expression for a passive image is obtained by solving the Rayleigh–Sommerfield integral, under the Fresnel approximation, and passive images were simulated. A 192-element array was used to create passive images, in real time, from 520-kHz ultrasound scattered by a 1-mm steel wire. Azimuthal positions of this target were accurately estimated from the passive images. Next, stable and inertial cavitation was passively imaged in saline solution sonicated at 520 kHz. Bubble clusters formed in the saline samples were consistently located on both passive images and B-scans. Passive images were also created using broadband emissions from bovine liver sonicated at 2.2 MHz. Agreement was found between the images and source beam shape, indicating an ability to map therapeutic ultrasound beams *in situ*. The relation between these broadband emissions, sonication amplitude, and exposure conditions are discussed.
 [DOI: 10.1121/1.3238260]

PACS number(s): 43.60.Fg, 43.35.Ei, 43.80.Vj [CCC] Pages: 3071–3083

I. INTRODUCTION

This paper presents a method for spatially sensitive passive cavitation detection with a conventional ultrasound imaging array. Here, microbubble position is ascertained passively by sensing acoustic emissions with a synthetically focused imaging array. The resulting two-dimensional (2D) maps are referred to as passive cavitation images. Acoustic cavitation may be defined as any observable activity of a bubble population under acoustic stimulation[1] and is often categorized into stable and inertial cavitation.[2] The overall motivation for this work is to map cavitation effects produced during ultrasound ablation. Therapeutic ultrasound fields induce nonlinear bubble oscillations causing subharmonic emissions related to stable cavitation, broadband emissions associated with inertial cavitation,[3] and kilohertz frequency emissions resulting from vaporous boiling.[4] Passive images of these acoustic emissions can provide spatial information about the location and dynamics of an active bubble population.

Microbubble activity is known to play an important role in several ultrasound-based therapeutic applications, some of which are briefly reviewed. In shock-wave lithotripsy, kidney stones are broken by inertially collapsing bubbles that generate shock waves and fluid micro-jets, leading to stone surface erosion and fragmentation.[5–7] Several researchers suggest that micro-streaming due to stable cavitation and micro-jetting due to inertial cavitation enhance thrombolytic mechanisms.[8–10] Ultrasound-induced cavitation also offers the exciting possibility of targeted drug and gene delivery using lipid-coated microbubbles which can encapsulate bioactive materials.[11–13] Cavitation activity occurring during ultrasound ablation of soft tissue[14,15] is postulated to increase ultrasound absorption.[16] This can result in faster treatments but can also complicate energy deposition and distort ablative lesion shapes.[17–20]

Passive cavitation detection methods have been employed to monitor cavitation arising from ultrasound-tissue interaction. During some recent experimental studies on ultrasound ablation with focused and unfocused, continuous-wave (CW) sources, subharmonic and broadband-emission signals, consistent with stable bubble oscillations and inertial collapse, respectively, have been detected passively.[21–23] Some researchers have found significant correlation between broadband-emission levels and tissue echogenicity following soft tissue ablation with high-intensity focused ultrasound (HIFU) sources *in vivo*[24] and unfocused intense ultrasound *in vitro*.[23] *In vitro* studies involving HIFU[25] and unfocused ultrasound[23] have indicated that increased broadband emissions corresponded with increase in the volume of coagulated or necrosed tissue. Another study with HIFU ablation of tissue-mimicking phantoms documents a sudden decrease in broadband-emission levels with the onset of boiling.[21] During tissue boiling, low-frequency (kilohertz) acoustic emissions have also been detected experimentally with both HIFU[4,22] and unfocused intense ultrasound[23] A previous *in vitro* study of bovine liver ablation with unfocused ultrasound found that subharmonic emission levels depended on tissue temperature as well as sonication amplitude.[23] Passive cavitation detection is a promising basis for ultrasound ablation guidance and control.

A typical passive cavitation detection system employs a single-element focused transducer, where the detector focus is aligned with the expected location of cavitation activity. In the case of HIFU ablation, acoustic cavitation and gas formation due to tissue boiling can cause the ablative lesion (tissue coagulation) to grow toward the ultrasound source.[17,20] However, a single-element focused passive detector cannot provide information about position-dependent

[a)]Author to whom correspondence should be addressed. Electronic mail: doug.mast@uc.edu

changes in cavitation activity during ablative lesion formation. An unfocused detector can be used to obtain a bulk recording of acoustic emissions, but at the cost of spatial sensitivity.[23] Multiple detectors interrogating the insonation volume may enable the measurement of position-dependent cavitation. A multi-element detector like an ultrasound array has the potential to provide 2D-resolved measurement of cavitational acoustic emissions. Passive imaging with arrays has been attempted in the detection of subterranean acoustic sources using back-projection and cross-correlation methods.[26] Cross-correlation methods are also under investigation for cavitation imaging during HIFU ablation.[27] These algorithms require a synchronous signal acquisition by array elements. Many conventional ultrasound imaging systems are not designed for passive data capture and their available beamforming methods are not suitable for this kind of imaging. The development of customized beamforming methods is needed for passive cavitation imaging using a diagnostic ultrasound system. Recently, passive imaging was applied in HIFU ablation of tissue phantoms, where a conventional ultrasound array was used in a passive mode by switching off the initial transmit pulse and spatial sensitivity was obtained in the azimuthal direction (normal to the array axis).[21]

This paper describes beamforming methods to scan array subapertures electronically and receive acoustic emission signals sequentially for each receive-focus position. The signal level detected at each focal position gives the brightness value for the corresponding point on a 2D map, called a passive cavitation image. Here, analytic expressions for passive cavitation image formation are described.[28] Simulations of passive cavitation images with different beamforming configurations are shown. Finally, experiments with passive cavitation imaging in phosphate-buffered saline (PBS) and *ex vivo* liver tissue are presented. Passive cavitation methods developed here have potential application in real-time, direct visualization of ultrasound-tissue interaction, particularly in guidance and monitoring of ultrasound ablation.

II. THEORY

In B-mode imaging, ultrasound arrays transmit a short acoustic pulse and the medium is imaged from the resulting echoes. The reflector distance is determined through time-of-flight calculations and the image brightness depends on medium properties such as acoustic impedance and attenuation.[29] In contrast, passive cavitation images as defined here resolve acoustic sources by dynamically focusing their passively detected acoustic emissions. Thus, for passive cavitation imaging, image characteristics such as spatial resolution, contrast, and dependence on imaging system parameters differ from conventional B-mode imaging. Below, analytic expressions are derived for passive cavitation images with aims to elucidate experimental results and optimize imaging system parameters through computer simulations.

A. Signal received by a rectangular array element

To ascertain the emission signal detected by a rectangular array element, a cavitating bubble is modeled as a point-source emitting spherical sound waves and the resulting pressure field is integrated over the element surface. For a pulsating sphere of finite radius a, the volume flow Q of fluid from the source relates to the particle velocity u as[30]

$$Q(t) = 4\pi a^2 \frac{\partial u(a,t)}{\partial r}. \tag{1}$$

In the small-radius limit $a \to 0$, the pressure field at a position $\mathbf{r}$ for the point-source (bubble) radiating in a medium with density ρ and sound speed c is given by[30]

$$p(\mathbf{r},t) = \frac{\rho \dot{Q}(t - |\mathbf{r}|/c)}{4\pi|\mathbf{r}|}. \tag{2}$$

The received emission signal $s_n(t)$ for an element n centered at $\mathbf{r}_0$ (area $\mathbf{S}_0$) due to a bubble at $\mathbf{r_s}$ is obtained by integrating the radiated pressure from Eq. (2) over the element surface. This results in the Rayleigh–Sommerfield integral[31]

$$s_n(t) = \oint \frac{\rho \dot{Q}(t - |\mathbf{r_0} - \mathbf{r_s}|/c)}{4\pi|\mathbf{r_0} - \mathbf{r_s}|} dS_0, \tag{3}$$

which can be expressed in the frequency domain as

$$S_n(\omega) = \frac{-i\omega\rho\dot{\mathcal{Q}}(\omega)}{4\pi} \oint \frac{e^{(i|\mathbf{r_0}-\mathbf{r_s}|\omega/c)}}{|\mathbf{r_0} - \mathbf{r_s}|} dS_0, \tag{4}$$

where ω denotes radial frequency and dS_0 is an area-element on the receiver surface. An approximate solution of the frequency-domain Rayleigh–Sommerfield integral can be obtained using the Fresnel approximation,

$$|\mathbf{r_0} - \mathbf{r_s}| \approx z_s + \frac{(x_s - x_0)^2}{2z_s} + \frac{(y_s - y_0)^2}{2z_s},$$

$$\frac{1}{|\mathbf{r_0} - \mathbf{r_s}|} \approx \frac{1}{z_s}. \tag{5}$$

For an array element with dimensions $2a \times 2b$ focused at depth F_x in elevation and F_y in azimuth, Eq. (4) can be written under the Fresnel approximation as[32]

$$\begin{aligned} S_n(\omega) &= \frac{-i\omega\rho\dot{\mathcal{Q}}(\omega)}{4\pi z_s} e^{ik(z_s^2 + |\mathbf{r_s}|^2)/2z_s} \\ &\quad \times \int_{-a}^{a} e^{ikx_0^2/2F_x} e^{ik(x_0^2 - 2x_0x_s)/2z_s} dx_0 \\ &\quad \times \int_{-b}^{b} e^{iky_0^2/2F_y} e^{ik(y_0^2 - 2y_0y_s)/2z_s} dy_0, \\ &= \frac{-i\omega\rho\dot{\mathcal{Q}}(\omega)\Phi(\mathbf{r_s},k)}{4\sqrt{\tilde{k}_x}\sqrt{\tilde{k}_y}} \left(\mathbf{F}\left[\frac{kx_s + \tilde{k}_x a}{\sqrt{\pi\tilde{k}_x z_s}} \right] \right. \\ &\quad \left. - \mathbf{F}\left[\frac{kx_s - \tilde{k}_x a}{\sqrt{\pi\tilde{k}_x z_s}} \right] \right) \times \left(\mathbf{F}\left[\frac{ky_s + \tilde{k}_y b}{\sqrt{\pi\tilde{k}_y z_s}} \right] \right. \\ &\quad \left. - \mathbf{F}\left[\frac{ky_s - \tilde{k}_y b}{\sqrt{\pi\tilde{k}_y z_s}} \right] \right), \end{aligned} \tag{6}$$

where the wave number $k=\omega/c$, $\tilde{k}_x=k(1-z_s/F_x)$, and $\tilde{k}_y=k(1-z_s/F_y)$. $\mathbf{F}$ is the complex Fresnel integral which is given by $\mathbf{F}(\zeta)=\int_0^\zeta e^{i\pi u^2/2}du$. The multiplicative term Φ is expressed as $\Phi(\mathbf{r_s},k)=\exp(ik[z_s^2+|\mathbf{r_s}|^2-k(x_s^2/\tilde{k}_x+y_s^2/\tilde{k}_y)]/(2z_s))$.

B. Analytic expression for passive images

In the passive cavitation imaging method presented here, acoustic emissions are spatially resolved through synthetic focusing of array subapertures at multiple depths. For simplicity, a subaperture may be approximated as a single, uniform, continuous rectangular transducer, which produces a voltage proportional to the received acoustic pressure, integrated over the element surface.[31] The brightness value for a passive image point (Y,Z) is defined as the total beamformed energy from a group of acoustic sources sensed by a subaperture which is focused at that point,

$$I(Y,Z)=\sum_{\forall\omega}|S(\omega,Y,Z)|^2, \tag{7}$$

where ω is the radial frequency of a single radiating frequency component and the symbol $\forall$ ("for all") indicates summation over all radiating frequencies. In the case of stably cavitating bubbles, these correspond to discrete harmonics, subharmonics, and ultraharmonics of the sonication frequency.

To obtain the point-spread function for a passive cavitation image, an expression for the received signal $S(\omega,Y,Z)$ due to a single source at (x_s,y_s,z_s) is obtained by idealizing the linear array to have small pitch, no edge limits, and continuous receive focusing, so that an N-element subaperture can be represented by a single rectangular transducer of length $2b\times N$:

$$\begin{aligned}S(\omega,Y,Z)=&\frac{-A_0 i\omega\rho\dot{Q}(\omega)\Phi(\mathbf{r_s},k)}{4\sqrt{\tilde{k}_x}\sqrt{\tilde{k}_y}}\\&\times\left(\mathbf{F}\left[\frac{kx_s+\tilde{k}_x a}{\sqrt{\pi\tilde{k}_x z_s}}\right]-\mathbf{F}\left[\frac{kx_s-\tilde{k}_x a}{\sqrt{\pi\tilde{k}_x z_s}}\right]\right)\\&\times\left(\mathbf{F}\left[\frac{ky_s-Y+\tilde{k}_y bN}{\sqrt{\pi\tilde{k}_y z_s}}\right]\right.\\&\left.-\mathbf{F}\left[\frac{ky_s-Y-\tilde{k}_y bN}{\sqrt{\pi\tilde{k}_y z_s}}\right]\right),\end{aligned} \tag{8}$$

where $\tilde{k}_y=k(1-z_s/Z)$ and Φ is as defined for Eq. (6).

C. Passive images by time-delay focusing

To model the point-spread function of a passive cavitation image created using time-delay receive beamforming, the signal received by an individual array element [Eq. (6) with $F_y\to\infty$] is subjected to an appropriate delay. Time-delayed, received signals from all elements in a subaperture are summed to obtain the beamformed emission signal. To focus a subaperture with N elements on its axis at depth Z, the time delay applied to the nth element is given by

$$\tau_n=\frac{-Z-\sqrt{(y_n-Y)^2+Z^2}}{c}. \tag{9}$$

The beamformed emission signal, corresponding to an image brightness value, is then expressed as

$$I(\omega,Y,Z)=\sum_{\forall\omega}\left|\sum_{\forall n}e^{i\omega\tau_n}S_n(\omega)\right|^2. \tag{10}$$

III. SIMULATIONS

The theory discussed above was used to simulate passive cavitation images using MATLAB 7.1 (Mathworks Inc., Natick, MA) software. Simulations were performed for the 192-element, 7.5-MHz center frequency linear array (L7, Ardent Sound, Mesa, AZ) employed in the experiments reported below. Each element on this array is 0.195 mm in azimuth and 7.0 mm in elevation, with pitch 0.22 mm and a constant focal depth of 25 mm in elevation. The maximum allowable subaperture dimension for the imaging system employed here is 64-elements, spanning 14.08 mm. The minimum subaperture size was set to 8-elements, spanning 1.76 mm. The imaging system allowed discrete receive-focusing to be performed in 16 distinct focal zones. Received signals due to cavitational emissions were computed from Eq. (6) using a rational approximation to the complex Fresnel integral.[33]

A. Simulated passive images of a point-source

Simulated passive images are first shown for a single point-source located on the array axis at 20-, 55-, or 90-mm depth. The simulated sources were modeled to emit acoustic energy at harmonics of 520-kHz between 5.2 and 9.36 MHz. This choice of fundamental frequency, harmonic components, and frequency bandwidth is consistent with the harmonic scattering experiments described below. Beamforming of emissions was simulated using Eq. (6) for a 14.08-mm subaperture and image brightness was computed by summing beamformed emission energy for each harmonic according to Eq. (7).

Images shown in the top row of Fig. 1 were simulated using Eq. (6) for an idealized subaperture with continuous receive-focusing and no truncation at the array edges. Images in the bottom row of Fig. 1 were created by time-delay beamforming the received signal energy, as given in Eq. (10). Receive foci were placed at 16 distinct depths separated by 6.2 mm. The total image depth was 101 mm and subaperture sizes were truncated at the array edges. For example, the 32nd A-line was formed by beamforming the received signals from elements 1–64, but the 4th A-line was comprised of signals received by elements 1–8. All these parameters were consistent with specifications of the ultrasound imaging system used in the experiments reported in Sec. IV. Figure 2 shows simulated images of a point-source created with a constant f-number (ratio of receive-focus depth to subaperture width) subaperture, while the other imaging parameters were same as Fig. 1. It can be seen from Fig. 1 that the brightness pattern narrows or converges at the point-source depth, while it broadens at greater or lesser depths.

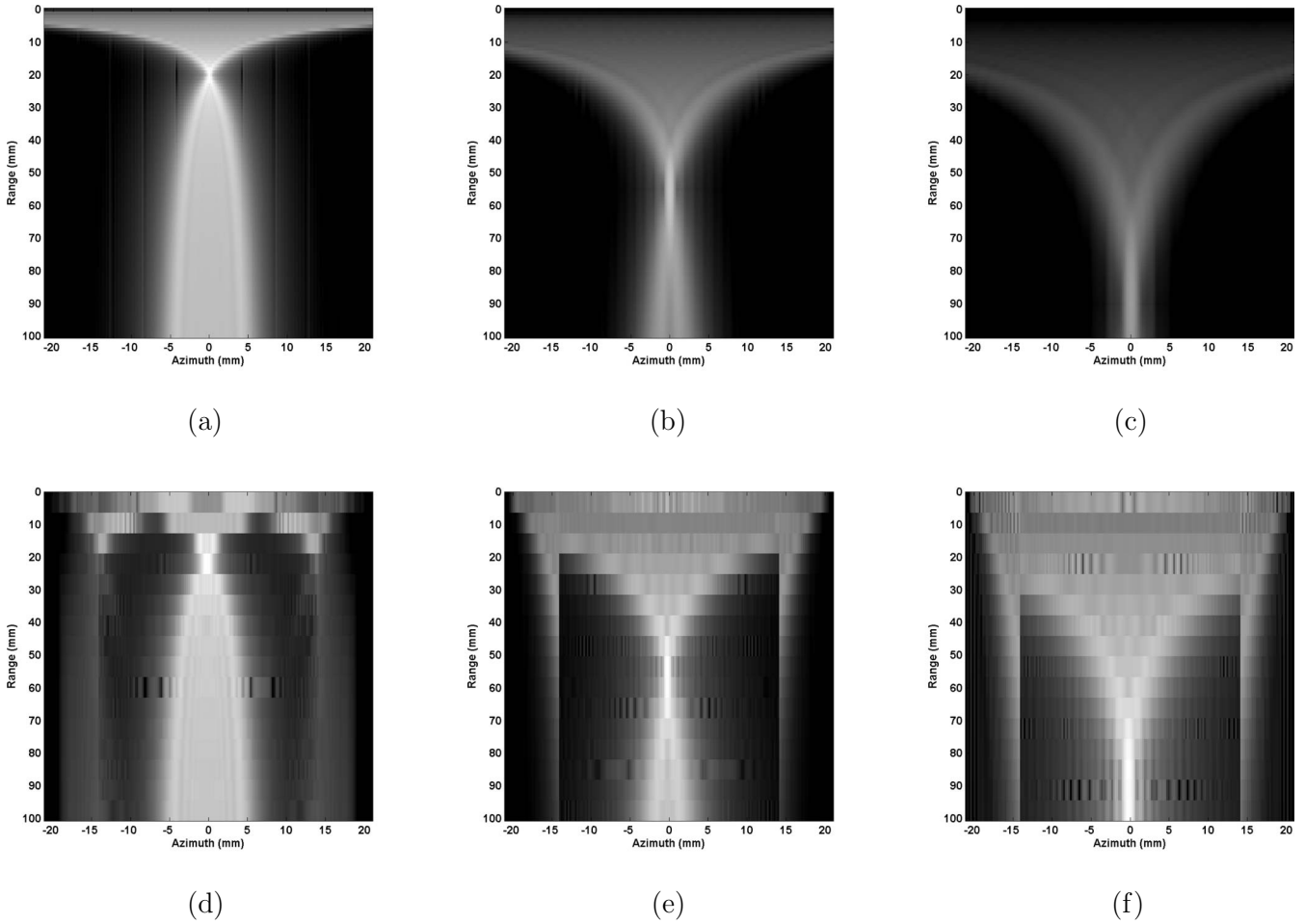


FIG. 1. Computer simulations for passive images created using 64-element (14.08 mm) subapertures are shown here. The top row contains examples of simulations where an "idealized" array approximation is used for each subaperture, with single point-sources at depths (a) 20 mm, (b) 55 mm, (c) and 90 mm. Simulated images in the bottom row were created through time-delay focusing at 16 equally spaced depths (6.2 mm) with single point-sources located at (d) 20 mm, (e) 55 mm (e), and (f) 90 mm. All images are plotted with a 40 dB dynamic range.

This is true for both "idealized" and time-delay approaches. With a constant f-number, the corresponding image shows similar characteristics but the brightness pattern on the image converges at a greater depth than the point-source range.

The beamforming configuration influences image resolution depending on the size of the synthetic focus. A good measure for this is the −6-dB beam width of the receiver sensitivity pattern of the subaperture in the azimuthal direction. The imaging array is assumed to have a fixed focus in elevation (similar to the L7 array employed in the experiments described below). The focus size in the azimuthal direction changes with subaperture size and focal depth, and can be approximated by the focal width of an unapodized line aperture.[29] For a wavelength λ, the −6-dB width for an N-element subaperture is thus approximated as

$$W_{-6\text{-dB}} \approx \frac{F_y \lambda}{2Nb}. \tag{11}$$

Hence, image resolution in azimuth is finer at higher frequencies. For constant subaperture widths, this resolution becomes coarser with increasing depth, as seen on Fig. 1. For a constant f-number subaperture, azimuthal width of the receive focus is unchanged with depth. Hence, azimuthal resolution stays the same even at greater depths, as seen in Fig. 2.

B. Simulated passive images of multiple bubbles

In a clinical or experimental application, an ability to image multiple bubbles is needed. Analytic expressions, derived above, can be used to assess passive imaging performance in the presence of emissions from multiple bubbles. Figure 3(a) is a simulated image of five acoustic sources with emissions in the frequency range 6.3–6.7 MHz (similar to experiments described below). For simplicity, it is assumed that all the sources continuously emit acoustic energy from their respective locations. The acoustic emission is assumed to be a stationary process, i.e., the temporal average of the emission energy stays constant during the formation of the entire image. These five sources are located at positions (azimuth mm, range mm): (−12,40), (−10,40), (0, 10), (0, 40), (0, 70). The source closest to the array at 10-mm depth can be resolved. However, the other two sources deeper on the axis cannot be seen very clearly on the image. The two off-axis sources, which are at the same depth and separated in the azimuth by 2 mm, can be clearly resolved. The azimuthal

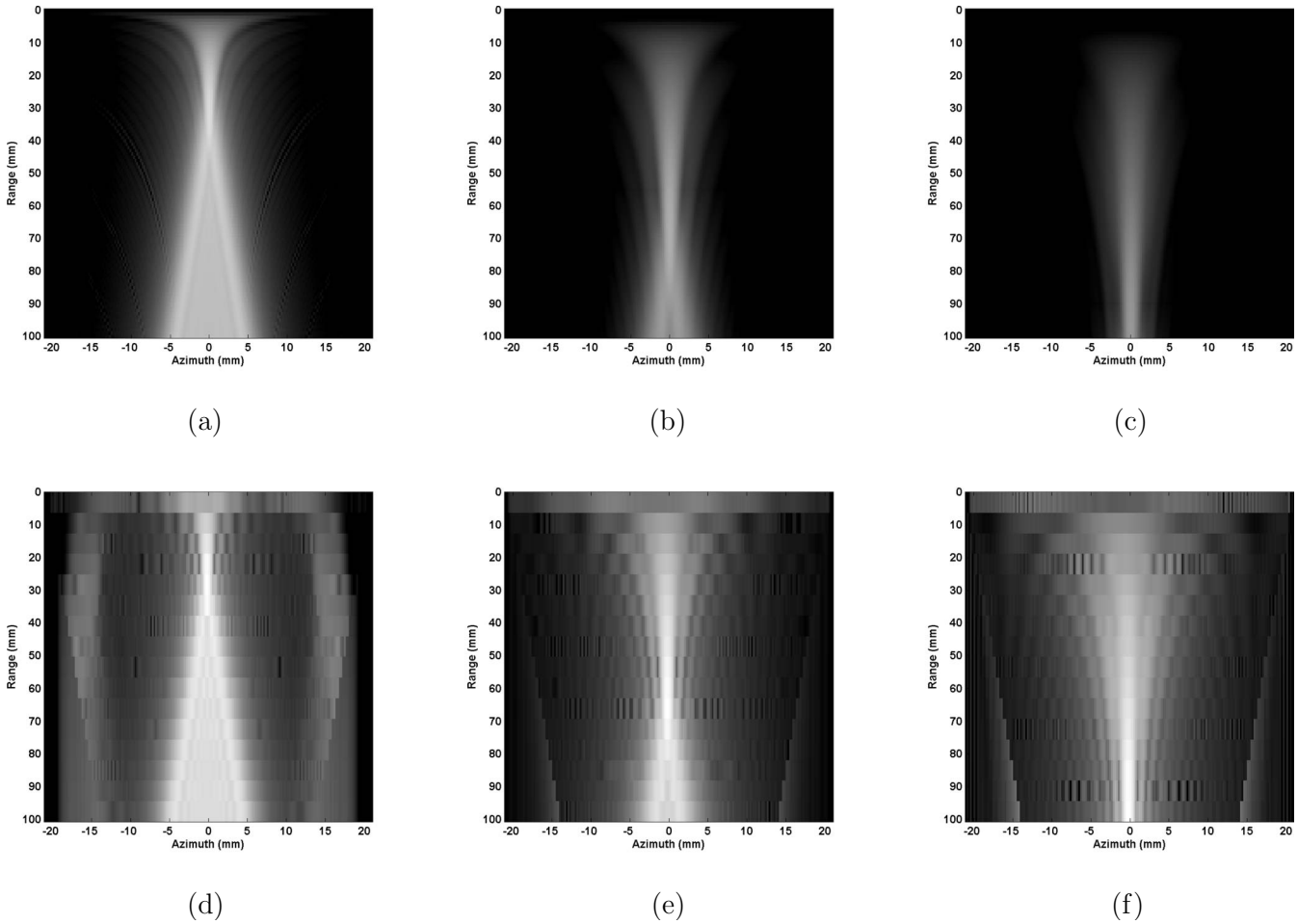


FIG. 2. Computer simulations for passive images created using a constant f-number (7.1) subaperture are shown here. Except for the subaperture sizes, images were created using the same simulation methods, focus positions, and source locations as the corresponding panels in Fig. 1.

width of the synthetic focus is smaller than the range width, especially at greater depths. This leads to better resolution in the azimuthal direction.

During HIFU ablation, cavitation activity is known to occur near the focal region, which may result in a cluster of active microbubbles.[16] It is important to determine whether passive imaging can be employed to estimate the location and size of such a cloud of multiple acoustic sources. Figure 3(b) has a simulated image of 25 sources emitting energy between 6.3 and 6.7 MHz. This frequency range is consistent with experiments in saline solution, sonicated with 520-kHz ultrasound. The 25 point-sources are randomly placed in the image plane (zero elevation) between −7 to −3 mm in azimuth and 31–41 mm in range. On this image, the azimuthal position of the bubble cluster can be identified more readily than its position in depth direction. To test whether the passive image is consistent with the location of the simulated bubble cluster, the brightness value of the passive image [Eq. (7)] was averaged between depths 31 and 41 mm for the entire image width. This yielded the azimuthal distribution of the averaged brightness value shown in Fig. 3(c). The number of point-sources at a given azimuthal location was represented by a histogram, shown in Fig. 3(c). For comparison, the azimuthal distribution of brightness value on the passive image and the bubble location histogram were normalized and scaled between 0 and 1. The azimuthal brightness distribution was consistent with this histogram. A similar analysis was performed in the range (or depth) direction. In Fig. 3(d), the passive image brightness value was averaged between −7 and −3 mm azimuth for all depths and compared with the histogram for range locations of the point-sources. Greater agreement is seen between the azimuth-dependent emission energy and source distribution in Fig. 3(c), compared to the corresponding depth-dependent distributions in Fig. 3(d).

IV. EXPERIMENTAL METHODS

Experiments were conducted to test passive cavitation imaging methods in free-field and tissue media. The overall experimental set up is shown Fig. 4. Details of experimental system components, experimental configurations, and data processing methods are presented below.

A. Ultrasound imaging system

Passive imaging was performed with a 192-element linear array with center frequency of 7.5 MHz, array element dimensions of 7×0.195 mm^2 (elevation × azimuth), and pitch of 0.22 mm (L7, Ardent Sound, Mesa, AZ), and controlled by the Iris 2 imaging system (Ardent Sound, Mesa,

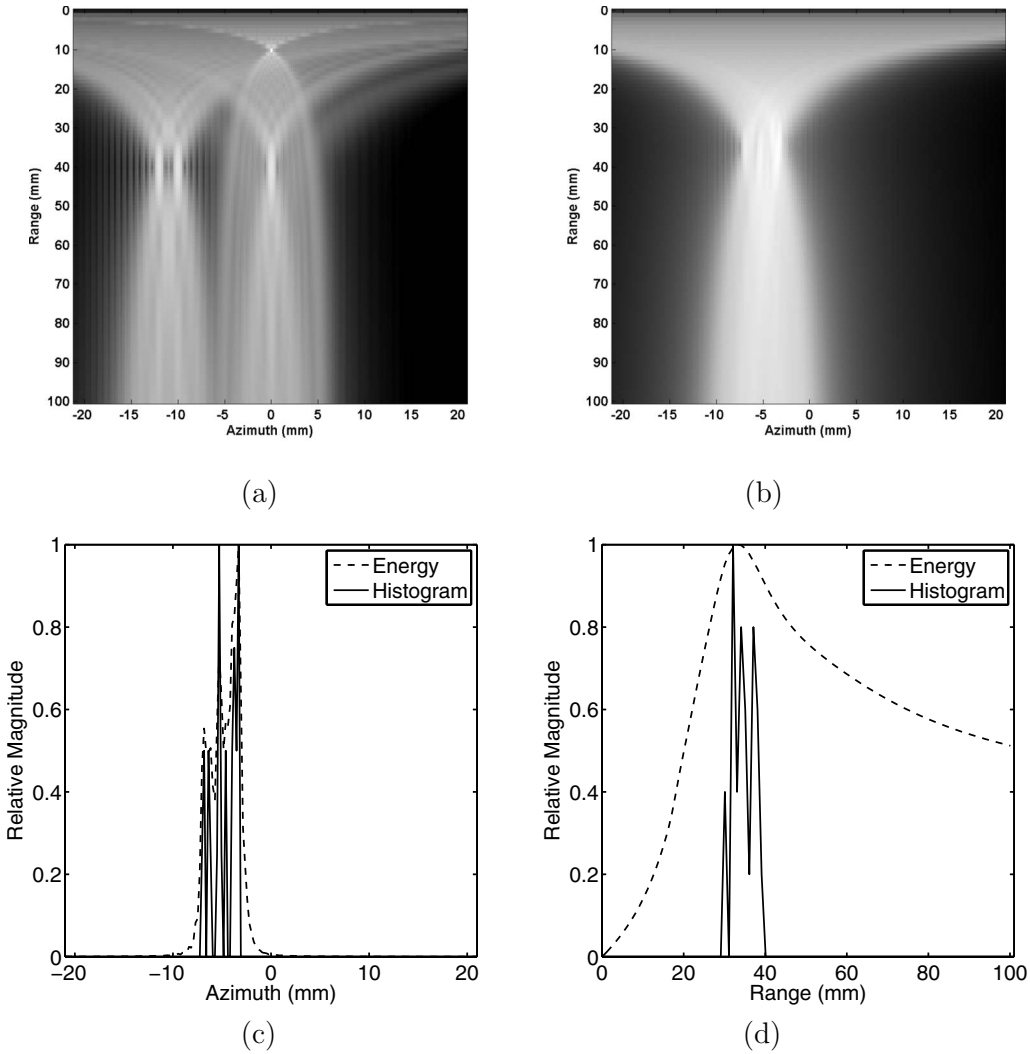


FIG. 3. Point-sources represented in this figure were assumed to emit acoustic energy between 6.3 and 6.7 MHz (64-element subapertures). (a) Simulated image of five point-sources at (azimuth mm, range mm): (−12,40), (−10,40), (0, 10), (0, 40), (0, 70). (b) Simulated image of a cluster of 25 point-sources placed randomly in the image plane between −7 and −3 mm in azimuth, and 3141 mm in depth. (c) Comparison between depth-integrated simulated image brightness as a function of azimuth, and the azimuthal distribution of point-sources. (d) Comparison between azimuth-integrated simulated image brightness as a function of range, and the range distribution of point-sources. Passive cavitation images are plotted with a 40 dB dynamic range.

AZ). The imaging array was focused in elevation (shorter array dimension) with an acoustic lens (focal length = 25 mm). This made it less sensitive to out-of-plane acoustic emission during cavitation experiments. This array was operated under constant width subaperture (64-elements) and constant *f*-number subaperture (7.1) beamforming configurations, with 16 receive focal zones (length of 6.2 mm) and a total image depth of 101 mm. Image frames were captured at

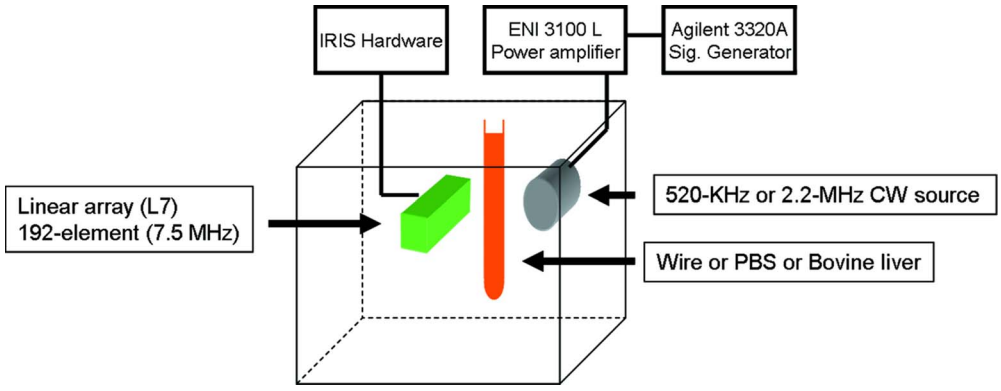


FIG. 4. (Color online) Experimental setup: CW ultrasound sources sonicate a 1-mm steel wire, PBS and bovine liver, while a 192-element linear array captures passive images.

a rate of 28 Hz and the beamformed rf lines were digitally recorded using a personal computer based A/D card (Compuscope CS14200, Gage Applied, Montreal, Canada) at a sampling frequency of 33.3 MHz.

B. Transducers

Two CW ultrasound sources were used for sonication. Their specifications were (1) C302, 520 kHz, 1 in. diameter, unfocused (Panametrics, Waltham, MA) and (2) IX327, 2.2 MHz, 4×15 mm^2 (UTX, Holmes, NY), cylindrically focused at a depth of $\approx$41 mm. They were driven with a signal generator (33220A, Agilent, Santa Clara, CA) and a power amplifier (3100L, ENI, Bell Electronics, Kent, WA).

During the initial experiments, dealing with cavitation detection in free-field, a single-element passive cavitation detector was used in conjunction with the imaging array to record bubble emissions. This was a cylindrically focused transducer, made from PVDF material, 0.75 in. diameter, center frequency of 10 MHz, and focal depth of 19 mm (Valpey Fischer 46654, Hopkinton, MA). The signals received by this detector were amplified by a low-noise preamplifier (SR 560, Stanford Research Systems, Sunnyvale, CA) and then digitally recorded (sampling rate = 50 MHz) using a Waverunner 6050A oscilloscope (LeCroy Corp., Chestnut Ridge, NY).

C. Transducer calibration

Each source transducer was calibrated using a scanning hydrophone system. This system consisted of a motorized three-axis translation assembly (NF90 series, Velmex Inc., Bloomfield, NY), with a 0.02 mm precision, to move needle hydrophones (0.5 mm diameter; SN 1239, Precision Acoustics Ltd., Dorchester, United Kingdom) through the acoustic fields. The signals measured by the hydrophone were recorded on a digital oscilloscope (Waverunner DSO, LeCroy Corp., Chestnut Ridge, NY) and stored on a desktop computer. The entire system was controlled by subroutines written in LABVIEW (National Instruments Corp., Austin, TX), executing on the same computer.[34]

D. Experiment configurations

Passive cavitation imaging was tested using three different experimental configurations. The source transducers were used to induce cavitation in PBS solution and *ex vivo* bovine liver. The ultrasound sources, imaging array and samples were suspended in a glass tank filled with de-ionized, degassed ($\%O_2 < 35$) water, filtered to keep the particle size less than 0.2 μm. To align the single-element transducers, they were driven in a pulse-echo mode by an ultrasound analyzer (5052UAX50, Panametrics, Waltham, MA) and a steel disk ($\approx$2 mm in diameter) was used as a target. The position of the steel disk, relative to the ultrasound array, was ascertained from B-mode images. These B-scans were obtained by the same array used for passive imaging.

1. Prediction of acoustic source position

Measurement of emissions from a single emission source could assist in validating the passive imaging models, described above. Instead of creating a single bubble or using a hydrophone as a point-source,[35] for simplicity, scattered ultrasound from a 1-mm steel wire, orthogonal to the image plane, was recorded passively. The wire was moved to 21 distinct positions in the image plane and passive images were captured. Scattered harmonics of the 520-kHz fundamental between 5.2 and 9.36 MHz (bandwidth of the L7 array) were resolved with a constant f-number configuration.

2. Cavitation imaging in PBS

To test passive imaging of cavitational emissions from a cluster of bubbles, cavitation activity was created by sonicating PBS solution in a 30-mm latex condom, placed orthogonal to the image plane. The PBS solution was exposed to 520-kHz, continuous-wave ultrasound between 0 and 0.15 MPa (peak-negative). During some initial experiments, passive cavitation detection was also performed with a 10-MHz transducer, along with passive imaging. Axes of the single-element detector and the source were perpendicularly aligned in the image plane.[23] Signals values recorded with the source transducer turned off were used as reference noise levels.

3. Cavitation imaging *in* ex vivo *bovine liver*

To assess passive cavitation imaging in sonicated soft tissue, *ex vivo* bovine liver was exposed to 2.2-MHz, continuous-wave, focused ultrasound. The source was aligned using a pulse-echo technique with the propagation direction orthogonal to the image plane and the source focus at 20 mm depth on the imaging array axis. The sonication amplitude was increased from 0–1.96 MPa peak-to-peak pressure, corresponding to 0–0.58 MPa peak-negative pressure. These pressures were measured in the free-field using the scanning hydrophone system mentioned above.

For the experiments described here, four samples of fresh bovine liver, less than 12 h *postmortem*, were used. Each sample was cut to a size of $7 \times 3 \times 3$ cm^3. One side (7×3 cm^2) of this sample had the liver capsule intact and this side faced the source transducer. During the four experiments, which were conducted over a three hour period, the bovine liver was placed in saline over ice, to minimize tissue decay and formation of any gas bubbles. Initial temperature of the tissue samples was between 17 °C and 21 °C. To prevent significant temperature increase due to absorption of the source ultrasound, the 2.2-MHz transducer was driven for single cycles of 4 s, and then turned off for at least 55 s. This on/off sequence was repeated for 16 increasing acoustic pressure levels. After each exposure the tissue sample was sliced systematically near the spatial position of the HIFU focus to confirm that there was no tissue coagulation.

No temperature measurements were made during the exposure. The temperature rise for these exposures, estimated through numerical methods, was 2.93 °C for one duty cycle at the highest sonication amplitude. Section VI contains

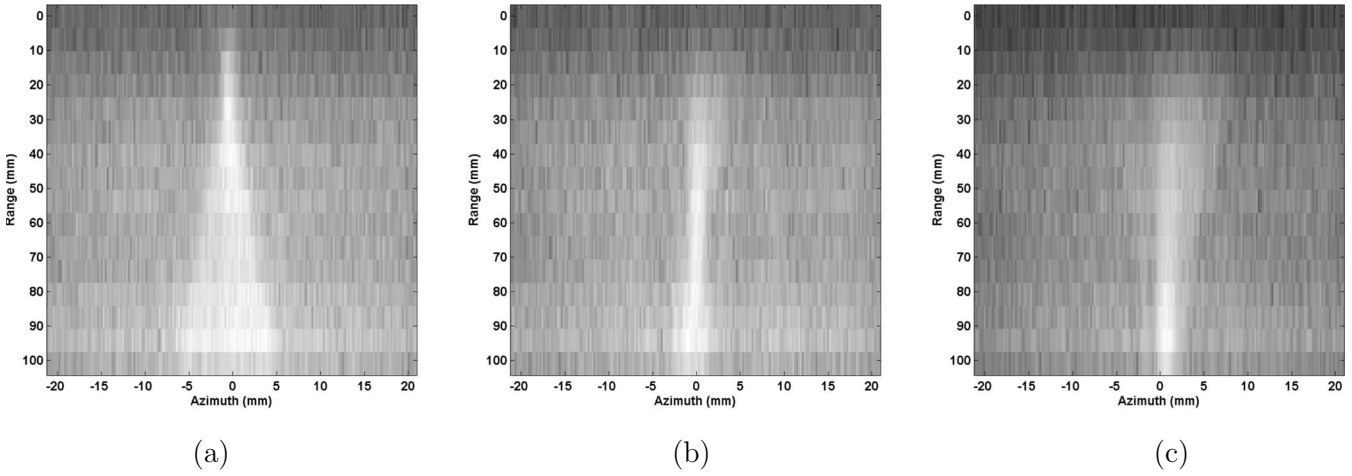


FIG. 5. (a) Passive images of 520-kHz ultrasound scattered from a 1-mm steel wire located at 20-mm depth, (b) 55 mm and (c) 90 mm. A constant *f*-number (7.1) subaperture was employed. Energy in the source harmonics (5.2–9.36 MHz) was integrated. Passive cavitation images are plotted with a 40 dB dynamic range.

some information about the effect of temperature on cavitation activity and calculations for the estimated temperature rise.

E. Power spectrum estimation

To form passive cavitation images, each rf *A*-line was beamformed using receive foci at multiple depths. The focus position corresponds to a passive image point and is at the center of a focal zone. The brightness value at an image point is equal to the integrated emission energy contained in the bounding focal zone. To calculate the emission energy, power spectra were estimated for the beamformed rf *A*-line signals within each focal zone. Energy in frequency bands consistent with inertial and stable cavitation was integrated to form separate images for the two regimes.

Estimation of power spectra was done over a total of 36 image frames, each consisting of 192 rf lines (101-mm image depth). Each rf line was beamformed with 16 equally spaced receive foci. Each focal zone consisted of 276 data points, with the synthetic focus positioned at the center of the zone, corresponding to a point on the passive image. Hence, every passive image had 192 × 16 points (number of rf lines per frame × number of focal zones; azimuth × range). For each focal zone, the sampled acoustic emission segments from nine consecutive frames were concatenated to form 2494-point signals, allowing power spectrum estimation with high frequency resolution. No temporal windowing function was applied. Power spectra were estimated by averaging the squared magnitude of the discrete Fourier transform for 4 of these 2484-point signals.

To obtain a single passive cavitation image (created only with subharmonic or broadband frequencies), 36 image frames (unfiltered images created from rf lines) were utilized. The frame rate for this capture was 28 Hz, so that the total data set required to form a single passive cavitation image was acquired over 1.3 s. The acquisition of each rf line is asynchronous, but the assumed stationarity of cavitational emissions allowed these individual lines to be juxtaposed, forming a meaningful passive cavitation image. Stationarity implies that the temporal average of cavitational emission energy radiating from a specific localized volume, which contains a bubble cluster, remains constant over the total data acquisition time. Hence, it is assumed that while the same bubbles do not necessarily cavitate over the entire acquisition time, the average energy of cavitational acoustic emissions remains constant over the entire acquisition time.

Power spectra were computed for acoustic emission signals recorded by the 10-MHz detector using the periodogram method. A total of 38 signal traces, each of length 100 μs sampled at 50 MHz, were analyzed. A discrete Fourier transform was computed for each trace, after applying a 2500-point rectangular window. Squared magnitudes of these Fourier transforms were averaged across all the traces.

V. EXPERIMENTAL RESULTS

Passive cavitation images captured using the three experimental configurations described above are presented in this section.

A. Prediction of acoustic source position

Images formed from ultrasound energy scattered by a steel wire are shown in Fig. 5 constant *f*-number subaperture). The wire was placed at depths of 20, 55, and 90 mm. These images are consistent with the simulation results shown in Fig. 2, for similar imaging parameters. This wire target was moved in the imaging plane to 21 known positions and its position was estimated from the passive images. To estimate the azimuthal location of the wire, the image grayscale value was averaged along the depth direction for all 192 azimuthal points. The position of maximum magnitude for this averaged value was taken as an estimate of the wire's azimuthal position. The location of maximum grayscale value in the range direction, at the estimated azimuth of the wire, was taken as an estimate of the wire's range position. The resulting rms error in predicting the azimuthal location was 0.9 mm and the corresponding range prediction error was 17.2 mm. The range position was consistently overesti-

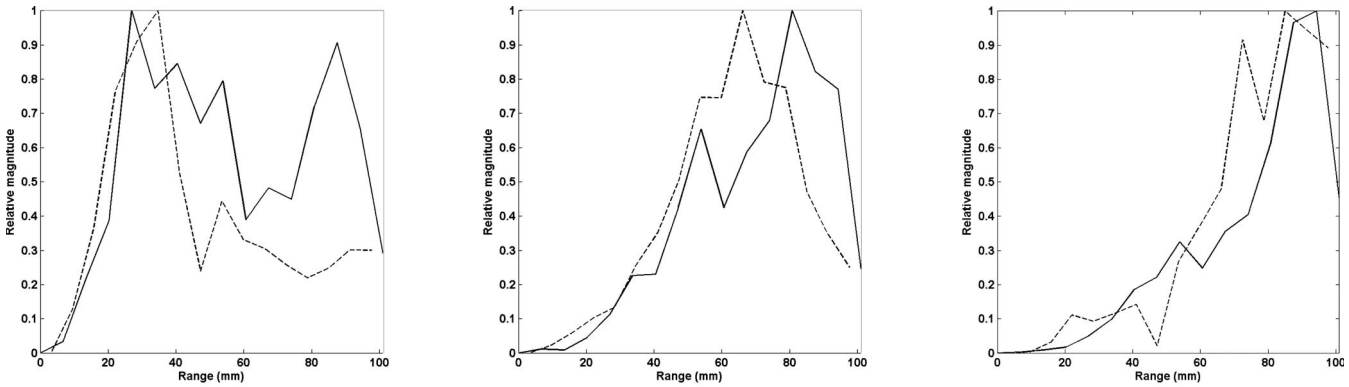


FIG. 6. Image magnitude along the array axis, plotted with respect to the range dimension, for representative simulation (---) and experiment (—) cases. The depth location of simulated point-sources and wire target (in experiment) were (a) 20 mm, (b) 55 mm, and (c) 90 mm.

mated in the passive images, as shown in Fig. 2. Also, consistent with simulations, the image resolution is coarser in range than azimuth.

The wire target employed in these experiments emulated a localized acoustic emission source. The L7 imaging array was focused in elevation (smaller dimension of the array) with an acoustic lens. Hence, the array was less sensitive to acoustic scattering from portions of the wire outside the imaging plane. The resulting effective source, although localized in space, was a rough approximation of a single point-source. This setup more closely mimicked a localized cluster of acoustic sources, similar to a cavitating bubble cloud, likely to be encountered by passive imaging methods during a scenario such as ultrasound ablation. The finite dimensions of this effective source could explain the overestimation of the wire's range position. The target acts as a group of sources and leads to a broader energy pattern along the range direction compared to the energy pattern of individual sources. This discrepancy may cause substantial errors when estimating the range location of spatially distributed acoustic emission sources. Figure 6 illustrates axial variation in passive image magnitude for simulated point-sources and acoustic scattering from the wire target.

B. Cavitation imaging in PBS

To image confirmed cavitation activity, bubbles were nucleated in PBS solution, using a 520-kHz source. During some initial experiments, acoustic emissions were recorded using a single-element detector (10 MHz) along with the passive images. Power spectra were computed for the emissions detected by the single-element transducer and the linear array. Broadband emissions consistent with inertial cavitation were observed in the data acquired by both methods (Fig. 7). These experiments served to verify that the imaging array was indeed detecting cavitation activity.

Following this initial confirmation, the single-element detector was not employed in further testing. Bubble activity was verified from B-mode imaging instead. When sonication amplitude was increased above 0.13 MPa (peak-negative pressure), echogenic bubbles were seen on B-mode images. During sonication at 520 kHz, echogenic bubbles would appear inside the PBS solution initially (away from the container walls). Due to acoustic radiation force, these bubbles would be pushed toward the distal wall of the container, where they would accumulate [see Fig. 8(a)] and cavitate for several minutes, as confirmed from B-scan images. The source pressure was increased from 0–0.15 MPa (peak-negative pressure). At each pressure value, a B-scan was captured, and then passive images were recorded using 64-element subapertures. Following these two steps, a B-scan was captured again. The initial and final B-scans did not show any significant change in the position or size of the bubble cloud, consistent with the assumption of stationary emission energy.

Power spectra were estimated in passive image focal zones, as explained above. For a given focal zone, energy contained in the 6.5-MHz bin, corresponding to an ultraharmonic frequency (12.5 × 520 kHz), was ascertained. The passive cavitation image point centered in this focal zone was then assigned the estimated energy level. This process was performed at every focal zone position to result in a 192 × 16 point passive cavitation image made from energy at 6.5 MHz. Similarly, separate images were formed from broadband emissions within 6.3–6.7 MHz (the 6.5-MHz ultraharmonic component was excluded). Representative images can

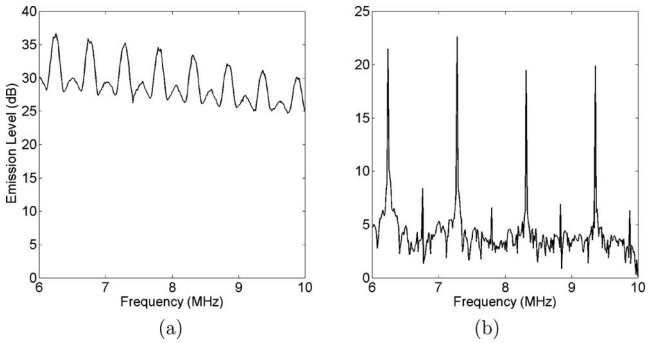


FIG. 7. Acoustic emission spectra from PBS solution sonicated with 520-kHz, cw ultrasound at 0.125 MPa (peak-negative pressure). Power spectra measured in dB relative to the measured noise floors were computed from (a) rf data acquired by the L7 array and (b) acoustic emission signals recorded by the single-element (10 MHz) detector.

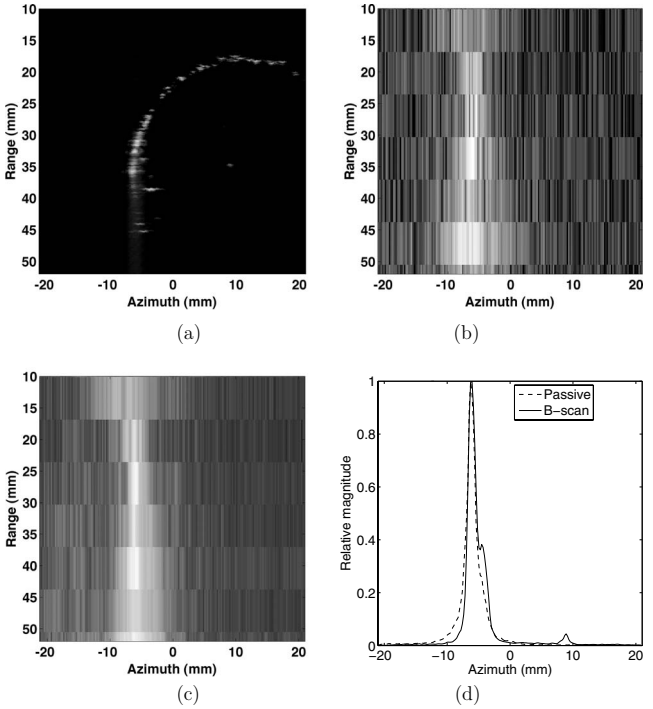


FIG. 8. Representative passive cavitation images in saline solution due to 520-kHz CW ultrasound 0–0.15 MPa (peak-negative): (a) B-scan showing a cavitating bubble cloud, (b) co-registered passive cavitation image formed from ultrahar-monic emissions (6.5 MHz), and (c) co-registered passive cavitation image formed from broadband emissions (6.3–6.7 MHz). (d) Comparison between B-scan and passive image brightness levels integrated between 31 and 41-mm depth, across all azimuths. Passive cavitation images are plotted with a 30 dB dynamic range.

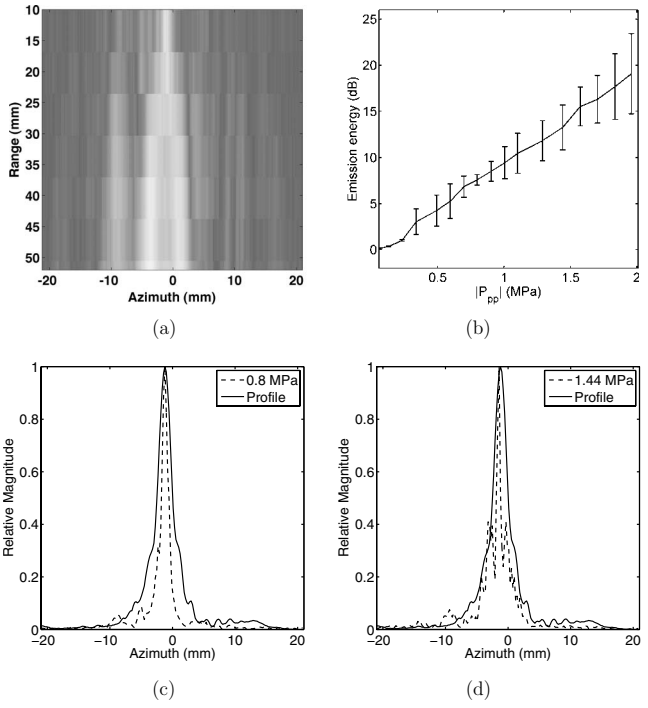


FIG. 9. (a) Representative passive cavitation image using broadband emissions (8–10 MHz) from bovine liver sonicated with 2.2-MHz, CW, focused ultrasound at 0.8 MPa peak-to-peak pressure amplitude (0.38 MPa peak-negative pressure). (b) Spatially integrated emission energy as a function of sonication amplitude, plotted as mean ± st. dev. (c) Comparison of emission amplitude at 20-mm depth with measured beam profile at 0.80 MPa (peak to peak) sonication pressure. (d) Comparison of emission amplitude at 20-mm depth with measured beam profile at 1.44 MPa (peak to peak) sonication pressure. Passive cavitation images are plotted with a 40 dB dynamic range.

be seen in Fig. 8. In 11 trials, where sonication pressure exceeded 0.125 MPa (peak-negative), half-order $(nf_0+f_0/2)$ ultraharmonic energy emissions were detected in one case, while one-third $(nf_0+f_0/3)$ and two-third order $(nf_0+2f_0/3)$ ultraharmonics were observed in several other runs. These could possibly result from stable cavitation activity.[3] Broadband emissions, possibly due to inertial cavitation, were recorded in all 11 trials.

A region-of-interest (ROI) was chosen between the 31-41-mm depth and across all azimuths (entire image width), on both the B-mode and broadband-emission images. A depth-integrated image value was computed as a function of azimuthal position by summing the image magnitude along depth in the ROI, for both images. B-mode image magnitudes were obtained from pulse-echo rf data. Broadband-emission image magnitudes were obtained from filtered, passive-rf signals. A representative comparison of azimuthal distributions of depth-integrated image magnitudes can be seen in Fig. 8(d). Pearson's correlation coefficient was computed for these two distributions for 11 trials. It was found to be statistically significant for all cases and the correlation coefficient value was always greater than 0.85. This confirms the possibility of predicting the azimuthal location of bubble clusters from passive images, consistent with computer simulations (Fig. 3). Here, the passively acquired rf signals were filtered to include emission energy in bands corresponding to stable and inertial cavitation (ultraharmonic or broadband frequencies) from mechanically active microbubbles. Conversely, harmonic scattering from larger bubbles could also be imaged by appropriately filtering the passively received rf signals.

C. Cavitation imaging in *ex vivo* bovine liver

Fresh bovine liver was exposed to a 2.2-MHz, focused source and passive images were captured with the L7 array. Broadband-emission energy was detected and spatially resolved using 64-element subapertures [Fig. 9(a)]. The broadband frequency range between 8–10 MHz, sufficiently different than the fundamental frequency of 2.2 MHz, was utilized. The imaging array has a center frequency of 7.5 MHz and is sensitive up to 10 MHz. To ensure that the fourth harmonic (8.8 MHz) was filtered out, acoustic energy in the band 8.63–8.93 MHz was removed. Figure 9(a) clearly indicates the ability of this imaging modality to obtain spatial information about acoustic emissions from the tissue. Broadband-emission energy integrated across the entire image increased monotonically with the sonication amplitude [Fig. 9(b)]. This increase in the broadband-emission energy could be related to inertial cavitation activity. However, broadband-emission levels estimated in this manner may also contain contributions from spectral side lobes of fourth-harmonic signals nonlinearly scattered from bubbles in the tissue samples.

Broadband-emission images were also employed to ascertain the azimuthal position of the source focus. The location of the maximum brightness value at 20-mm depth was

assigned the value of source focus azimuth. Source focus azimuths (mean ± st. dev.) for the four samples, based on 16 sonication levels per sample, were −1.02 ± 0.5, −1.05 ± 0.46, −3.69 ± 0.46, and −1.57 ± 0.9 mm. The position estimate from sample 3 is different, but the other cases show fair agreement. B-mode images of sample 3 showed accumulation of echogenic gas bubbles at several locations away from the source focus. It was observed that the largest emissions for sample 3 emanated from these bubbles localized outside the sonication focal region, causing the discrepancy in source focus localization from the passive image. This result could be significant for an application such as HIFU ablation monitoring, where tissue coagulation may also happen away from the source focus. It should be noted that the standard deviation for all samples was less than 1 mm, indicating good image resolution in the azimuthal direction. Range location of the acoustic emissions cannot be determined as reliably as their azimuthal position. However, if visually compared, the passive image in Fig. 9(a) is more similar to the simulated image of a point-source at 20-mm depth, as seen in Fig. 1(d) than to simulated images of point-sources at greater depth in Figs. 1(e) and 1(f). This implies that broadband acoustic emissions emanated from within the tissue samples close to the 20-mm depth on the image, corresponding to the source focus position (a more quantitative analysis follows in Sec. VI).

Azimuthal broadband image brightness distribution was compared with the source beam shape, as measured in free-field by a scanning hydrophone system. Figures 9(c) and 9(d) show a comparison between relative magnitudes of source pressure and broadband-emission energy with respect to azimuth, both measured for the same spatial region at 20-mm depth in the image plane. Azimuthal brightness distribution on the broadband-emission image is consistent with the measured beam shape. With an increase in sonication amplitude, the brightness pattern on broadband-emission images broadens azimuthally. At lower sonication pressures, broadband emissions occur mainly in the main lobe, resulting in a broadband emission pattern narrower than the source beam pattern [Fig. 9(c)]. For higher sonication amplitudes, the emission pattern broadens and is closer in width to the source beam pattern [Fig. 9(d)].

These results suggest that passive cavitation imaging may be useful for imaging or mapping therapeutic ultrasound beams. The relation between broadband emissions and source pressure amplitude, as seen in Fig. 9(b), may be utilized to obtain a mathematical function between the two quantities. It may also be possible to ascertain in situ pressure distributions based on the passive cavitation image brightness by exploiting such a numerical mapping.

VI. DISCUSSION

Passive cavitation imaging is examined here with a future goal of application in real-time guidance and monitoring of ultrasound ablation. Hence, the acoustic emissions detected from sonicated bovine liver samples are discussed. Similarly, strengths, limitations, and possible improvements for passive cavitation imaging are also discussed.

A. Tissue emissions and the inertial cavitation threshold

Experimental results presented above indicate a monotonic rise in cumulative broadband emissions energy with increasing sonication amplitudes [Fig. 9(b)]. In other experimental studies with sonicated Optison solutions[36,37] and ablated tissue *in vitro*,[23] a similar trend has been observed. Other researchers also report a rise in the number of macroscopic bubbles in muscle[14] and gel phantoms[38] with increasing acoustic pressures. During the experiments reported in Sec. V, broadband emissions were recorded over a prolonged period. This can result from pre-existing gas and microbubbles in the tissue samples, as evidenced in some earlier publications where similar CW sonications were employed.[14,39] Such cavities can grow through rectified diffusion[40] to a size where they undergo violent inertial collapse.[41] The fragments resulting from such a transient event may produce smaller bubble nuclei, "providing cavitation with a self-enhancing mechanism of positive feedback"[3] due to repeated collapse and coalescence of the bubbles.[42,43]

Cavitation-related broadband emissions occur above threshold pressure amplitude that depends on factors such as viscoelastic properties of the medium, prevalence of pre-existing nuclei and dissolved gas content, among others. An upper bound for this threshold, given an optimally sized bubble and single cycle excitation at 2.2 MHz, is close to 0.47 MPa (peak-negative).[44] It is also documented that inertial cavitation thresholds could be 3–40 times higher in tissue than in water.[45] However, for the CW exposure conditions employed here, inertial cavitation may occur at lower pressure values due to the presence of pre-existing nuclei and their growth through rectified diffusion. Also, the liver samples used in the experiments contained blood and saline solution around the tissue, with several possible cavitation nucleation sites.

It is estimated through simulations that small (<3 μm) bubbles may grow through rectified diffusion.[46] Pre-existing nuclei under 3 μm size, subjected to multiple cycle exposures with frequencies between 2 and 3 MHz, may have the threshold for rectified diffusion as low as 0.09 MPa (peak-negative).[46,47] Hence, bubbles grown by rectified diffusion may either undergo inertial collapse or cause nonlinear scattering of source harmonics. A peak-negative pressure of 0.09 MPa was eclipsed in the main lobe of the focused source beam for most exposure conditions, presented above. The pressure values achieved in the first side lobe, about 1/7 of the main lobe. It is likely that pressure in the side lobes was close to this threshold value, especially at higher sonication amplitudes such as 1.44 MPa (peak-to-peak). This could lead to broadband emissions in the side lobes at such pressure levels (see Fig. 9).

Temperature changes in the tissue may alter the acoustic pressure threshold values for inertial cavitation or rectified diffusion, due to either the dissolved gas coming out of solution or coagulation of liver samples due to hyperthermia. An estimate of the temperature rise was calculated for the exposure conditions employed during the tissue experiments (reported above). This was achieved by solving a simplified Pennes bio-heat transfer equation at the location of the 2.2-

MHz source focus, with acoustic nonlinear propagation, thermal diffusion, and perfusion losses ignored.[19] This yielded a "worst case" estimate for temperature rise. The value of temperature increase was 16.24 °C at the end of 16 sonications of 4 s on/55 s off, where the pressure amplitude was sequentially increased between 0 and 1.96 MPa (peak-to-peak pressure measured in free-field). For the duty cycle with the highest sonication amplitude, the temperature increase was 2.93 °C. The actual temperature rise during the experiments would be less than this estimate, given the conductive heat dissipation. With an initial tissue temperature close to 20 °C, it was likely that the temperature elevation caused by the HIFU source would not result in thermal coagulation.[19] While no coagulation was observed when the tissue samples were sliced after exposure, the source ultrasound beam could possibly have caused localized temperature elevations in the tissue. These can result in increased nucleation events and a reduced inertial cavitation threshold.[45] The relation between cavitation in tissue, the concomitant emissions, and ultrasound pressure needs further elucidation.

B. Ablation monitoring

The strong dependence of bubble activity on temperature and physical properties of the medium make passive cavitation imaging a potential tool for ablation monitoring. The authors have previously explored some preliminary statistical frameworks that relate these quantities.[48] Spatially resolved bubble emissions may enable real-time tracking of cavitation-related phenomena such as "tadpole" lesioning during HIFU ablation.[17] Future studies may include ultrasound ablation experiments with passive imaging where position-dependent cavitation will be used as a measure of local heat deposition and tissue coagulation.

Cavitation activity, as well as vaporization due to boiling, may occur during ultrasound ablation. Passive imaging may be used for ablation monitoring by tracking both these phenomena. Experimental evidence suggests that low-frequency emissions arise due to tissue boiling. Resolution in passive images depends on the emission frequency and images from such kilohertz-frequency emissions will probably have low resolution. However, tissue boiling may be monitored by imaging the source harmonics scattered by the resulting vapor bubbles.

Based on the consistent agreement between passive images and HIFU beam shapes, as presented in Sec. IV, passive imaging technique may potentially be employed to align therapeutic ultrasound beams during clinical procedures.

C. Improving cavitation imaging performance

Cavitation imaging performance can be improved if the range location of the acoustic emissions can be ascertained. Although range resolution of passive cavitation images is limited, spatial correlation between acquired passive images and simulated images may be exploited to estimate source range locations more accurately. A possible metric R for this may be based on the Pearson product-moment correlation coefficient:[49]

$$R(y_s,z_s) = \frac{\sum_{\forall Y}\sum_{\forall Z}(I_a(Y,Z)-I_a')(I_s(y_s,z_s,Y,Z)-\bar{I}_s)}{\sqrt{\sum_{\forall Y}\sum_{\forall Z}(I_a(Y,Z)-\bar{I}_a)^2}\sqrt{\sum_{\forall Y}\sum_{\forall Z}(I_s(Y_s,z_s,Y,Z)-\bar{I}_s)^2}}, \quad (12)$$

where I_a is the acquired image and I_s is the simulated image of a point-source at (y_s,z_s). The value of (y_s,z_s) that maximizes R can be used as an estimate of the source location. If this function R is computed for Fig. 9(a) and each of the simulated images in Figs. 1(d)–1(f), the value of R comes out to be 0.54 for point-source location (0,20), 0.33 for point-source location (0, 55), and 0.37 for point-source location (0, 90). This result is consistent with the location of the HIFU focus, which was aligned to position (0, 20) by the pulse-echo method mentioned above. Development of such algorithms is beyond the scope of this paper.

While modeling-based approaches may help in improving cavitation imaging performance, modifications to the ultrasound imaging hardware may potentially achieve better results. In the Iris 2 system described above, the beamformed rf A-lines are acquired sequentially. If multiple A-lines were recorded with parallel channels, it would be possible to receive cavitational emissions synchronously. This would enable the use of cross-correlation based methods[26] or inverse source reconstruction techniques.[50] Such parallel receive hardware would also make it possible to image short-lived cavitational events that occur during procedures such as lithotripsy. A lithotripter produces a shock wave which leads to a single transient cavitational event. Passive sensing of such an event with multiple receive channels would potentially provide effective range gating.

VII. CONCLUDING REMARKS

In this paper, passive cavitation imaging was introduced and the point-spread function of a cavitation image was derived. Passive images were simulated for both constant subaperture width and constant f-number beamforming, which indicated that the bubble emissions could be resolved accurately along the array azimuth. Passive imaging was implemented on a 192-element ultrasound array and was successfully tested by imaging harmonic scattering from a wire. Cavitational emissions from microbubbles nucleated in saline solution and broadband emissions from bovine tissue were also imaged. The distribution of azimuthal energy in the broadband emission images of *ex vivo* tissue was compared with the beam shape of the ultrasound source. The consistent agreement between the two indicates a strong possibility of mapping therapeutic ultrasound beams *in situ*, though several quantitative questions dealing with the direct relationships between sonication amplitude, tissue temperature, and passive image brightness need to be resolved.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R21 EB008483 and the University of Cincinnati College of Medi-

cine Dean's Bridge Funding Program. Jonathan Kopechek is thanked for help with transducer calibration.

[1]R. E. Apfel, "Acoustic cavitation inception," Ultrasonics **22**, 167–173 (1984).
[2]H. G. Flynn, "Physics of acoustic cavitation in liquids," Phys. Acoust. **1**, 57–172 (1964).
[3]T. G. Leighton, *The Acoustic Bubble* (Academic, London/Academic, San Diego, CA, 1994), Chap. 4.
[4]A. Anand, L. Byrd, and P. J. Kaczkowski, "*In situ* thermal parameter estimation for HIFU therapy planning and treatment monitoring," Proc.-IEEE Ultrason. Symp. **1**, 137–140 (2004).
[5]L. A. Crum, "Cavitation microjets as a contributory mechanism for renal calculi disintegration in ESWL," J. Urol. (Baltimore) **140**, 1587–1590 (1984).
[6]R. K. Zemen, W. J. Davros, B. S. Garra, and S. C. Horri, "Cavitation effects during lithotripsy Part 1. Results of *in vitro* experiments," Radiology **177**, 157–161 (1990).
[7]R. O. Cleveland, "Acoustics of shock wave lithotripsy," in Renal Stone Disease 1st Annual International Urology Research Symposium (2007), pp. 311–316.
[8]S. B. Olsson, B. Johansson, A. M. Nilsson, C. Olsson, and A. Roijer, "Enhancement of thrombolysis by ultrasound," Ultrasound Med. Biol. **20**, 375–382 (1994).
[9]S. Datta, C. C. Coussios, L. E. McAdory, J. Tan, T. Porter, G. De Courten-Myers, and C. K. Holland, "Correlation of cavitation with ultrasound enhancement of thrombolysis," Ultrasound Med. Biol. **32**, 1257–1267 (2006).
[10]E. C. Everbach and C. W. Francis, "Cavitational mechanisms in ultrasound-accelerated thrombolysis at 1 MHz," Ultrasound Med. Biol. **26**, 1153–1160 (2000).
[11]E. C. Unger, T. Porter, W. Culpd, R. Labella, T. Matsunagaa, and R. Zutshi, "Therapeutic applications of lipid-coated microbubbles," Adv. Drug Delivery Rev. **56**, 1291—1314 (2004).
[12]K. Hynynen, "Ultrasound for drug and gene delivery to the brain," Adv. Drug Delivery Rev. **60**, 1209–1217 (2008).
[13]J. A. Kopechek, T. M. Abruzzo, B. Wang, S. M. Chrzanowski, D. A. B. Smith, P. H. Kee, S. Huang, J. H. Collier, D. D. McPherson, and C. K. Holland, "Ultrasound-mediated release of hydrophilic and lipophilic agents from echogenic liposomes," J. Ultrasound Med. **27**, 1597–1606 (2008).
[14]G. R. ter Haar and S. Daniels, "Evidence for ultrasonically induced cavitation *in vivo*," Phys. Med. Biol. **26**, 1145–1149 (1981).
[15]K. Hynynen, "The threshold for thermally significant cavitation in dog's thigh muscle *in vivo*," Ultrasound Med. Biol. **17**, 157–169 (1991).
[16]C. C. Coussios, C. H. Farny, G. R. ter Haar, and R. A. Roy, "Role of acoustic cavitation in the delivery and monitoring of cancer treatment by high-intensity focused ultrasound (HIFU)," Int. J. Hyperthermia **23**, 105–120 (2007).
[17]N. A. Watkin, G. R. ter Haar, and I. H. Rivens, "The intensity dependence of the site of maximal energy deposition in focused ultrasound surgery," Ultrasound Med. Biol. **22**, 483–491 (1996).
[18]L. Curiel, F. Chavrier, B. Gignoux, S. Pichardo, S. Chesnais, and J. Y. Chapelon, "Experimental evaluation of lesion predicting modeling in prescence of cavitation bubbles: Intended for high-intensity focused ultrasound prostate treatment," Med. Biol. Eng. Comput. **42**, 44–52 (2004).
[19]T. D. Mast, I. R. S. Makin, W. Faidi, M. M. Runk, P. G. Barthe, and M. H. Slayton, "Bulk ablation of soft tissue with intense ultrasound: Modeling and experiments," J. Acoust. Soc. Am. **118**, 2715–2724 (2005).
[20]V. Khokhlova, M. Bailey, J. Reed, B. Cunitz, P. Kaczkowski, and L. Crum, "Effects of nonlinear propagation, cavitation, and boiling in lesion formation by high intensity focused ultrasound in a gel phantom," J. Acoust. Soc. Am. **119**, 1834–1848 (2006).
[21]C. H. Farny, "Identifying and monitoring the roles of cavitation in heating from high-intensity focused ultrasound," Ph.D. thesis, Boston University, Boston, MA 2007).
[22]J. McLaughlan, I. H. Rivens, and G. R. ter Haar, "A study of cavitation activity in *ex vivo* tissue exposed to high intensity focused ultrasound," AIP Conf. Proc. **911**, 178–184 (2007).
[23]T. D. Mast, V. A. Salgaonkar, C. Karunakaran, J. A. Besse, S. Datta, and C. K. Holland, "Acoustic emissions during 3.1 MHz ultrasound bulk ablation *in vitro*," Ultrasound Med. Biol. **34**, 1434–1448 (2008).
[24]B. A. Rabkin, V. Zedric, and S. Vaezy, "Hyperecho in ultrasound images of HIFU therapy: Involvement of cavitation," Ultrasound Med. Biol. **31**, 947–956 (2005).
[25]C. C. Coussios, J. R. T. Collins, and A. P. Muckle, "Noninvasive monitoring and control of inertial cavitation dynamics during HIFU exposure *in vitro*," AIP Conf. Proc. **911**, 164–170 (2007).
[26]S. J. Norton and B. J. Carr, "Passive imaging of underground acoustic sources," J. Acoust. Soc. Am. **119**, 2840–2847 (2006).
[27]M. Gyongy, M. Arora, A. Noble, and C. C. Coussios, "A passive array technique for cavitation mapping during HIFU treatment," J. Acoust. Soc. Am. **123**, 3223 (2008).
[28]V. A. Salgaonkar, S. Datta, C. K. Holland, and T. D. Mast, "Passive imaging of cavitational acoustic emissions with ultrasound arrays," AIP Conf. Proc. **1113**, 73–77 (2009).
[29]T. L. Szabo, *Diagnostic Ultrasound Imaging: Inside Out* (Elsevier, Burlington, MA, 2004), Chap. 6.
[30]D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000), Chap. 10.
[31]A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications*, 2nd ed. (Acoustical Society of America, Melville, NY, 1989), Chap. 5.
[32]T. D. Mast, "Fresnel approximations for ultrasonic fields of rectangularly symmetric sources," J. Acoust. Soc. Am. **121**, 3311–3322 (2007).
[33]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, DC, 1972), Chap. 7.
[34]T. M. Porter, D. A. B. Smith, and C. K. Holland, "Acoustic techniques for assessing the optison destruction threshold," J. Ultrasound Med. **25**, 1519–1529 (2006).
[35]R. O. Cleveland, O. A. Sapozhnikov, M. R. Bailey, and L. A. Crum, "A dual passive cavitation detector for localized detection of lithotripsy-induced cavitation in vitro," J. Acoust. Soc. Am. **107**, 1745–1758 (2000).
[36]D. M. Hallow, A. D. Mahajan, T. E. Mccutchen, and M. R. Prausnitz, "Measurement and correlation of acoustic cavitation with cellular bioeffects," Ultrasound Med. Biol. **32**, 1111–1122 (2006).
[37]J. Tu, J. Hwang, T. Matula, A. Brayman, and L. A. Crum, "Intravascular inertial cavitation activity detection and quantification in vivo with Optison," Ultrasound Med. Biol. **32**, 1601–1609 (2006).
[38]S. Daniels, D. Blondel, L. A. Crum, G. R. ter Haar, and M. Dyson, "Ultrasonically induced gas bubble production in agar based gets: Part I, experimental investigation," Ultrasound Med. Biol. **13**, 527–539 (1987).
[39]P. F. Lewin and L. Bjorno, "Thresholds for rectified diffusion and acoustic microstreaming by bubbles in biological tissue," App. Science Reviews **38**, 25–35 (1982).
[40]A. I. Eller and H. G. Flynn, "Rectified diffusion during nonlinear pulsations of cavitation bubbles," J. Acoust. Soc. Am. **37**, 493–503 (1965).
[41]E. T. Ayme-Bellegarda, "Collapse and rebound of a gas-filled spherical bubble immersed in a diagnostic ultrasound field," J. Acoust. Soc. Am. **88**, 1054–1060 (1990).
[42]C. C. Church and E. L. Carstensen, "'Stable' inertial cavitation," Ultrasound Med. Biol. **27**, 1435–1437 (2001).
[43]D. F. Gaitan, L. A. Crum, C. C. Church, and R. A. Roy, "Sonoluminescence and bubble dynamics for a single, stable cavitation bubble," J. Acoust. Soc. Am. **91**, 3166–3183 (1992).
[44]R. E. Apfel and C. K. Holland, "Gauging the likelihood of cavitation from short-pulse low-duty cycle diagnostic ultrasound," Ultrasound Med. Biol. **17**, 179–185 (1991).
[45]C. C. Church and X. Yang, "A theoretical study of gas bubble dynamics in tissue," AIP Conf. Proc. **838**, 217–224 (2006).
[46]S. Datta, "The role of cavitation in enhancement of rt-PA thrombolysis," Ph.D. thesis, University of Cincinnati, Cincinnati, OH (2008).
[47]C. C. Church, "Prediction of rectified diffusion during nonlinear bubble pulsations at biomedical frequencies," J. Acoust. Soc. Am. **83**, 2210–2217 (1988).
[48]V. A. Salgaonkar, C. P. Karunakaran, J. A. Besse, G. Heinlein, S. Datta, C. K. Holland, and T. D. Mast, "Image-guided ex vivo liver ablation by unfocused ultrasound using passive cavitation detection," Proc. SPIE **6440**, Q1–Q10 (2007).
[49]J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," Am. Stat. **42**, 59–66 (1988).
[50]Y. Kim and P. A. Nelson, "Optimal regularisation for acoustic source reconstruction by inverse methods," J. Sound Vib. **275**, 463–487 (2004).

Automatic estimation of position and orientation of an acoustic source by a microphone array network

Alberto Yoshihiro Nakano,[a] Seiichi Nakagawa, and Kazumasa Yamamoto
Department of Information and Computer Sciences, Toyohashi University of Technology, Toyohashi 441-8580, Japan

(Received 24 March 2009; revised 29 September 2009; accepted 2 October 2009)

A method which automatically provides the position and orientation of a directional acoustic source in an enclosed environment is proposed. In this method, different combinations of the estimated parameters from the received signals and the microphone positions of each array are used as inputs to the artificial neural network (ANN). The estimated parameters are composed of time delay estimates (TDEs), source position estimates, distance estimates, and energy features. The outputs of the ANN are the source orientation (one out of four possible orientations shifted by 90° and either the best array which is defined as the nearest to the source) or the source position in two dimensional/three dimensional (2D/3D) space. This paper studies the position and orientation estimation performances of the ANN for different input/output combinations (and different numbers of hidden units). The best combination of parameters (TDEs and microphone positions) yields 21.8% reduction in the average position error compared to the following baselines and a correct orientation ratio greater than 99%. Position localization baselines consist of a time delay of arrival based method with an average position error of 34.1 cm and the steered response power with phase transform method with an average position error of 29.8 cm in 3D space.
 [DOI: 10.1121/1.3257548]

PACS number(s): 43.60.Jn, 43.60.Np, 43.66.Qp [EJS] Pages: 3084–3094

I. INTRODUCTION

Microphone arrays[1,2] have received increasing attention in the past few years, especially in spatial filtering (beamforming)[3] and sound source localization[4–6] for speech, audio, and acoustic processing. Microphone array techniques depend on many factors, including placement, geometrical configuration, number of microphones, as well as the conditions and the number of active acoustic sources in the environment under investigation. This is the main reason why a distributed microphone array network, comprised of eight T-shaped microphone arrays, is used in this study to estimate the position and orientation of a directional acoustic source in an actual enclosed environment.

Acoustic localization is an important task in many practical applications such as videoconferencing,[7] hands-free communication system,[8] hearing aids,[9] and human-machine interaction.[10] Different source localization methods were proposed in the literature[1,6,11–13] and they can be separated into two groups; indirect and direct estimation methods. The difference consists of whether a set of time delay estimates (TDEs) of microphone pair signals is estimated or not and used in the localization method.

Average magnitude difference function (AMDF),[14] adaptive eigenvalue decomposition (AED),[11] information theory (IT),[15] and generalized cross-correlation[16] (GCC) methods are some examples of time delay estimation techniques. In the AMDF approach, an average magnitude difference function is calculated from the microphone signals, and the time delay corresponds to the time lag which minimizes this function. In this approach, when the signals are time aligned, the average magnitude difference between them becomes minimum. The AED approach focuses directly on the estimation of the impulse responses between the acoustic source and a pair of microphones. It is assumed that the relative delay between signals corresponds to the time lag between direct path components of estimated impulse responses. In the IT approach, it is shown that the time delay that maximizes the mutual information between pairs of signals is equivalent to the time delay that maximizes the cross-correlation between these signals. The GCC approach has been the traditional method for time delay estimation, where the optimum time delay corresponds to the time lag which maximizes the cross-correlation function obtained for a pair of microphone signals. A variant of the GCC approach, the GCC with phase transform (GCC-PHAT), is employed in this research. In the GCC-PHAT method, each component of the cross-power spectrum of microphone pair signals is equally emphasized for generating a prominent peak in the cross-correlation function. But, due to the noise and reverberation conditions in the test environment, a more robust GCC-PHAT is derived in Sec. II C and used for both the indirect and direct position estimation methods employed in this research.

In indirect localization methods, a set of time delays is initially estimated from different microphone pair signals, and the optimal source position is found by geometrical derivation, that is, by solving a set of equations that compute the intersection point of hyperplanes in the space, with each hyperplane defined for each TDE. Examples of these methods can be found in Ref. 12, originally derived for global posi-

[a]Author to whom correspondence should be addressed. Electronic mail: alberto@slp.ics.tut.ac.jp

tioning system but useful for acoustic localization when the array geometry is unknown, and in Ref. 6 when the array configuration is known *a priori*, which permits simplifications in the formulation of the position estimation method. Indirect methods are highly dependent on the correct delay estimation, and either inaccurate estimates or small variations in the estimates to the optimum values imply an unreliable position estimate that can be outside of the studied space. To avoid mismatched microphones due to production tolerances, aging effects, and unbalanced microphone array geometry, array and microphone compensation methods were proposed[2,17,18] in order to reduce the estimate variance. In this work, it is assumed that all microphones and arrays are well balanced, disregarding mismatch errors.

Direct localization methods are space exploration techniques that search the studied space for the point with highest spatial likelihood. Steered response power phase transform[1] (SRP-PHAT), based on the maximization of power obtained by steering the microphone array to all potential source positions, and global coherence field (GCF),[19] similar to SRP-PHAT approach, are examples of these methods. Direct methods do not depend on the array geometry but require more computation compared to indirect methods.

The source orientation[19–23] also plays an important role in acoustic localization because a directional source does not radiate uniformly in all directions, and the quality of signals recorded by distant microphones is affected not only by environmental noise and reverberation but also by the speaker's relative orientation.[21] Sachar *et al.*[20] proposed the *energy method*, where differences in the source radiation pattern can be detected and used to predict the source orientation. Signals obtained by a huge microphone array (HMA), composed of 448 microphones, are processed (to compensate the inverse-square-law attenuation of the source signal, to reduce the background noise and masking effect of reverberation, and to enhance directional components of the signal) and used to estimate the source radiation pattern. An analysis of this pattern indicates the most likely source orientation having higher gain in the estimated radiation pattern compared to the other orientations. This method presented 60% correct orientation ratio at $\pm 5°$ tolerance level or nearly 100% correct performance at $\pm 10°$ tolerance. Abad *et al.*[21] presented the SRP-PHAT orientation estimation and the high/low band ratio (HLBR) orientation estimation methods. In the SRP-PHAT orientation approach, the position and orientation are estimated together, and in addition to the theoretical time delays computed from each spatial position to each microphone pair, weights representing the influence of each cross-correlation in terms of the relative orientation are computed from a normalized source directivity pattern. For this method, an average error about 44° was obtained.[21] The ability of the system to correctly classify the source orientation within eight classes shifted by 45° was about 37% and admitting a classification error of ± 1 adjacent class a value of 73% was obtained. HLBR is defined as the ratio between energies of high (3500–4500 Hz) and low (200–400 Hz) frequency bands of the radiation pattern. It is shown that HLBR maintains the same characteristics of the radiation pattern and is more robust to array calibration mismatches and propagation losses. The angle which maximizes the correlation between a mathematical HLBR model and an estimated HLBR model corresponds to the orientation of the source. Using the same standard metrics of the SRP-PHAT orientation method, i.e., the average error, the ability of the system to correctly classify the source orientation within eight classes separated by 45°, and assuming correct classification error of ± 1 adjacent class, 51°, 30%, and 68% were obtained, respectively. Brutti *et al.*[19,23] extended the GCF position localization method to consider the source orientation. The new method was named the oriented global coherence field. In this method, the exploration space is composed of weights corresponding to the position of the source and weights corresponding to the orientation of the source. The point in the space with highest sum of weights will show the estimated source position and the estimated orientation. It was reported that more than 99% of estimates were within an error of 45°.

In the studies on position and orientation estimation methods surveyed above, it is assumed that either the position is known *a priori* for orientation estimation or that the sound source is direct to the array so that the orientation is known *a priori* for position estimation. A few studies, as can be found in Refs. 21 and 23, estimate the position and the orientation together.

In this work, artificial neural networks (ANNs) are used in the position and orientation estimation. ANN (Ref. 24) is a massively parallel distributed structure with input-output mapping, it has the ability to learn from and adapt to certain conditions, and it can invoke assumptions about the underlying physical phenomena responsible for the generation of the input data. Employing ANNs, it is expected that combining estimated parameters of each array of the network could be sufficient to predict the source orientation and could give a more accurate source position than that estimated by an individual array. Here in this work, we define representation level as the compactness or aggregation of information of a given process, in our case the estimation of the position and the orientation, such that the higher the representation is, the smaller the number of parameters necessary to describe the same information will be. Figure 1 shows the representation level of the estimated parameters used at the ANN input. The low representation is given by TDEs and microphone positions, the intermediate representation is given by source position estimates, and the high representation parameter is given by distance estimates derived from source position estimates. There is a reduction in the number of estimated parameters from the lower to the higher representation level. In this work, for each array TDEs are represented by three values, microphone positions are represented by 12 values, a source position estimate is represented by three values, and a distance estimate is represented by one value. In Sec. IV E 2, we show that the lower the representation level is, the better are the ANN results.

The outline of this paper is organized as follows. In Sec. II classical and robust GCC-PHAT methods for TDEs are presented as well as two baseline position estimation methods; the time delay of arrival (TDOA)-based and the SRP-PHAT. In Sec. III, the proposed method that jointly estimates

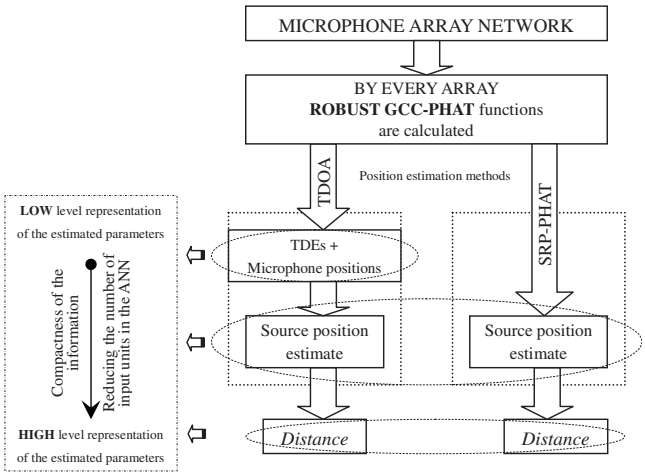


FIG. 1. Description of the estimation of different parameters (TDEs, source position estimate, and distance estimate) by every array of the microphone array network used in the proposed position and orientation estimation method.

the position and orientation of a directional acoustic source using ANNs is presented. In Sec. IV, experimental setup and results are presented, and in Secs. V and VI, discussions and conclusions are presented.

II. BACKGROUND

A. Signal model

Consider P identical arrays, each one with Q microphones, where each microphone is defined as m, for $m=1,\ldots,Q$. Given a signal source $s(t)$, the signal at each microphone can be represented as

$$\mathbf{x}_{p_m}(t) = h_{p_m}(t) * s(t) + n(t), \tag{1}$$

where $p \in \{1,\ldots,P\}$, "$*$" denotes convolution, $h_{p_m}(t)$ is the reverberation impulse response that describes the propagation path between the source $s(t)$ and the mth microphone of the pth array, and $n(t)$ is the additive background noise.

B. GCC-PHAT method and time delay estimation

The GCC-PHAT function calculated from two signals $\mathbf{x}_{p_m}(t)$ and $\mathbf{x}_{p_n}(t)$, $m \neq n$, of the pth array is defined as

$$R(\tau_{p_{mn}}) = \int_{-\infty}^{+\infty} \frac{\mathbf{X}_{p_m}(f)\mathbf{X}^*_{p_n}(f)}{|\mathbf{X}_{p_m}(f)\mathbf{X}^*_{p_n}(f)|} e^{-j2\pi f \tau_{p_{mn}}} df, \tag{2}$$

where $R(\tau_{p_{mn}})$ is a function of the time delay $\tau_{p_{mn}}$, and $\mathbf{X}_{p_m}(f)$ and $\mathbf{X}_{p_n}(f)$ are the spectral representations of the signals $\mathbf{x}_{p_m}(t)$ and $\mathbf{x}_{p_n}(t)$, respectively. In this method, a phase dependent equalized cross-power spectral function becomes a time difference dependent cross-correlation function by Fourier transform. Therefore, the TDE $\hat{\tau}_{p_{mn}}$ between pair of signals corresponds, ideally, to the time difference that maximizes $R(\tau_{p_{mn}})$ as

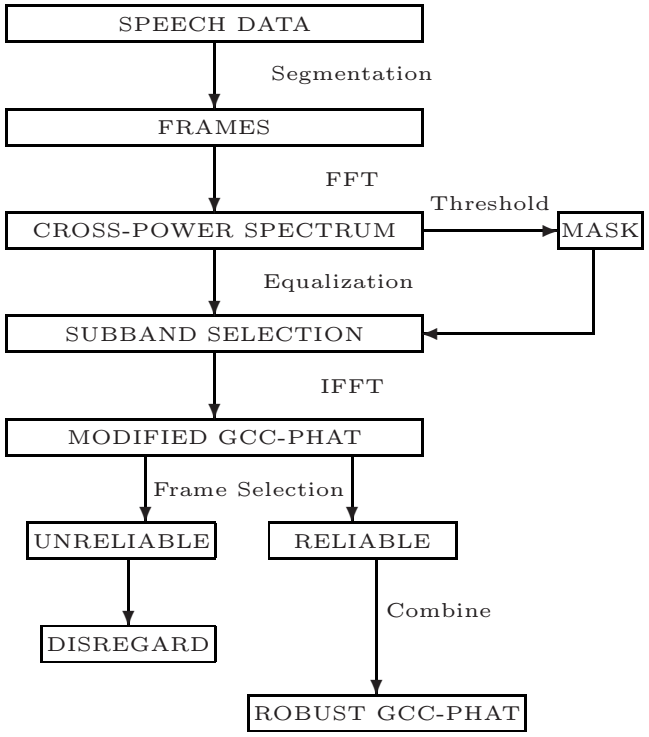


FIG. 2. Diagram block which illustrates the improvement of GCC-PHAT method.

$$\hat{\tau}_{p_{mn}} = \max_{\tau_{p_{mn}}}\{R(\tau_{p_{mn}})\}. \tag{3}$$

In the equalization procedure in Eq. (2), all frequency components are equally weighted generating a prominent peak in time domain. However, components of the spectrum with weak signal-to-noise ratio (SNR) are also accentuated, reducing the efficiency of this method under high noise conditions.[1] In Sec. II C, this fact is explored to generate a robust GCC-PHAT function. The TDEs are used in the localization algorithm presented in this work.

C. Improving GCC-PHAT function

Steps to improve the GCC-PHAT function are described in the block diagram of Fig. 2. Initially, signals $\mathbf{x}_{p_m}(t)$ and $\mathbf{x}_{p_n}(t)$ are segmented into L frames of N samples. Each pair of frames is converted to its frequency domain representation, and the cross-power spectral function is calculated. In the next stage, a subband selection method, where only high SNR spectrum components are selected in the cross-power spectral function, is performed. In this method a threshold γ is defined and a binary mask is created (see Fig. 3) as follows:

$$\text{mask} = \begin{cases} 1, & b > \gamma \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

where b is the magnitude of the cross-power spectrum component value. γ is defined as

$$N.\text{level} \ll \gamma < N.\text{level} + G \times \text{peak}, \tag{5}$$

where N.level, G, and peak are the average noise level, a gain factor $0<G<1$, and the highest cross-power peak

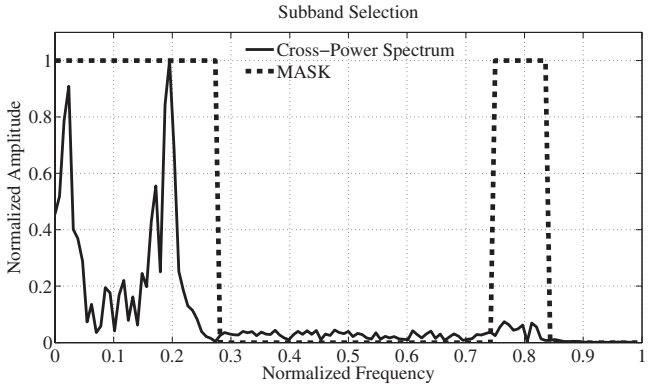


FIG. 3. Subband selection method. Binary mask (dashed line) selects the cross-power spectrum subbands (continuous line) with high SNR.

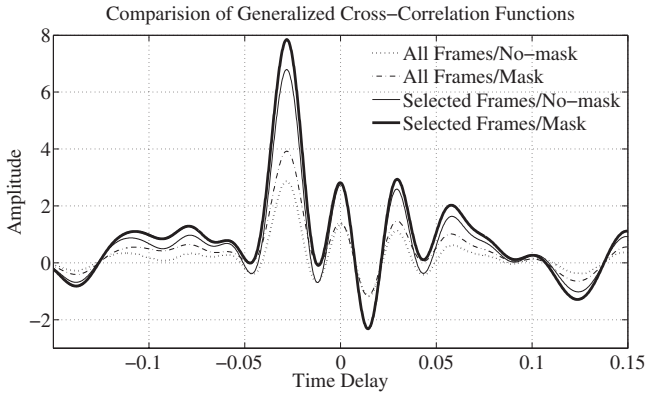


FIG. 4. GCC-PHAT function estimation with/without masking and/or frame selection. The dotted, dashed-dotted, thin solid, and thick solid lines represent the combination of all frames without masking, all frames with masking, only reliable frames without masking, and only reliable frames with masking of a segment of speech in the GCC-PHAT function estimation, respectively.

value in the frame, respectively. A value of $G=0.05$, used in the experiments, was determined experimentally by analyzing our recorded files. Then, the generated mask is applied to the equalized cross-power spectral function to reduce parts of undesirable contributions of noise and reverberation, and the inverse Fourier transform converts the masked cross-power spectral function to a modified cross-correlation function. At this point, the combination of the estimated cross-correlation function of different frames could generate a robust cross-correlation function; however, not all frames will contribute positively in this procedure due to the low SNR and strong and long reverberation effects which vary from frame to frame. A way to select reliable frames, i.e., frames that improve the GCC-PHAT function, is needed. In the procedure adopted in this work, the TDEs are calculated by frames using aforementioned modified cross-correlation functions for all microphone pairs of an array. If all TDEs have zero values, then it is assumed that the TDEs were generated by *unreliable frames*. If the time delays have zero values over all microphone pairs, then the propagated wave arrives in phase at the array. This situation is only possible when the source is located far away, although, in reality the source is at a relative near distance from the considering microphone array. Unreliable frames were due to the contribution of the undesirable correlated noise at the array. Finally, disregarding *unreliable frames* and combining only *reliable frames*, a robust GCC-PHAT function is created

$$R'(\tau_{p_{mn}}) = \sum_{k=1}^{K} R(\tau_{p_{mn}},k), \tag{6}$$

where k is the index of the set of reliable frames, $k \in [1,\ldots,K]$, so that K, $K<L$, is the total number of reliable frames, $R(\tau_{p_{mn}},k)$ is the modified GCC-PHAT function of the kth reliable frame, and $R'(\tau_{p_{mn}})$ is the robust GCC-PHAT function. The TDE $\hat{\tau}_{p_{mn}}$ can be obtained as in Eq. (3) by maximizing $R'(\tau_{p_{mn}})$. An interpolation procedure[25] can be applied to the GCC-PHAT function to increase the resolution time lag, thereby improving the TDE accuracy. The search of the highest peak in GCC-PHAT function can be restricted by the array geometry and the studied space[25,26] to look only for physically possible delays, that is,

$$\tau_{p_{mn}}^{\min} \leqslant \hat{\tau}_{p_{mn}} \leqslant \tau_{p_{mn}}^{\max}, \tag{7}$$

where $\tau_{p_{mn}}^{\min}$ and $\tau_{p_{mn}}^{\max}$ are the minimum and maximum possible values of $\tau_{p_{mn}}$. Figure 4 illustrates the evolution of the GCC-PHAT function considering the masking effect and/or reliable frame selection using the same segment of speech. It can be noted that with masking and frame selection a more prominent peak can be obtained.

D. TDOA-based and SRP-PHAT localization methods

TDOA based localization methods are performed in two steps: first, the TDEs of microphone pair signals are determined; and second, the source position estimate is found by the intersection point of hyperplanes in the space, with each one defined for each TDE. The intersection point is found by solving a set of defined equations. We used the position localization method from Wang *et al.*[6] tailored for the T-shaped microphone array shown in Fig. 5. The set of equations can be found in more detail in the cited reference. Thus, in the initial step, using the robust GCC-PHAT function described in Sec. II C, a set of three TDEs, $\{\tau_{12},\tau_{13},\tau_{14}\}$, is estimated for microphone pairs {1,2}, {1,3}, and {1,4}, taking microphone 1 as the reference, and in the second step, the source position estimate is found.

The SRP-PHAT is a robust position localization method that explores the space, searching for the region with the highest spatial likelihood obtained by a cumulative voting process involving cross-correlation functions of microphone

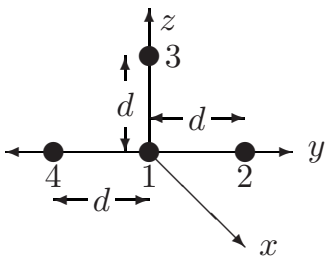


FIG. 5. T-shaped microphone array composed of microphones {1,2,3,4}. d is the distance between adjacent microphones.

pair signals. While in Ref. 1 this method is described in detail, here, we give only the basic concepts. In the SRP-PHAT method, the space is divided into small regions, and the theoretical delays between these regions and microphone pairs are pre-computed and stored. In our experiments, we divide the space into small regions of $5.0\times5.0\times5.0$ cm^3. Thus, each small region l, characterized by a point in the space $\boldsymbol{\alpha}_l=(x_l,y_l,z_l)$, is associated with a vector of time delays

$$\boldsymbol{\tau}(\boldsymbol{\alpha}_l)=[\tau_{12}(\boldsymbol{\alpha}_l),\tau_{13}(\boldsymbol{\alpha}_l),\ldots,\tau_{1Q}(\boldsymbol{\alpha}_l)], \tag{8}$$

where $m,n=1,\ldots,Q$ for $m\neq n$. After the cross-correlation functions have been calculated by the robust GCC-PHAT function described in Sec. II C, a search-and-sum procedure is performed. For each small region l, the cross-correlation values corresponding to the theoretical time delays $\boldsymbol{\tau}(\boldsymbol{\alpha}_l)$ are found and accumulated. Once all regions have been swept, an acoustic map is created in the space. Finally, it is assumed that the most likely source position $\hat{\boldsymbol{\alpha}}$ will be the region with the highest spatial likelihood. The SRP-PHAT method can be mathematically formulated as

$$\hat{\boldsymbol{\alpha}}=\arg\max_{\boldsymbol{\alpha}_l}\sum_{m\neq n}R'(\tau(\boldsymbol{\alpha}_l)). \tag{9}$$

In this work, we denote SRP-PHAT$_{\text{array}}$ as the method in which one source position estimate is obtained by a single array, and SRP-PHAT$_{\text{all}}$ as the method in which one source position estimate is obtained by the entire array network.

III. PROPOSED SOURCE'S POSITION AND ORIENTATION ESTIMATION METHOD

The ANN is a structure composed of units (input, hidden, and output) connected by weighted links between them. If the unit has only output connections, then it is an input unit. If the unit has only input connections, then it is an output unit. If it has both types of connection, then it is a hidden unit. Each unit computes the sum of weighted outputs of prior units with connection leading to this unit, v, and applies to the activation function $f_{\text{act}}(\cdot)$, which limits the output amplitude. The logistic function

$$f_{\text{act}}(v)=\frac{1}{1+e^{-v}} \tag{10}$$

was assumed as the activation function whose output is in the interval [0, 1]. The output function takes the activation function output to generate the output of the unit. Here, the output function is the identity function. In this work, a gating two-stage ANN is studied in the estimation of the position and orientation.

Two-stage fully connected feedforward ANNs, illustrated in Fig. 6, are used in the joint estimation of the position and orientation of an acoustic source. Different combinations of the estimated parameters from the received signals and microphone positions of each array of the array network are used as input of different ANN configurations with different numbers of hidden units, as defined in Table III. We tested different numbers of hidden layers for each input/output configuration. By increasing the number of hidden units we could obtain better results but at the cost of increasing the adaptation time. Then, we opted to define the number of hidden units at most twice the number of the input units. The estimated parameters, defined in detail in Sec. III A, consist of energy features, TDEs, source position estimates, and distance estimates. We aimed to obtain a more accurate source position estimate when source position estimates from the TDOA-based or SRP-PHAT localization method are used as the source position starting values for the ANN algorithm. In the proposed method, the use of a HMA (Ref. 20) configuration, formed by hundreds of microphones, and analyses of theoretical radiation pattern models are avoided at the expense of a training phase in the ANN. A final consideration must be pointed out, here two different situations are investigated: the *best array selection* with orientation estimation and *position estimation* with orientation estimation.

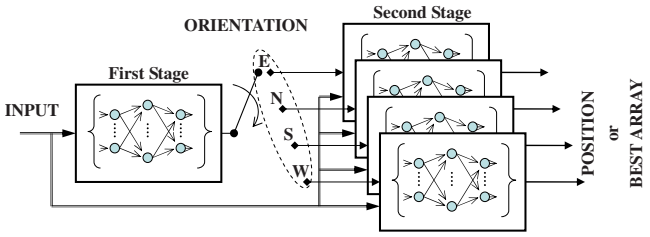


FIG. 6. (Color online) Two-stage artificial neural network studied topology. The configurations ANN$_1$, ANN$_2$, and ANN$_3$ are defined in Table III. The **INPUT**(ANN$_1$)={P+C+D} for the ANN$_1$ configuration comprises the set of energy features and distance estimates. The **INPUT**(ANN$_2$)={P+C+{$\hat{x},\hat{y}$}/{$\hat{x},\hat{y},\hat{z}$}} for the ANN$_2$ configuration comprises the set of energy features and source position estimates. The **INPUT**(ANN$_3$)={P+C+{$\hat{\tau}$}+[x,y,z]} for the ANN$_3$ configuration comprises the set of energy features, TDEs, and microphone positions. The **first STAGE OUT.**={Orientation(E,N,S,W)} gives the orientation estimation at the first stage output and the **second STAGE OUT.**={(ANN$_1$)→BEST ARRAY; (ANN$_2$,ANN$_3$)→POSITION} gives either the best array or the source position estimate at the second stage output.

A. ANN input/output definitions

Energy features correspond to power and correlation estimates calculated by the recorded data. The maximum power at the array is defined as the highest power value over all signals of an array, that is,

$$P_p=\max_m\{P_{p_m}\}, \tag{11}$$

where P_{p_m} is the power calculated for signal $\mathbf{x}_{p_m}(t)$ of the pth array.

The maximum signal correlation across the array is defined as the highest correlation value over all signals of an array, that is,

$$C_p=\max_{m\neq n}\{C_{p_m,p_n}\}, \tag{12}$$

where C_{p_m,p_n} is the correlation calculated between signals $\mathbf{x}_{p_m}(t)$ and $\mathbf{x}_{p_n}(t)$ of the pth array.

Distance is defined by the Euclidean distance between the source position estimate and the corresponding array position which generated that estimate and is given by

$$D_p = \sqrt{(\hat{x} - x_p)^2 + (\hat{y} - y_p)^2 + (\hat{z} - z_p)^2}, \quad (13)$$

where $(\hat{x},\hat{y},\hat{z})_p$ and (x_p,y_p,z_p) are the source position estimate by the pth array and the true position of the pth array, respectively. The position error, used to define the *best array*, is defined by the Euclidean distance between the source position estimate and the real position of the source given by

$$\text{Error}_p = \sqrt{(\hat{x} - x_0)^2 + (\hat{y} - y_0)^2 + (\hat{z} - z_0)^2}, \quad (14)$$

where (x_0,y_0,z_0) is the real position of the source. In order to simplify the notation at tables and figures, $\mathrm{P}=\{P_1,\ldots,P_P\}$, $\mathrm{C}=\{C_1,\ldots,C_P\}$, $\mathrm{D}=\{D_1,\ldots,D_P\}$, $\{\hat{\mathrm{x}},\hat{\mathrm{y}},\hat{\mathrm{z}}\}=\{(\hat{x},\hat{y},\hat{z})_1,\ldots,(\hat{x},\hat{y},\hat{z})_P\}$, $\{\hat{\mathrm{x}},\hat{\mathrm{y}}\}=\{(\hat{x},\hat{y})_1,\ldots,(\hat{x},\hat{y})_P\}$, [x,y,z] and $\{\hat{\boldsymbol{\tau}}\}=\{\hat{\boldsymbol{\tau}}_1,\ldots,\hat{\boldsymbol{\tau}}_P\}$ are adopted for the set of power estimates, correlation estimates, distance estimates, source position estimates in three dimensional (3D), source position estimates in two dimensional (2D), microphone positions, and set of TDEs for arrays $p=1,\ldots,8$ and microphones $m=1,2,3,4$, respectively.

B. Orientation and *best array* selection using distance estimates

Best array is the array which yields the most reliable source position estimate, it yields the smallest position error, and it is the closest to the true source position. In an attempt to automatically select the best array, the distance estimate D_p of the pth array was used. The set of distance estimates D (a total of eight values) with the set of energy features P (power) and C (correlation) was used in the input of the ANN whose output corresponded to the orientation and the best array. The tested ANN corresponds to the ANN_1 configuration defined in Sec. IV C. This experiment was evaluated to verify if the distance estimates could be used to select the best array.

C. Orientation and position estimation using source position estimates or TDE sets

At the input of the ANN, the set of energy features P and C and the set of source position estimates $\{\hat{x},\hat{y}\}$ or $\{\hat{x},\hat{y},\hat{z}\}$ are directly used in a mapping process to find a more accurate source position estimate and estimate the orientation using the entire microphone array network information. This approach has the advantage of exploring the spatial information given by the set of source position estimates compared to one-dimensional information generated by the distance estimate set. The tested ANN corresponds to the ANN_2 configuration defined in Sec. IV C. When using TDE set $\{\hat{\boldsymbol{\tau}}\}$, an additional information given by the microphone position set [x,y,z] is used in the input. In this case, the tested ANN corresponds to the ANN_3 configuration defined in Sec. IV C.

IV. EXPERIMENTS AND RESULTS

A. Automatic estimation: Setup and conditions

All experiments were conducted in a 5 m wide $\times$ 6.4 m long $\times$ 2.65 m high room containing eight T-shaped microphone arrays (see Fig. 7), with one array fixed to each wall (arrays A, B, C, and D) and four arrays fixed to the ceiling

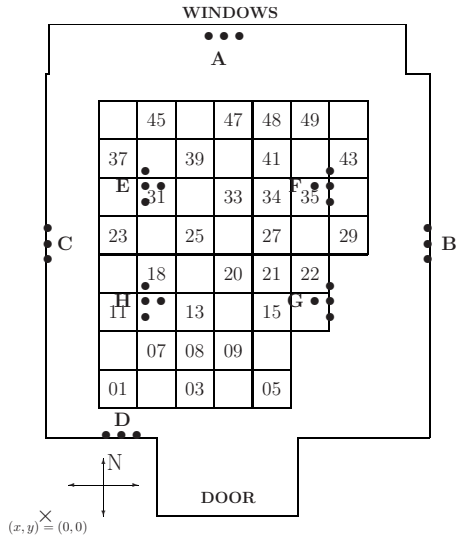


FIG. 7. The environmental setup used in the automatic estimation experiment. The areas considered in the study are numbered.

(arrays E, F, G, and H). Each array was mounted on a structure with an acoustic absorber to reduce reflection effects near the microphones. The distance between adjacent microphones at each array was set to 20 cm. The position of the center microphone in each array is given in centimeters as follows: A(236.5, 619.0, 206.0); B(497.0, 354.5, 200.0); C(3.0, 354.5, 200.0); D(98.5, 105.0, 200.0); E(130.0, 423.5, 255.0); F(370.0, 423.5, 255.0); G(370.0, 273.5, 255.0); and H(130.0, 273.5, 255.0). The room was divided into 50 areas, each 50×50 cm^2, but only 29 areas, suitably distributed and covering the entire room, were considered in our analyses. The array positions and areas are depicted in Fig. 7.

A loudspeaker was set-up over a stand fixed 140 cm above the floor to simulate an acoustic source. The stand was put in the center of each numbered area indicated in Fig. 7, and 300 Japanese words from two adult male speakers were played, with the average duration of an utterance being 0.6 s. In each studied area, the loudspeaker was turned to reflect four different orientations shifted by 90°: east (E), north (N), south (S), and west (W) orientations were considered, resulting in 116 study cases (29×4) and 34 800 samples ($29 \times 4 \times 300$). Utterances were recorded at 48 kHz by a 32 channel acquisition system and downsampled to 12 kHz. The acquisition and the playback equipment are separated, each one running on different computers with Linux operating system. The acquisition board is manufactured by Tokyo Electron Device Limited, the loudspeaker is an Entry S type manufactured by ALR Jordan, and the microphones are ECM-C10 model produced by Sony Corporation. The small loudspeaker emits the sound energy by its front part and attenuates at its rear. Thus, in this work we assumed that the loudspeaker could roughly model the human head. The background noise was around 35 dBA measured by the Ono Sokki (model LA-4350) sound meter. The estimated SNR estimated by the recorded signals was 15 dB, sufficient to reduce the performance of the GCC-PHAT estimation method when the estimates were determined by frames, which was verified experimentally. In the GCC-PHAT analysis, a frame length of 256 samples, frame shift of 128

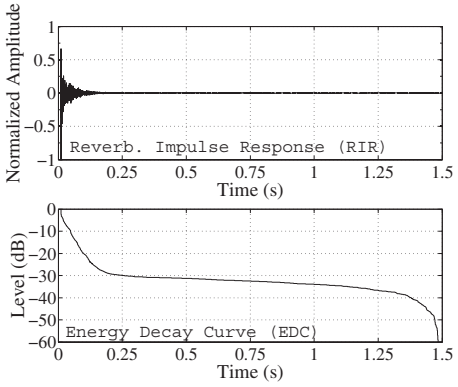


FIG. 8. RIR and EDC measured from the center of area 20 (see Fig. 7) to the center microphone of array A.

samples, and Hamming window were considered. In the experiments, one source position estimate was obtained per utterance. The reverberation impulse response (RIR) and the energy decay curve (EDC) are measured from the center of position 20 (see Fig. 7) to the center microphone in array A when a loudspeaker was turned to the north. The RIR and EDC are illustrated in Fig. 8. The reverberation time T_{60} corresponds to the time lag when the EDC decays 60 dB; here it is longer than 1 s.

B. Automatic estimation: Baselines for position estimation

Tables I and II present the results of the baseline position estimation methods in terms of the average position error at each array (arrays A, B, C, D, E, F, G, and H) in 2D/3D spaces. An analysis of Table I shows that the SRP-PHAT$_{\text{array}}$ is more robust than the TDOA-based method, and arrays fixed at the ceiling (arrays E, F, G, and H) yield better results than arrays fixed at the walls (arrays A, B, C, and D). In Table II, the *oracle* selection is performed for the best array selection for the TDOA-based and the SRP-PHAT$_{\text{array}}$ methods, i.e., the array with the best source position estimate is always selected. In the SRP-PHAT$_{\text{all}}$ method, in which one source position estimate is obtained by the entire array network, we had the best baseline results.

C. Automatic estimation: Two-stage ANNs for position and orientation estimation

ANNs were studied using the Stuttgart neural network simulator (SNNS).[27] For the ANN training/testing phase, the recorded data from each of the 29 areas were divided into two complementary sets with 80% of the data used in the training phase and 20% in the testing phase and with no overlap between both data sets. In the results, CLOSED and OPEN tests refer to the results obtained by the trained ANN evaluated using the training and testing data sets, respectively. Cross-validation was performed to generalize the results of the proposed position and orientation estimation method. This process was performed five times, but permuting the original data set. In the experiments, the average position error was calculated in 2D (x,y) and 3D (x,y,z) spaces. The following measures were used to express the results in the evaluation phase of the ANNs: the correct orientation ratio (%), which expresses the agreement between the estimated and the true source orientation, the average orientation error (°), which denotes the angle mismatch between the estimated and the true source orientation, the best array selection ratio (%), which expresses the agreement between the chosen array and the array nearest to the source position, and the average position error (centimeters), which denotes the mismatch between the estimated and the true source position. In the ANN training phase, a sufficient number of iterations assures convergence to a steady state or to avoid overtraining.

Figure 6 illustrates the gating two-stage ANN, and Table III presents the studied configurations. In the training phase, at the first stage, the correct orientation was used as the target value, while at the second stage either the best array or the actual source position was used. The ANN$_1$ inputs consist of the combination of the sets of energy features and distance estimates {P+C+D}, which estimates the orientation at the first stage (one out of four output values) and selects the best array at the second stage (one out of eight output values). The ANN$_2$ inputs consist of the combination of the sets of energy features and source position estimates {P+C+$\{\hat{x},\hat{y}\}$/$\{\hat{x},\hat{y},\hat{z}\}$}, which estimates the orientation in the first stage and the position in 2D/3D space in the second stage. Finally, the ANN$_3$ inputs consist of the combination of the sets of energy features, microphone positions, and TDEs {P+C+[x,y,z]+$\{\hat{\boldsymbol{\tau}}\}$}, which estimates the orientation in the first stage and the position in 3D space in the second stage.

TABLE II. The average position error of the baselines: the TDOA-based, SRP-PHAT$_{\text{array}}$, and SRP-PHAT$_{\text{all}}$.

Method	Space 2D (cm)	Space 3D (cm)
TDOA-based (best array—oracle)	21.2	34.1
SRP-PHAT$_{\text{array}}$ (best array—oracle)	18.6	31.7
SRP-PHAT$_{\text{all}}$	16.0	29.8

TABLE I. The average position error by array considering the baseline TDOA-based and SRP-PHAT$_{\text{array}}$ estimation methods. The position of arrays {A, B, C, D, E, F, G, and H} is defined in Sec. IV A and illustrated in Fig. 7.

Method	Average position error by array—2D/3D spaces (cm) A	B	C	D	E	F	G	H
TDOA-based	659.9/678.2	396.6/407.6	372.4/384.4	585.4/600.6	160.0/207.8	191.1/242.7	223.9/270.2	154.2/209.4
SRP-PHAT$_{\text{array}}$	105.0/114.9	112.2/123.5	95.4/107.5	117.6/125.8	63.4/79.0	76.5/92.2	70.0/86.3	79.3/97.9

TABLE III. Two-stage ANN configurations. The results were obtained considering the hidden unit numbers in bold. (1) and (2) in output denote first and second stages, respectively.

Topology	No. of ANN units		
	Input	Hidden	Output
ANN_1	24 (P+C+D)	**48**	4(1),8(2)
ANN_2	32 (P+C+$\{\hat{x},\hat{y}\}$)	**80**	4(1),2(2)
ANN_2	40 (P+C+$\{\hat{x},\hat{y},\hat{z}\}$)	**80**	4(1),3(2)
ANN_3	120 ([x,y,z]+$\{\hat{\boldsymbol{\tau}}\}$)	**240**	4(1),3(2)
ANN_3	136 (P+C+[x,y,z]+$\{\hat{\boldsymbol{\tau}}\}$)	**272**	4(1),3(2)

The input set $\{P+C+[x,y]+\{\hat{\boldsymbol{\tau}}\}\}$ is not presented here because the results were closed to the values obtained using the 3D results disregarding the z-dimension.

D. Automatic estimation: Orientation estimation using only energy features

In Ref. 20, the orientation of a directional acoustic source was estimated using 448 microphones which were used to estimate the radiation pattern of the source. Here, an experiment was evaluated to verify if the ANN can roughly model the source radiation pattern like of Sachar *et al.* using a small number of microphones and thus predict the source orientation. The set of energy features P and C, as defined in Sec. III A, was used in the orientation estimation. A simple one-stage fully connected feedforward ANN, like the first stage of the proposed two-stage ANN, illustrated in Fig. 6, with 16 input units {P+C}, 48 hidden units, and four (E, N, S, W) output units was studied. The number of hidden units was determined experimentally. The results of this experiment are discussed in Sec. IV E 3.

E. Automatic estimation: Results

1. Best array selection by individual criteria

In this section, an experiment was used to determine whether signal power, signal correlation, and distance estimate could be used separately to choose the best array. Three individual criteria were defined: the best array has maximum power estimate ($\mathcal{P}$), it has maximum correlation estimate ($\mathcal{C}$), and it has minimum distance estimate ($\mathcal{D}$) among all arrays. Each criterion must match the array which yields the minimum position error in a total of 34 800 (300 utterances $\times$ 4 orientations $\times$ 29 areas) input patterns. The results are in Table IV, where the correct selection ratio (L_1) is the relation between the number of matches and the total input patterns, and the average position error (L_2) is calculated by taking the selected arrays. Using $SRP\text{-}PHAT_{array}$ position estimates we obtained better results in terms of L_2 than using TDOA-based position estimates, but these results are worse compared to baselines in Table II. Finally, it is clear by analyzing L_1 results that the selection of the best array by each individual criterion is not a suitable approach. This leads us to investigate a method, in our case exploring neural networks, to obtain more confident results.

TABLE IV. The correct selection ratio in percent (L_1) and the average position error in centimeters (L_2) when the minimum position error calculated in 2D/3D spaces matches the criteria defined in Sec. IV E 1. Source position estimates by the TDOA-based and $SRP\text{-}PHAT_{array}$ methods were used to calculate position errors.

Criteria	TDOA-based		$SRP\text{-}PHAT_{array}$	
	L_1	L_2	L_1	L_2
$\mathcal{P}$ and min. error in 2D	36.4	161.1	32.4	51.9
$\mathcal{C}$ and min. error in 2D	37.1	249.9	33.6	49.8
$\mathcal{D}$ and min. error in 2D	21.7	122.5	31.9	53.2
$\mathcal{P}$ and min. error in 3D	35.5	223.5	31.6	67.6
$\mathcal{C}$ and min. error in 3D	35.7	331.0	31.6	66.3
$\mathcal{D}$ and min. error in 3D	5.4	175.0	16.5	85.9

2. Proposed position and orientation estimation method

Table V presents the results of the proposed position and orientation estimation method, with configuration ANN_1 defined in Table III, using the sets of energy features P and C, and distance estimates D derived by TDOA-based and $SRP\text{-}PHAT_{array}$ source position estimates, respectively. The results show that the $SRP\text{-}PHAT_{array}$ has slightly better performance than the TDOA-based method in terms of the used measures. Table VI presents the results of the configuration ANN_2 defined in Table III using the set of energy features P and C and the set of source position estimates $\{\hat{x},\hat{y}\}$ or $\{\hat{x},\hat{y},\hat{z}\}$ obtained by the TDOA-based and $SRP\text{-}PHAT_{array}$ methods, respectively. The results show that the $SRP\text{-}PHAT_{array}$ is better than the TDOA-based method in terms of correct orientation ratio and average position error in this approach. Table VII presents the results of the configuration ANN_3 defined in Table III using the set of energy

TABLE V. Results obtained by the neural network topology ANN_1. The distance estimate set D was calculated by the TDOA-based (TDOA) and $SRP\text{-}PHAT_{array}$ (SRP) source position estimates. The average position error value was calculated using the position estimates from the selected arrays for the given dimensional space.

Inputs	Space	Test	Corr. orient. ratio (%)		Avg. orient. error (deg)		Corr. select. ratio (%)		Avg. pos. error (cm)	
			TDOA	SRP	TDOA	SRP	TDOA	SRP	TDOA	SRP
24(P+C+D)	2D	CLOSED	80.7	81.7	23.8	22.5	60.3	59.2	30.9	27.3
24(P+C+D)	2D	OPEN	77.8	79.3	27.1	25.5	56.6	55.6	32.6	28.5
24(P+C+D)	3D	CLOSED	78.9	79.4	25.8	25.4	63.1	62.3	43.9	41.8
24(P+C+D)	3D	OPEN	76.0	76.8	29.3	28.6	59.2	58.7	45.8	43.4

TABLE VI. Results obtained by the neural network topology ANN_2. Source position estimates $\{\hat{x},\hat{y}\}/\{\hat{x},\hat{y},\hat{z}\}$ in 2D/3D were estimated by the TDOA-based (TDOA) and SRP-$PHAT_{array}$ (SRP) methods. The average position error value was calculated using the position estimates for the given dimensional space, and the value (*) was the average position error in 2D directly calculated from the 3D estimates disregarding the z-dimension.

			Corr. orient. ratio (%)		Avg. orient. error (deg)		Avg. pos. error (cm)	
Inputs	Space	Test	TDOA	SRP	TDOA	SRP	TDOA	SRP
32(P+C+$\{\hat{x},\hat{y}\}$)	2D	CLOSED	87.4	97.6	15.7	2.9	29.9 (27.9*)	23.1 (23.1*)
32(P+C+$(\hat{x},\hat{y})$)	2D	OPEN	83.9	94.7	19.7	6.5	31.1 (29.1*)	23.5 (23.5*)
40(P+C+$\{\hat{x},\hat{y},\hat{z}\}$)	3D	CLOSED	93.4	99.7	8.2	0.4	32.9	27.6
40(P+C+$\{\hat{x},\hat{y},\hat{z}\}$)	3D	OPEN	90.3	98.3	12.0	2.2	33.9	27.9

features P and C, the set of TDEs $\{\hat{\tau}\}$, and the set of microphone positions [x, y, z]. The columns marked by "△" assume the training and testing conditions in Sec. IV C. However, in the columns marked by "□," the training and testing conditions were modified. The modification consisted of separating training and testing data sets by areas. Six areas were randomly chosen inside the room, avoiding border areas. (Areas 1, 3, 5, 11, 22, 23, 29, 37, 43, 45, 47, and 49 are border areas in Fig. 7. Although area 48 is also a border area, it is surrounded by other areas and can thus be disregarded.) The data from these areas were used as the testing data set, while the data from the other 23 areas formed the training data set. There was no overlap between the training and testing data sets, and five different data permutations were simulated for cross-validation. Based on this separation, approximately 80% and 20% of the data were used as training and testing data sets, respectively. CLOSED and OPEN tests were performed using the training and testing data sets, and the data from all areas were common to both data sets in △ characterizing *CLOSED POSITONS TEST*, while in □ the data from areas in the training data set were not in the testing data set characterizing *OPEN POSITONS TEST*. Comparing results of Table VII to Tables V and VI, the best results in terms of correct orientation ratio and average position error were obtained, where *CLOSED POSITONS TEST* yielded better results than *OPEN POSITONS TEST*.

3. Orientation estimation by only energy features

In the CLOSED/OPEN test results of the one-stage ANN using only the set of energy features P and C, we had correct orientation ratios of 56.2%/57.4%, respectively, and even increasing the number of iteration in the ANN training phase, there was not a significant improvement in the ratio. This behavior seemed to be strange since it was expected that the ANN could model the source radiation pattern in order to predict the source orientation. A possible cause of this divergence could be due to small amount of energy feature samples, which were restricted to 16 samples {P+C}. The orientation estimation task by using only energy features showed to be difficult even using the features of all array of the array network.

V. DISCUSSION

A. Two-stage ANNs for position and orientation estimation

As shown in Table V, the set of distance estimates D derived from the TDOA-based source position estimates yielded better results in terms of the best array selection than using the SRP-$PHAT_{array}$ source position estimates. However, for the correct orientation ratio and average position error, the opposite was observed, for instance, in the OPEN results in terms of the correct orientation ratio, the best array selection, and the average position error were 77.8% (76.0%), 56.6% (59.2%), and 32.6 cm (45.8 cm) in 2D (3D) space for the TDOA-based method and were 79.3% (76.8%), 55.6% (58.7%), and 28.5 cm (43.4 cm) in 2D (3D) space for the SRP-$PHAT_{array}$ method. The results in terms of the average position error were worse than the baseline presented in Table II, but encourage us to continue our investigation seeking for better results.

From Table VI, it is evident that SRP-$PHAT_{array}$ source position estimates yielded better results than the TDOA-

TABLE VII. Results obtained by the neural network topology ANN_3. TDEs and microphone positions of every array were employed. "△" denotes *CLOSE POSITONS TEST* and "□" denotes *OPEN POSITONS TEST*. The average position error value was calculated using the position estimates for the given dimensional space, and the value (*) was the average position error in 2D directly calculated from the 3D estimates disregarding the z-dimension.

		Corr. orient. ratio (%)		Avg. orient. error (deg)		Avg. pos. error (cm)	
Inputs	Test	△	□	△	□	△	□
120([x,y,z]+$\{\hat{\tau}\}$)	CLOSED	99.9	96.8	0.1	4.2	24.4 (20.9*)	24.9 (21.6*)
120([x,y,z]+$\{\hat{\tau}\}$)	OPEN	99.4	87.0	0.7	17.0	24.6 (21.0*)	25.3 (22.5*)
136(P+C+[x,y,z]+$\{\hat{\tau}\}$)	CLOSED	99.9	96.3	0.1	4.5	23.2 (20.3*)	25.7 (22.1*)
136(P+C+[x,y,z]+$\{\hat{\tau}\}$)	OPEN	99.5	84.2	0.6	18.5	23.3 (20.5*)	28.2 (24.8*)

based, for instance, in OPEN results, the correct orientation ratio and the average position error were 83.9% (90.3%) and 31.1 cm (33.9 cm) in 2D (3D) space for TDOA-based method and were 94.7% (98.3%) and 23.5 cm (27.9 cm) in 2D (3D) spaces for SRP-PHAT$_{array}$ method. Comparing Tables V and VI, the improvement in the orientation estimation changing the set of distance estimates D by the set of source position estimates $\{\hat{x},\hat{y}\}/\{\hat{x},\hat{y},\hat{z}\}$ can be explained by the fact that a 2D/3D coordinate has more spatial information than a one-dimensional value defined by D. The ANN$_2$ configuration yielded a correct orientation ratio of 90.3% using source position estimates by TDOA-based and 98.3% using source position estimates by SRP-PHAT$_{array}$, while the ANN$_1$ configuration yielded a correct orientation ratio of 76.0% using distance estimates derived from TDOA-based and 76.8% using distance estimates derived from SRP-PHAT$_{array}$ for OPEN results in 3D space. In Table VI, the source position estimates by SRP-PHAT$_{array}$ yielded better results (27.9 cm) than the baseline SRP-PHAT$_{all}$ (29.8 cm) for the 3D case in terms of the average position error, but for the 2D case, the SRP-PHAT$_{all}$ is still the best. Values (*), in Table VI, are the 2D estimates obtained directly by the 3D estimates disregarding the z dimension, which showed to be better estimates than those obtained by training the ANNs using 2D source position estimates.

Comparing the results in Table VII with those in Tables V and VI, we noted that the set of TDEs and the set of microphone positions (low level parameters) were more suitable for the orientation and position estimation than the set of source position estimates (intermediate level parameters) or the set of distance estimates (high level parameters). In *CLOSED POSITONS TEST* in Table VII, a correct orientation ratio of 99.5% and an average position error of 23.3 cm in OPEN results in 3D case were obtained, a reduction of 21.8% compared to the baseline SRP-PHAT$_{all}$. However, SRP-PHAT$_{all}$ (16.0 cm) is still better in 2D case compared to the estimation method (20.5 cm). In the *OPEN POSITONS TEST*, the objective was to jointly estimate the position and orientation in the untrained areas. The results show that the correct orientation ratio ranges from 84% to 87% and the average position error is 25.3 cm, which is still better than the baseline SRP-PHAT$_{all}$ in 3D space.

In *OPEN POSITONS TEST*, adding the set of energy features P and C in the input data set, we had an increase in the reliability of the positions used in the training phase. Thus, the ANNs were more dependent on these positions. In the testing phase, the input pattern tends to get closer to the positions used in the training phase, and the orientation is affected by this estimation because of overtraining. Therefore, there is an increase in the position error and decrease in the orientation ratio due to the confidence given by the set of energy features in the trained areas, which reduce the performance in the untrained areas in *OPEN POSITONS TEST*, as observed in Table VII.

B. Additional comments: Automatic estimation

For different representation levels of the estimated parameters, an improvement in the results and in the number of input units in the ANN was observed by moving from the high to low level of representation, which can be observed by comparing the results of Tables VII (using high representation), V (using intermediate representation), and VI (using low representation). Although the analyses presented in this paper concentrate on using all eight arrays, other analyses could be performed by individual array or combining different arrays, such as using only walls or ceiling arrays. Ceiling arrays are especially important because they can perceive signals directly and it is more difficult physically to block them than wall arrays. For instance, considering energy features P and C, TDEs $\{\hat{\boldsymbol{\tau}}\}$, and microphone positions [x,y,z] sets in the experiments, the correct orientation ratio (%) and average position error (2D/3D) by using only array **A** (wall array) were 67.9% and 48.7/53.0 cm, using only array **E** (the ceiling array yielded the best average position error in Table I) were 76.0% and 38.6/45.4 cm, and the combination of arrays **E** and **G** (ceiling arrays) was 88.5% and 26.4/30.7 cm in training and testing conditions of Sec. IV C.

The estimated position and orientation by the proposed method could be used in some practical application, for example, in automatic speech or speaker recognition. In Ref. 6, the source position estimate was used to compensate the microphone signals due to the channel effect. It was shown that in a squared region of 60×60 cm^2 around the sound source there was no much effect in recognition results; thus the estimated position by our proposed method could be used in recognition tasks. The estimated orientation could be used in a compensation method too. As a directional sound source does not radiate equally in all directions, suitable weights can be applied to microphone signals that are not in front of the speaker, compensating attenuation and distortion due to the relative orientation prior to a signal processing step.

C. Additional comments: Could the proposed automatic estimation method model the human auditory perception system?

Here, a short discussion about the proposed method and the human auditory perception system is presented. A subjective experiment with blindfolded listeners was performed in the same environment. After hearing a spoken phrase by a human speaker, the listener was allowed to remove the blindfold and indicate the speaker facing angle (E, N, S, or W) and the speaker position inside the room. Six listeners participated in the experiment. The results showed correct orientation ratios of 76.7% and 79.9% with average position errors of 65.4 and 60.5 cm in 2D (Ref. 28) before and after a training phase, respectively. In the training phase, the listener was allowed to remove the blindfold and verify the speaker position and orientation. Comparing these results with the best results of the proposed automatic method, we noted that the proposed method using the entire microphone array network outperforms the human auditory perception ability to discriminate the position and orientation of an acoustic source. However, it is unfair to compare the array network, which is distributed in the space, with the auditory system, which is concentrated in the head. A more fair comparison would be with an individual array. It is clear that arrays **A** (wall array near the window) and **E** (ceiling array), individu-

ally, do not have the same performance as the array network, as noted by the results in Sec. V B. The average position errors obtained by individual arrays **A** and **E** were better than those obtained in the subjective experiment, but in terms of the correct orientation ratio the auditory system yielded better estimates. Thus, one array cannot model the auditory perception system accurately in terms of orientation estimation. Additionally, exploring the spatial diversity by combining two ceiling arrays **E** and **G** showed to be sufficient to surpass the auditory perception system in terms of the considered measures.

VI. CONCLUSIONS

In this paper, a method which automatically estimates the position and orientation of a directional acoustic source by neural networks was proposed. Different strategies for the automatic estimation changing the type of the input parameters were compared. The research was restricted to fully connected feedforward neural networks, but other ANN types and topologies could also be tested. The proposed method using low level parameter (TDEs and microphone positions) input set yielded better results than intermediate (source position estimates) and high parameter (distance values) input sets.

In terms of position estimation, the proposed method was better than the conventional position estimation methods (TDOA-based and SRP-PHAT) presented in this paper at least in 3D space. In terms of correct orientation ratio, accurate estimation was obtained using low or intermediate level parameters. As the position information is important in speech recognition, it is expected that employing the orientation information will be useful in this task, for instance, either selecting the nearest array facing the speaker or restricting the vocabulary that must be recognized in a given direction in a voiced command application, supposing that the speaker faces the device before uttering the command. Finally, as presented in last examples, speech recognition application will be in our next research step.

ACKNOWLEDGMENTS

We would like to thank the Global COE program "Frontiers of Intelligent Sensing" and MEXT for supporting our research.

[1]M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, New York, 2001).
[2]E. Hansler and G. Schmidt, *Speech and Audio Processing in Adverse Environments* (Springer, New York, 2008).
[3]J. G. Ryan and R. A. Goubran, "Optimum near-field performance of microphone arrays subject to a far-field beampattern constraint," J. Acoust. Soc. Am. **108**, 2248–2255 (2000).
[4]J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," IEEE Trans. Signal Process. **50**, 1843–1854 (2002).
[5]X. Chen, Y. Shi, and W. Jiang, "Speaker tracking and identifying based on indoor localization system and microphone array," in 21st International Conference on Advanced Information Networking and Applications Workshops (2007), Vol. **2**, pp. 347–352.
[6]L. Wang, N. Kitaoka, and S. Nakagawa, "Robust distance speaker recognition based on position-dependent CMN by combining speaker-specific GMM with speaker-adapted HMM," Speech Commun. **49**, 501–513 (2007).
[7]H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in video conferencing," in Proceedings of the ICASSP (1997), Vol. **I**, pp. 187–190.
[8]S. Fischer and K. U. Simmer, "An adaptive microphone array for hands-free communication," in Proceedings of the 4th International Workshop on Acoustic Echo and Noise Control, IWAENC-95 (1995), pp. 44–47.
[9]M. R. Bai and C. Lin, "Microphone array signal processing with application in three-dimensional spatial hearing," J. Acoust. Soc. Am. **117**, 2112–2121 (2005).
[10]K. Nakadai, H. Nakajima, M. Murase, S. Kaijiri, K. Yamada, T. Nakamura, Y. Hasegawa, H. G. Okuno, and H. Tsujino, "Robust tracking of multiple sound sources by spatial integration of room and robot microphone arrays," Proceedings of the ICASSSP (2006), Vol. **IV**, pp. 929–932.
[11]J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," J. Acoust. Soc. Am. **107**, 384–391 (2000).
[12]R. Bucher and D. Misra, "A synthesizable vhdl model of the exact solution for three-dimensional hyperbolic positioning system," VLSI Des. **15**, 507–520 (2002).
[13]A. Brutti, M. Omologo, and P. Svaizer, "Speaker localization based on oriented global coherence field," Proceedings of the Interspeech (2006), pp. 2606–2609.
[14]J. Chen, J. Benesty, and Y. Huang, "Performance of gcc-and amdf-based time-delay estimation in practical reverberant environment," EURASIP J. Appl. Signal Process. **1**, 25–36 (2005).
[15]F. Talantzis, A. G. Constantinides, and L. C. Polymenakos, "Estimation of direction of arrival using information theory," IEEE Signal Process. Lett. **12**, 561–564 (2005).
[16]C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-24**, 320–327 (1976).
[17]V. C. Raykar, "Automatic position calibration of multiple microphones," in Proceedings of the ICASSSP (2004), Vol. **IV**, pp. 69–72.
[18]I. McCowan, M. Lincoln, and I. Himavan, "Microphone array shape calibration in diffuse noise fields," IEEE Trans. Audio, Speech, Lang. Process. **16**, 666–670 (2008).
[19]A. Brutti, M. Omologo, and P. Svaizer, "Oriented global coherence field for the estimation of the head orientation in smart room equipped with distributed microphone arrays," in Proceedings of the Interspeech (2005), pp. 2337–2340.
[20]J. M. Sachar and H. F. Siverman, "A baseline algorithm for estimating talker orientation using acoustical data from a large-aperture microphone array," in Proceedings of the ICASSSP (2004), Vol. **IV**, pp. 65–68.
[21]A. Abad, C. Segura, C. Nadeu, and J. Hernando, "Audio-based approaches to head orientation estimation in a smart-room," in Proceedings of the Interspeech (2007), pp. 590–593.
[22]C. Segura, C. Canton-Ferrer, A. Abad, J. R. Casas, and J. Hernando, "Multimodal head orientation towards attention tracking in smartrooms," Proceedings of the ICASSP (2007), Vol. **II**, pp. 681–684.
[23]A. Brutti, M. Omologo, P. Svaizer, and C. Zieger, "Classification of acoustic maps to speaker position and orientation from a distributed microphone network," in Proceedings of the ICASSSP (2007), Vol. **IV**, pp. 493–496.
[24]S. Haykin, *Neural Networks: A Comprehensive Foundation* (Macmillan, New York, 1994).
[25]K. Varma, T. Ikuma, and A. A. Beex, "Robust tde-based doa estimation for compact audio arrays," in 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop (2002), pp. 214–218.
[26]R. Parisi, A. Cirrillo, M. Panella, and A. Uncini, "Source localization in reverberant environments by consistent peak selection," in Proceedings of the ICASSSP (2007), Vol. **I**, pp. 37–40.
[27]http://www.ra.cs.uni-tuebingen.de/SNNS/ (Last viewed 3/24/2009).
[28]A. Y. Nakano, K. Yamamoto, and S. Nakagawa, "Auditory perception of speaker's position, distance and facing angle in a real enclosed environment," in Acoustical Society of Japan—Autumn Meeting (2008), pp. 525–526.

Exploitation of symmetries for image reconstruction in linearized variable density diffraction tomography

Daxin Shi
Toshiba Medical Research Institute USA, Vernon Hills, Illinois 60061

Mark A. Anastasio[a)]
Department of Biomedical Engineering, Medical Imaging Research Center, Illinois Institute of Technology, Chicago, Illinois 60616

(Received 15 April 2009; revised 19 August 2009; accepted 21 August 2009)

The problem of reconstructing an object's weakly varying compressibility and density distributions in three-dimensional (3D) acoustic diffraction tomography is studied. Based on the Fourier diffraction projection theorem for acoustic media, it is demonstrated that the 3D Fourier components of an object's compressibility and density distributions can be decoupled algebraically, thereby providing a method for separately reconstructing the distributions. This is facilitated by the identification and exploitation of tomographic symmetries and the rotational invariance of the imaging model. The developed reconstruction methods are investigated by use of computer-simulation studies. The application of the proposed image reconstruction strategy to other tomography problems is discussed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238157]

PACS number(s): 43.60.Lq, 43.60.Pt, 43.60.Rw, 43.60.Uv [WMC] Pages: 3095–3105

I. INTRODUCTION

Diffraction tomography [1,2] (DT) is a three-dimensional (3D) linearized inverse scattering technique that represents a generalized form of x-ray tomography that is applicable to diffracting wavefields. In the classic DT measurement configuration, the object is illuminated by a time-harmonic scalar wavefield, and the amplitude and phase of the forward scattered wavefield are measured on a detector plane behind the object. From knowledge of such measurements obtained at a collection of tomographic view angles, DT reconstruction algorithms[3–6] can produce estimates of the object's complex-valued refractive index distribution.

Although it was originally proposed and continues to be developed[7–9] in the optics community, since the late 1970s (Refs. 10–17) ultrasound-based DT and related methods have concurrently been developed as diagnostic imaging methods. Because of their potential for quantitative and reproducible ultrasound imaging, the development of ultrasound DT systems for biomedical imaging applications remains an active and important research area.[14–20] Moreover, the mathematical formulation of DT is relevant to several other modern wave-based imaging modalities.[8,21–25] Conventional formulations of DT that are based on the Helmholtz equation are valid for acoustic radiation only if the object's density is constant and equal to the fluid medium in which it is embedded. When applied to acoustic media, conventional DT reconstruction algorithms produce a composite image that reflects both the compressibility and density variations in the object.[26] In certain cases, however, it may be useful to reconstruct separate images of the object's compressibility and density variations.

Over the past 2 decades, several ultrasound DT algorithms have been developed[26–30] for reconstruction of separate images of an object's compressibility and density variations. Such algorithms are referred to as *variable density* DT reconstruction algorithms, and they reduce to conventional DT reconstruction algorithms for constant density objects. In Ref. 26, Devaney developed a variable density DT reconstruction theory for weakly scattering two-dimensional (2D) nondispersive objects assuming the classic DT measurement configuration. The reconstruction theory required that the object be insonified, at each tomographic view angle, by use of two plane-waves having distinct temporal frequencies. Filtered backpropagation reconstruction algorithms were proposed and validated by use of noiseless simulation data. Although not demonstrated, it was predicted that the performance of the algorithms would degrade significantly in the presence of data noise. Using identical physical assumptions, Witten *et al.*[28] developed a variable density DT method for use with cylindrical measurement geometries. A related method was proposed by Moghaddam and Chew,[29] which required an assumption that the measurements were acquired in the far-field. Both of these methods require use of explicit multi-dimensional Fourier domain interpolation operations, which can compromise image fidelity. The methods developed by Norton[27] and Mensah and Lefebvre[30] are distinct from those described above in that they utilize finite-bandwidth broadband insonifying sources and different data-acquisition strategies.

In this article, novel reconstruction methods for 3D variable density DT are developed and investigated. The physical assumptions regarding the scattering object and measurement geometry utilized by Devaney[26] are assumed and applied to the 3D case. Specifically, our method requires that a quasi-nondispersive and weakly scattering object is insoni-

[a)]Author to whom correspondence should be addressed. Electronic mail: anastasio@iit.edu

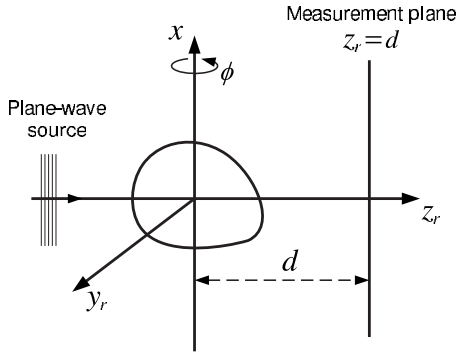


FIG. 1. The classic 3D DT measurement geometry. A plane-wave insonifies the object, and the forward scattered pressure wavefield is measured on the detector plane $z_r = d$. Two such measurements are acquired by use of insonifying plane-waves that have different temporal frequencies. Tomographic scanning is achieved by simultaneously rotating the source and detector plane about the x-axis.

fied, in turn, by use of plane-waves at two distinct temporal frequencies, and near field measurements of the transmitted wavefields are recorded. By a *quasi-nondispersive* object, we refer to one that is characterized by a compressibility distribution $\kappa(\vec{r},\omega)$ that is approximately independent of temporal frequency ω over the interval in which the multi-frequency measurements are acquired. We demonstrate that the Fourier components of a weakly scattering object's compressibility and density variations can be decoupled without use of *ad-hoc* Fourier domain interpolation operations or explicit multi-dimensional spatial domain deconvolution operations. This is accomplished by identification and exploitation of tomographic symmetries and the rotational invariance of the imaging model, and results in two mathematically equivalent 3D reconstruction formulas. From knowledge of the Fourier components, computationally efficient conventional DT reconstruction algorithms can be employed for determination of the two separate object properties. This image reconstruction approach is distinct from those previously proposed. The proposed reconstruction methods are implemented numerically, and the results of a preliminary 3D computer-simulation study are presented.

This article is organized as follows. In Sec. II, we review the imaging model of 3D variable density acoustic DT that provides the basis for our analysis. In Sec. III, tomographic Fourier space symmetries are identified and formulated mathematically. These symmetries are utilized for the development and analysis of two novel reconstruction formulas in Sec. IV. A computer-simulation study is presented in Sec. V to validate our reconstruction algorithms and includes a preliminary numerical investigation of their noise properties. Finally, a discussion of the algorithms and a summary of the article is provided in Sec. VI.

II. BACKGROUND

The tomographic scanning geometry we utilize is shown in Fig. 1. The rotated Cartesian coordinate system $\vec{r} = (x, y_r, z_r)$ is related to a reference system $\vec{r} = (x, y, z)$ by a rotation about the x-axis, which represents the axis of tomographic scanning. The rotated and reference coordinate systems are related as $y_r = y\cos\phi + z\sin\phi$ and $z_r = -y\sin\phi + z\cos\phi$, where the angle $\phi \in [0, 2\pi)$ is measured from the positive y-axis and represents the tomographic view angle. The unit vectors $\hat{s}_1$, $\hat{s}_{2,r}(\phi)$, and $\hat{s}_{0,r}(\phi)$ are used to indicate the directions of the positive x-, y_r-, and z_r-axes, respectively. Hereafter, unless necessary, the ϕ dependence of $\hat{s}_{2,r}(\phi)$ and $\hat{s}_{0,r}(\phi)$ will be suppressed. The physical model of 3D variable density DT is reviewed below. This material is a direct generalization of the 2D model described by Devaney.[26]

A. Imaging model

As shown in Fig. 1, consider that an object contained in a finite volume V is immersed in a homogeneous fluid bath and insonified at view angle ϕ by a plane-wave $P_{\text{in}}(\vec{r},\omega) = e^{ik\hat{s}_{0,r}\cdot\vec{r}}$ $(i \equiv \sqrt{-1})$. The time-dependence $e^{-i\omega t}$ is assumed, and $k \equiv \omega / C_0$ denotes the wavenumber, where C_0 is the acoustic speed in the background fluid. The total pressure $P(\vec{r},\omega)$ transmitted through the object is measured on the detector plane $z_r = d$ that is perpendicular to the direction of the insonifying plane-wave. When the tomographic view angle ϕ is varied, the insonifying wavefield and detector plane rotate simultaneously about the x-axis. The pressure field $P(\vec{r},\omega)$ satisfies the equation[31]

$$(\nabla^2 + k^2)P(\vec{r},\omega) = -k^2\gamma_\kappa(\vec{r},\omega)P(\vec{r},\omega) + \nabla\cdot[\gamma_\rho(\vec{r})\nabla P(\vec{r},\omega)], \tag{1}$$

where ∇^2 is the 3D Laplacian operator, and

$$\gamma_\kappa(\vec{r},\omega) \equiv [\kappa(\vec{r},\omega) - \kappa_0]/\kappa_0 \tag{2}$$

and

$$\gamma_\rho(\vec{r}) \equiv [\rho(\vec{r}) - \rho_0]/\rho(\vec{r}) \tag{3}$$

describe the compressibility and density variations, respectively, of the object. The constants κ_0 and ρ_0 denote the compressibility and density of the background medium, respectively. The object properties $\gamma_\kappa(\vec{r},\omega)$ and $\gamma_\rho(\vec{r})$ are bounded functions with compact support. It should be noted that $\gamma_\kappa(\vec{r},\omega)$ can depend on the temporal frequency ω for dispersive objects, while $\gamma_\rho(\vec{r})$ is always frequency-independent. Note that when $\rho(\vec{r}) = \rho_0$, Eq. (1) reduces to the Helmholtz equation that describes the conventional DT imaging problem.

The transmitted pressure field can be expressed as

$$P(\vec{r},\omega) = P_{\text{in}}(\vec{r},\omega)e^{\psi(\vec{r},\omega)}, \tag{4}$$

where $\psi(\vec{r},\omega)$ is a complex-valued phase function. As in most treatments of DT, we will assume that the object is weakly scattering in the sense that the first-Rytov approximation (e.g., see Ref. 32, Chap. 13) to the wave equation is valid.[33] However, as discussed in Sec. VI, the resulting algorithms possess characteristics that will render them well-suited for extension to the non-linear reconstruction problem.[34–36] Without loss of generality, $\psi(\vec{r},\omega)$ can be expressed as

$$\psi(\vec{r},\omega) = e^{-ik\hat{s}_{0,r}\cdot\vec{r}}F(\vec{r},\omega), \tag{5}$$

where $F(\vec{r},\omega)$ is some function on $\mathbb{R}^3$. On substitution from Eqs. (4) and (5) into Eq. (1) and linearizing the resulting equation by use of the first-Rytov approximation, one obtains

$$(\nabla^2 + k^2)F(\vec{r},\omega) = \{-k^2[\gamma_\kappa(\vec{r},\omega) + \gamma_\rho(\vec{r})] + ik\hat{s}_{0,r}\cdot\nabla\gamma_\rho(\vec{r})\}e^{ik\hat{s}_{0,r}\cdot\vec{r}}. \quad (6)$$

The solution of this equation yields

$$\psi(\vec{r},\omega) = -\frac{1}{4\pi}\int_V \{-k^2[\gamma_\kappa(\vec{r}',\omega) + \gamma_\rho(\vec{r}')] + ik\hat{s}_{0,r}\cdot\nabla'\gamma_\rho(\vec{r}')\}e^{-ik\hat{s}_{0,r}\cdot(\vec{r}-\vec{r}')}G(\vec{r},\vec{r}')d\vec{r}', \quad (7)$$

where ∇' denotes the gradient operator acting on the primed coordinates and $G(\vec{r},\vec{r}') = e^{ik|r'-r'|}/|\vec{r}-\vec{r}'|$ is the 3D free space Green's function. The goal of variable density DT, which we address in this article, is to reconstruct separately both $\gamma_\kappa(\vec{r},\omega)$ and $\gamma_\rho(\vec{r})$ via inversion of Eq. (7). This will require knowledge of $P(\vec{r},\omega)$, or equivalently, $\psi(\vec{r},\omega)$, on detector planes at a collection of tomographic view angles for two temporal frequencies ω.

B. Fourier diffraction projection theorem for acoustic media

Let $\psi(x,y_r,\phi;z_r,\omega)$ denote $\psi(\vec{r},\omega)$ evaluated on a detector plane of constant z_r at tomographic view angle ϕ. The function

$$\hat{\psi}(u,v_r,\phi;z_r,\omega) = \int\int_{\mathbb{R}^2}\psi(x,y_r,\phi;z_r,\omega)e^{-i(ux+v_r y_r)}dxdy_r \quad (8)$$

will denote the 2D Fourier transform of $\psi(x,y_r,\phi;z_r,\omega)$ on the detector plane, while

$$\hat{\gamma}_\kappa(\vec{\nu},\omega) = \int_{\mathbb{R}^3}\gamma_\kappa(\vec{r},\omega)e^{-i\vec{\nu}\cdot\vec{r}}d\vec{r} \quad (9)$$

and

$$\hat{\gamma}_\rho(\vec{\nu}) = \int_{\mathbb{R}^3}\gamma_\rho(\vec{r})e^{-i\vec{\nu}\cdot\vec{r}}d\vec{r} \quad (10)$$

define the 3D Fourier transforms of $\gamma_\kappa(\vec{r},\omega)$ and $\gamma_\rho(\vec{r})$, respectively. Note that, according to these definitions, the frequency variables u and v_r are the Fourier conjugates to the detector plane coordinates x and y_r, respectively, and the vector $\vec{\nu}=(u,v_r,\nu_z)$ describes a point in the 3D Fourier space that is conjugate to $\vec{r}=(x,y_r,z_r)$.

A Fourier diffraction projection (FDP) theorem for 2D acoustic media has previously been derived.[26] That result can be extended readily to the 3D case as

$$\hat{\psi}(u,v_r,\phi;z_r,\omega) = \frac{ik^2 e^{im(u,v_r,k)z_r}}{2\sqrt{k^2-u^2-v_r^2}}\left\{\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k),\phi;\omega] - \left[\frac{u^2+v_r^2+m(u,v_r,k)^2}{2k^2}-1\right]\times\hat{\gamma}_\rho[u,v_r,m(u,v_r,k),\phi]\right\}, \quad (11)$$

where

$$\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k),\phi;\omega] \equiv \hat{\gamma}_\kappa(u\hat{s}_1 + v_r\hat{s}_{2,r}(\phi) + m(u,v_r,k)\hat{s}_{0,r}(\phi),\omega), \quad (12)$$

$$\hat{\gamma}_\rho[u,v_r,m(u,v_r,k),\phi] \equiv \hat{\gamma}_\rho(u\hat{s}_1 + v_r\hat{s}_{2,r}(\phi) + m(u,v_r,k)\hat{s}_{0,r}(\phi)), \quad (13)$$

and

$$m(u,v_r,k) = \sqrt{k^2-u^2-v_r^2} - k. \quad (14)$$

Throughout this article, it will be assumed that the spatial frequencies u and v_r satisfy

$$u^2 + v_r^2 \le k^2. \quad (15)$$

The condition stated in Eq. (15) reflects that we do not attempt to utilize information carried by evanescent wavemodes, which we assume are not measurable, thereby ensuring that $m(u,v_r,k)$ remains real-valued. It should be noted that the points $\vec{\nu}=(u,v_r,m(u,v_r,k))$ define the surface of a shifted hemi-sphere (Ewald sphere) of radius k in the Fourier space. Equation (11) therefore relates the 2D Fourier transform of the phase function on a detector plane at view angle ϕ to a linear combination of the 3D Fourier transforms of $\gamma_\kappa(\vec{r},\omega)$ and $\gamma_\rho(\vec{r})$ evaluated on a hemi-sphere of radius k that is centered at $(-k,0,0)$ in the corresponding rotated coordinate system. Equation (11), which we will refer to as the acoustic FDP theorem, provides a starting point for the development of image reconstruction methods [26,29,37] in variable density acoustic DT, including the ones presented in this article. Equation (11) demonstrates that, in general, the Fourier components of $\gamma_\kappa(\vec{r},\omega)$ and $\gamma_\rho(\vec{r})$ cannot be independently reconstructed from knowledge of $\hat{\psi}(u,v_r,\phi;z_r,\omega)$ acquired by use of an insonifying wave with a fixed temporal frequency ω. However, this difficulty can be circumvented [26] if knowledge of $\hat{\psi}(u,v_r,\phi;z_r,\omega)$ is obtained at an additional distinct temporal frequency.

III. FOURIER SPACE SYMMETRIES FOR QUASI-NONDISPERSIVE OBJECTS

A. Quasi-nondispersive objects

We define a *quasi-nondispersive* object as one that is characterized by a compressibility distribution $\kappa(\vec{r},\omega)$ that is approximately independent of temporal frequency ω over a certain finite closed interval. If this temporal frequency interval is chosen short enough, a wide-class of acoustic objects can satisfy this definition. Because our imaging method does not involve the use of a broadband insonifying source, we do not require the more restrictive assumption that the object is strictly nondispersive.

Consider that at each tomographic view angle ϕ the object is insonified, in turn, by use of two probing wavefields $P_{\text{in}}(\vec{r},\omega_1)$ and $P_{\text{in}}(\vec{r},\omega_2)$, where $\omega_1<\omega_2$. This results in determination of the phase functions $\psi(x,y_r,\phi;d,\omega_1)$ and $\psi(x,y_r,\phi;d,\omega_2)$, respectively, on the detector plane $z_r=d$. We assume that the frequencies ω_1 and ω_2 reside within a quasi-nondispersive temporal frequency interval of the object, and therefore the compressibility distribution can be regarded as temporal frequency-independent, i.e.,

$$\gamma_\kappa(\vec{r}) \equiv \gamma_\kappa(\vec{r},\omega_1) = \gamma_\kappa(\vec{r},\omega_2). \tag{16}$$

Because they represent the 3D Fourier transform of the *same* ω-independent function $\gamma_\kappa(\vec{r})$ evaluated on two shifted hemispheres of radii $k_1=\omega_1/C_0$ and $k_2=\omega_2/C_0$, the ω-dependence of functions $\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi;\omega_1]$ and $\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_2),\phi;\omega_2]$, respectively, will be omitted hereafter.

According to Eq. (11), the phase measurements satisfy

$$\hat{\psi}(u,v_r,\phi;d,\omega_1) = \frac{ik_1^2 e^{im(u,v_r,k_1)d}}{2\sqrt{k_1^2-u^2-v_r^2}}\left\{\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi] - \left[\frac{u^2+v_r^2+m(u,v_r,k_1)^2}{2k_1^2}-1\right]\times\hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi]\right\} \tag{17}$$

and

$$\hat{\psi}(u,v_r,\phi;d,\omega_2) = \frac{ik_2^2 e^{im(u,v_r,k_2)d}}{2\sqrt{k_2^2-u^2-v_r^2}}\left\{\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_2),\phi] - \left[\frac{u^2+v_r^2+m(u,v_r,k_2)^2}{2k_2^2}-1\right]\times\hat{\gamma}_\rho[u,v_r,m(u,v_r,k_2),\phi]\right\}. \tag{18}$$

Equations (17) and (18) represent a system of two equations with four unknowns; the unknowns are the 3D Fourier components of $\gamma_\kappa(\vec{r})$ and $\gamma_\rho(\vec{r})$ that reside on the surfaces of shifted Ewald hemi-spheres of radii k_1 [Eq. (17)] and k_2 [Eq. (18)]. In Sec. III B, we describe Fourier space symmetries that will subsequently facilitate the transformation of this 2×4 system into a 2×2 one that can be solved uniquely.

B. Fourier space symmetries

It will prove useful to explicitly identify symmetries that relate the Fourier components of the object properties that reside on the different Ewald hemi-spheres. It can be verified that

$$\hat{\gamma}_\kappa[u,v_r',m(u,v_r',k_2),\phi] = \hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi+\phi'] \tag{19}$$

and

$$\hat{\gamma}_\rho[u,v_r',m(u,v_r',k_2),\phi] = \hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi+\phi'], \tag{20}$$

where

$$v_r' = \mathrm{sgn}(v_r)\left[R^2-\frac{(R^2+u^2)^2}{4k_2^2}\right]^{1/2}, \tag{21}$$

with $\mathrm{sgn}(v_r)\equiv1$ for $v_r\geq0$ and -1 otherwise, and

$$R^2 \equiv v_r^2 + m(u,v_r,k_1)^2. \tag{22}$$

The angle ϕ' is defined as

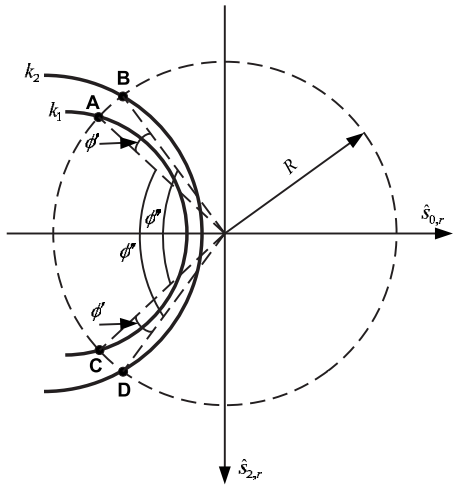


FIG. 2. The intersection of plane of constant u with the semi-Ewald k_1-surfaces and k_2-surfaces. This illustration is used to describe the Fourier space symmetry properties in Sec. III B.

$$\phi' = -\,a\tan\left(\frac{v_r'}{m(u,v_r,k_1)}\right) + a\tan\left(\frac{v_r}{m(u,v_r',k_2)}\right) \tag{23}$$

and satisfies $\phi'(-v_r)=-\phi'(v_r)$. It is also true that v_r' satisfies

$$v_r'^2 + m(u,v_r',k_2)^2 = R^2. \tag{24}$$

Equations (19) and (20) state simply that the vectors $\vec{\nu}_1 = u\hat{s}_1(\phi)+v_r\hat{s}_{2,r}(\phi+\phi')+m(u,v_r,k_1)\hat{s}_{0,r}(\phi+\phi')$ and $\vec{\nu}_2 = u\hat{s}_1(\phi)+v_r'\hat{s}_{2,r}(\phi)+m(u,v_r',k_2)\hat{s}_{0,r}(\phi)$ specify the same location in 3D Fourier space. Let the Ewald surfaces of radii k_1 and k_2 centered at $(-k_1,0,0)$ and $(-k_2,0,0)$ in the rotated coordinate system be denoted as the k_1-surface and k_2-surface, respectively. The symmetries described by Eqs. (19) and (20) can also be understood by consideration of Fig. 2, which displays the intersection of an arbitrary plane of constant u in 3D Fourier space with the k_1- and k_2-surfaces. Equations (19) and (20) indicate that by rotating the k_1-surface about the u-axis (the direction $\hat{s}_1$), it can be made to intersect the k_2-surface at prescribed locations. For example, point "B" on the k_2-surface can be made to overlap with point "A" on the k_1-surface by rotating the k_1-surface by the angle $|\phi'|$ in the clockwise direction. Similarly, point "C" on the k_1-surface will coincide with point "D" on the k_2-surface when the k_1-surface is rotated by the angle $|\phi'|$ in the counterclockwise direction.

A second set of Fourier space symmetries can also be formulated. It can be verified that

$$\hat{\gamma}_\kappa[u,-v_r',m(u,v_r',k_2),\phi] = \hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi-\phi''] \tag{25}$$

and

$$\hat{\gamma}_\rho[u,-v_r',m(u,v_r',k_2),\phi] = \hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi-\phi''], \tag{26}$$

where

$$\phi'' = -a\tan\left(\frac{v_r'}{m(u,v_r',k_2)}\right) - a\tan\left(\frac{v_r}{m(u,v_r,k_1)}\right). \quad (27)$$

The angle ϕ'' is a function of u and v_r and satisfies $\phi''(-v_r) = -\phi''(v_r)$. The symmetries in Eqs. (25) and (26) reflect that the vectors $\vec{v}_1' = u\hat{s}_1 + v_r\hat{s}_{2,r}(\phi - \phi'') + m(u,v_r,k_1) \times \hat{s}_{0,r}(\phi - \phi'')$ and $\vec{v}_2' = u\hat{s}_1 - v_r'\hat{s}_{2,r}(\phi) + m(u,v_r',k_2)\hat{s}_{0,r}(\phi)$ describe the same location in 3D Fourier space. Again, the symmetries can be visualized by reference to Fig. 2. Equations (25) and (26) state that point B on the k_2-surface can be made to coincide with point C on the k_1-surface by rotating the k_1-surface by an angle $|\phi''|$ in the clockwise direction about the u-axis. Similarly, point D on the k_2-surface can be made to coincide with point A on the k_1-surface by rotating the k_1-surface by an angle $|\phi''|$ in a counterclockwise direction about the u-axis.

IV. RECONSTRUCTION METHODS

Below we derive two novel reconstruction methods for variable density DT that can reconstruct separately the Fourier components $\hat{\gamma}_\kappa[\cdot]$ and $\hat{\gamma}_\rho[\cdot]$. This task is facilitated by use of the symmetries identified in Sec. III. From knowledge of these Fourier data, conventional DT reconstruction methods[6,38] can be employed for reconstruction of images that depict $\gamma_\kappa(\vec{r})$ and $\gamma_\rho(\vec{r})$.

For convenience, we define a modified data function as

$$\hat{D}_l(u,v_r;\phi) \equiv -2ik_l^{-2}e^{-im(u,v_r,k_l)d}(k_l^2 - u^2 - v_r^2)^{1/2} \times \hat{\psi}(u,v_r,\phi;d,\omega_l) \quad \text{for } l = 1,2. \quad (28)$$

Here, the index l specifies the choice of temporal frequency ω_l, or equivalently wavenumber k_l, of the insonifying plane-wave. It should be noted that $\hat{D}_l(u,v_r;\phi)$ is determined completely from knowledge of the complex phase function $\psi(x,y_r,\phi;d,\omega_l)$, and can therefore be regarded as a measurable quantity. In terms of the modified data function, Eqs. (17) and (18) can be re-expressed as

$$\hat{D}_1(u,v_r;\phi) = \hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi] - \left[\frac{u^2+v_r^2+m(u,v_r,k_1)^2}{2k_1^2} - 1\right] \times \hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi] \quad (29)$$

and

$$\hat{D}_2(u,v_r;\phi) = \hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_2),\phi] - \left[\frac{u^2+v_r^2+m(u,v_r,k_2)^2}{2k_2^2} - 1\right] \times \hat{\gamma}_\rho[u,v_r,m(u,v_r,k_2),\phi]. \quad (30)$$

A. Reconstruction method 1

The first reconstruction method is developed by use of the symmetries described by Eqs. (19) and (20). The modified data function $\hat{D}_2(u,v_r;\phi)$ in Eq. (30) when evaluated at the frequencies (u,v_r') yields

$$\hat{D}_2(u,v_r';\phi) = \hat{\gamma}_\kappa[u,v_r',m(u,v_r',k_2),\phi] - \left[\frac{u^2+R^2}{2k_2^2} - 1\right]\hat{\gamma}_\rho[u,v_r',m(u,v_r',k_2),\phi], \quad (31)$$

where v_r' and R^2 were defined in Eqs. (21) and (24), respectively. On substitution from Eqs. (19) and (20) into Eq. (31), one obtains

$$\hat{D}_2(u,v_r';\phi) = \hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi+\phi'] - \left[\frac{u^2+R^2}{2k_2^2} - 1\right] \times \hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi+\phi']. \quad (32)$$

Equations (29) and (32) constitute a system of equations that relate the two phase measurements on the detector plane $z_r = d$ at view angle ϕ to two distinct frequency components of γ_κ and two distinct frequency components of γ_ρ; i.e., they represent a system of two equations with four unknowns. It should be noted that all of the unknown frequency components reside on Ewald surfaces that have a common radius k_1. However, the Fourier components that are present in Eq. (32) reside on Ewald surfaces whose orientations have been offset by the angular shift ϕ'. To circumvent this fact, we can exploit the rotational invariance of the problem.

Because $\hat{D}_l(u,v_r;\phi)$, with $l = 1,2$,

$$\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi], \quad \text{and} \quad \hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi]$$

are 2π-periodic functions of ϕ, they can be expressed as the Fourier series

$$\hat{D}_l(u,v_r;\phi) = \sum_{n=-\infty}^{\infty} \hat{D}_{l,n}(u,v_r)e^{in\phi}, \quad (33)$$

$$\hat{\gamma}_k[u,v_r,m(u,v_r,k_1),\phi] = \sum_{n=-\infty}^{\infty} \hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)]e^{in\phi}, \quad (34)$$

and

$$\hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi] = \sum_{n=-\infty}^{\infty} \hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)]e^{in\phi}, \quad (35)$$

where

$$\hat{D}_{l,n}(u,v_r) = \frac{1}{2\pi}\int_0^{2\pi} \hat{D}_l(u,v_r;\phi)e^{in\phi}d\phi, \quad (36)$$

$$\hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)] = \frac{1}{2\pi}\int_0^{2\pi} \hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi] \times e^{in\phi}d\phi, \quad (37)$$

and

$$\hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)] = \frac{1}{2\pi}\int_0^{2\pi} \hat{\gamma}_{\rho}[u,v_r,m(u,v_r,k_1),\phi] \times e^{in\phi} d\phi. \quad (38)$$

On substituting Eqs. (33)–(35) into Eqs. (29) and (32), one obtains

$$\hat{D}_{1,n}(u,v_r) = \hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)] - \left[\frac{u^2+R^2}{2k_1^2} - 1\right]\hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)] \quad (39)$$

and

$$\hat{D}_{2,n}(u,v_r') = \hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)]e^{in\phi'} - \left[\frac{u^2+R^2}{2k_2^2} - 1\right]\hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)]e^{in\phi'}, \quad (40)$$

respectively. Equations (39) and (40) represent a system of two equations with two unknowns, where the unknowns are the Fourier series expansion coefficients of the desired Fourier components. The solution of this system yields

$$\hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)] = \frac{k_2^2(2k_1^2-R^2-u^2)\hat{D}_{2,n}(u,v_r')e^{-in\phi'} - k_1^2(2k_2^2-R^2-u^2)\hat{D}_{1,n}(u,v_r)}{(k_1^2-k_2^2)(R^2+u^2)}, \quad (41)$$

and

$$\hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)] = \frac{2[\hat{D}_{1,n}(u,v_r) - \hat{D}_{2,n}(u,v_r')e^{-in\phi'}]k_1^2k_2^2}{(k_1^2-k_2^2)(R^2+u^2)}. \quad (42)$$

Equation (41) coupled with Eq. (34) provides the desired method for reconstruction of the Fourier data $\hat{\gamma}_{\kappa}[u,v_r,m(u,v_r,k_1),\phi]$ for $u^2+v_r^2 \le k_1^2$. Similarly Eqs. (42) and (35) permit reconstruction of the Fourier data $\hat{\gamma}_{\rho}[u,v_r,m(u,v_r,k_1),\phi]$ for $u^2+v_r^2 \le k_1^2$. From knowledge of the Fourier data $\hat{\gamma}_{\kappa}[u,v_r,m(u,v_r,k_1),\phi]$ and $\hat{\gamma}_{\rho}[u,v_r,m(u,v_r,k_1),\phi]$, separate estimates of $\gamma_{\kappa}(\vec{r})$ and $\gamma_{\rho}(\vec{r})$ can be obtained by use of conventional DT reconstruction algorithms.[2,6,38]

B. Reconstruction method 2

By use of the symmetry properties described in Eqs. (25) and (26), an alternative method for reconstruction of $\gamma_{\kappa}(\vec{r})$ and $\gamma_{\rho}(\vec{r})$ can be derived. Consider the modified data function $\hat{D}_2(u,v_r;\phi)$, defined in Eq. (30), evaluated at frequencies $(u,-v_r')$:

$$\hat{D}_2(u,-v_r';\phi) = \hat{\gamma}_{\kappa}[u,-v_r',m(u,v_r',k_2),\phi] - \left[\frac{u^2+R^2}{2k_2^2} - 1\right]\hat{\gamma}_{\rho}[u,-v_r',m(u,v_r',k_2),\phi], \quad (43)$$

where R^2 is defined in Eq. (24). On substitution from Eqs. (25) and (26) into Eq. (43), one obtains

$$\hat{D}_2(u,-v_r';\phi) = \hat{\gamma}_{\kappa}[u,v_r,m(u,v_r,k_1),\phi-\phi''] - \left[\frac{u^2+R^2}{2k_2^2} - 1\right] \times \hat{\gamma}_{\rho}[u,v_r,m(u,v_r,k_1),\phi-\phi'']. \quad (44)$$

Following the same procedure outline in Sec. IV A, computation of the Fourier series expansion of both sides of Eq. (44) indicates that

$$\hat{D}_{2,n}(u,-v_r') = \hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)]e^{-in\phi''} - \left[\frac{u^2+R^2}{2k_2^2} - 1\right] \times \hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)]e^{-in\phi''}, \quad (45)$$

where $\hat{D}_{2,n}(u,-v_r')$ is defined in Eq. (36). Equations (45) and (39) constitute a system of two equations with two unknowns that can be solved as

$$\hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)] = \frac{k_2^2(2k_1^2-R^2-u^2)\hat{D}_{2,n}(u,-v_r')e^{in\phi''} - k_1^2(2k_2^2-R^2-u^2)\hat{D}_{1,n}(u,v_r)}{(k_1^2-k_2^2)(R^2+u^2)} \quad (46)$$

and

$$\hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)] = \frac{2[\hat{D}_{1,n}(u,v_r) - \hat{D}_{2,n}(u,-v_r')e^{in\phi''}]k_1^2k_2^2}{(k_1^2-k_2^2)(R^2+u^2)}. \tag{47}$$

Equation (46), together with Eq. (34), provides an alternative method for reconstruction of the Fourier data $\hat{\gamma}_\kappa[u,v_r,m(u,v_r,k_1),\phi]$ for $u^2+v_r^2 \leq k_1^2$. Similarly, Eqs. (47) and (35) permit reconstruction of the Fourier data $\hat{\gamma}_\rho[u,v_r,m(u,v_r,k_1),\phi]$ for $u^2+v_r^2 \leq k_1^2$. From knowledge of these Fourier data, conventional DT reconstruction methods[2,6,38] can be employed for reconstruction of estimates of $\gamma_\kappa(\vec{r})$ and $\gamma_\rho(\vec{r})$.

C. Features of the reconstruction formulas

1. Consistency conditions

In a mathematical sense, the two reconstruction methods developed in Secs. IV A and IV B are identical. In other words, if applied to consistent measurement data that are free of noise, finite sampling effects, or other errors, they would produce identical reconstructions. Such data must satisfy the consistency conditions [39]

$$\hat{D}_{2,n}(u,-v_r') = \hat{D}_{2,n}(u,v_r')e^{-in[\phi'(v_r)+\phi''(v_r)]}, \tag{48}$$

where we have made the v_r-dependence of ϕ' and ϕ'' explicit. These consistency conditions can be interpreted graphically by considering again the Fourier space illustration in Fig. 2. Equation (48) states simply that point D on the k_2-surface and point B on a version of the k_2-surface that has been rotated by the angle $|\phi'(v_r)+\phi''(v_r)|$ counterclockwise define the same location in 3D Fourier space. One can verify that substitution from Eq. (48) into Eqs. (46) and (47) yields, Eqs. (41) and (42), respectively. However, when Eq. (48) is not satisfied, this mathematical equivalence of the two reconstruction methods does not hold true.

2. Exploitation of statistically complementary information

In the presence of data noise, reconstruction methods 1 and 2 will produce distinct estimates of $\gamma_\kappa(\vec{r})$ and $\gamma_\rho(\vec{r})$. A natural strategy for exploiting statistically complementary information inherent in the two estimates of each object property is to linearly combine them. Let $\hat{\gamma}_{\kappa,n}^{(1)}[u,v_r,m(u,v_r,k_1)]$ and $\hat{\gamma}_{\kappa,n}^{(2)}[u,v_r,m(u,v_r,k_1)]$ denote the estimates of $\hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)]$ produced by use of Eqs. (41) and (46), respectively. Similarly, let $\hat{\gamma}_{\rho,n}^{(1)}[u,v_r,m(u,v_r,k_1)]$ and $\hat{\gamma}_{\rho,n}^{(2)}[u,v_r,m(u,v_r,k_1)]$ denote the estimates $\hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)]$ produced by use of Eqs. (42) and (47), respectively. Each set of estimates can be combined linearly as

$$\hat{\gamma}_{\kappa,n}^{[\alpha]}[u,v_r,m(u,v_r,k_1)] = \alpha\hat{\gamma}_{\kappa,n}^{(1)}[u,v_r,m(u,v_r,k_1)] + (1-\alpha)\hat{\gamma}_{\kappa,n}^{(2)}[u,v_r,m(u,v_r,k_1)] \tag{49}$$

and

$$\hat{\gamma}_{\rho,n}^{[\beta]}[u,v_r,m(u,v_r,k_1)] = \beta\hat{\gamma}_{\rho,n}^{(1)}[u,v_r,m(u,v_r,k_1)] + (1-\beta)\hat{\gamma}_{\rho,n}^{(2)}[u,v_r,m(u,v_r,k_1)], \tag{50}$$

where $\hat{\gamma}_{\kappa,n}^{[\alpha]}[u,v_r,m(u,v_r,k_1)]$ and $\hat{\gamma}_{\rho,n}^{[\beta]}[u,v_r,m(u,v_r,k_1)]$ are the final estimates of $\hat{\gamma}_{\kappa,n}[u,v_r,m(u,v_r,k_1)]$ and $\hat{\gamma}_{\rho,n}[u,v_r,m(u,v_r,k_1)]$, respectively. The combination coefficients $\alpha=\alpha(u,v_r)$ and $\beta=\beta(u,v_r)$ can be complex-valued and, in principle, can be chosen to minimize the variance of the final estimate.[5] A detailed analysis of the statistical properties of this reconstruction strategy remains a topic for future study.

3. Singularities and noise amplification

Let the denominators of the reconstruction formulas in Eqs. (41), (42), (46), and (47) be denoted by

$$M \equiv (k_1^2-k_2^2)(R^2+u^2) = (k_1^2-k_2^2)[v_r^2 + m(u,v_r,k_1)^2 + u^2]. \tag{51}$$

It can be verified readily that $M=0$ when $u=0$ and $v_r=0$, indicating that all of the reconstruction formulas contain a pole at the origin of Fourier space. It can also be confirmed that this is the only singularity that the reconstruction formulas possess within the domain $u^2+v_r^2<k_1^2$. This indicates that it can be difficult to reconstruct accurately certain low-frequency components of $\gamma_\kappa(\vec{r})$ and $\gamma_\rho(\vec{r})$ from noisy measurement data. This behavior is intrinsic to the variable density DT problem and has been observed in other studies.[26,29] In practice, regularization methods can be employed [40,41] to mitigate the effects of the poles in the reconstruction formulas. Equation (51) also indicates that $M \to 0$ as $k_1 \to k_2$. Accordingly, it is desirable to make $|k_1-k_2|$ sufficiently large, without violating the quasi-nondispersive object assumption, to improve the numerical stability of the problem.

V. NUMERICAL RESULTS

We performed computer-simulation studies to validate and demonstrate the 3D variable density acoustic DT reconstruction methods.

A. Simulation studies

Scanning geometry: The 3D scanning geometry shown in Fig. 1 was employed in our simulation studies. The insonifying plane-wave propagates in the positive z_r direction, and the forward scattered pressure field is recorded on a 2D detector array placed at a distance $d=0.60$ m behind the object. The detector array had a dimension of 0.64×0.64 m^2 and contained 256×256 discrete elements. This corresponded to a detector pixel size of 0.0025×0.0025 m^2. At each view angle ϕ, two plane-waves of wavelengths $\lambda_1=0.002$ m and $\lambda_2=0.001$ m, or equivalently, wavenumbers $k_1=1000\pi$ m^{-1} and $k_2=2000\pi$ m^{-1}, were used to insonify the object. The angular scanning consisted of sampling ϕ at 360 views that were evenly spaced over the interval $[0,2\pi)$.

Phantom object: A 3D mathematical phantom was utilized to represent the object to be imaged. The phantom was comprised of a uniform sphere containing two uniform ellip-

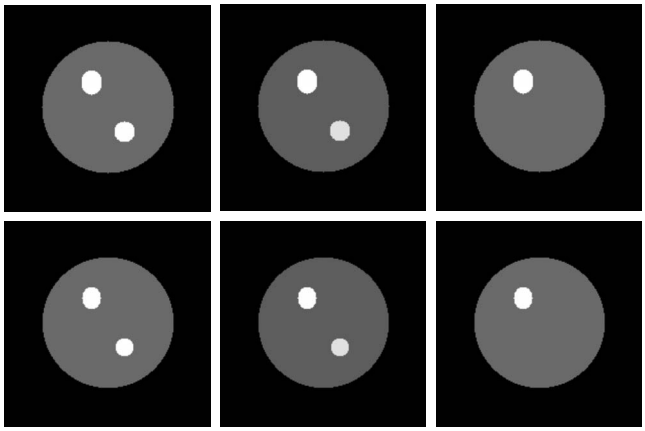


FIG. 3. The 3D mathematical phantom that was used in the simulation studies. The top row, from left to right, displays images depicting the $z=0$ slice of $\mathrm{Re}\{\gamma_\kappa(\vec{r})\}$, $\mathrm{Im}\{\gamma_\kappa(\vec{r})\}$, and $\gamma_\rho(\vec{r})$, respectively. The corresponding images of the $z=0.0075$ m slice are shown in the bottom row.

soids. The sphere was centered at (0, 0, 0), had a radius of 0.08 m, and was assigned the values $\gamma_\kappa(\vec{r})=0.01+i0.001$ and $\gamma_\rho(\vec{r})=0.01$. The imaginary component of $\gamma_\kappa(\vec{r})$ was included to incorporate the effects of attenuation. This model can be understood by noting that $\gamma_\kappa(\vec{r})$ can be expressed as

$$\gamma_\kappa(\vec{r}) = \frac{C_0^2}{C^2(\vec{r})} - 1, \tag{52}$$

where the object's speed of sound $C(\vec{r})$ can be defined as complex-valued in the presence of attenuation. One ellipsoid was centered at $(-0.02,0.03,0)$, had semiaxes of 0.0125, 0.015, and 0.0175 m, and was assigned the values $\gamma_\kappa(\vec{r})=0.02+i0.0022$ and $\gamma_\rho(\vec{r})=0.02$. The second ellipsoid was centered at $(0.02, -0.03,0)$, had semiaxes of 0.0125, 0.0125, and 0.015 m, and was assigned the values $\gamma_\kappa(\vec{r})=0.02+i0.002$ and $\gamma_\rho(\vec{r})=0.01$. Two-dimensional slices through the phantom object are displayed in Fig. 3. The top row of images, from the left to the right, contains images of $\mathrm{Re}\{\gamma_\kappa(\vec{r})\}$, $\mathrm{Im}\{\gamma_\kappa(\vec{r})\}$, and $\gamma_\rho(\vec{r})$, respectively, corresponding to the slice $z=0$. Here, $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary components, respectively, of a complex-valued quantity. The images corresponding to the slice $z=0.0075$ m are shown in the bottom row of Fig. 3.

Simulation data: At each view angle ϕ, the 2D Fourier transformed phase functions $\hat{\psi}(u,v_r,\phi;z_r,\omega_1)$ and $\hat{\psi}(u,v_r,\phi;z_r,\omega_2)$ were calculated analytically from knowledge of $\hat{\kappa}_\kappa[u,v_r,m(u,v_r,k),\phi]$ and $\hat{\gamma}_\rho[u,v_r,m(u,v_r,k),\phi]$, respectively, by use of Eq. (11). It was assumed that ω_1 and ω_2 resided within a quasi-nondispersive interval of the object so that the ω-dependence of $\gamma_\kappa(\vec{r})$ could be neglected. Subsequently, $\psi(x,y,\phi;z_r=d,\omega_1)$ and $\psi(x,y,\phi;z_r=d,\omega_2)$ were calculated by use of the 2D inverse fast Fourier transform (FFT) algorithm, and the total wavefields $P(x,y_r,\phi;\omega_1)$ and $P(x,y_r,\phi;\omega_2)$ on the detector planes were determined by use of Eq. (4). Note that the generation of the simulation data in this way presumes the validity of the first-Rytov weak scattering condition. The deleterious effects of strong scattering were not considered in our simulations.

Let $P(x,y_r,\phi;\omega_l)=P_R(x,y_r,\phi;\omega_l)+iP_I(x,y_r,\phi;\omega_l)$, $l=1,2$, denote the pressure wave-field complex amplitude measured on the detector planes. To model the effects of measurement noise and other data inconsistencies, we treated $P(x,y_r,\phi;\omega_l)$ as an uncorrelated circular Gaussian stochastic process.[42] For each frequency ω_l, this stochastic process was defined by the mean and standard deviation of $P_R(x,y_r,\phi;\omega_l)$, denoted by μ_R and σ_R, respectively, and by the mean and standard deviations of $P_I(x,y_r,\phi;\omega_l)$, denoted by μ_I and σ_I, respectively. In our simulations, the means μ_R and μ_I were set equal to the simulated noiseless data $P_R(x,y_r,\phi;\omega_l)$ and $P_I(x,y_r,\phi;\omega_l)$, respectively, at each detector location. The standard deviations were set according to the rules $\sigma_R/\mu_R=\sigma_I/\mu_I=0.1\%$. Noisy versions of the measured pressure wave-field data $P(x,y_r,\phi;\omega_l)$, $l=1,2$, were created by sampling realizations of this stochastic process.

B. Reconstruction procedure

Reconstruction method 1 described in Sec. IV A was implemented numerically as follows. From the noiseless and noisy versions of $P(x,y_r,\phi;\omega_l)$, the complex phases were determined via Eq. (4) and Fourier transformed by use of the 2D FFT algorithm. The modified data functions $\hat{D}_1(u,v_r;\phi)$ and $\hat{D}_2(u,v_r;\phi)$ were computed subsequently by use of Eq. (28). From the set of uniformly spaced values of (u,v_r) at which $\hat{D}_2(u,v_r;\phi)$ was evaluated, the set of non-uniformly spaced values (u,v_r') defined by Eq. (21) was computed by use of a one-dimensional (1D) linear interpolation operation. Prior to performing the interpolation, the sampling density of the uniformly spaced data $\hat{D}_2(u,v_r;\phi)$ was increased by a factor of 16 via zero-padding in the spatial domain.[4] The Fourier series expansion coefficients $\hat{D}_{1,n}(u,v_r)$ and $\hat{D}_{2,n}(u,v_r)$ were calculated by application of the 1D FFT to the sampled angular coordinate of the data functions. The reconstruction formulas given in Eqs. (41) and (42), along with Eqs. (34) and (35), were then employed for determination of the Fourier data $\hat{\gamma}_\kappa^{(1)}[u,v_r,m(u,v_r,k_1),\phi;\omega]$ and $\hat{\gamma}_\rho^{(1)}[u,v_r,m(u,v_r,k_1),\phi]$, respectively. A simple regularization strategy was utilized to mitigate the effects of the singularities in the formulas located at the origin of Fourier space. Specifically, the frequency components residing within a small sphere of radius ϵ, centered at the origin, were set to 0. In this work, the value of ϵ was set to 0.001.

Reconstruction method 2 described in Sec. IV B was implemented in an analogous fashion. In that case, Eqs. (46) and (47), along with Eqs. (34) and (35), were employed to determine the Fourier data $\hat{\gamma}_\kappa^{(2)}[u,v_r,m(u,v_r,k_1),\phi;\omega]$ and $\hat{\gamma}_\rho^{(2)}[u,v_r,m(u,v_r,k_1),\phi]$, respectively. Finally, from knowledge of $\hat{\gamma}_{\kappa,n}^{(l)}[u,v_r,m(u,v_r,k_1)$ and $\hat{\gamma}_{\rho,n}^{(l)}[u,v_r,m(u,v_r,k_1)]$, $l=1,2$, the DT reconstruction method in Ref. 6 was employed to obtain images of the object properties $\gamma_\kappa(\vec{r})$ and $\gamma_\rho(\vec{r})$. The matrix size of the reconstructed images was $256\times256\times256$, where the voxel size was $0.0025\times0.0025\times0.0025$ m^3.

C. Reconstructed images

The top row of Fig. 4, from the left to the right, contains images depicting the $z=0$ slice of $\mathrm{Re}\{\gamma_\kappa(\vec{r})\}$, $\mathrm{Im}\{\gamma_\kappa(\vec{r})\}$, and

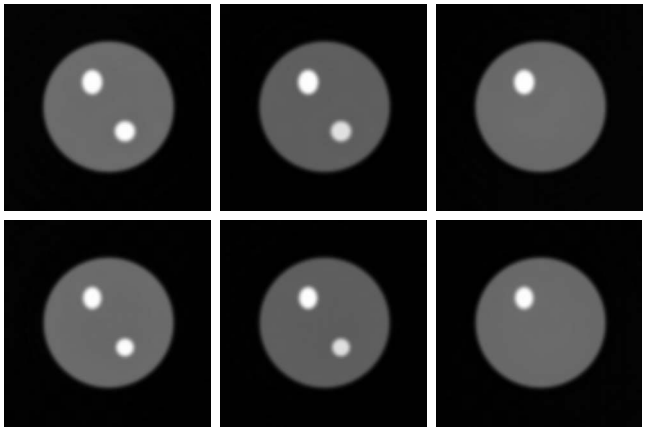


FIG. 4. Images reconstructed from noiseless simulation data by use of reconstruction method 1. The top row, from left to right, displays images depicting the $z=0$ slice of $\mathrm{Re}\{\gamma_\kappa(\vec{r})\}$, $\mathrm{Im}\{\gamma_\kappa(\vec{r})\}$, and $\gamma_\rho(\vec{r})$, respectively. The corresponding images of the $z=0.0075$ m slice are shown in the bottom row.

$\gamma_\rho(\vec{r})$, respectively, reconstructed from the noiseless simulation data by use of reconstruction method 1. The reconstructed images corresponding to the slice $z=0.0075$ m are shown in the bottom row of Fig. 4. Profiles through the upper-left ellipsoid in the images in the top row of Fig. 4 are superimposed on the true object profiles in Fig. 5, which reveals that the noiseless images are reconstructed with high fidelity. Although not shown, the profiles corresponding to the images of the $z=0.0075$ m slice in the bottom row of Fig. 4 were also found to represent the true object profiles accurately. Additionally, the images reconstructed from the noiseless simulation data by use of reconstruction method 2 were found to be virtually identical to those reconstructed by method 1. This is expected because the two reconstruction methods are identical mathematically and should therefore produce identical images from consistent measurement data in the absence of finite sampling effects.

The top row of Fig. 6, from the left to the right, contains images depicting the $z=0$ slice of $\mathrm{Re}\{\gamma_\kappa(\vec{r})\}$, $\mathrm{Im}\{\gamma_\kappa(\vec{r})\}$, and $\gamma_\rho(\vec{r})$, respectively, reconstructed from the noisy simulation data by use of reconstruction method 1. The images corresponding to the slice $z=0.0075$ m are shown in the bottom row of Fig. 6. The corresponding images reconstructed by use of method 2 are displayed in Fig. 7. As expected, the noisy images reconstructed by use of the two reconstruction methods have similar appearances, but are distinct. Although not shown, this was confirmed by examination of profiles through the noisy images. The texture of the noise patterns in the noisy images reflects that the low-frequency components of the object properties are more susceptible to noise contamination than are the high-frequency ones, which is a result of the singularities in the reconstruction formulas at the origin of Fourier space. We expect that the noise properties of the images can be modified favorably when more sophisticated regularization methods[40,41] are employed to mitigate the effects of the singularities.

VI. SUMMARY

When applied to acoustic media, conventional DT reconstruction algorithms produce a composite image that reflects both the compressibility and density variations in the object. In certain applications, however, it may be useful to reconstruct separate images depicting the compressibility and density variations because knowledge of such individual intrinsic tissue properties may reveal valuable diagnostic information that is not conveyed directly in the composite image. Several algorithms have been developed[26–30] to accomplish such reconstructions under various approximations. In this article, we have revisited the variable density DT reconstruction problem and proposed novel 3D reconstruction methods. Our methods require that a quasi-nondispersive and weakly scattering object is insonified, in turn, by use of plane-waves at two distinct temporal frequencies, and measurements of the transmitted wavefields are recorded at a collection of view angles that span a 2π angular range. We demonstrated that the individual Fourier components of a scattering object's compressibility and density variations can be decoupled algebraically through the identification and ex-

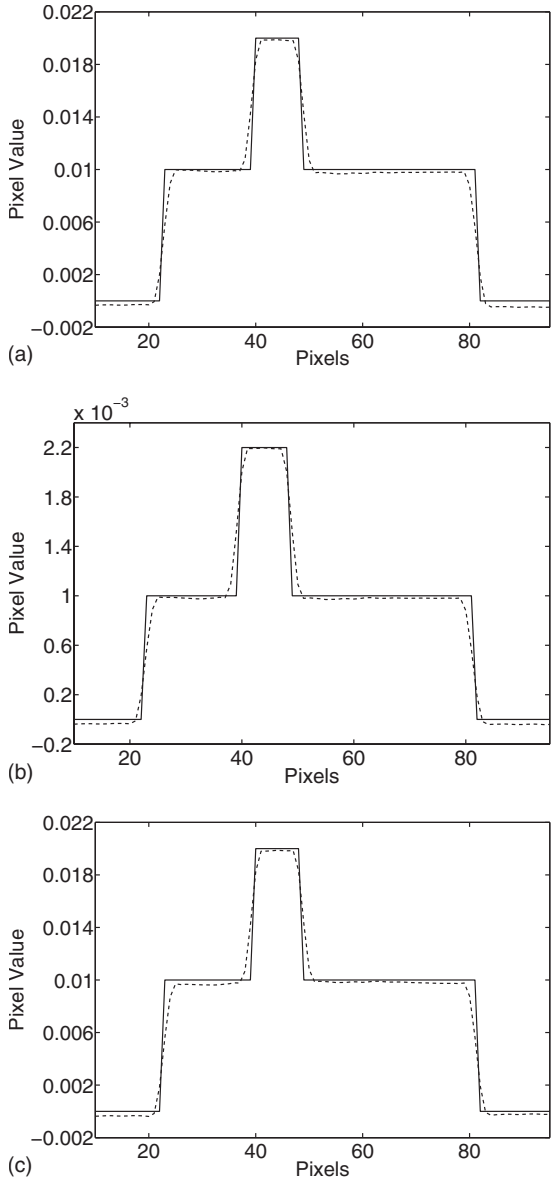


FIG. 5. Profiles through the upper-left ellipsoid in the noiseless images shown in the top row of Fig. 4. The true and reconstructed profiles correspond to the solid and dashed lines, respectively.

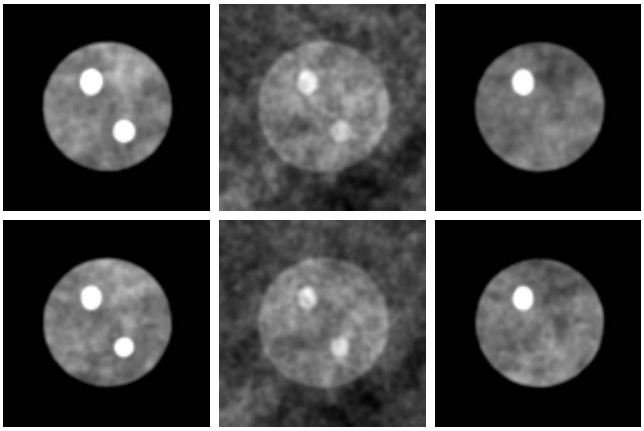


FIG. 6. Images reconstructed from noisy simulation data by use of reconstruction method 1. The noise levels in the pressure data were chosen, as described in Sec. V A. The top row, from left to right, displays images depicting the $z=0$ slice of $\mathrm{Re}\{\gamma_\kappa(\vec{r})\}$, $\mathrm{Im}\{\gamma_\kappa(\vec{r})\}$, and $\gamma_\rho(\vec{r})$, respectively. The corresponding images of the $z=0.0075$ m slice are shown in the bottom row.

ploitation of tomographic symmetries and the rotational invariance of the imaging model. The physical reason that the density and compressibility distributions can be separated in this way is that the Rytov approximation is used to linearize the forward scattering problem. From the decoupled Fourier data, conventional DT reconstruction algorithms can be employed for determination of the two separate object properties.

We expect that this general image reconstruction strategy may be applicable to other linearized inverse scattering problems in acoustics and optics where Fourier space symmetries[21] can be exploited. For example, it may permit the development of image reconstruction algorithms for holographic acoustic tomography in which only the intensity of the transmitted wavefield is recorded.[43,44]

A majority of DT reconstruction algorithms developed to date, including the ones developed in this article, have utilized weak scattering approximations. The first-Rytov weak scattering approximation utilized in this work may not remain valid in many medical imaging applications at diagnostic ultrasound frequencies.[45] However, the developed algorithms represent a fundamentally different and computationally efficient approach to the 3D variable density DT reconstruction problem, which may provide a natural framework for the incorporation of higher-order scattering approximations[34,35] into the algorithms. Because of the simple algebraic nature of the reconstruction formulas, it may also be possible to incorporate into them the effects of frequency-dependent attenuation[44] and phase-velocity. These represent important and interesting topics for future study.

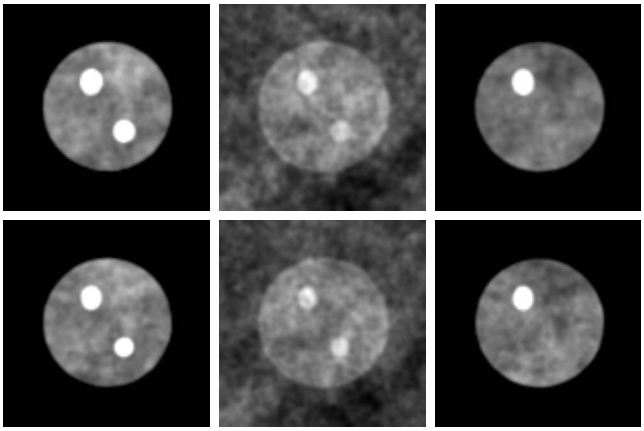


FIG. 7. Images reconstructed from noisy simulation data by use of reconstruction method 2. The noise levels in the pressure data were chosen, as described in Sec. V A. The top row, from left to right, displays images depicting the $z=0$ slice of $\mathrm{Re}\{\gamma_\kappa(\vec{r})\}$, $\mathrm{Im}\{\gamma_\kappa(\vec{r})\}$, and $\gamma_\rho(\vec{r})$, respectively. The corresponding images of the $z=0.0075$ m slice are shown in the bottom row.

ACKNOWLEDGMENTS

The authors thank Mr. Thomas Deffieux for his assistance in their preliminary investigations of the 2D version of the variable density DT reconstruction problem. This research was supported in part by the National Science Foundation CAREER Award No. 0546113.

[1] E. Wolf, "Principles and development of diffraction tomography," in *Trends in Optics*, edited by A. Consortini (Academic, San Diego, CA, 1996), pp. 83–110.
[2] A. J. Devaney, "Reconstructive tomography with diffracting wavefields," Inverse Probl. **2**, 161–183 (1986).
[3] A. J. Devaney, "A filtered backpropagation algorithm for diffraction tomography," Ultrason. Imaging **4**, 336–350 (1982).
[4] S. Pan and A. Kak, "A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation," IEEE Trans. Acoust., Speech, Signal Process. **31**, 1262–1275 (1983).
[5] X. Pan, "Unified reconstruction theory for diffraction tomography, with consideration of noise control," J. Opt. Soc. Am. A **15**, 2312–2326 (1998).
[6] M. A. Anastasio and X. Pan, "Computationally efficient and statistically robust image reconstruction in 3D diffraction tomography," J. Opt. Soc. Am. A **17**, 391–400 (2000).
[7] V. Lauer, "New approach to optical diffraction tomography yielding a vector equation of diffraction tomography and a novel tomographic microscope," J. Microsc. **205**, 165–176 (2001).
[8] P. S. Carney and J. C. Schotland, "Inverse scattering for near-field microscopy," Appl. Phys. Lett. **77**, 2798–2800 (2000).
[9] G. Gbur and E. Wolf, "Hybrid diffraction tomography without phase information," J. Opt. Soc. Am. A **19**, 2194–2202 (2002).
[10] R. Muller, M. Kaveh, and G. Wade, "Reconstructive tomography and applications to ultrasonics," Proc. IEEE **67**, 567–587 (1979).
[11] J. F. Greenleaf, "Computerized tomography with ultrasound," Proc. IEEE **71**, 330–337 (1983).
[12] M. Kaveh, R. Mueller, R. Rylander, T. Coulter, and M. Soumekh, "Experimental results in ultrasonic diffraction tomography," Acoust. Imaging **9**, 433–450 (1979).
[13] S. J. Norton and M. Linzer, "Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical, and spherical apertures," IEEE Trans. Biomed. Eng. **28**, 202–220 (1981).
[14] N. Duric, P. Littrup, E. Holsapple, A. Babkin, R. Duncan, A. Kalinin, R. Pevzner, and M. Tokarev, "Ultrasound tomography of breast tissue," Proc. SPIE **5035**, 24–32 (2003).
[15] R. Leach, S. Azevedo, J. Beryman, H. Bertete-Aquirre, D. Chambers, J. Mast, P. Littrup, N. Duric, S. Johnson, and F. Wuebbeling, "Comparison of ultrasound tomography methods in circular geometry," Proc. SPIE **4687**, 362–377 (2002).
[16] T. D. Mast, "Wideband quantitative ultrasonic imaging by time-domain diffraction tomography," J. Acoust. Soc. Am. **106**, 3061–3071 (1999).
[17] I. Akiyama, N. Takizawa, and A. Ohya, "Narrow aperture phased array computed tomography," Acoust. Imaging **25**, 163–170 (2000).
[18] M. P. Andre, H. S. Janee, P. J. Martin, G. P. Otto, B. A. Spivey, and D. A. Palmer, "High-speed data acquisition in a diffraction tomography system employing large-scale toroidal arrays," Int. J. Imaging Syst. Technol. **8**, 137–147 (1997).
[19] F. Natterer and F. Wubbeling, "A propagation-backpropagation method for ultrasound tomography," Inverse Probl. **11**, 1225–1232 (1995).
[20] N. Sponheim, L. J. Gelius, I. Johansen, and J. J. Stamnes, "Quantitative results in ultrasonic tomography of large object using line sources and

curved detector arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**, 370–379 (1991).
[21]N. Baddour, "Theory and analysis of frequency-domain photoacoustic tomography," J. Acoust. Soc. Am. **123**, 2577–2590 (2008).
[22]M. Xu, Y. Xu, and L. V. Wang, "Time-domain reconstruction algorithms and numerical simulations for thermoacoustic tomography in various geometries," IEEE Trans. Biomed. Eng. **50**, 1086–1099 (2003).
[23]Y. Huang and M. A. Anastasio, "Statistically principled use of in-line measurements in intensity diffraction tomography," J. Opt. Soc. Am. A **24**, 626–642 (2007).
[24]M. A. Anastasio and D. Shi, "On the relationship between intensity diffraction tomography and phase-contrast tomography," Proc. SPIE **5535**, 361–368 (2004).
[25]C.-Y. Chou, Y. Huang, D. Shi, and M. A. Anastasio, "Image reconstruction in quantitative x-ray phase-contrast imaging employing multiple measurements," Opt. Express **15**, 10002–10025 (2007).
[26]A. J. Devaney, "Variable density acoustic tomography," J. Acoust. Soc. Am. **78**, 120–130 (1985).
[27]S. J. Norton, "Generation of separate compressibility and density images in tissue," Ultrason. Imaging **5**, 240–252 (1983).
[28]A. Witten, J. Tuggle, and R. Waag, "A practical approach to ultrasonic imaging using diffraction tomography," J. Acoust. Soc. Am. **83**, 1645–1652 (1988).
[29]M. Moghaddam and W. C. Chew, "Variable density linear acoustic inverse problem," Ultrason. Imaging **15**, 255–266 (1993).
[30]S. Mensah and J. P. Lefebvre, "Enhanced compressibility tomography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1245–1252 (1997).
[31]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[32]M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Cambridge University Press, Cambridge, UK, 1999).
[33]B. Chen and J. J. Stamnes, "Validity of diffraction tomography based on the first-Born and first-Rytov approximations," Appl. Opt. **37**, 2996–3006 (1998).
[34]M. Moghaddam and W. C. Chew, "Simultaneous inversion of compressibility and density in the acoustic inverse problem," Inverse Probl. **9**, 715–730 (1993).
[35]Z.-Q. Lu and Y.-Y. Zhang, "Acoustical tomography based on the second-order Born transform perturbation approximation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 296–302 (1996).
[36]G. Tsihrintzis and A. Devaney, "Higher order (nonlinear) diffraction tomography: Inversion of the Rytov series," IEEE Trans. Inf. Theory **46**, 1748–1761 (2000).
[37]M. A. Anastasio, D. Shi, and T. Deffieux, "Image reconstruction in variable density acoustic tomography," Proc. SPIE **5750**, 326–331 (2005).
[38]Z. Lu, "Multidimensional structure diffraction tomography for varying object orientation through generalised scattered waves," Inverse Probl. **1**, 339–356 (1985).
[39]X. Pan, "Consistency conditions and linear reconstruction methods in diffraction tomography," IEEE Trans. Med. Imaging **19**, 51–54 (2000).
[40]M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging* (Institute of Physics, Bristol, 1998).
[41]R. Bernier and H. H. Arsenault, "Deconvolution of two-dimensional images with zeros in the transfer function," Appl. Opt. **30**, 5163–5168 (1991).
[42]N. Takai, H. Kadono, and T. Asakura, "Statistical properties of the speckle phase in image and diffraction fields," Opt. Eng. (Bellingham) **25**, 627–635 (1986).
[43]M. A. Anastasio, D. Shi, Y. Huang, and G. Gbur, "Image reconstruction in spherical wave intensity diffraction tomography," J. Opt. Soc. Am. A **22**, 2651–2661 (2005).
[44]M. A. Anastasio, Q. Xu, and D. Shi, "Multispectral intensity diffraction tomography: Single material objects with variable densities," J. Opt. Soc. Am. A **26**, 403–412 (2009).
[45]Q. Zhu and B. Steinberg, "Wavefront amplitude distortion and image sidelobe levels: Part I—Theory and computer simulations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 747–753 (1993).

Adaptive extraction of modulation for cavitation noise

Fei Bao,[a] Xinlong Wang, and Zhiyong Tao
Key Laboratory of Acoustics and Institute of Acoustics, Nanjing University, Nanjing 210093, China

Qingfu Wang and Shuanping Du
State Key Laboratory of Ocean Acoustics, Hangzhou Applied Acoustics Research Institute, Hangzhou 310012, China

(Received 10 June 2009; revised 16 September 2009; accepted 17 September 2009)

Modulation analysis is an important issue in target classification and identification for ship-radiated noise. However, the modulated cavitation noise sought for analyzing is always submerged under strong ambient noise and difficult to be separated out. In this paper, an approach is proposed to extract the modulated cavitation noise adaptively by combining empirical mode decomposition and singular value decomposition. The results for both synthetical and practical signals demonstrate the practicability and effectivity of the approach.
 [DOI: 10.1121/1.3244987]

PACS number(s): 43.60.Mn, 43.60.Lq, 43.60.Wy [EJS] Pages: 3106–3113

I. INTRODUCTION

Classification of ship-radiated noise has attracted a lot of attention in past few decades. Unfortunately, several factors contribute to making this aim difficult to accomplish, which include complex sources that induce noise of a moving vessel, strong disturbance from ambient noise, surface and bottom reverberation effects, and lack of any *a priori* knowledge of the observed targets.[1–5,16,17] Therefore, robust and representative features are needed for discriminating different vessels.

The ship-radiated noise mainly consists of a broadband component, generated by propeller cavitation, and a number of narrow-band lines generated by both propeller cavitation and internal machinery (engine, generators, pumps, etc.).[6–8] Analysis of these parameters can provide valuable information regarding vessel classification and identification. However, it is difficult to obtain these spectrum characters from ship-radiated noise, which is typically non-stationary, and perhaps, nonlinear. Usually, in the power spectrum of one ship-radiated noise, the lines that indicate the propeller rotation and the internal machinery mix with each other and are difficult to discriminate. As the broadband cavitation noise is modulated by sum of a few harmonics due to propeller rotation, the information of propeller rotation can be acquired by a demodulating procedure, namely, detection of envelope modulation on noise (DEMON) analysis. Based on this instrument, Lourens[5] determined the propeller speed for classifying different ships, and Rajagopal *et al.*[9] fused DEMON analysis into an expert system for identification. Both of their operations require a preprocessing to filter out a signal, which is considered to behave better characteristic of the modulated cavitation noise, with certain frequency range. Therefore, when there is lack of or sometimes none of *a priori* knowledge about the cavitation noise, it is difficult to design an appropriate filter. Much recently, Shi and Hu[10,11] discussed the theoretical basis of cavitation noise systemically, and pointed out that the result of demodulation spectrum (DMS) for cavitation noise is influenced by the stability of ship-radiated noise. By adopting wavelet packet decomposition, they compared the DMSs of different subspaces and found their differences, then, selected one with the most information of modulation. The wavelet decomposition applied here just worked like a dyadic filter bank and cannot fulfill the optimal filtering. Consequently, an efficient and adaptive scheme is expected for extracting the amplitude-modulated cavitation noise.

In this paper, we attempt to combine two decomposition methods, empirical mode decomposition (EMD) and singular value decomposition (SVD), to achieve the extraction of the modulated cavitation noise from complex ship-radiated noise. In Sec. II, we introduce the modulation features of the cavitation noise and the operation for demodulation analysis. Next, in Sec. III, we apply a recently developed EMD for extracting the component of modulation and show the adaptivity of this method. Then, in Sec. IV, we present the theory of the SVD-based filter and propose an approach to add this technique to make the extraction more efficient. In Sec. V we examine the practicability and performance of our approach by processing ship-radiated noise. Finally, the conclusion of this work is provided in Sec. VI.

II. CAVITATION NOISE AND DEMON ANALYSIS

Cavitation noise is engendered by the propeller of a sailing ship. The pressure on a blade surface varies due to a nonuniform wake inflow velocity created by the hull and also due to the change in hydrostatic pressure. As the pressure is lowest in the upper region of the motional propeller blade, a cavity will be formed when the rotational speed is sufficiently high and it collapses rapidly when pressure increases as the blade turns downward. Since this collapse occurs every time when a blade passes through the region of low pressure, noise will be generated at blade rotational speed and its harmonics. Some tonal lines indicate that it can be observed

[a]Author to whom correspondence should be addressed. Electronic mail: baofeinju@gmail.com

from power spectrum in low-frequency range.[6,7] However, these tonal lines are often mixed with a series of tonal components produced by internal machinery of a ship and are difficult to distinguish.

As the cavitation noise $n_c(t)$ is modulated by the rotation of the propeller blades, it can be modeled as an amplitude-modulated broadband noise source $m(t)$ as follows:

$$m(t) = \left[1 + \sum_{i=1}^{N} A_i \cos(2\pi i f_r + \varphi_i)\right] n_c(t), \tag{1}$$

where if_r is the modulation frequency, A_i is the amplitude, and φ_i is the random initial phase. The basic modulation frequency f_r indicates the rotational frequency of a propeller. The number of propeller blades can be determined from the differences in height between the amplitudes of A_i. Let B be the number of blades; in general, the amplitude of each Bth characteristic frequency should be dominant.[1,12,2] Comparing it to the modulation frequencies, the cavitation noise $n_c(t)$ distributes in much higher frequency range.[13]

For this amplitude-modulated signal $m(t)$, we calculate its real envelope by Hilbert transform, pass it through a low-pass filter, and then compute the power spectrum of the filtered signal. The power spectrum is recognized as the DMS of $m(t)$. We choose the cutoff frequency of 100 Hz for the low-pass filter because it is suitable for both of the synthetical and practical signals in our experiments. Therefore, for each DMS in our paper, the sampling frequency of the filtered signal for calculating is 200 Hz. Essentially, the filtered signal, which is expected to represent the amplitude-modulating component $\Sigma_{i=1}^{N} A_i \cos(2\pi i f_r)$, describes the variation of the envelope of $m(t)$. Therefore, by estimating the frequency components of the envelope of $m(t)$, a DMS can reflect the modulation frequencies. The upper procedure is considered as the DEMON analysis.

Further, we adopt information entropy as a quantitative measurement to evaluate whether a DMS creates some distinct peaks, which possibly denote the modulation peaks we seek for. In statistical description, information entropy is a measurement of uncertainty of a system. With the use of the distribution of probabilities $p=p_i$, $0 \le p_i \le 1$, $i=1,2,\ldots,N$ of the system,[14] the information entropy can be presented as

$$H = -\sum_{i}^{N} p_i \ln p_i. \tag{2}$$

A smaller entropy value indicates greater certainty of a system and a larger one indicates uncertainty. By analyzing the probability distribution of the spectral coefficients, we can calculate the information entropy of a DMS and obtain the spectral certainty. For a DMS, the spectral certainty relates to its spectral structure; that is, a DMS with jumbled spectral structure has little certainty and that which exhibits one or more distinct spectral peaks has large one. Therefore, if the information entropy of a DMS is small enough, it must have a better structure and must be suitable for modulation analysis.

III. ADAPTIVE EXTRACTION VIA EMD

A. Algorithm of EMD

EMD, which can decompose adaptively a signal into some intrinsic oscillatory modes, has become rather ubiquitous in nonlinear and non-stationary signal processing. Being a novel signal analysis tool, EMD is intuitive, direct, and adaptive.[15]

The essence of EMD algorithm is to decompose a complex signal into several elementary AM-FM models, called intrinsic mode functions (IMFs), which represent the oscillation modes embedded in the original signal. An IMF is a function that satisfies two conditions.

(1) In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one.
(2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

In order to obtain these IMFs, original signal has to get through a certain principle, named sifting process, which is described as follows. First, all the local extrema are identified and all the local maxima (minima) are connected to obtain the upper (lower) envelope; second, the difference $h_1(t)$ between the data $X(t)$ and the local mean of the two envelopes is calculated; third, the previous sifting procedure is continually repeated k times until $h_k(t)$ is an IMF. Then, $h_k(t)$ as the IMF_1 is separated from the data and the residue $r_1(t)$ is gained. The next IMFs are extracted from $r_1(t)$ successively until the residue $r_n(t)$ reaches a predetermined criterion. Finally, the original signal is decomposed into n empirical modes, and a residue

$$X(t) = \sum_{i=1}^{n} \mathrm{IMF}_i + r_n(t). \tag{3}$$

EMD does not utilize, in decomposition, any *a priori* mathematical functions for orthogonal bases, but those specific and pertinent to current data to be processed.[15] This feature makes the EMD method particularly suitable for analyzing complex signal or time series that are usually non-stationary and nonlinear. Thus, wide applications of EMD have been submitted.[16,17]

According to the traditional definition, EMD is considered as a bandpass filter bank.[18,19] Flandrin *et al.*[19] claimed that, in the case of structured broadband stochastic processes such as fractional Gaussian noise, the built-in adaptivity of EMD makes its behavior spontaneous as a dyadic filter bank.

B. Extraction by EMD

In practical terms, EMD method decomposes a complex signal not only according to its frequency property, but also much referring to its peculiar internal structure such as modulation. When decomposing an amplitude-modulated signal by EMD, it will not be separated into a series of monochromatic components, but will be preserved as an independent mode.

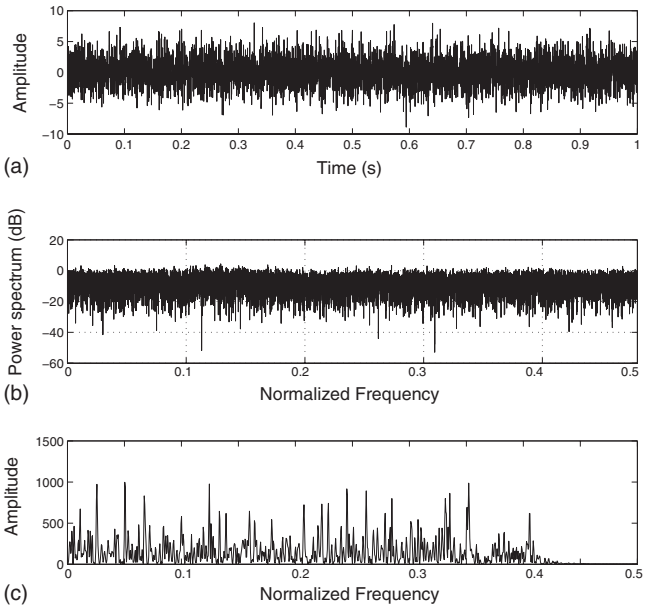


FIG. 1. (a) Time domain waveform of the synthetical signal $s(t)=m(t)+n(t)$, (b) its power spectrum, and (c) the DMS.

Further, we consider the decomposition under noise disturbance. When we apply EMD to extract an amplitude-modulated signal, whose structure is much similar to the modulated cavitation noise, from strong background noise, this component would be obtained integrally in certain IMF. Let us analyze a synthetical signal

$$s(t) = m(t) + n(t), \tag{4}$$

which consists of a modulated signal in the form of Eq. (1) and a white noise $n(t)$ (with zero mean and $\sigma_n^2=4.97$) as background noise, where $\Sigma_{i=1}^{N} A_i \cos(2\pi i f_r)=\cos(20\pi t)+0.6\cos(40\pi t)+0.55\cos(60\pi t)$ and $n_c(t)$ is a broadband stochastic signal (with zero mean and $\sigma_{n_c}^2=1.25$) ranging from 5280 to 7680 Hz. The coefficients of A_i set here are used to simulate the intensity of modulation frequency for a typical category of cargo ship, which has one propeller with three blades. All these signals are with the sampling frequency of 48 kHz and the length of 4 s. Here, as we consider the modulated stochastic signal to be the signal we are concerned about, the signal-to-noise ratio is quite of low value −28.58 dB. Therefore, as shown in Fig. 1, the power spectrum of $s(t)$ is flat in whole frequency domain, which exhibits the spectral characteristic of white noise. In addition, its DMS shows several peaks with similar amplitude and tells nothing about the information of modulation.

Now, by invoking EMD technique, we decompose $s(t)$ into a series of IMFs. To each of them, we calculate its DMS and then compute the information entropy H of the DMS. We find, as shown in Table I, that the information entropy of the DMS for the second IMF is obviously smaller than the others. In Fig. 2(b), it can be observed apparently that the DMS of IMF_2 presents three remarkable peaks (located at normalized frequencies 0.050, 0.100, and 0.150), indicating the modulation frequencies, which reflect the variation of the envelope of IMF_2. For comparison, as displayed in Fig. 2, the DMSs corresponding to IMF_1 and IMF_3, which present the disordered spectral structure, are useless for modulation analysis. Therefore, by comparing the information entropy of each IMF, we can affirm that IMF_2 contains the modulated component $m(t)$ and name it as target IMF.

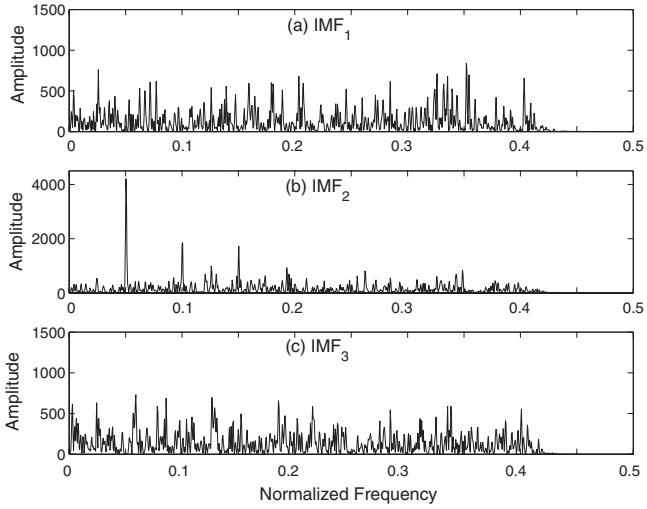


FIG. 2. (a)–(c) are the DMSs of the first three IMFs decomposed from $s(t)=m(t)+n(t)$.

C. Adaptivity

When the frequency range of a modulated component does not distribute within one certain pass-band of a dyadic filer, by using the ordinary dyadic filter or a wavelet filter, this modulated component will be broken into two or more parts. Thus, the modulation analysis will not be accurate. Next, we will test if EMD would overcome this shortcoming.

EMD is usually considered as a dyadic filter bank for decomposing noise signal with Gaussian distribution, the effective bandwidth of each equivalent filter is confirmed by the overlapping positions of the envelopes of power spectra of neighbor IMFs.[19] Figure 3 (the upper plot) shows the spectral structures of these IMFs. It can be read from the figure that the spectral envelopes of the first six IMFs intersecting, successively, with each other at the normalized frequencies f_n 0.242, 0.135, 0.065, 0.037, and 0.021 perform the behavior similar to a dyadic filter bank. Here the frequency is normalized to the sampling frequency. Whereas, the modulated signal $m(t)$ we are inspecting has a frequency band ranging from 5280 to 7680 Hz (or from 0.110 to 0.160 by normalization), which is not contained exactly within the one pass-band of the equivalent filters. In Fig. 1(b), the

TABLE I. Information entropy H of DMS for different IMFs.

	IMF_1	IMF_2	IMF_3	IMF_4	IMF_5	IMF_6	IMF_7	IMF_8
H	3.38	2.25	3.42	3.44	3.39	3.56	3.52	3.61

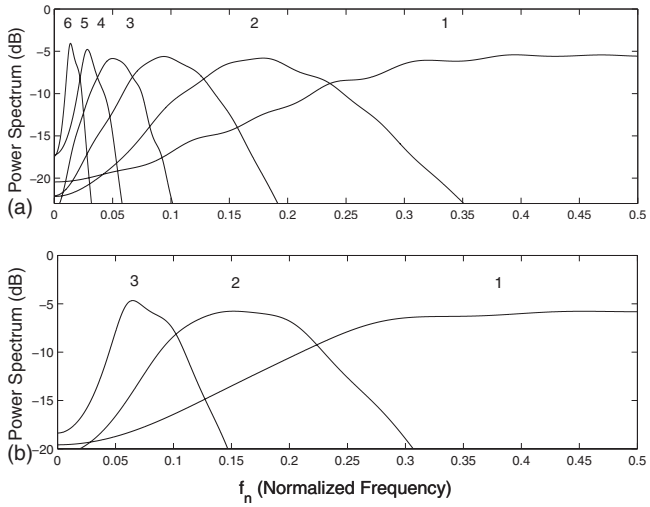


FIG. 3. (a) Upper: EMD equivalent filters for Gaussian white noise. In the case of Gaussian noise, EMD can be interpreted as a dyadic filter bank of overlapping bandpass filters. (b) Lower: EMD equivalent filters for synthetical signal $s(t)$, which do not behave the character of an ordinary dyadic filter bank.

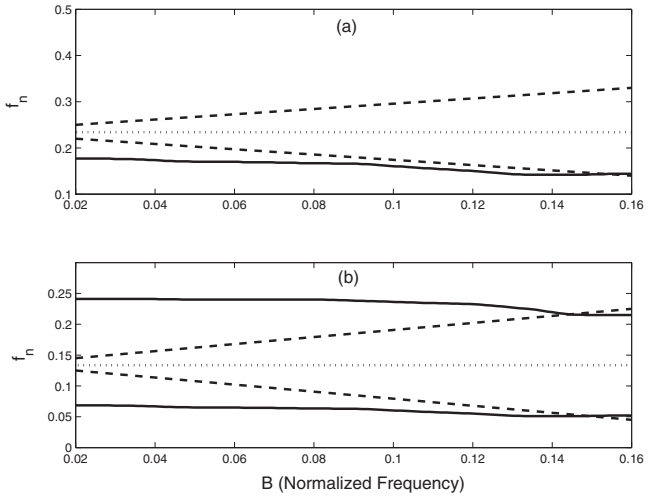


FIG. 4. Equivalent band structures of (a) IMF_1 and (b) IMF_2 versus the bandwidth B of the modulated component $m(t)$. The dashed lines indicate the bounds of bandwidth of the modulated component $m(t)$, the solid lines indicate the bounds of the equivalent bandwidth of the IMF, and the dotted lines indicate the central frequency of $m(t)$.

power spectrum of the noise-tainted modulated signal $s(t)$ performs similar to that of the Gaussian white noise; if EMD works just like an ordinary dyadic filter bank, $m(t)$ should be broken into pieces. However, comparing to those of the Gaussian white noise, the decomposed IMFs of $s(t)$ have different frequency ranges. In the lower plot of Fig. 3 they are the envelopes of power spectra of the first three IMFs, IMF_1, IMF_2, and IMF_3, the second of which has been identified to be the target IMF, which contains the modulated component. Here, we observe that the normalized frequency band of the equivalent filter for IMF_2 shifts from [0.135,0.242] in Fig. 3(a) down to [0.109,0.226]. With the change in bandwidth of this equivalent filter, the modulated component, whose frequency range distributes within the normalized frequency of [0.110,0.160], is entirely contained in IMF_2.

To examine the general adaptivity observed above, we set the central frequency of the modulated component $m(t)$ to the edge value where the spectral envelopes of the equivalent filters shown in Fig. 3 (the upper plot) intersect with each other, and then, change the bandwidth B of $m(t)$ from the normalized value 0.020 to 0.160, with each step of 0.010. First, we set the normalized central frequency to 0.242, the normalized frequency of the two power spectra of the first and second equivalent filters. To construct a modulated signal with the certain bandwidth, we obtained the carrier by filtering from a white stochastic signal, which is of zero mean and variance $\sigma_{n_c}^2 = 1.1$, with the predefined bandwidth B. The background noise is a white noise of zero mean and variance $\sigma_n^2 = 2.9$. After decomposing this noise-like complex signal into a few IMFs, we calculated the bandwidth for each IMF. The results are graphically shown in Fig. 4(a). In the figure, the dashed lines display the variation of the bounds of the frequency band of the modulated component $m(t)$ as its bandwidth B varies, the dotted line indicates the central frequency of $m(t)$, and the area above the solid line represents the practical effective band of the IMF, which embodies $m(t)$. Therefore, we discover that, when $B < 0.150$, the modulated component is always preserved in a single IMF, here (IMF_1).

Similarly, we set the central frequency to the normalized frequency of 0.135, approximately the edge frequency between the second and third equivalent filters as shown in Fig. 3 (the upper plot); the results are demonstrated in Fig. 4(b). In the figure, with the variation of bandwidth $B < 0.140$, the modulated component $m(t)$ is also found to be contained in the area limited by two solid lines, which denotes the equivalent bandwidth of IMF_2.

From the above examples, we learnt that, when a noise-like signal contains a modulated component, EMD does not work simply as a dyadic filter bank and actually owns the property to protect the modulation structure from being broken. Therefore, for the situation that a broadband modulated component $m(t)$ is buried in noise, when its bandwidth is not too wide, the modulation-bearing IMF will adjust itself adaptively to cover nearly the spectrum of $m(t)$, thus ensuring the integrity of the modulated component in the decomposition.

IV. SVD-BASED FILTERING FOR TARGET IMF

A. SVD-based filter

SVD is a stable and computational effective decomposition. Besides it provides an accurate method for determining the effective rank of a matrix, SVD has been extensively applied, too, in subspace estimation, spectral analysis, and system identification problems.[20–27] Moreover, it has been adopted as a denoising instrument in various signal processing fields.[28–32]

In the theory of SVD, any $m \times n (m \geq n)$ real-valued matrix A, produced by a time series $x(t)$, can be decomposed uniquely as

$$A = U \sum V^t = \sum_{i=1}^{n} \alpha_i \vec{u}_i \vec{v}_i^{\,t}, \tag{5}$$

where $U=[\vec{u}_1,\vec{u}_2,\ldots,\vec{u}_n]\in R^{m\times n}$ (named the "left singular vectors") and $V=[\vec{v}_1,\vec{v}_2,\ldots,\vec{v}_n]\in R^{n\times n}$ (the "right singular vectors") both are two column-orthogonal matrices, and V^t denotes the transpose of V. Σ is a diagonal matrix with descending diagonal elements α_i $(i=1,\ldots,n)$, namely, the singular values of A, which represent the energy of matrix A projected on each subspace. In theory, the rank of A is decided by the number of its nonzero singular values.

If $x(t)$ is regular, it usually has good mathematical properties, such as the continuity and the differentiability, and contains large redundant information. In this term, A is not a full rank matrix. We suppose that the rank of A is r, where $r\leq\min(m,n)$, then the singular values should be

$$\alpha_1\geq\alpha_2\geq\cdots\geq\alpha_r>\alpha_{r+1}=\cdots=\alpha_n=0. \tag{6}$$

When $x(t)$ is superimposed by noise, we observe the matrix B generated from the noise polluted signal. B can be taken as the summation of A and a perturbation matrix E, expressed as

$$B=A+E. \tag{7}$$

By the influence of noise, matrix B will be of full rank, and the singular values of B are

$$\beta_1\geq\beta_2\geq\cdots\geq\beta_{n-1}\geq\beta_n>0. \tag{8}$$

Here, $(\beta_{r+1},\ldots,\beta_n)$ are usually small but not necessarily zero. Therefore, the dominant singular values $\beta_1,\beta_2,\ldots,\beta_r$ represent the subspaces of signal $x(t)$, while the insignificant small ones indicate that of the noise.

By setting the insignificant singular values, which are of $(\beta_{r+1},\ldots,\beta_n)$, to zero, a new matrix

$$\hat{B}=\sum_{i=1}^{r}\beta_i\vec{u}_i\vec{v}_i^{\,t} \tag{9}$$

is obtained. Then, a new time series $\hat{x}(t)$ is reconstructed by $\hat{B}$ and considered approximate to the original signal $x(t)$. The procedure of truncating the last few insignificant singular values works here and resembles to a filter, which essentially retains the energy in signal subspace and discards the energy in the noise subspace. Therefore, the SVD-based filter is accomplished.

B. Parameter selection

For practical operation, two problems are needed to be considered: (1) the constructing method for the data matrix B, and (2) the threshold for the truncated singular values.

To construct a data matrix for SVD, usually there are two ways. One is to produce a Hankel matrix by a sliding window. For a time series $x(t)$ of length N, if the window width is chosen to be L, the Hankel matrix will be of the dimension of $L\times(N-L+1)$. The other one is to form a matrix by partitioning the signal into segments of width L and placing each one as a row. By this method, supposing N is the multiple of L, the dimension of the data matrix is decreased to $L\times(N/L)$. Therefore, the computational cost will be considerably saved, especially when N is large. Kanjilal and Palit[23] found that, by adopting this method to produce a data matrix for SVD, when the row length of the data matrix is the multiple of a noisy signal period, α_1/α_2 (the ratio between the first and second singular values) is remarkable large, and the erroneous selection of row length would not produce a large α_1/α_2, due to the misalignment. However, for this situation, the first few singular values are still significantly large and can represent the energy of the signal. Hence, as we just need to preserve the singular values that indicate the energy of the signal, there is no need to accurately select the row length of the data matrix. In this paper, in order to reduce the computational complexity, we choose the second method to construct the data matrix and select the suitable row length to make the data matrix approximately square.

Estimation of the threshold for truncating singular values is critical to the effectiveness of the algorithm. If we do not retain enough singular values, the recovered signal will lose some information of the original signal. If we reserve redundant singular values, some noise would be kept. Thus, we have to estimate which of the singular values are authentically significant and could represent the energy of signal component. To effectively determine the rank of a perturbed matrix, Konstantinides and Yao[25] derived tight threshold bounds for perturbation due to the independent identically distributed Gaussian noise. Based on the theories of perturbations of singular values and statistical significance test, the bounds of the threshold ξ for the singular values of data matrix B can be determined as

$$\sqrt{c}\leq\xi\leq\sqrt{mn}\sigma, \tag{10}$$

where σ is the variance of E and c is a parameter determined from the statistics of signal and noise. Whereas, for practical application, it is hard to estimate the variance of noise, thus we cannot employ this kind of criterion to select the threshold.

In theory, if we select an appropriate threshold, the recovered signal should perform the characteristic of the target component we pursue. In this paper, the target component is the amplitude-modulated component. Thus, for a recovered signal obtained by any possible threshold, we can judge whether the threshold is appropriate via the information entropy of its DMS. If the information entropy reaches minimum value, it means that the DMS has the most possibility to present some distinct spectral lines and supply useful information of modulation.

C. Promotion of discrimination for DMS

In Sec. III, by using EMD, we have extracted a target IMF, which contains the modulated component. Further, the information of modulation can be accurately acquired from this IMF. Nevertheless, with the enhancement of background noise, the effect of modulation analysis based on EMD will be degenerated. Therefore, we try to promote the performance of the extraction with the help of the SVD-based filter.

For signal $s(t)$ in Eq. (4), we increase the background noise $n(t)$, and repeat the procedure of modulation analysis. For this time, the variance of $n(t)$ is $\sigma_n^2=6.63$. As a result, through information entropy of the DMS for each IMF, we

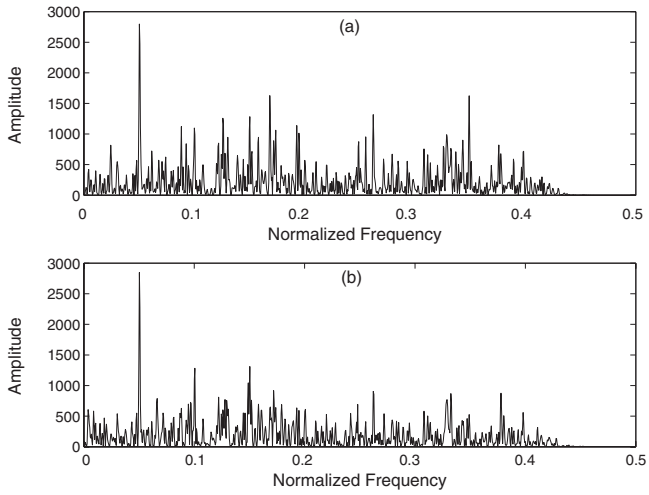


FIG. 5. DMSs of (a) IMF_2, and (b) the signal filtered from IMF_2 by SVD-based method.

can confirm that the target IMF is IMF_2. Although the information entropy of IMF_2 is still obviously smaller than other IMFs, comparing it to the result in Fig. 2 it has increased from 2.25 to 2.54. Corresponding to the enhancement of information entropy, the quality of the DMS degenerates. In Fig. 5 (the upper one), only one distinct peak located at normalized frequency of 0.050 can be observed, and the peaks located at normalized frequency of 0.100 and 0.150 are both submerged by the peaks that are caused by noise. As this DMS just represents one obvious peak, which indicates the modulation frequency of 0.050 by normalization, it fails to supply us the entire information of the modulation.

Following, we apply the SVD-based filter to extract better modulated component. With sampling frequency of 48 kHz and length of 4 s, IMF_2 has a total of 192 000 sampling points, so we build a 400×480 data matrix for SVD. After SVD, we obtain 400 singular values. The last 350 ones, which are all close to zero, are apparently insignificant. So we select the threshold by just considering the first 50 ones. As a result, for each of the corresponding thresholds, we recover the signal and calculate the information entropy of its DMS. We discover that, as displayed in Fig. 6 (the upper one), while choosing the threshold of 8, the information entropy of the DMS for the recovered signal reaches the minimum value. So that, we consider 8 as the ideal threshold for truncating the singular values.

Next, we will examine whether the threshold we have chosen is really ideal. For an appropriate threshold, the recovered signal $r(t)$ should represent similar properties, e.g., the waveform and the DMS, to the modulated component $m(t)$. Accordingly, by analyzing the similarity of these properties, we can judge the quality of the recovered signal. R_{wave} is applied to measure the similarity of the waveforms between $m(t)$ and $r(t)$, which is defined as their correlation coefficient

$$R_{\text{wave}} = \frac{\text{cov}[m(t),r(t)]}{\sqrt{\text{cov}[m(t),m(t)]\text{cov}[r(t),r(t)]}}, \tag{11}$$

where cov is the calculation of covariance.

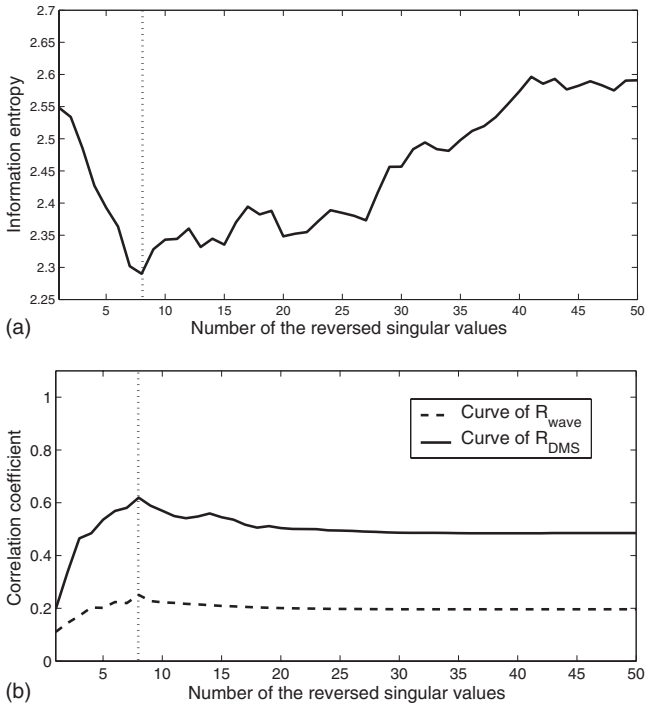


FIG. 6. (a) Variation of information entropy with different selections of the threshold, (b) variation of R_{wave} with different selections of the threshold in dashed curve, and that of R_{DMS} in solid curve.

As a DMS is computed by the discrete Fourier transform, it can be described by the Fourier coefficients. Then, the similarity of the DMSs between $m(t)$ and $r(t)$ can be estimated by the Fourier coefficients F_m and F_r, of their DMSs, which is expressed as

$$R_{\text{DMS}} = \frac{\text{cov}(F_m,F_r)}{\sqrt{\text{cov}(F_m,F_m)\text{cov}(F_r,F_r)}}. \tag{12}$$

In Fig. 6 (the lower one), the variations of both R_{wave} and R_{DMS} for the 50 possible thresholds are shown. The maxima of two parameters both occur at the threshold of 8, which testifies that the threshold we have selected is appropriate and accurate.

For the DMS of the SVD-based filtered signal, displayed in Fig. 5 (the lower one), its information entropy is calculated to be 2.29. Comparing with the upper DMS in Fig. 5, whose entropy is 2.54, three distinct peaks can be observed at normalized frequencies of 0.050, 0.100, and 0.150. Thus, with the decrease of 0.250 in information entropy, the quality of the DMS is promoted well.

For different intensities of $n(t)$, we repeat the above calculation. We find, for each $s(t)$ with different noise levels, after EMD, that the information entropy of IMF_2 is smaller than that of the others; meanwhile, the frequency range of $m(t)$ is contained in that of IMF_2 too. Therefore, we consider IMF_2 as the target IMF. Then, for the DMS of the signal filtered from IMF_2 by the SVD-based filter, we compare its information entropy to that of IMF_2; all results are displayed in Fig. 7. From the figure we can observe that the entropy for the DMS of the SVD-based filtered signal is always smaller than the one for IMF_2. Moreover, when $\sigma_n^2>7$, the entropy of the DMS for the SVD-based filtered signal descends

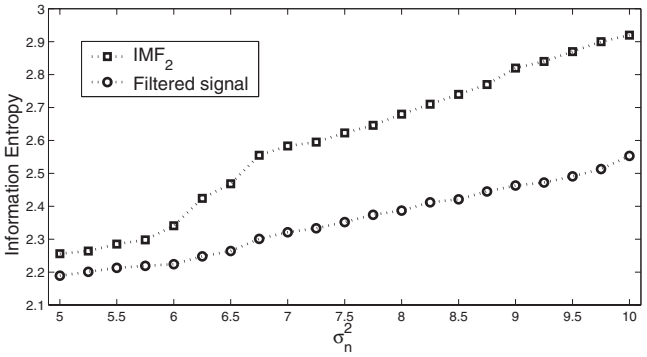


FIG. 7. Variation of information entropy versus noise variance by the two methods, (a) the upper curve: for the DMS of IMF_2, and (b) the lower curve: for the DMS of the signal filtered from IMF_2 by SVD-based method.

nearly 0.300 or so from that of IMF_2, which means the quality of the DMS is much heightened. Therefore, by adding the SVD-based filter, we can purify the component of the target IMF further.

Consequently, we propose an approach to extract the modulated component from background noise adaptively and efficiently, which is described as follows: First we obtain the target IMF by EMD and then, for the target IMF, we extract the component that behaves better characteristic of modulation further by adding the SVD-based filter.

V. APPLICATION TP SHIP-RADIATED NOISE

As we have proposed an approach to separate a modulated signal from wide band background noise by combing with EMD and SVD, in this section, we apply it to practical ship-radiated noise for the purpose of target recognition by analyzing the propeller rotational frequency and the number of blades. We recorded an underwater acoustic signal radiated from a cargo ship sailed in oceanic environment, spanning 80 s in time and being sampled at 48 kHz. The ship has one propeller with three blades, and the sailing speed when recording is 10 kn. For this sailing speed, the rotational speed of propeller is approximately 246 RPM (rotation per minute), corresponding to the propeller rotational frequency of 4.10 Hz. We divide the signal into 20 samples with length of 4 s each. Then, for each of them, we calculate the DMS and attempt to analyze the propeller rotational frequency and the blade number. Theoretically, in the DMS of each signal, there should exist three distinct peaks located at 4.10, 8.20, and 12.3 Hz corresponding to the normalized 0.032, 0.041, and 0.062, which indicate the propeller rotational frequency and its harmonics caused by blades.

However, caused by the disturbance of ambient noise, the DMS calculated directly from the recorded signal always supplies us worthless information. For instance, in Fig. 8, the DMS for the first sample (lasting from 1 to 4 s of the recorded signal) is displayed without any obvious peak.

Then we extract the modulated cavitation noise for the first sample by the proposed approach and also by EMD for comparison. For this signal, the target IMF is IMF_3 and its DMS is displayed in Fig. 9 (the upper plot). In this DMS, whose information entropy is 2.58, only an obvious peak indicates that the basic modulation frequency can be viewed

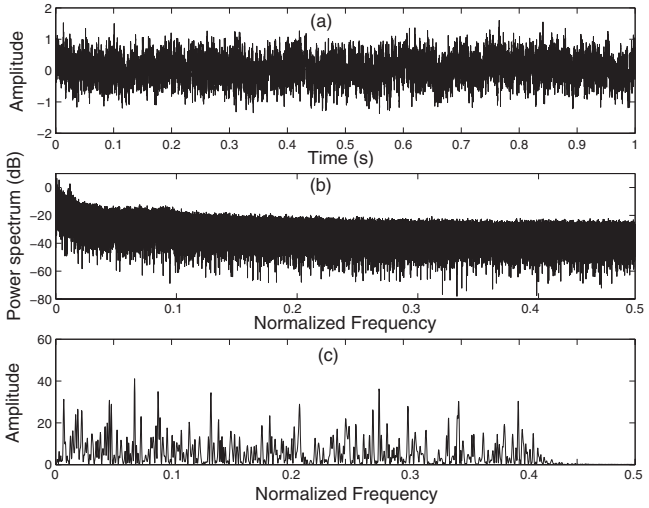


FIG. 8. (a) Time domain waveform of the ship-radiated signal, (b) its power spectrum, and (c) the DMS.

at normalized frequency of 0.021; the peaks, which denote the harmonics, are all submerged in the peaks caused by noise. For the approach we proposed, in the DMS of the SVD-based filtered signal, which is displayed in Fig. 9 (the lower plot), its entropy descends to 2.33 and in the spectrum all three modulation peaks we expect can be observed clearly.

For each of the 20 samples, we compare the DMSs of original samples, IMF_3 decomposed from the samples and the signal filtered by the SVD-based method from IMF_3, and the information entropies of these DMSs are exhibited in Table II. In the table, H_{original}, H_{IMF_3}, and H_{filtered} indicate the information entropies of the DMSs for the three kinds of signals mentioned above. By comparing those entropies, we discover that H_{original} is always larger than 3, H_{IMF_3} is much smaller, and H_{filtered} is the smallest. The corresponding man-made judgments via visual inspecting are summarized as follows: None of the DMSs of the original samples can supply useful modulation information; for the DMSs those of IMF_3 decomposed from the samples, we can find a distinct basic

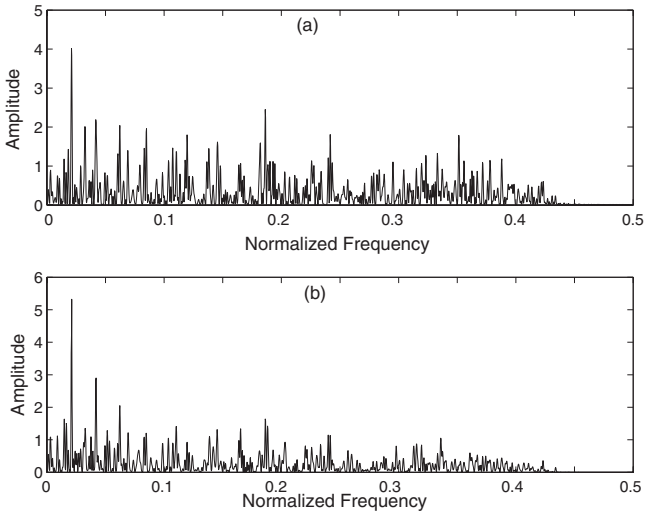


FIG. 9. DMSs of (a) IMF_3, and (b) the signal filtered from IMF_3 by SVD-based method.

TABLE II. Information entropy H of DMS for different samples.

Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
H_{original}	3.32	3.44	3.42	3.44	3.52	3.13	3.53	3.29	3.28	3.18	3.43	3.23	3.53	3.17	3.23	3.19	3.25	3.24	3.37	3.36
H_{IMF_3}	2.58	2.81	2.64	2.54	2.34	2.28	2.32	2.31	2.64	2.47	2.20	2.24	2.31	2.74	2.63	2.65	2.84	2.43	2.63	2.45
H_{filtered}	2.33	2.43	2.31	2.27	2.10	2.11	2.09	2.14	2.32	2.22	2.12	2.10	2.13	2.52	2.43	2.32	2.44	2.15	2.31	2.23

modulation peak for 15 samples and observe all the three modulation peaks for 7 samples; for the DMSs those of the signal filtered by the SVD-based method from IMF_3, we can find a distinct basic modulation peak for each sample and observe all the three modulation peaks for 16 samples. Therefore, we conclude that, by adopting EMD in the modulation analysis for the ship-radiated noise, the modulated cavitation noise can be extracted facilely and adaptively, and then, by combining the SVD-based filter, we can further enable the approach more stable and efficient.[32]

VI. CONCLUSIONS AND REMARKS

In this paper, by utilizing EMD and SVD, we presented an adaptive and efficient approach to extract the modulated component from a complex signal. With the application of EMD, we first separated a target IMF that contains the modulated component, then, based on SVD, we further filtered out the component that behaves better characteristic of modulation. The experiments showed that the method represents outstanding performance and the application for extracting the modulated cavitation noise also achieved satisfied results.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation of China under Grant No. 10574070 and the State Key Laboratory Foundation of China under Grant No. 9140C240207060C24.

[1]D. Ross, *Mechanics of Underwater Noise* (Pergamon, New York, 1976).
[2]M. A. Deaett, "Signature modeling for acoustic trainer synthesis," IEEE J. Ocean. Eng. **12**, 143–147 (1987).
[3]S. Yang, Z. S. Li, and X. L. Wang, "Vessel radiated noise with fractal features," Electron. Lett. **36**, 923–925 (2000).
[4]S. Yang, Z. S. Li, and X. L. Wang, "Ship recognition via its radiated sound: The fractal based approaches," J. Acoust. Soc. Am. **112**, 172–177 (2002).
[5]J. G. Lourens, "Classification of ships using underwater radiated noise," Communications and Signal Processing, COMSIG **88**, 13–18 (1988).
[6]P. T. Arveson and D. J. Vendittis, "Radiated noise characteristics of a modern cargo ship," J. Acoust. Soc. Am. **107**, 118–129 (2000).
[7]S. C. Wales and R. M. Heitmeyer, "An ensemble source spectra model for merchant ship-radiated noise," J. Acoust. Soc. Am. **111**, 1211–1231 (2002).
[8]M. V. Trevorrow, B. Vasiliev, and S. Vagle, "Directionality and maneuvering effects on a surface ship underwater acoustic signature," J. Acoust. Soc. Am. **124**, 767–778 (2008).
[9]R. Rajagopal, B. Sankaranarayanan, and P. R. Rao, "Target classification in a passive sonar—An expert system approach," IEEE Trans. Acoust., Speech, Signal Process. **5**, 2911–2914 (1990).
[10]G. Z. Shi and J. C. Hu, "Ship noise demodulation line spectrum fusion feature extraction based on the wavelet packet," Wavelet Analysis and Pattern Recognition, ICWAPR **7**, 846–850 (2007).
[11]G. Z. Shi and J. C. Hu, "Theoretical analysis of the ship noise demodulation spectrum harmonic clan feature," Acta Acust. **33**, 19–25 (2007).
[12]A. Kummert, "Fuzzy technology implemented in sonar systems," IEEE J. Ocean. Eng. **18**, 483–490 (1993).
[13]R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983), pp. 423–424.
[14]J. N. Kapur and H. K. Kesavan, *Entropy Optimization Principles With Applications* (Academic, San Diego, CA, 1992).
[15]N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear non-stationary time series analysis," Proc. R. Soc. London, Ser. A **454**, 903–995 (1998).
[16]L. Zhang, J. Huang, and Q. Zhang, "Spectral analysis of noise characteristics caused by ship propeller cavitation," OCEANS **2007**, 1–5 (2007).
[17]M. Grenie, "Acoustic detection of propeller cavitation," IEEE ICASSP **90**, 2907–2910 (1990).
[18]Z. H. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," Proc. R. Soc. London, Ser. A **460**, 1597–1611 (2004).
[19]P. Flandrin, G. Rilling, and P. Gonçalvés, "Empirical mode decomposition as a filter bank," IEEE Signal Process. Lett. **11**, 112–114 (2004).
[20]F. J. Alonso, J. M. Del Castillo, and P. Pintado, "Application of singular spectrum analysis to the smoothing of raw kinematic signals," J. Biomech. **38**, 1085–1092 (2005).
[21]D. Baratta, G. Cicionib, F. Masulli, and L. Studer, "Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting," Neural Networks **16**, 375–387 (2003).
[22]F. Kleibergen and R. Paap, "Generalized reduced rank tests using the singular value decomposition," J. Econometr. **133**, 97–126 (2006).
[23]P. P. Kanjilal and S. Palit, "Singular value decomposition applied to the modeling of quasiperiodic processes," Signal Process. **35**, 257–267 (1994).
[24]P. P. Kanjilal and S. Palit, "On multiple pattern extraction using singular value decomposition," IEEE Trans. Signal Process. **43**, 1536–1540 (1995).
[25]K. Konstantinides and K. Yao, "Statistical analysis of effective singular values in matrix rank determination," IEEE Trans. Acoust., Speech, Signal Process. **36**, 757–763 (1988).
[26]R. Kakarala and P. O. Ogunbona, "Signal analysis using a multiresolution form of the singular value decomposition," IEEE Trans. Image Process. **10**, 724–735 (2001).
[27]Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," IEEE Trans. Signal Process. **49**, 457–469 (2001).
[28]Z. Hou, "Adaptive singular value decomposition in wavelet domain for image denoising," Pattern Recogn. **36**, 1743–1763 (2003).
[29]M. Bydder and J. Du, "Noise reduction in multiple-echo data sets using singular value decomposition," Magn. Reson. Imaging **24**, 849–856 (2006).
[30]K. Konstantinides, B. Natarajan, and G. S. Yovanof, "Noise estimation and filtering using block-based singular value decomposition," IEEE Trans. Image Process. **6**, 479–483 (1997).
[31]Y. Shim and Z. Cho, "SVD pseudoinversion image reconstruction," IEEE Trans. Acoust., Speech, Signal Process. **29**, 904–909 (1981).
[32]V. D. Vrabie, J. I. Mars, and J. L. Lacoume, "Modified singular value decomposition by means of independent component analysis," Signal Process. **84**, 645–652 (2004).

Comparison of *in-situ* calibration methods for quantifying input to the middle ear

James D. Lewis,[a] Ryan W. McCreery, Stephen T. Neely, and Patricia G. Stelmachowicz
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 4 June 2009; revised 12 September 2009; accepted 14 September 2009)

Sound pressure level *in-situ* measurements are sensitive to standing-wave pressure minima and have the potential to result in over-amplification with risk to residual hearing in hearing-aid fittings. Forward pressure level (FPL) quantifies the pressure traveling toward the tympanic membrane and may be a potential solution as it is insensitive to ear-canal pressure minima. Derivation of FPL is dependent on a Thevenin-equivalent source calibration technique yielding source pressure and impedance. This technique is found to accurately decompose cavity pressure into incident and reflected components in both a hard-walled test cavity and in the human ear canal through the derivation of a second sound-level measure termed *integrated pressure level* (IPL). IPL is quantified by the sum of incident and reflected pressure amplitudes. FPL and IPL were both investigated as measures of sound-level entering the middle ear. FPL may be a better measure of middle-ear input because IPL is more dependent on middle-ear reflectance and ear-canal conductance. The use of FPL in hearing-aid applications is expected to provide an accurate means of quantifying high-frequency amplification. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3243310]

PACS number(s): 43.64.Ha, 43.58.Vb, 43.20.Ks [BLM] Pages: 3114–3124

I. INTRODUCTION

In-situ ear-canal sound-level measurements are used to describe acoustic input to the middle ear (ME) in a variety of different clinical and research applications including otoacoustic emission stimulus level calibration and hearing-aid output validation. *In-situ* estimates of ME input have advantages over coupler calibration methods because ear-canal acoustic impedance is variable and, therefore, the impedance of individual ears differs widely from any standard coupler. Unfortunately, standing waves in the ear canal due to reflections from the tympanic membrane (TM) may have adverse effects on *in-situ* estimates of input to the ME, especially at frequencies above 4 kHz. The current study demonstrates how prior determination of sound-source characteristics provides a means to overcome most of the standing-wave problems usually associated with *in-situ* measurements.

In-situ sound-level measurements are commonly used to quantify the amplification provided by hearing aids and the subsequent input to the ME. In this procedure, a probe microphone is used to measure the amplified sound-level in the ear canal. Ear-canal sound-level is then compared to a prescribed amplification target and hearing-aid gain is adjusted to meet the target value. The use of real-ear measurements is generally considered to provide a more accurate estimation of ME input than 2-cm^3 coupler approaches since sound pressure generated in an ear canal may vary significantly from that in a coupler (Sachs and Burkhard, 1972; Larson *et al.*, 1977; Nelson Barlow *et al.*, 1988; Feigin *et al.*, 1989).

The pressure measured in the ear canal by a probe microphone (P_{probe}) can be modeled as the sum of forward- and reverse-traveling pressure waves:

$$P_{probe} = P_{forward} + P_{reverse}, \quad (1)$$

where $P_{forward}$ and $P_{reverse}$ represent the complex forward-traveling pressure wave from the sound source and the reverse-traveling wave reflected by the TM, respectively, at any particular frequency. Across the length of the ear canal, the forward and reverse pressure waves combine both constructively and destructively depending on the phase relationship between the pressure components. For frequencies with $\frac{1}{4}$ wavelengths more than twice the length of the canal, constructive phase interactions between pressure components result in near uniform pressure distribution along the length of the canal. Pressure measured by a probe microphone at these frequencies may be interpreted as the input to the ME regardless of the distance between the probe microphone and the TM. On the other hand, at the frequency for which the distance between the probe microphone and the TM is $\frac{1}{4}$ wavelength, a pressure null is created by an anti-phase interaction between forward and reverse pressure components (Wiener and Ross, 1946; Khanna and Stinson, 1985; Gilman and Dirks, 1986). The pressure measured by a probe microphone for this frequency can no longer be assumed to represent the pressure at the TM and may deviate from TM pressure by as much as 24 dB (Stinson *et al.*, 1982; Gilman and Dirks, 1986; Dirks and Kincaid, 1987; Chan and Geisler, 1990). Frequencies immediately surrounding the frequency of a pressure null will be similarly affected by destructive phase interactions, although to a lesser degree. Probe pressure for these frequencies does not accurately estimate the pressure at the TM (Gilman and Dirks, 1986).

At the entrance of an un-occluded ear canal 27 mm in length, a pressure minimum at a frequency of approximately 3 kHz may be expected ($f=v/4\lambda$, where v is the speed of sound in air and λ is the length of the ear canal). Research and clinical applications involving *in-situ* measures are gen-

[a]Author to whom correspondence should be addressed. Electronic mail: james-lewis@uiowa.edu

erally performed in an ear canal occluded by a probe assembly, earmold (EM), or hearing aid. Depending on the depth of the occluding object, the ear canal will be effectively shortened thus broadening the bandwidth over which uniform pressure distribution can be expected. Pertinent to hearing-aid applications, with an earmold (or hearing aid) and probe microphone extending to the second bend of the canal (approximately 1/3 the length of the canal), pressure above approximately 4–5 kHz will vary as a function of distance from the TM. Traditionally, hearing aids have provided limited gain above 4 kHz and the importance of accurately quantifying pressure at the TM at higher frequencies has been unnecessary. However, recent studies suggest that an extended bandwidth (out to 9 kHz) can improve the perception of fricatives for children with hearing-impairment (Stelmachowicz *et al.*, 2001, 2004, 2007) which may be beneficial for early speech and language development. As technological advances widen the bandwidth of wearable hearing aids, it will be necessary to accurately quantify high-frequency hearing-aid output in individual ears in order to optimize both audibility and comfort.

One potential solution to quantify input to the ME at higher frequencies is placement of the probe microphone in close proximity to the TM (Siegel, 1994). Distances from the TM expected to be most insensitive to pressure minima have ranged from 1–9 mm (Dirks and Kincaid, 1987; Chan and Geisler, 1990; Hellstrom and Axelsson, 1993); however, this distance is dependent on the highest frequency of interest for the given application. Regarding amplification extending to 9 kHz, however, a study by McCreery *et al.* (2009) demonstrated that pressure measured at a probe microphone situated within 2 mm of the TM across repeated trials is variable for frequencies above 4 kHz. Both Stinson and Shaw (1982) and Stinson (1985) similarly demonstrated that spatial differences in sound pressure across the dimensions of the TM exist; however, their results suggest that this does not occur until approximately 10 kHz. This non-uniform pressure distribution across the TM is attributed to the $\frac{1}{4}$ wavelength at 10 kHz compared to the cross-sectional dimensions of the TM as well as the observation that the plane of reflectance changes above 10 kHz. A final issue with probe placement in close proximity to the TM is that of clinical feasibility especially with young children who may not tolerate the procedure. Accordingly, placement of the probe in close proximity to the TM does not appear to solve the problem of non-uniform pressure distribution in the canal.

An alternative solution to circumvent the problem of standing-wave pressure cancellations in the ear canal is to decompose total ear-canal pressure into its respective forward- and reverse-traveling components. Input to the ME can then be quantified as the pressure incident on the TM, devoid of any standing-wave cancellations. This decomposition of total pressure is achieved by prior determination of the Thevenin-equivalent source impedance and pressure of the transducer (Sivian and White, 1933; Rabinowitz, 1981; Allen, 1986; Keefe *et al.*, 1992; Neely and Gorga, 1998; Hudde *et al.*, 1999; Farmer-Fedor and Rabbitt, 2002). Knowing the source impedance (Z_s) and pressure (P_s), the impedance of any unknown load (Z_L), whether the ear canal or another cavity, is defined as

$$Z_L = \frac{Z_s P_L}{P_s - P_L}, \tag{2}$$

where P_L is the pressure response of the load. Forward-traveling pressure (P_{forward}) is then defined as

$$P_{\text{forward}} = \frac{1}{2} P_L \left(1 + \frac{Z_0}{Z_L}\right) \tag{3}$$

and reverse-traveling pressure (P_{reverse}) as

$$P_{\text{reverse}} = \frac{1}{2} P_L \left(1 - \frac{Z_0}{Z_L}\right), \tag{4}$$

where Z_0 is the characteristic impedance of a set of brass cavities used during determination of the Thevenin-equivalent characteristics of the transducer. The derivation of the above quantities is more completely outlined in Scheperle *et al.*, 2008 and will not be expanded upon in this report.

The logarithmic equivalent (reference=20 μPa) of the quantity expressed in Eq. (3) is identified in the literature as the *forward pressure level* (FPL) and has been used in several different applications to quantify input to the ME. Scheperle *et al.* (2008) compared the variability in otoacoustic emission (OAE) levels across probe insertion depths for stimuli calibrated in terms of dB sound pressure level (SPL) and dB FPL. As probe depth was varied, it was hypothesized that SPL stimulus calibration would adjust levels to account for standing-wave pressure minimum effects at the location of the probe. Such adjustments would result in variable stimulus levels for each depth and subsequently variable OAE levels. As anticipated, the calibration method had a significant effect on variability of OAE level with FPL calibration resulting in less variability across insertion depth at frequencies above 2 kHz, where standing-wave pressure minima were expected.

Withnell *et al.* (2009) and McCreery *et al.* (2009) expressed *in-situ* sound-level entering the ear at behavioral threshold in terms of both SPL and FPL. Two significant differences were observed between these threshold measures. At frequencies below that of the standing-wave pressure minima, SPL thresholds were approximately 6 dB higher than FPL thresholds. Conversely, at the frequencies of the pressure minima, FPL thresholds exceeded SPL thresholds. These differences were attributed to the constructive and destructive phase interactions which occur at each respective location in the ear canal given the length of the occluded canal. Both of these studies suggest that FPL measures provide a more valid measure of input to the middle ear.

The current paper introduces an additional measure to FPL, termed *integrated pressure level* (IPL), to express sound-level entering the middle ear. Within a hard-walled test cavity, IPL is hypothesized to describe the sound pressure generated at a microphone positioned flush with the terminating surface of the cavity, where forward and reverse pressures waves are believed to be in-phase (for frequencies with $\frac{1}{4}$ wavelengths longer than the diameter of the terminat-

ing surface). Accordingly, the forward- and reverse-traveling pressure wave magnitudes sum and contribute to IPL:

$$P_I = |P_{\text{forward}}| + |P_{\text{reverse}}|. \quad (5)$$

A benefit of IPL is that it permits assessment of the validity of the Thevenin-equivalent source calibration method. The difference between IPL and FPL can be described by the pressure reflectance ($\Re$) and, therefore, power reflectance ($\Re^2$) properties of the test cavity:

$$\Re^2 = \left| \frac{P_{\text{reverse}}}{P_{\text{forward}}} \right|^2. \quad (6)$$

This relationship permits the accuracy with which forward and reverse pressure components have been isolated to be determined. In the current study, this was investigated for a hard-walled test cavity and the human ear canal. For a hard-walled cavity with a rigid termination, complete reflection ($\Re^2 = 1$) of the forward wave was expected to result in a difference between IPL and FPL of approximately 6 dB:

$$\text{IPL} - \text{FPL} = 20 \log_{10}(1 + \Re). \quad (7)$$

The relationship between FPL and IPL described by Eq. (7) should also be applicable to measurements in the ear canal. However, because the canal is terminated by the TM, which is not rigid, complete reflectance at all frequencies is not expected. Rather, maximal power transfer from the ear canal to the ME occurs around 4 kHz while the transfer efficiency is less for both higher and lower frequencies, evidenced by minimal power reflectance at 4 kHz and greater reflectance at surrounding frequencies (Keefe *et al.*, 1993; Voss and Allen, 1994; Feeney and Sanford, 2004; Liu *et al.*, 2008), respectively. This relationship was investigated by measuring *in-situ* behavioral thresholds in SPL from which equivalent FPL and IPL thresholds were derived. The difference between FPL and IPL was hypothesized to mirror the reflectance of the ME, assuming that the validity of the calibration is demonstrated in the hard-walled cavity and will generalize to the ear canal. This generalization is complicated by several factors including the non-uniform plane wave propagation in the canal as well as complex vibratory patterns of the TM at high frequencies where forward and reverse pressure waves may not combine constructively.

Each threshold measure (FPL, IPL, and SPL) was also investigated as a means of quantifying input to the ME system. SPL thresholds are known to be susceptible to standing-wave pressure minima, thereby disqualifying SPL as an accurate measure of sound entering the ME. To compare FPL and IPL, correlations between each measure and properties of sound transmission from the ear canal to the ME, conductance, and power reflectance were performed. Considering the ME as the system and the sound pressure at threshold the input to that system, the measure least dependent on the properties of the ME (power reflectance and conductance) can be considered the preferable means of quantifying input.

II. METHODS

A. Subjects

Twenty-two normal hearing subjects (ages 9–21 years; mean age: 15 years, 3 months) participated in the current study. All subjects passed a hearing screen at 20 dB hearing level across the octave frequencies from 0.5–8 kHz in the test ear. Otoscopy was performed to rule out the presence of excessive cerumen, TM perforations, and pressure equalization tubes. No subjects reported a history of middle-ear surgery. All testing was performed in a sound-treated booth.

Ten of the subjects had participated in a previous study and had custom EMs which were used in the present study. For subjects without an EM, an ear impression was made using Westone BLEND™ impression powder and liquid and subsequently modified to serve as an EM by inserting a size 13 thick TRS tube through the sound bore to terminate flush with the medial end. All testing took place immediately after construction of the temporary EM to ensure that the material did not shrink. All earmolds extended to approximately the second bend of the ear canal. If the earmold contained a vent, it was occluded prior to measurements.

B. Thevenin-equivalent source calibration

An ER-2A (Etymotic Research) tube-phone coupled to the subject's EM was used to deliver stimuli. Pressure responses were measured with an ER-7C (Etymotic Research) probe microphone. Specifically, a silicone probe tube from the ER-7C microphone was taped to the anterior side of the canal portion of the EM to extend 4 mm beyond the termination of the sound bore. This distance was selected to approximate that used in clinical hearing-aid probe microphone measurements and to avoid possible evanescent modes affecting frequencies below 10 kHz (Burkhard and Sachs, 1977).

The Thevenin-equivalent source characteristics of the tube-phone and EM assembly were determined using five brass tubes [8.7 mm outside diameter, 8 mm inside diameter (i.d.)] of varying lengths (83.0, 54.3, 40.0, 25.6, and 18.5 mm) with known impedances (Allen, 1986; Keefe *et al.*, 1992). A 30-mm section of brass tubing was fabricated to function as a coupling device between the source assembly and the five calibration cavities. The coupler was composed of two sections: a 10-mm portion with a slightly larger inner diameter relative to the outer diameter of the brass cavities and a 20-mm portion with the same inner diameter as the brass cavities. Putty was used to couple the subject's EM to the smaller diameter end of the coupler. This process was used to minimize variations that would have occurred if the EM had to be re-seated in each of the five calibration cavities. The pressure response to a broadband-noise stimulus (256 ms, 61 dB SPL re: 2-cc coupler) was measured at the 4-mm probe depth in each cavity. A sound card (Digital Audio Laboratories, Inc. CardDeluxe) and personal computer were used for simultaneous playback and recording of the stimulus and pressure response, respectively, with a sampling rate of 32 kHz. Initial estimates of cavity length were based on the location of the first pressure null for each cavity. Five simultaneous equations were subsequently solved to arrive at

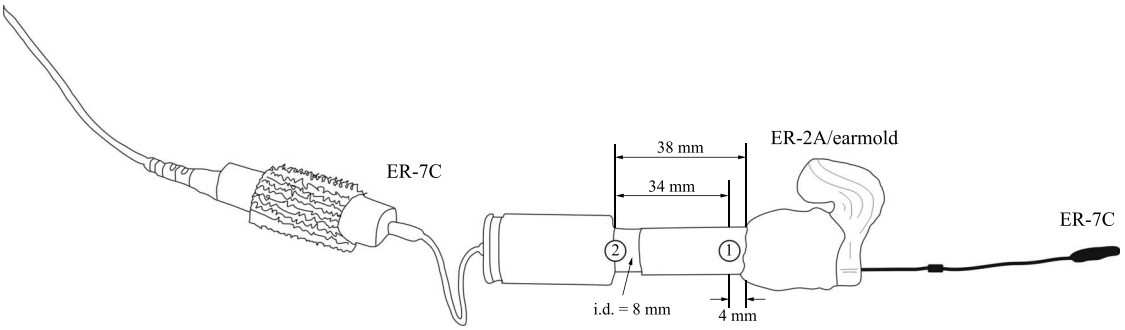


FIG. 1. The apparatus used for source calibration validation permitted measurement of cavity pressure at two locations, represented by (1) and (2), corresponding to the point 4 mm beyond the medial end of the earmold (1) and at the distal end of the cavity (2).

initial estimates of the source impedance (Z_s) and pressure (P_s). Final values of Thevenin source parameters were obtained by iteratively adjusting the cavity-length estimates until a minimum error was achieved. The Thevenin-equivalent source parameters were calculated using each subject's earmold as differences in earmolds were expected to affect both source impedance and pressure.

For each subject, the acoustic load impedance (Z_L) of two cavities (a test cavity and the ear canal) were each separately calculated according to Eq. (2), where Z_s and P_s were the Thevenin source characteristics for the subject's earmold-source assembly. Load pressure was measured in the appropriate load (test cavity and ear canal) at a probe tube 4 mm from the medial end of the earmold (P_{probe}) for the broadband-noise stimulus. Signal presentation and response acquisition were identical to that previously described. From the load pressure response (P_{probe}), integrated (P_I), forward (P_{forward}), and reverse (P_{reverse}) pressures were derived from Eqs. (3)–(5), respectively, in each cavity.

C. Measurement and analysis

1. Test cavity

a. Measurement. A brass tube (8 mm i.d., 17.5 mm length) was used to validate the source calibration for each subject's earmold. The tube was terminated at one end by the source assembly (ER-2A, earmold, brass coupler) and by a plastic probe-tube adapter at the other end. Total length of the test cavity was approximately 38 mm. Pressure measurements within the cavity were taken at two locations. The first location was at the probe-tube placed 4 mm beyond the sound bore exit of the earmold. The second location was at the opposite (*distal*) end of the cavity where an ER-7C probe-tube was routed through a plastic probe-tube adapter that provided a hard, reflective surface. The distal location was expected to provide a SPL measure of pressure primarily resulting from constructive as opposed to destructive phase interactions between incident and reflected pressure components. The test-cavity configuration is shown in Fig. 1.

For each subject's earmold, the load impedance (Z_L) of the test cavity was calculated using the Thevenin characteristics of the source-earmold assembly. Load pressure was measured at the probe microphone 4 mm beyond the termination of the earmold (P_{probe}) for a broadband-noise stimulus. Integrated pressure (P_I), forward pressure (P_{forward}), and reverse pressure (P_{reverse}) were subsequently derived. Following measurement of probe pressure, the ER-7C microphone was attached to the probe tube positioned at the distal end of the cavity and distal pressure (P_{distal}) was measured.

b. Analysis. To determine the accuracy by which pressure components in the test cavity were isolated, integrated cavity pressure (P_I dB IPL) was compared with distal cavity pressure (P_{distal} dB SPL). The difference between these measures was defined as the *validation error* and was calculated as

$$\bar{\Delta}(f) = \frac{1}{22}\sum_{n=1}^{22} 20\log_{10}\left|\frac{P_I(f)}{P_{\text{distal}}(f)}\right|, \tag{8}$$

where f is frequency (Hz) and n is the subject.

To further investigate the accuracy of the pressure decomposition in the test cavity, the relationship between forward cavity pressure (P_{forward} dB FPL) and integrated cavity pressure (IPL) was quantified. It was hypothesized that the difference between IPL and FPL would be approximately 6 dB across all frequencies corresponding to complete reflection of the forward-traveling pressure wave at the termination of the cavity [Eq. (7)].

2. In-situ behavioral thresholds

a. Measurement. The load impedance of the ear canal (Z_L) was calculated for each subject using the Thevenin characteristics of their custom earmold-source assembly and load pressure measured in the ear canal at 4 mm beyond the earmold (P_{probe}) for a broadband-noise stimulus. Load conductance of the ear canal (G_L) was calculated as the real part of the inverse ear-canal impedance (Z_L). The frequency corresponding to maximal load conductance was identified as the subject's *notch frequency*. This frequency corresponds to a pressure null in the ear-canal broadband pressure response (P_{probe}) at the location of the probe microphone where forward- and reverse-traveling pressure waves are in antiphase.

In-situ behavioral thresholds measured at the 4-mm probe microphone location were obtained at 14 frequencies using an automated method of limits with a 5 dB step-size until the standard deviation of the threshold response was less than or equal to 2.5 dB at each frequency. Thresholds (dB SPL) were initially measured at the octave frequencies from 0.5–8 kHz and at 9 and 10 kHz. Additional thresholds

were then obtained at the subject's notch frequency and frequencies one octave below to $\frac{1}{2}$ octave above the notch frequency in $\frac{1}{4}$ octave increments. Thresholds in units of dB FPL were derived by adding the difference between the ear-canal load pressure (P_{probe} dB SPL) and the forward pressure (P_{forward} dB FPL) at each frequency to the corresponding dB SPL threshold. Thresholds in units of dB IPL were derived by adding the difference between the ear-canal load pressure (P_{probe} dB SPL) and the integrated pressure (P_I dB IPL) at each frequency to the corresponding dB SPL threshold. This procedure required thresholds to be measured only once thereby eliminating the variability typically associated with clinical behavioral threshold measurements across multiple trials.

b. Analysis. The sensitivity of each threshold measure to standing-wave pressure minima was investigated by calculating the magnitude of the pressure cancellation at the subject's notch frequency, referred to as *notch depth*. Notch depth was quantified as the difference between (1) the average of the thresholds $\frac{1}{2}$ octave above and below the notch frequency and (2) the threshold at the notch frequency. This frequency range approximates the width of the pressure null in the immediate vicinity of the probe microphone. Threshold curves in dB SPL were expected to exhibit large notch depths, while minimal notch depths were expected for either dB FPL or IPL threshold curves.

The dependency of both FPL and IPL thresholds on ME power reflectance and ear-canal conductance was investigated using linear regression analyses. For power reflectance, correlations were performed as follows:

(1) individually for the 14 frequencies where thresholds were measured,
(2) across two broad frequency ranges with the first including frequencies from 0.5–4 kHz and the second including frequencies >4 kHz, and
(3) across the entire frequency range including frequencies from 0.5–10 kHz.

Correlations with conductance were performed as follows:

(1) individually for 4 frequencies (0.5, 1, 2, and 4 kHz) and
(2) across a broad frequency range including frequencies from 0.5–4 kHz.

The interaction between power reflectance and conductance was also examined across the same groups as regressions performed for conductance, using multivariate linear regressions. Both power reflectance and conductance describe the transmission of sound energy to the middle ear. In quantifying input to the middle ear, the threshold measure least dependent on either of these properties of sound transmission is the preferable choice.

III. RESULTS

A. Test cavity

Figure 2(a) shows cavity pressure at the 4-mm probe (P_{probe} dB), at the distal end (P_{distal} dB), and expressed in IPL (P_I dB). A pressure maximum is present in all responses at the frequency roughly equivalent to the $\frac{1}{2}$ wavelength resonant frequency for the 38-mm tube (measured=4.42 kHz, expected=4.55 kHz). Two pressure minima are present in the 4-mm probe response at 2.56 and 7.42 kHz corresponding to frequencies with $\frac{1}{4}$ and $\frac{3}{4}$ wavelengths of 34 mm (the distance between the probe tube and distal end of the cavity), respectively. At these frequencies, the 4-mm probe pressure underestimates distal pressure by 24 and 11 dB, respectively, as shown in Fig. 2(b). Conversely, IPL is on average within 1 dB of distal pressure from 0.125–10 kHz. IPL accurately estimated distal pressure to within 2 dB from 0.25–10 kHz for 19 of the 22 subjects [illustrated by the shaded region in Fig. 2(b)].

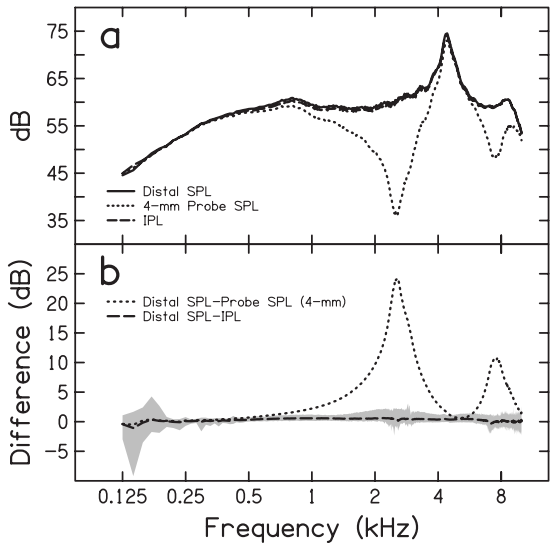


FIG. 2. The top panel shows the mean broadband-noise pressure responses corresponding to SPL measured at the distal terminating end of the cavity, SPL measured at the 4-mm probe microphone position, and IPL. The bottom panel is the difference between distal SPL and the estimated distal pressure. Estimated pressure is equivalent to either the pressure measured at the 4-mm probe position or IPL. For the latter of these, the shaded area represents the deviation from distal pressure between the 5th and 95th percentiles.

Figure 3(a) shows the mean relative contributions to IPL from the forward (FPL) and reverse (RPL) pressure components. FPL and RPL are nearly equivalent through 6 kHz, which corresponds to near unity power reflectance [see Fig. 3(b)]. The resulting IPL is essentially 6 dB greater than FPL across this frequency range. Above 6 kHz, the reflectance at the distal end of the cavity decreases (less of the forward-traveling pressure wave is reflected), resulting in a deviation between FPL and RPL and a difference of less than 6 dB between IPL and FPL.

B. *In-situ* behavioral thresholds

Subject's pure-tone thresholds (dB SPL, FPL, and IPL) are presented in Figs. 4 and 5, mean data are shown in Figs. 6 and 7. Figure 4 presents data from subjects showing distinct threshold notches in the SPL threshold curve, while Fig. 5 present data from subjects showing less distinct threshold notches. These notches are not equally apparent in the corresponding FPL or IPL threshold curves. The frequencies at which these notches occurred are reported in Table I. There

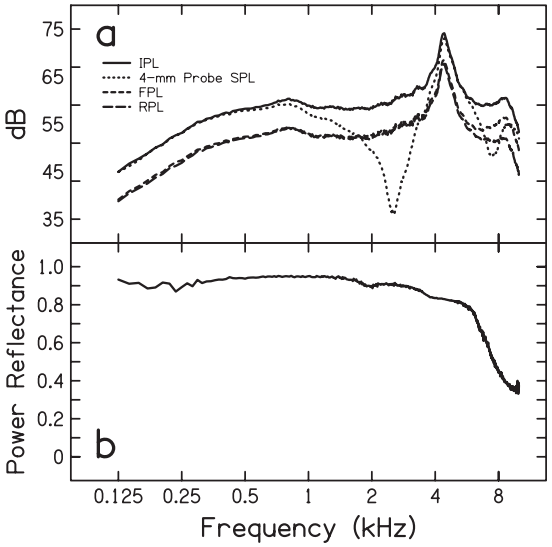


FIG. 3. The top panel presents the incident and reflected pressure components, FPL and RPL, respectively, contributing to the integrated pressure (IPL). SPL at the 4-mm probe position is provided for comparison. The bottom panel is the power reflectance of the ear-canal model derived from the incident and reflected pressures.

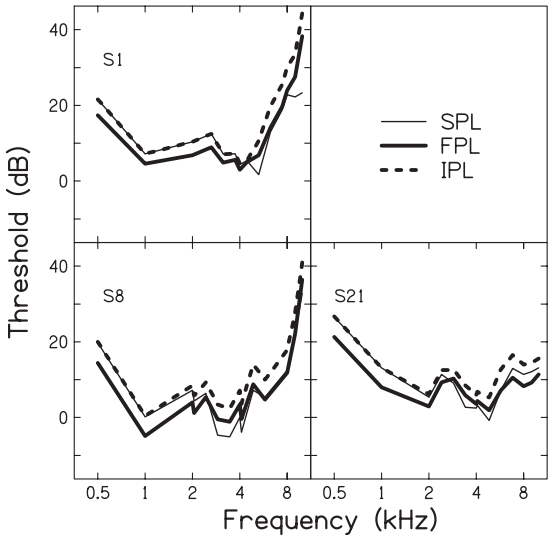


FIG. 5. Individual subject thresholds for subjects with minimal notch depths. Thresholds are expressed in terms of dB SPL, FPL, and IPL.

was considerable variability across subjects regarding notch frequency (mean=5.94 kHz, σ=0.994 kHz). The lowest notch frequency occurred at 4 kHz and the highest at 7.66 kHz. The pressure cancellation at these frequencies was also variable across subjects and is reported in Table I as notch depth. The mean SPL threshold notch depth was −11.18 dB (σ=5.15 dB) and ranged from −0.35 dB for subject S4 to −19.60 dB for subject S6. FPL and IPL mean threshold notch depths were −0.78 dB (σ=4.99 dB) and −0.46 dB (σ=5.07 dB), respectively. Notch depths ranged from −10.05 to 6.95 dB for FPL and from −9.85 to 7.51 dB for IPL. The occurrence of positive notch depths for FPL and

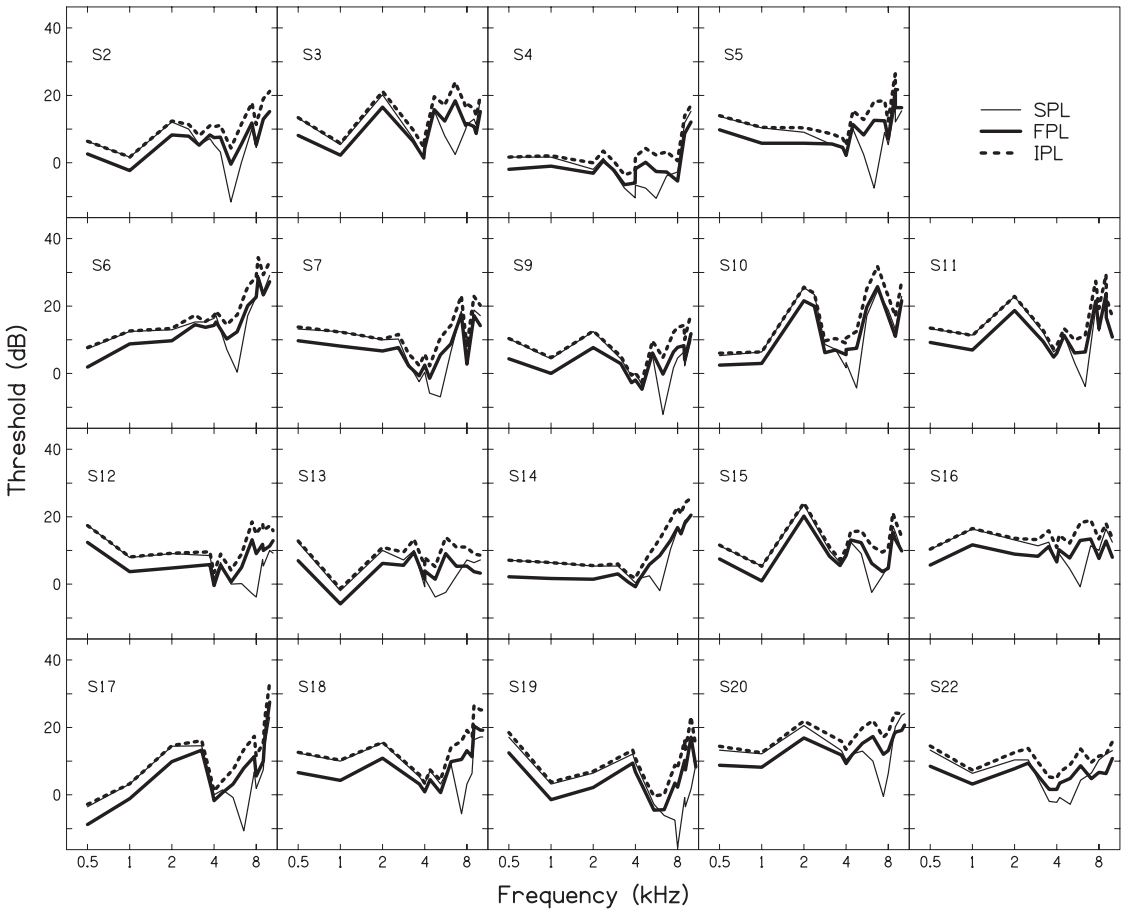


FIG. 4. Individual subject thresholds for subjects with maximal notch depths. Thresholds are expressed in terms of dB SPL, FPL, and IPL.

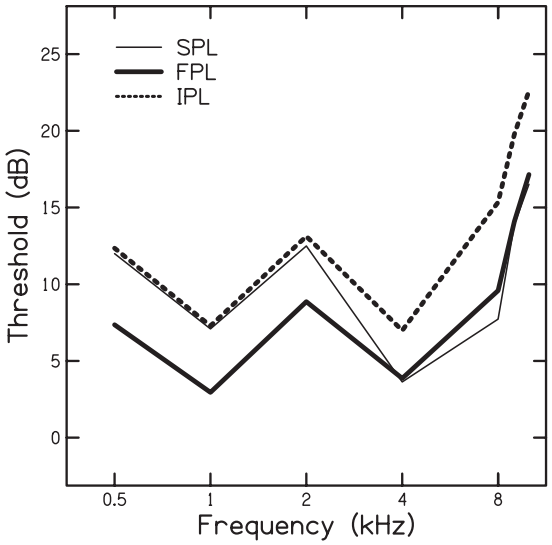


FIG. 6. Mean dB SPL, FPL, and IPL thresholds for the octave frequencies 0.5–8 kHz and 9 and 10 kHz.

IPL is a by-product of the calculation and indicates that the threshold pressure at the notch frequency was greater than the threshold pressures at the surrounding frequencies. These instances may be attributed to either genuine changes in auditory sensitivity across the frequency range over which notch depth was calculated or the subject changing their detection criteria.

A one-way analysis of variance was performed to determine the effects of threshold measure (SPL, FPL, and IPL) on notch depth. A significant effect was demonstrated across all measures [$F(2,63)=31.81$, $p<0.001$, and $\eta^2=0.5024$]. *Post-hoc* testing using Bonferroni correction for multiple comparisons suggested that ($p<0.05$) SPL notch depth was significantly different from both FPL and IPL notch depths, but the difference between FPL and IPL notch depth was not significant. At frequencies below that of the notch frequency, SPL and IPL thresholds were nearly identical. Across this same range, FPL thresholds were on average 4–6 dB less than either SPL or IPL thresholds. Across the entire frequency range, FPL thresholds were always lower than corresponding IPL thresholds; however, the difference did vary across frequency. In general, the smallest difference between measures (2–3 dB) was around 3–4 kHz with larger differences (4–6 dB) for adjacent frequencies. SPL and FPL thresholds approached each other at frequencies above the notch frequency.

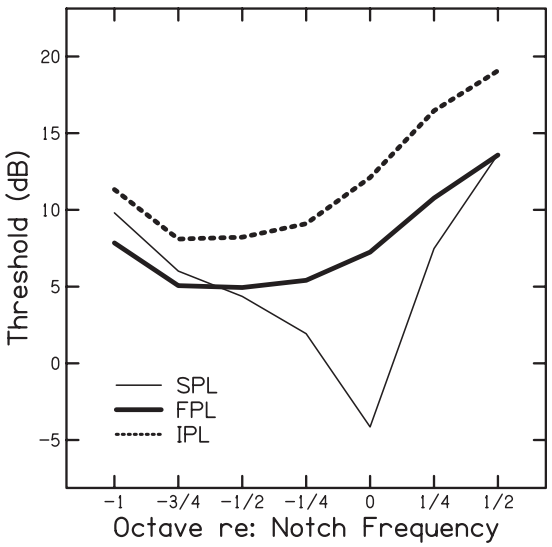


FIG. 7. Mean dB SPL, FPL, and IPL thresholds for frequencies surrounding the notch frequency. The exact values of these frequencies varied across subjects and were dependent on the pressure minimum frequency.

TABLE I. Each subject's notch frequency and notch depths for the threshold measures (SPL, FPL, and IPL) are provided. From the notch frequency, distance between the 4-mm probe microphone and TM was estimated as the $\frac{1}{4}$ wavelength distance.

ID	Notch (Hz)	SPL notch depth (dB)	FPL notch depth (dB)	IPL notch depth (dB)
S1	5279	−9.90	−4.55	−4.54
S2	5282	−19.30	−10.05	−9.85
S3	6610	−10.90	4.65	5.35
S4	4734	−0.35	4.55	3.86
S5	6352	−15.20	3.35	4.31
S6	5856	−19.60	−6.60	−6.20
S7	5173	−16.30	−4.50	−3.87
S8	4097	−4.05	−2.20	−2.64
S9	6296	−12.75	−1.75	−0.91
S10	4749	−17.45	−8.20	−7.88
S11	6392	−15.90	−5.90	−5.02
S12	7484	−6.45	6.95	7.51
S13	5657	−6.85	4.45	4.46
S14	5992	−10.95	−0.50	0.17
S15	6105	−11.35	−2.70	−1.80
S16	5884	−10.35	4.10	4.23
S17	6524	−12.40	4.75	5.57
S18	7346	−10.40	4.50	5.13
S19	7658	−5.70	0.65	0.91
S20	7405	−16.2	−4.65	−4.65
S21	4828	−8.55	−5.15	−5.94
S22	4993	−4.95	1.7	1.63
Mean	5940.73	−11.18[a]	−0.78	−0.46
Std. dev.	993.56	5.15	4.99	5.07

[a]SPL notch depth (dB) was significantly different ($p<0.05$) from both FPL and IPL notch depths (dB).

Mean and individual power reflectance functions are shown in Fig. 8. Considerable variability is apparent across subjects at all frequencies. The mean data demonstrate a gradual decrease in reflectance until a minimum reflectance of 0.2 is reached at 4 kHz. Reflectance increases steeply across the frequency range 4–7 kHz until near complete reflectance. Above 7 kHz, reflectance decreases to 0.8 by 10 kHz. As mentioned previously, correlations between thresholds (FPL and IPL) and power reflectance were performed to determine the dependency of each measure on the middle-ear system. The measure best suited for quantifying input to the ME would be least dependent on power reflectance. Linear regressions between thresholds (FPL and IPL) and power reflectance (dB) at each of the 14 frequencies where thresholds were measured revealed that power reflectance was a significant predictor of FPL ($p=0.0032$, $r=0.5989$) and IPL ($p=0.0007$, $r=0.6659$) thresholds only at the notch fre-

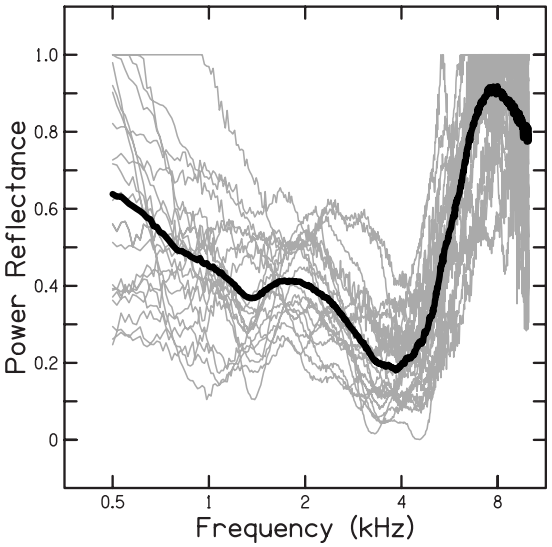


FIG. 8. Individual subject's power reflectance curves and mean curve. Individual reflectance curves are represented by the thin gray lines. Mean data are represented by the thick black line.

quency. Fisher's z-test ($p<0.05$) revealed no significant difference between power reflectance as a predictor of FPL compared to IPL thresholds.

Across the frequency range including frequencies from 0.5–4 kHz, power reflectance was a significant predictor of FPL ($p=0.0494$, $r=0.1664$) and IPL ($p=0.0001$, $r=0.3313$) threshold. Fisher's z-test revealed a significant difference between correlations ($p=0.036$). For the frequency range including all frequencies above 4 kHz, power reflectance was a significant predictor of both FPL ($p<0.001$, $r=0.3286$) and IPL ($p<0.001$, $r=0.4353$) thresholds. Fisher's z-test ($p<0.05$) revealed no significant difference between correlations. When the frequency range over which the regression was performed was expanded to include frequencies from 0.5–10 kHz, power reflectance was a significant predictor of both FPL ($p<0.001$, $r=0.3872$) and IPL ($p<0.001$, $r=0.5095$) thresholds. Fisher's z-test ($p<0.05$) revealed no significant difference between the two correlations. Based on correlations with power reflectance alone, both FPL and IPL were significantly correlated across several frequency ranges; however, a significantly stronger correlation was exhibited between IPL and power reflectance for frequencies from 0.5–4 kHz suggesting that FPL is better suited for quantifying ME input.

Correlations between thresholds and ear-canal conductance were also performed to investigate the dependency of each measure on sound transmission properties from the ear canal to the middle ear. Similar to correlations with power reflectance, the measure least dependent on conductance would be best suited for quantifying ME input. A significant correlation ($p<0.05$) between ear-canal conductance and either FPL or IPL threshold was not demonstrated at 0.5, 1, 2, or 4 kHz (correlations above 4 kHz were not examined). Conductance was significantly correlated with both FPL ($p=0.0023$, $r=-0.2556$) and IPL ($p<0.0001$, $r=-0.3368$) when the regression was performed across the broader frequency range including frequencies from 0.5–4 kHz. Fisher's z-test ($p<0.05$) revealed no significant difference between the two correlations. From correlations with conductance alone, neither FPL nor IPL can be predicted to be the better measure for quantifying ME input.

As a final test to determine the dependency of threshold measure on the ME system, the interaction between power reflectance and conductance as a predictor of threshold was determined. This interaction was not a significant predictor ($p<0.05$) of either threshold measure at 0.5, 1, 2, or 4 kHz (correlations above 4 kHz were not examined). However, when the frequency groups were combined to form a single group from 0.5–4 kHz, reflectance and conductance were significant predictors of IPL threshold ($p<0.0001$) accounting for 14.9% of the variance. Accordingly, FPL appears to be less dependent on the ME and a better measure for specifying ME input.

IV. DISCUSSION

A. Validation of the source calibration

The first objective of the study was to investigate the validity of the Thevenin-equivalent source calibration procedure that decomposes total pressure into forward and reverse components. IPL, derived from the pressure response measured at 4 mm into the test cavity, accurately estimated the pressure at the distal end of the cavity to within 2 dB from 0.25–10 kHz. The agreement between these measures affirmed the initial hypothesis that pressure at the distal end of the cavity, across the terminating surface, would be equivalent to an in-phase interaction between the forward- and reverse-traveling pressure components and would, therefore, be equivalent to IPL. As mentioned previously, IPL is a derivation of pressure achieved by summing the magnitudes of the forward and reverse pressure components.

Within the test cavity, FPL and RPL demonstrated that complete reflectance at the distal end of the cavity did not occur. This was unexpected since the cavity was constructed with a uniform diameter, hard walls, and a hard, flat terminating surface. Nearly complete reflectance was observed for frequencies extending to 4 kHz; however, reflectance dropped sharply above this frequency range. It is likely that the termination of the cavity was not as rigid as originally anticipated. The coupling between the brass tube and the surface of the probe-tube adapter (used to terminate the cavity) was achieved using putty and may have resulted in the tube not being pressed flush against the adapter's surface. Additionally, it is possible that the change in diameter of the transmission tube from the 8 mm cavity to the 1 mm distal probe tube could have contributed to the less than complete reflectance above 4 kHz. Although the reflectance of the cavity at these frequencies was unexpected, IPL and distal pressure still showed excellent agreement.

If the test-cavity results are extended to application in the human ear canal, IPL offers a needed alternative to direct measures of SPL at the TM. Current SPL measures made at locations in the ear canal away from the TM are traditionally interpreted as the pressure generated at the TM. At frequencies sufficiently below the $\frac{1}{4}$ wavelength of the canal, an in-phase interaction between pressure components occurs

and is considered an accurate estimate of TM pressure. Within the test cavity, this type of interaction was seen at all frequencies when a probe microphone was flush with the distal end of the cavity. It is tempting to think of the test cavity as a model of the ear canal with the distal end of the cavity being synonymous with the TM. IPL might therefore be interpreted as the pressure generated at the TM where forward and reverse pressure components constructively interact. This assumes, however, that the TM acts as the ER-7C microphone, which may be too simplistic. The ER-7C has a flat frequency response through 10 kHz; however, the eardrum is known to exhibit various modes of vibration across its surface in response to different frequencies (Tonndorf and Khanna, 1972). Khanna and Stinson (1985) further demonstrated that the spatial distribution of pressure across the surface of the cat TM is not uniform, particularly above 10 kHz, but will be dependent on the location of the probe.

B. Quantifying input to the middle ear

The second goal of the study was to determine whether SPL, FPL, or IPL is best for quantifying input to the middle ear. *In-situ* SPL measures have consistently been shown to misrepresent the pressure entering the middle ear due to standing-wave pressure minima in the ear canal. The effects of these pressure minima were evidenced in the current study by the large notches present in the SPL threshold curves. These same notches in threshold were not present when threshold was expressed in either FPL or IPL, demonstrating the insensitivity of these measures to standing-wave pressure minima in the canal. Many subjects had apparent notches in FPL and IPL measures; however, the SPL notch was considerably greater when quantified in terms of notch depth. FPL and IPL notches imply that there is either a genuine change in hearing sensitivity or variability in the subject's threshold identification criteria. Results also demonstrate that the effect of standing-wave pressure cancellation is not localized to one specific frequency. SPL and IPL thresholds were essentially identical until the notch frequency where SPL threshold curves deviated from IPL due to the pressure minima. Above the frequency of the minima, however, SPL continued to deviate from IPL. This finding is in agreement with that of previous investigators showing that pressure cancellation is not only localized to the $\frac{1}{4}$ wavelength frequency but also to adjacent frequencies (Gilman and Dirks, 1986).

Neely and Gorga (1998), Withnell *et al.* (2009), and McCreery *et al.* (2009) similarly showed the susceptibility of *in-situ* SPL behavioral thresholds to standing waves and the subsequent misrepresentation of sound entering the middle ear. Withnell *et al.* (2009) and McCreery *et al.* (2009) also demonstrated the insensitivity of FPL to standing-wave pressure minima. Both conclude that expressing input to the ME in FPL is preferable to traditional SPL. Results from the current study agree with those of others in that FPL is preferable to SPL. Unique to this study is the use of IPL which also appears to be preferable to SPL in quantifying middle-ear input.

The relationship between FPL and IPL was discussed earlier and modeled in Eq. (7): IPL thresholds will exceed FPL thresholds according to the power reflectance properties of the ME. Figure 9 shows the average power reflectance from the current study compared to that measured by other investigators. It was shown earlier that power reflectance was variable across subjects; however, the average data agrees relatively well with that expected for the ME. Voss *et al.* (2008) commented on the variability in reflectance measures and attributed some of it to differences in middle-ear cavity volume across individuals. Despite the variability, power reflectance suggests an increasing difference between FPL and IPL thresholds for frequencies adjacent to 4 kHz, as observed in the results. Accordingly, IPL thresholds reflect the contribution of the magnitude of the reverse-traveling pressure wave.

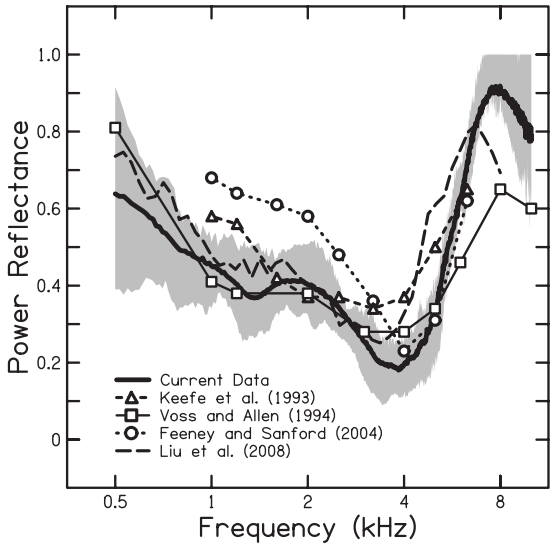


FIG. 9. Mean power reflectance of the current study compared to that in the literature. Shaded area indicates data between the 25th and 75th percentiles. Data are shown from Keefe *et al.*, 1993; Voss and Allen, 1994; Feeney and Sanford, 2004; and Liu *et al.*, 2008.

To determine whether FPL or IPL was better suited to quantify ME input, correlations between each measure and both power reflectance and conductance were performed. The rationale for this is that when specifying the input to a system, that input should not be dependent on the properties of the system. Both power reflectance and conductance describe the transmission of sound energy from the ear canal to the middle ear and the measure least dependent on these will be a more accurate indicator of the input to the middle-ear system. Based on this, FPL appears to be the preferable measure. Specifically, the correlation between power reflectance and IPL was significantly higher than that with FPL across the frequency range 0.5–4 kHz. Additionally, the interaction between power reflectance and conductance from 0.5–4 kHz was a significant predictor of IPL thresholds but not FPL thresholds. Both of these suggest that IPL is more dependent on the properties of the middle-ear system than is FPL.

C. Clinical implications

The motivation for the present work came from recent studies demonstrating the importance of high-frequency amplification (above 4–5 kHz) for the perception of fricatives

by children with hearing impairment (Stelmachowicz *et al.*, 2001, 2004, 2007). The effective bandwidth of current hearing aids, however, extends only to approximately 6 kHz. As the development of hearing aids with extended bandwidths to 10 kHz is expected, real-ear verification of high-frequency gain becomes complicated. As demonstrated in the current study and in the work of others, traditional SPL probe microphone measures will underestimate the pressure at the TM when the probe is in the vicinity of a standing-wave pressure minimum. Compensation for this pressure cancellation at the probe microphone by increasing hearing-aid gain to meet amplification targets has the potential to result in overamplification.

Underestimation in pressure at the TM can be expected for frequencies as low as 4 kHz when a typical probe-tube insertion depth is used. Accordingly, current hearing-aid fitting procedures are susceptible to deleterious effects resulting from standing-wave pressure minima. From the present data, the magnitude of this underestimation of pressure can be quantified as the difference between IPL and SPL thresholds (see Fig. 7) and approaches 16 dB at the notch frequency. Hearing-aid gain might therefore be expected to exceed the recommended amplification target by a similar amount if the clinician mistakenly compensates for this underestimation. It is more appropriate to quantify this underestimation as the difference between IPL and SPL as opposed to FPL and SPL since the probe microphone is sensitive to the interaction of the incident and reflected pressure waves. Had the cavity length been shorter such that no pressure cancellations were seen across the frequency range in this study, IPL and SPL would have been equivalent since both would have resulted from constructive phase interactions. Additionally, the difference between FPL and SPL thresholds at frequencies where SPL and IPL were equivalent (low frequencies) should not be interpreted as SPL overestimating pressure since SPL takes into account both the forward and reflected pressure waves while FPL only considers the forward wave. Rather, the agreement between IPL and SPL at the low frequencies suggests that SPL accurately estimates input to the ME at those frequencies. Besides underestimating ME input at the notch frequencies, SPL also underestimates, to a lesser extent, input pressure at frequencies surrounding the notch.

The use of either FPL or IPL in clinical real-ear probe microphone measurements offers a means of verifying high-frequency amplification since both are insensitive to ear-canal standing-wave pressure minima. FPL represents the sound-level delivered to the TM by the hearing aid. IPL represents the sound-level assumed to exist across the surface of the TM. Future considerations before implementation of either of these measures in a clinical procedure include the terms in which amplification targets and thresholds are expressed. That is, if ear-canal sound-level is quantified in FPL (or IPL), it would be most accurate to also express both amplification targets and thresholds using the same sound-level measure.

V. CONCLUSIONS

The source calibration used in the present study and the earlier study by Scheperle *et al.* (2008) can be expected to accurately decompose incident and reflected pressure waves in the ear canal. FPL and IPL are derived sound-level measures resulting from the source calibration and, more specifically, the Thevenin-equivalent source characteristics (impedance and pressure). Both are equally insensitive to standing-wave pressure minima in cavities and therefore more accurately estimate sound-level entering the middle ear than traditional SPL probe measurements. FPL quantifies the incident pressure wave propagating toward the TM. IPL quantifies the summation of the incident and reflected pressure amplitudes and is assumed to reflect the pressure integrated across the surface of the TM. Alternatively, FPL may be a preferred means of quantifying input to the ME since it is less dependent on sound transmission properties from the ear canal to the middle ear. The use of either method in verification of hearing-aid output is recommended over traditional SPL measurements. It is likely that implementation of either measure in clinical hearing-aid applications will necessitate conversion of SPL thresholds and amplification targets into equivalent sound-level measures (i.e., FPL or IPL). These conversions could be easily implemented in hearing-aid analysis and fitting software.

ACKNOWLEDGMENTS

The authors thank Skip Kennedy for his assistance in the preparation of Fig. 1. This study was supported by grants from the NIH-NIDCD (Grant Nos. T35-DC8757, R01-DC8318, and R01-DC4300). The first author (J.D.L.) is a fourth-year student at the University of Iowa.

Allen, J. B. (**1986**). "Measurement of eardrum acoustic impedance," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 44–51.

Burkhard, M. D., and Sachs, R. M. (**1977**). "Sound pressure in insert earphone coupler and real ears," J. Speech Hear. Res. **20**, 799–807.

Chan, J. C. K., and Geisler, C. D. (**1990**). "Estimation of eardrum acoustic pressure and of ear canal length from remote points in the canal," J. Acoust. Soc. Am. **87**, 1237–1247.

Dirks, D. D., and Kincaid, G. E. (**1987**). "Basic acoustic considerations of ear canal probe measurements," Ear Hear. **8**, 60S–67S.

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.

Feeney, M. P., and Sanford, C. A. (**2004**). "Age effects in the human middle ear: Wideband acoustical measures," J. Acoust. Soc. Am. **116**, 3546–3558.

Feigin, J. A., Kopun, J. G., Stelmachowicz, P. G., and Gorga, M. P. (**1989**). "Probe-tube microphone measures of ear-canal sound pressure levels in infants and children," Ear Hear. **10**, 254–258.

Gilman, S., and Dirks, D. D. (**1986**). "Acoustics of ear canal measurement of eardrum SPL in simulators," J. Acoust. Soc. Am. **80**, 783–793.

Hellstrom, P.-A., and Axelsson, A. (**1993**). "Miniature microphone probe tube measurements in the external auditory canal," J. Acoust. Soc. Am. **93**, 907–919.

Hudde, H., Engel, A., and Lodwig, A. (**1999**). "Methods for estimating the sound pressure at the eardrum," J. Acoust. Soc. Am. **106**, 1977–1992.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). "Ear-canal impedance and reflection coefficient in human infants and adults," J. Acoust. Soc. Am. **94**, 2617–2637.

Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). "Method to measure acoustic impedance and reflection coefficient," J. Acoust. Soc. Am. **91**, 470–485.

Khanna, S. M. and Stinson, M. R. (**1985**). "Specification of the acoustical

input to the ear at high frequencies," J. Acoust. Soc. Am. **77**, 577–589.
Larson, V. D., Studebaker, G. A., and Cox, R. M. (**1977**). "Sound-levels in a 2-cc cavity, a zwislocki coupler, and occluded ear canals," J. Am. Aud. Soc. **3**, 63–70.
Liu, Y., Cohn, E., Ellison, J. C., Fitzpatrick, D. F., Gorga, M. P., Gortemaker, M., Sanford, C. A., and Keefe, D. H. (**2008**). "Wideband acoustic ear-canal reflectance, including wideband tympanometry and acoustic-reflex thresholds: System development and results on children with middle-ear fluid and adults," poster session presented at the 31st Midwinter Research Meeting of the Association for Research in Otolaryngology, Phoenix, AZ.
McCreery, R. W., Pittman, A. L., Lewis, J. D., Neely, S. T., and Stelmachowicz, P. G. (**2009**). "Use of forward pressure level (FPL) to minimize the influence of acoustic standing waves during probe-microphone measurements," J. Acoust. Soc. Am. **126**, 15–24.
Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound-level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.
Nelson Barlow, N. L., Auslander, M. C., Rines, D., and Stelmachowicz, P. G. (**1988**). "Probe-tube microphone measures in hearing-impaired children and adults," Ear Hear. **9**, 243–247.
Rabinowitz, W. M. (**1981**). "Measurement of the acoustic input immittance of the human ear," J. Acoust. Soc. Am. **70**, 1025–1035.
Sachs, R. M., and Burkhard, M. D. (**1972**). "Insert earphone pressure response in real ears and couplers," J. Acoust. Soc. Am. **52**, 183.
Scheperle, R. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2008**). "Influence of *in-situ*, sound-level calibration on distortion-product otoacoustic emission variability," J. Acoust. Soc. Am. **124**, 288–300.
Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.
Sivian, L. J., and White, S. D. (**1933**). "On minimum audible sound fields," J. Acoust. Soc. Am. **4**, 288–321.
Stelmachowicz, P. G., Lewis, D. E., Choi, S., and Hoover, B. (**2007**). "Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children," Ear Hear. **28**, 483–494.
Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., and Lewis, D. E. (**2001**). "Effect of stimulus bandwidth on the perception of /s/ in normal- and hearing-impaired children and adults," J. Acoust. Soc. Am. **110**, 2183–2190.
Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., Lewis, D. E., and Moeller, M. P. (**2004**). "The importance of high-frequency audibility in the speech and language development of children with hearing loss," Arch. Otolaryngol. Head Neck Surg. **130**, 556–562.
Stinson, M. R. (**1985**). "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 Shaw, E. A. G. (**1982**). "Wave effects and pressure distribution in the human ear canal," J. Acoust. Soc. Am. **71**, S88.
Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.
Tonndorf, J., and Khanna, S. M. (**1972**). "Tympanic-membrane vibrations in human cadaver ears studied by time-averaged holography," J. Acoust. Soc. Am. **52**, 1221–1233.
Voss, S. E., and Allen, J. B. (**1994**). "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.
Voss, S. E., Horton, N. J., Woodbury, R. R., and Sheffield, K. N. (**2008**). "Sources of variability in reflectance measurements on normal cadaver ears," Ear Hear. **29**, 651–665.
Wiener, F. M., and Ross, D. A. (**1946**). "The pressure distribution in the auditory canal in a progressive sound field," J. Acoust. Soc. Am. **18**, 401–408.
Withnell, R. H., Jeng, P. S., Waldvogel, K., Morgenstein, K., and Allen, J. B. (**2009**). "An *in situ* calibration for hearing thresholds," J. Acoust. Soc. Am. **125**, 1605–1611.

Postnatal development of sound pressure transformations by the head and pinnae of the cat: Binaural characteristics

Daniel J. Tollin[a] and Kanthaiah Koka
Department of Physiology and Biophysics, University of Colorado Health Sciences Center, Aurora, Colorado 80045

(Received 11 August 2009; revised 28 September 2009; accepted 1 October 2009)

There are three acoustical cues to sound location: Interaural time differences (ITDs), interaural level differences (ILDs), and monaural spectral shape cues. During development, the increasing interaural distance and pinnae size associated with a growing head and pinnae result in localization cues that change continuously until maturation is complete. Here the authors report measurements of both the physical dimensions of the head and pinnae, as well as acoustical measurements of the binaural localization cues of cats aged 1.3 weeks to adulthood. For a given source location, ILD magnitude tended to increase with both frequency and age. Moreover, the range of significant ILD production (~10 dB) shifted with age from higher to lower frequencies. ITD magnitude increased with age. Partial correlation analyses revealed that increasing pinnae size accounted for ~31% of the variance in the development of ILDs while increasing head size accounted for virtually none. On the other hand, increases in both the head and pinnae sizes contributed to the development of the ITD cues accounting for ~71% and ~25% of the variance, respectively. ILD and ITD cues in cats reach maturity by ~16 and ~22 weeks, respectively, which match the time period over which the pinnae and head dimensions reach maturity.
 [DOI: 10.1121/1.3257234]

PACS number(s): 43.64.Ha, 43.66.Pn, 43.66.Qp [JCM] Pages: 3125–3136

I. INTRODUCTION

The head and the pinnae are fundamental in shaping the spatial-location dependence of the spectral and temporal aspects of sounds that ultimately arrive at the tympanic membrane (Tollin and Koka, 2009; Kuhn, 1987). An important consequence of the acoustic directionality of the head and pinnae is their role in establishing the cues to sound source location. The three primary cues for location are generated by the spatial- and frequency-dependent reflections and diffractions of the propagating sound waves by the head and pinnae (Tollin, 2008). Interaural time differences (ITDs) arise because the distance of the path of sound to the two ears differs. Interaural level differences (ILDs) result jointly from the amplification effects of the pinnae ipsilateral to the sound source and the acoustic shadowing effect of the head and contralateral pinna that occurs primarily for high-frequency sounds. Finally, monaural spectral shape cues arise from differential reflection and diffraction of pressure waveforms from sounds originating from different directions by the head, torso, and pinnae.

The spatial- and frequency-dependence of the monaural and binaural cues to location are well documented in the adult cat (Wiener *et al.*, 1966; Middlebrooks and Pettigrew, 1981; Calford and Pettigrew, 1984; Irvine, 1987; Martin and Webster, 1989; Musicant *et al.*, 1990; Rice *et al.*, 1992; Young *et al.*, 1996; Xu and Middlebrooks, 2000; Phillips *et al.*, 1982; Tollin and Koka, 2009). Moreover, regardless of the species, the magnitudes of the cues to location and the manner in which they change with location are dependent on the physical size and dimensions of the head and pinnae (Shaw, 1974; Middlebrooks, 1999; Xu and Middlebrooks, 2000; Schnupp *et al.*, 2003; Maki and Furukawa, 2005; Tollin and Koka, 2009). Inter-individual differences in head and pinnae size and morphology are the basis for individual differences in the cues to location. These facts also create a challenge during development where the growing sizes of the head and pinnae in mammals increase dramatically from birth, changing not only the magnitude of the acoustical transformations but also the relationship between the cues and sound location.

Cats have been a common model for anatomical, physiological, and behavioral studies of auditory system development [see reviews by Kitzes (1990) and Walsh and McGee (1986)]. Their auditory system is relatively immature at birth and their physical size relative to other species (e.g., rat, mouse, and gerbil) permits good access to the neural structures of interest. A wealth of knowledge exists on the anatomy, physiology, and behavior of the adult cat binaural auditory system to which developmental data can be compared (Irvine, 1986). However, aside from some spatially and spectrally sparse measurements of the development of the ILD cues in kittens by Moore and Irvine (1979) there has been no systematic study of the development of the complete complement of acoustical cues and their relationship to the development of the linear dimensions of the head and pinnae. In this paper we investigate the development of the physical dimensions of the head and pinnae in the cat from the onset of hearing through adulthood and the concomitant changes in the binaural acoustical transformations of sound pressure at the ear. The development of the monaural transformations based on these experiments has been published (Tollin and Koka, 2009).

[a]Author to whom correspondence should be addressed. Electronic mail: daniel.tollin@ucdenver.edu

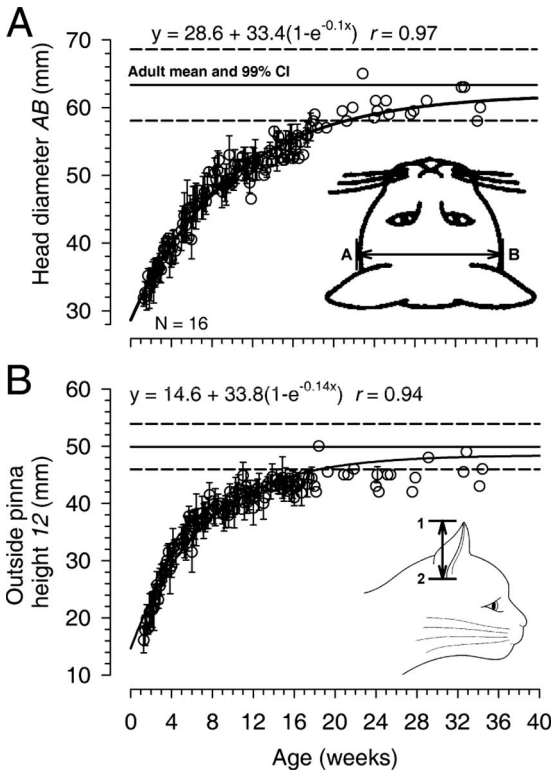


FIG. 1. Developmental growth of the head and pinnae of the cat. The two measured dimensions relevant to the development of the binaural acoustical cues to location are shown in the insets of (A) and (B). (A) Head diameter *AB*. (B) Outside pinna height 1-2. The measured data are from 16 animals. Data points with error bars indicate the across-animal mean ±SD of the measured dimension at that age. Data points without error bars indicate single animal measurements at that age. In each panel, the solid and dashed horizontal lines indicate the mean and 99% confidence interval, respectively, of the measured dimension taken from 12 adult animals. The parameters of the best fitting growth curve for each measured dimension is displayed in each panel along with the correlation coefficient (*r*) for the fit. Figures are replotted from Tollin and Koka (2009) with permission.

II. METHODOLOGY

A. Animal preparation and experimental setup

Twenty-nine domestic short-hair cats (Liberty Research, Waverly, NY) were used in this study. Most animals were female (5/29 were male). All animals had clean external ears and ear canals. The weight and linear measurements of head diameter and pinnae height and width of each animal were taken immediately before each experiment and were reported in our companion study [parts of Fig. 1 in this paper are replotted from Tollin and Koka, 2009; see insets, Figs. 1(A) and 1(B)]. Methods for these growth measurements and the acoustic measurement procedures are detailed in Tollin and Koka, 2009. All acoustic measurements reported here were computed from those measured in both the left and right ears from 289 to 625 different spatial locations. All surgical and experimental procedures complied with the guidelines of the University of Colorado Health Sciences Center Animal Care and Use Committees and the National Institutes of Health.

B. Data processing and data analysis

For data collected here, the free-field-to-eardrum impulse responses for each ear and each location were calculated as described in Tollin and Koka, 2009. These impulse responses were then truncated to 512 points (5.12-ms duration) by a 512-point Hanning window where the center of the window was set to coincide with the point of maximum amplitude in the impulse response. This windowing procedure removes the small-amplitude reflections that may be contained in the impulse response. Moreover, these raw impulse responses contain not only the desired spectral and temporal modifications of the sounds due to the head and pinnae, but also the loudspeaker, microphone, and measurement-system frequency responses. These equipment-related frequency responses were removed from each in-ear measurement by dividing the Fourier transforms of each of the impulse responses by that of appropriate calibration measurements made for each loudspeaker by placing the microphone tips at the location corresponding to where the center of the head would be. These resulting functions are referred to as the head-related transfer functions (HRTFs), as they capture for each ear and loudspeaker location the acoustical gains and delays introduced by the head and the pinnae. However, the resulting HRTF can be highly dependent on the exact placement of the tip of the probe tube microphone in the ear canal relative to the tympanic membrane. To reduce the confounding effects of the probe tube placement in the ear canal, for each ear the directional transfer functions (DTFs) were then calculated from the HRTFs by dividing the HRTF made at each spatial location by the geometrical mean of all the measured HRTFs across all measurement locations for that ear. The spectral features resulting from the exact placement of the probe tube microphone in the ear canal are expected to be similar for all measurement locations (i.e., they are not dependent on spatial location), so this "common" spectral feature is removed from the HRTFs, resulting in the DTFs (Middlebrooks and Green, 1990). In essence, the DTFs are the sound source direction-dependent components of HRTFs.

The amplitude spectra of the DTFs were calculated after the spectra were passed through a bank of 350 bandpass filters, the center frequencies of which were spaced at intervals of 0.0143 octave spanning from 1 to 32 kHz. The 3 dB bandwidth of filters was held constant across all frequencies at 0.12 octaves, and the upper and lower slopes of the filters fell off at ~105 dB/octave. These filters have properties similar to the bank of bandpass filters that have been used elsewhere to filter DTFs (Middlebrooks, 1999; Xu and Middlebrooks, 2000; Schnupp *et al.*, 2003). Only data up to 32 kHz were used here as the signal-to-noise ratio was poor for higher frequencies in some animals, particularly for sources contralateral to the ear being measured. This latter aspect was especially critical when computing the ILD cues as described below.

Two binaural cues to sound location were studied here. The ILD spectrum was derived by computing the differences (in decibels) in the DTFs, frequency by frequency between right and left ears at all elevations and for all azimuth angles. A positive ILD indicates that the decibel level at the right ear was higher than the decibel level at the left ear. The ILDs for particular frequencies and locations were extracted from the ILD spectra. The low-frequency ongoing ITDs in the fine structure of the sounds in this paper were measured by com-

puting for each spatial location the cross-correlation of the raw head-related impulse responses of the left and right ears (i.e., before conversion to DTFs) and by finding the delay corresponding to the maximum point in the cross-correlation function. These impulse responses were first low-pass filtered at 3.5 kHz before cross-correlation. This process was repeated for each location.

For spatial plotting purposes, the data were displayed as Mollweide projections (Bugayevskiy and Snyder, 1995). In each of these projections the nose of the animal is considered to be projecting out of the page at 0° azimuth and 0° elevation, as if the animal were looking at the reader. The Mollweide projections were plotted for elevations from −30° to +90° and all azimuths from −180° to +180°. Unweighted linear regressions were performed using the curve fitting package in SIGMAPLOT (V10, Systat Software, Inc., Richmond, CA).

III. RESULTS

Results are based on data from 29 animals (5/29 were male). Acoustical measurements were obtained in 20 of these animals, 9 of which were adults and 11 that were different ages ranging from 1.3 to 22.1 weeks (for convenience, the age in days was converted to weeks by dividing by 7 and the quotient rounded to the nearest 1/10 of 1 week; for example, 9 days divided by 7 is equivalent to 1.3 weeks). The latter 11 animals came from 5 different litters. Nine additional adults were used only for measurements of head and pinnae dimensions. Adult acoustic measurements were obtained in conjunction with physiological experiments that utilized the acoustical measurements for virtual space stimulus presentation. In this paper we show detailed data from four animals from different age groups spanning development: 1.3 weeks (K009), 5 weeks (K008), 10 weeks (K012), and adult (adult). Summary data, when shown, were computed from all animals.

A. Growth of the head and pinnae

Figure 1 replots from Tollin and Koka (2009) the linear measurements of head diameter *AB* and outside pinnae height 1-2 as a function of age in weeks for 16 cats starting at 1.3 weeks. The horizontal lines represent the mean and 99% confidence interval for these values in 12 adults (>52 weeks). To quantify the growth rate a three-parameter exponential rise to maximum function was fitted to the data of the form $y=y_o+a(1-e^{-bx})$, where x is age in weeks, y_o is the extrapolated dimension at birth (0 week), a is the amount by which that dimension increases during development, (y_o+a) is the asymptotic value at full development, and b is the rate of growth. This equation accurately characterized the growth of each dimension (based on F-test $p<0.0001$ for all fitted equations; correlation coefficients r are reported in the figure). The fitted parameters of the equations are shown in each panel in Fig. 1.

Based on the fitted growth curve, head diameter more than doubled increasing from 28.6 at birth to 62 mm, reaching 90% of adult value by 23 weeks. In contrast, the growth of the pinnae was much more rapid. Outside pinnae height 1-2 increased from 14.6 to 48.4 mm, reaching 90% at 16 weeks. Bodyweight (not shown) increased from 0.2 ± 0.06 kg at ~1.5 weeks and asymptotes at 2.6 ± 1.1 kg by ~18 weeks.

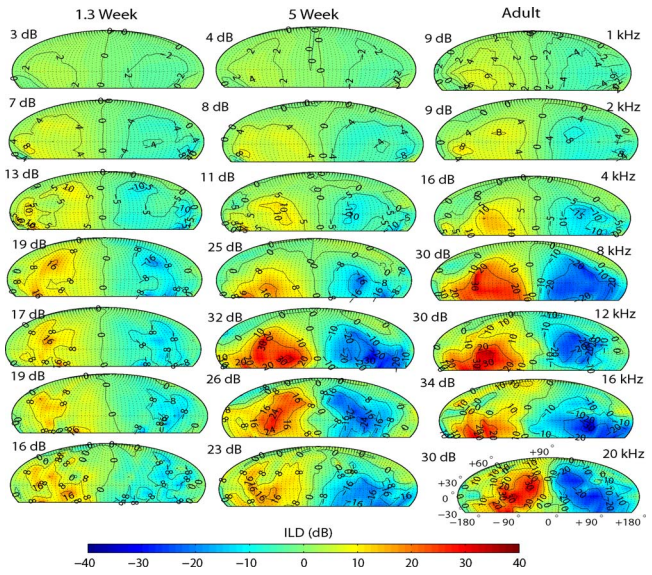


FIG. 2. (Color online) Spatial distributions of ILD for seven frequencies (upper right, right column) in three animals aged 1.3 weeks, 5 weeks, and adult. Maximum ILD for each animal and each frequency is indicated at the upper left-hand side of each plot. Color bar (bottom) indicates the ILD in decibels.

During development, the dimensions of the head and pinnae were highly and significantly correlated with the weight and age of the animal (mean $r^2=0.91\pm0.05$, $n=15$ pairwise comparisons), as expected because both structures were increasing in size. Tollin and Koka (2009) also reported that even in a group of 18 adults significant correlations were found: Outside pinna height 1-2 was significantly correlated with weight ($p<0.01$), but the head diameter *AB* was not. We did not track gender differences in development because the five males in the study were utilized for the acoustical measurements at very young ages.

B. Development of the ILD cues

1. Spatial distributions of ILDs

The difference between right and left ear DTF gains results in the ILD spectra. ILD cues varied with frequency and source location and with the age of the animals. Positive and negative ILDs indicate higher DTF gain for right and left ears, respectively. Figure 2 shows the spatial distributions of ILD cues computed from 1 to 20 kHz (upper right of right-hand column) in animals at three different ages: 1.3 and 5 weeks and adult. Frequencies higher than ~20 kHz produced complicated plots of ILD. The plots have been appropriately normalized so that they all use a common scale to indicate the value of the ILD (Fig. 2, bottom, color bar). The maximum ILD for each animal and each frequency is indicated at the upper left-hand side of each plot. Two major developmental trends are apparent regarding ILD for any source location. First, for a given age, the maximum ILDs tended to increase with increasing frequency. Second, for a given frequency, maximum ILDs increased with animal age.

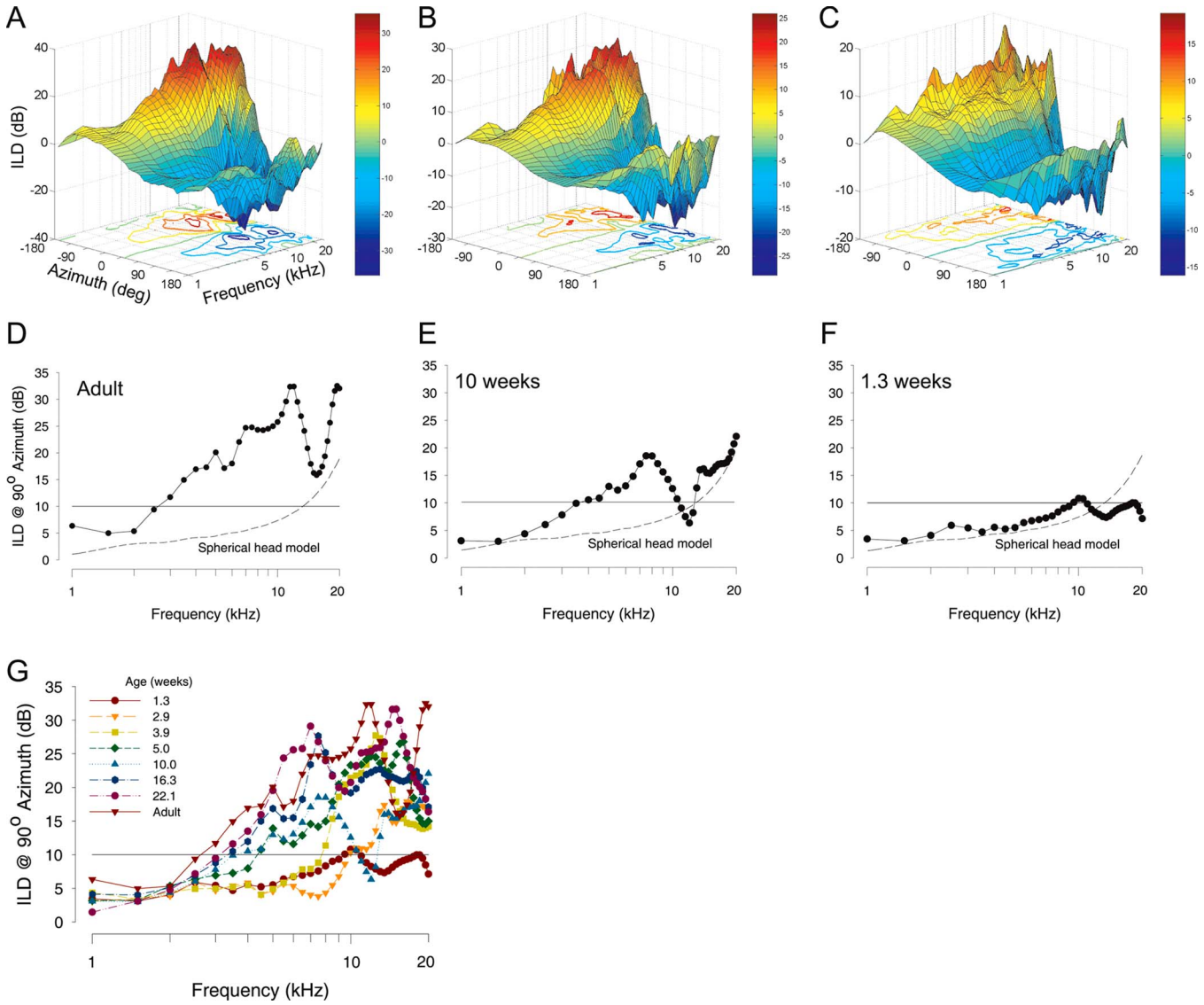


FIG. 3. (Color online) Spatial and frequency distributions of ILDs along the horizontal plane (0° elevation) for animals aged 1.3 weeks (C), 10 weeks (B), and adult (A). [(D)–(F)] ILD spectra (ILD vs frequency) for a source at 90° azimuth for the same animals in (A)–(C). Predicted ILD based on a spherical head model with an adult head diameter as input (dashed line). Solid line indicates a constant 10 dB ILD, which was the maximum ILD observed for the youngest animal in panel (F). (G) ILD spectrum at 90° azimuth for animals aged 1.3 weeks through adult.

2. Spatial and frequency distributions of ILDs along the horizontal plane

Spatial plots such as those shown in Fig. 2 are difficult to interpret over fine spatial position and frequency increments. Because the ILD is likely to be most useful as a cue to sound location for sources varying in azimuth, Fig. 3 plots the ILD as a joint function of frequency and source azimuth for sources along the horizontal plane (i.e., 0° elevation). These functions are shown for animals of three different ages: 1.3 (C) and 10 (B) weeks and adult (A). Regardless of age, ILDs are a complex function of azimuth and frequency. At low frequencies, ILDs vary little with azimuth while at high frequencies ILDs can vary by as much as ±35–40 dB in adults. Two general trends were observed in Figs. 3(A)–3(C). As the animals aged, for a given frequency, the ILDs increased in magnitude [note the change in the scale of the *z*-axis in panels (A)–(C)]. Moreover, the ILD vs azimuth functions as a whole shifted toward lower frequencies.

To show these two points more clearly, Figs. 3(D)–3(F) show the ILD spectra (i.e., ILD vs frequency) for one location (+90°) for the three different aged animals. Figure 3(F) shows data for the 1.3 week animal. Over the range of frequencies plotted, the maximum ILD at 1.3 weeks for this source location was only ~10 dB, which occurred first at a frequency of ~10 kHz. In each plot of the ILD spectrum for the animals of different ages [(D)–(F)] this value of ILD (10 dB) is indicated by a horizontal line. A 10 dB ILD is also approximately the ILD at which the ILD-vs-frequency functions show an inflection and begin to rise substantially above the relatively flat ILD-vs-frequency relation for low frequencies [see adult data in Fig. 3(D)]. Also plotted in these figures [(D)–(F)] is the ILD spectrum at an azimuth of 80° computed from a spherical head model (Duda and Martens, 1998) with an adult-sized head radius taken from Fig. 1(A) (32 mm). An azimuth of 80° instead of 90° was chosen for this comparison because 80° in the spherical head model yielded large

ILDs vs azimuth and frequencies compared to 90° (at 90° the maximum ILD in the head model over the frequency ranges plotted in Fig. 3 was ~6 dB). Moreover, the developmental trends apparent in the empirical ILD spectra differed little between azimuths of 80° and the 90° plotting in Fig. 3 (see discussion for rationale behind choosing 90° for this analysis). The absolute maximum ILD for any frequency and azimuth in the spherical head model was also produced for an azimuth of 80°, which was 22 dB at 22.8 kHz. Figure 3(E) shows the ILD spectrum for a 10 week old. Here the ILD magnitudes for most frequencies have increased substantially relative to the 1.3 week old. Additionally, the lowest frequency at which the ILD surpasses the 10 dB mark as shown in Fig. 3(F) was shifted toward a lower frequency, 4 kHz, than that in the younger (and smaller) animal. Finally, in an adult animal [Fig. 3(D)], the ILDs increased more and were shifted to even lower frequencies (2.5 kHz). Note that for the adult animal, over the frequency range up to 20 kHz, the head by itself (according to the spherical head model) does not produce much ILD (<~15 dB). Rather, the increase in the magnitude of the ILDs and the frequency ranges over which substantial ILDs occur (>10 dB) appear to result entirely from the development of the size of the pinnae. Finally, Fig. 2(G) summarizes the development of the ILD cue as a function of age and frequency for a source at 90° azimuth showing the general increase in ILD magnitude for a given frequency and the systematic shift of the ILD spectra toward lower frequencies. By the time the animals reached 16–22 weeks of age, the patterns of ILD vs frequency and azimuth were adult-like.

3. Development of the rate of change in ILD cues with changes in source azimuth

In addition to the development of the magnitude and frequency ranges of the ILD cue, another important attribute is the frequency-dependent rate of change in the ILD cues with changes in source azimuth, or the ILD-azimuth slope (in units of dB/deg). The rate of change in binaural cues such as ILD is an important determinant of spatial acuity, as typically assessed by behavioral measurement of the minimum audible angle (MAA) (Mills, 1958; Martin and Webster, 1987; Casseday and Neff, 1973; Huang and May, 1996). Here, the slopes of the ILD functions were computed as a function of frequency between ±10° about the midline (0° azimuth) for animals from 1.3 weeks through adult. This azimuth was chosen because ILDs vary most about this point and most psychophysical studies of sound location acuity are performed about 0° azimuth. Figure 4 shows the ILD slopes over a frequency range 3.5–16 kHz, which was the range where the slope was largest and varied most dramatically across ages. Also, the behavioral MAA in adult cats varies most substantially (e.g., both high and low MAAs) over this range (Martin and Webster, 1987; Casseday and Neff, 1973).

For comparison, the ILD slopes were averaged across 11 adult cats and the mean and associated 95% confidence intervals shown in Fig. 4. In the two youngest animals tested, 1.3 and 2.9 weeks, the ILD slopes were virtually flat over the frequency range plotted at ~0.2 dB/deg. As a comparison, the ILD slope computed from a spherical head model (Duda and Martens, 1998) with an adult head radius (32 mm) is shown (dashed line). As the animals aged, the ILD slopes showed a region of dramatic increase between 6 and 16 kHz. The frequency at which the local maximum of the ILD slope occurred decreased systematically with age, being ~16 kHz at 2.9 weeks and ~10 kHz by adulthood. As with the magnitude of the ILD in Fig. 3, the ILD slope reached adult values by 16–22 weeks, as indicated by slope values falling consistently within the 95% confidence intervals of the adult slopes.

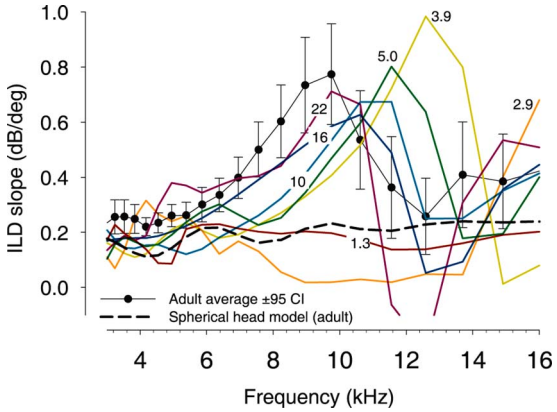


FIG. 4. (Color online) Development of the ILD-azimuth slope (dB/deg) for sources ±10° about the midline. The mean and associated 95% confidence interval for 11 adults are indicated by the filled circles and error bars, respectively. Data corresponding to animals of different ages (in weeks) are indicated by the parameter. The dashed line indicates the slope of the ILD based on a spherical head model with an adult head diameter as input.

4. Growth of the pinnae contributes to the increase in magnitude and frequency range of ILDs

In both Figs. 3(D)–3(F) and 4 estimates of ILDs and their slopes are plotted, respectively, from a spherical head model (Duda and Martens, 1998). It is clear in both data sets that the ILDs produced by a simple spherical head of an adult size cannot account for the large developmental changes in the magnitude of the ILDs and frequency ranges over which these ILDs are produced, at least over the frequency range examined here (up to 32 kHz). Aside from the head, the pinnae are the only other structures on the head that can account for these results.

We propose here that the pinnae, and not the head, are the most important factor in determining the ILD cues for location in small-headed mammals such as the cat because the head, by itself, cannot physically generate the large ILDs over the range of frequencies examined here. A direct test of this hypothesis would involve carefully removing the pinnae from both ears and repeating the measurements of ILD in a "head-only" condition. This was done in only one animal (see Fig. 7 in Tollin and Koka, 2009), so the results cannot be related to the developmental increase in size of the head and pinnae. As an alternative test, we assessed which aspect of developmental growth, head diameter [*AB* in Fig. 1(A)] or pinnae height [1-2 in Fig. 1(B)], could account for the most variance in the ILD data. Figure 5(A) plots the frequency at which the ILD first exceeded 10 dB at a location of 90° azimuth (0° elevation), as plotted in Figs. 3(D)–3(G), as a

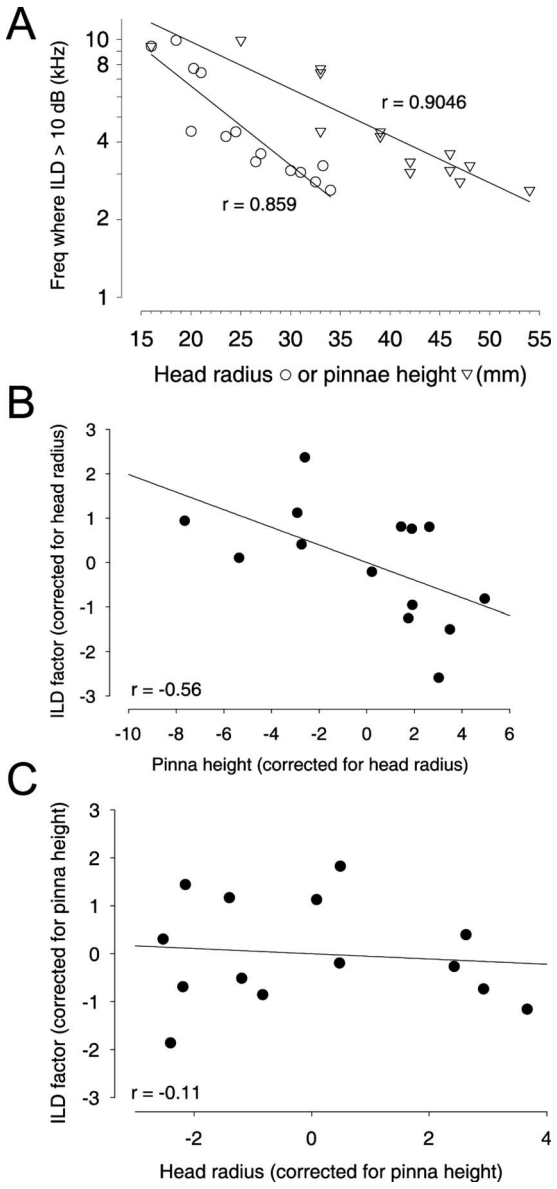


FIG. 5. (A) Frequency where the ILD cue first reached 10 dB for a source at 90° azimuth [as in Figs. 3(D)–3(G)] as a function of head radius (circles) and outside pinnae height (triangles). Solid lines indicate the linear regression through the respective data points. The correlation coefficients are also indicated for the regression. (B) Partial correlation of the ILD metric and pinnae height with the contribution of head radius held constant. (C) Partial correlation of ILD metric and head radius with the contribution of pinnae height held constant. In (B) and (C) the solid line indicates the linear regression and the associated correlation coefficient is shown.

function of both head diameter and outside pinnae height. This ILD factor was significantly correlated with both head $r=-0.859$, $p<0.0001$, and $n=14$) and pinnae ($r=-0.9046$, $p<0.0001$, and $n=14$) dimensions. In both cases, as the head and pinnae increased in size, the frequency at which the ILD reached 10 dB decreased systematically.

As pointed out by Tollin and Koka (2009), the developmental increase in the head and pinnae dimensions was highly and positively correlated. For the animals shown in Fig. 5, the correlation between pinnae and head dimensions (not shown) was highly significant ($r=0.9311$, $p<0.0001$, and $n=14$). To control for this confounding correlation, we computed the partial correlations (Hays, 1988) to examine which of these two factors, head or pinnae growth, explained the most variance in the development of ILD. Figure 5(C) shows that when the contribution of the pinnae size was held constant, the partial correlation between the ILD factor (i.e., the lowest frequency at which the ILD just reached 10 dB) and head radius, both of which were corrected for their respective correlations with pinnae size, no longer reached significance ($r=-0.11$ and $p=0.73$). The partial correlation of the ILD factor and head diameter adjusted for pinnae height is simply the correlation between the residuals from regressing ILD on pinnae height and the residuals from regressing head diameter on pinnae height [Fig. 5(C)]. Thus, when the correlation with the pinnae size was factored out, the increase in head diameter alone accounted for ~1% ($r^2=0.012$) of the variance in the ILD factor. On the other hand, when the contribution of head size was held constant, the partial correlation between the ILD factor and outer pinnae height (both corrected for their correlations with head radius) was significant and negative ($r=-0.56$ and $p=0.046$). The partial correlation of ILD and pinnae height adjusted for head diameter is the correlation between the residuals from regressing ILD on head diameter and the residuals from regressing pinnae height on head diameter. In other words, as the pinnae height increased during development, the frequency at which the ILD reached 10 dB (at 90° azimuth) decreased systematically. The pinnae size, by itself, accounted for ~31% of the variance in this particular aspect of ILD during development, while the increasing head size accounted for virtually none. Thus, the hypothesis that the increasing pinnae size, and not the increasing head size, was responsible for the increasing magnitudes of ILDs and the frequency range of substantial ILDs (>10 dB) during development over the frequency range examined here (up to 32 kHz) was supported.

C. Development of the ITD cue

1. Spatial distributions of ITDs

Figure 6(A) shows the spatial distributions of the low-frequency (<3.5 kHz) ongoing ITDs in animals aged 1.3 and 5 weeks and adult. As expected, the ITDs increased with changes in source azimuth away from the midline, but were relatively constant with changes in elevation for a given azimuth. For any one azimuth, the magnitudes of the ITDs also increased with animal age. To show this more plainly, Fig. 6(B) shows the ITD as a joint function of both azimuth along the horizontal plane (0° elevation) and animal age, for animals aged 1.3 weeks through adult. For source locations at the poles (±90° azimuth), the magnitude of the ITDs increases systematically with animal age. Although it is known that ongoing ITDs vary as a function of frequency in cats (Roth *et al.*, 1980), as predicted by the spherical head model of Kuhn (1977), this particular aspect was not investigated in this paper.

It was expected that as the head radius increased during development, the magnitudes of the ITDs would systematically increase. As one way of testing this hypothesis, Fig. 6(C) plots the maximum ITD as a function of the head radius. Maximum ITDs increased systematically from ~170 to

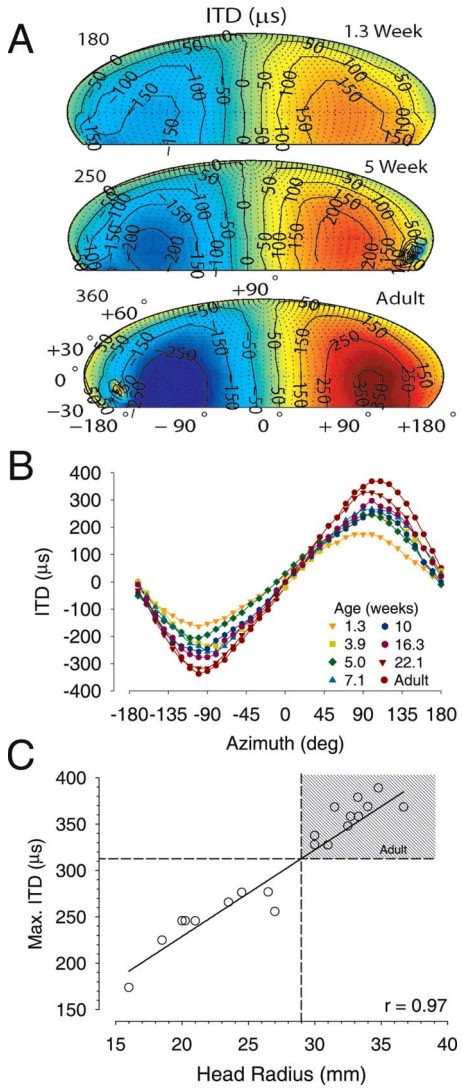


FIG. 6. (Color online) (A) Spatial distributions of the ITD for animals aged 1.3 weeks, 5 weeks, and adult. (B) ITD as a function of source azimuth along the horizontal plane (0° elevation) for animals aged 1.3 week through adult (age given by the parameter). (C) Maximum ITD as a function of head radius (n=20). Shaded region indicates ranges of ITD and head radius in adults. Solid line indicates the linear regression of maximum ITD on head radius; the associated correlation coefficient is also shown.

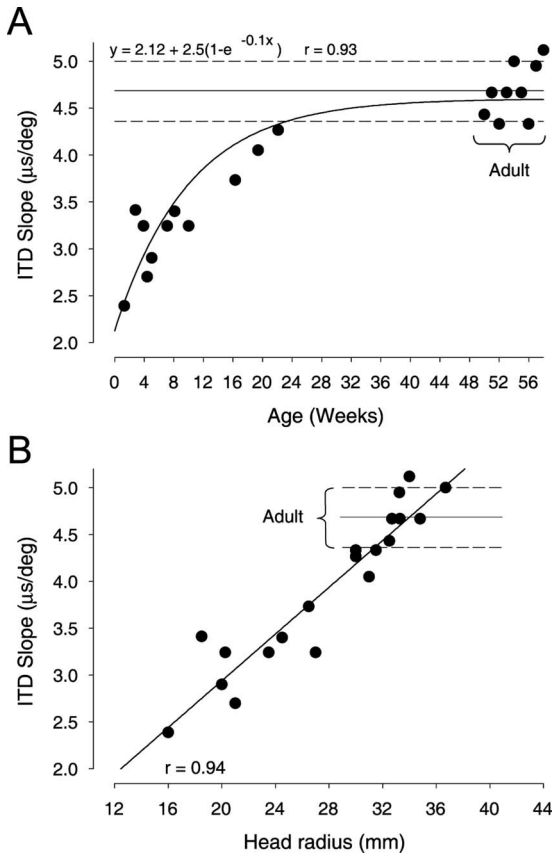


FIG. 7. (A) Development of the ITD-azimuth slope (μs/deg) as a function of age in weeks. Solid and dashed horizontal lines indicate the mean and 95% confidence interval of the ITD-azimuth slope based on measurements in nine adults. Curved line corresponds to the fitted function of the form given in the upper left. (B) Development of the ITD-azimuth slope as a function of head radius. Horizontal lines as in (A). Solid line indicates the linear regression of the ITD-azimuth slope on head radius; the corresponding correlation coefficient is indicated.

nearly 400 μs, a factor of 2.35, as the head radius increased from 16 to 38 mm, a factor of 2.38. The linear regression of maximum ITD on head diameter was significant (r=0.965, $p<0.0001$, y=9.76+34.4x, and n=20). As expected, head size was directly correlated with the magnitude of the ITD cue to source location. This is tested more directly in Sec. III C 3.

2. Development of the rate of change in ITD with changes in source azimuth

As with ILD above, one important characteristic of the ITD cue as it pertains to psychophysical sound location acuity is the rate of change in the ITD cue with changes in source azimuth, the ITD-azimuth slope (in units of μs/deg). Figure 7(A) plots the ITD vs azimuth slope computed from ±30° around 0° azimuth for 11 infants and juveniles (ages 1.3–22 weeks) and 9 adults. The mean ITD slope and associated 95% confidence interval for the 11 adults are shown as the solid and dashed lines, respectively, in Fig. 7(A). ITD slope increased from 2.4 to 4.7 μs/deg from 1.3 weeks of age through adult. The ITD slope about the midline increased by a factor of ~2, which is comparable to the factor by which head diameter (or radius) increased during development. A three-parameter exponential rise to maximum function of the form shown in the inset of Fig. 1(A) was fitted to the ITD slope vs age data with the exponent of the function fixed to the same value as that for the function describing the growth of the head in Fig. 1(A) (−0.1). The fit of this function was significant (r=0.93, $p<0.0001$, F-test) with the parameters shown at the top-left of Fig. 7(A). The youngest age at which the fitted function first surpasses the 95% confidence interval for the adult data (lower dashed line) was 23 weeks.

The magnitude of the ITD for a given azimuth increased with head radius as shown in Fig. 6(C), so the slopes of the ITD-azimuth functions were also expected to increase with radius. To show this, Fig. 7(B) replots the development of the ITD slope as a function of the associated head radius. The linear regression of ITD slope on radius was significant (r=0.94, $p<0.0001$, y=0.45+0.125x, and n=20). The smallest head radius at which the ITD slope first surpassed the lower

95% confidence interval for the adult data was ~31 mm (or a 62-mm diameter), which falls squarely in the range of adult head growth shown in Fig. 1(A). By ~23 weeks, the ITD-azimuth slope is adult-like.

3. Growth of the head and pinnae contributes to the developmental increase in the magnitude of ITDs

It is typically assumed that the magnitude of the ITD cue is determined directly by the interaural distance. Thus, the larger the separation between the two ears, as occurs during developmental growth of the head, the larger the ITD is presumed to be. However, in mammals such as cat, the length and physical size of the pinnae are also quite substantial. The height of the pinnae in adult cats is nearly the same as the head diameter [Figs. 1(A) and 1(B)]. Like the head, then, the pinnae also present as a substantial obstacle for the sound waves to propagate around, thus potentially increasing the effective diameter of the head. As such, the pinnae themselves may contribute to the magnitude of the ITD, as has been demonstrated recently by Koka *et al.* (2008) in the rat.

Figure 8(A) shows the maximum ITD as a function of head radius and outside pinnae height in 20 animals. In the analysis of the data in Fig. 8(A), we noticed that a single data point from one animal exerted a disproportionate influence on the regression results for both maximum ITD vs head radius and pinnae height. This data point (one for ITD vs head radius and one for ITD vs pinnae height) is shaded in Fig. 8(A). This single data point also disproportionately influenced the resulting partial correlation analyses below. For this animal, the measured maximum ITD was smaller than expected and both the head radius and pinnae height were larger than expected. We suspect that the head and pinnae sizes for this animal were overestimated. In the regression analyses, this data point produced standardized residual values (raw residual divided by the population standard error, Hays, 1988) that exceeded an absolute value of 2.5 indicative of an outlier. No other data point in any of the results in this paper produced such large standardized residuals. Because of the undue influence of this single data point, we removed this data point from subsequent analysis below. This data point did not overly influence any of the preceding analyses and was not omitted in those cases (i.e., omission would not have changed the overall results). After removal of the outlaying data point, the linear regressions of maximum ITD on pinnae height ($r=0.943$, $p<0.0001$, and $n=19$) and head radius ($r=0.978$, $p<0.0001$, and $n=19$) were still significant.

To test the hypothesis that the head diameter was the primary determinant of the increasing ITDs during development, and not the pinnae height, a partial correlation analysis (Hays, 1988) was conducted with maximum ITD, head diameter, and pinnae height. Partial correlation analysis is necessary here because there was a significant positive correlation between head diameter and pinnae height (not shown) in the 19 animals measured for this test ($r=0.923$ and $p<0.0001$). When the contribution of the pinnae height was held constant, the partial correlation between the maximum ITD and head radius (both corrected for their respective correlations with pinna height) was positive and significant ($r=0.84$ and $p<0.0001$). The partial correlation of the maximum ITD and head radius adjusted for pinna height is simply the correlation between the residuals from regressing maximum ITD on pinna height and the residuals from regressing head radius on pinnae height [Fig. 8(C)]. Increases in head radius accounted for only ~71% of the variance in the increase in maximum ITD during development. Interestingly, when the contribution of the head radius was held constant, the partial correlation between the maximum ITD

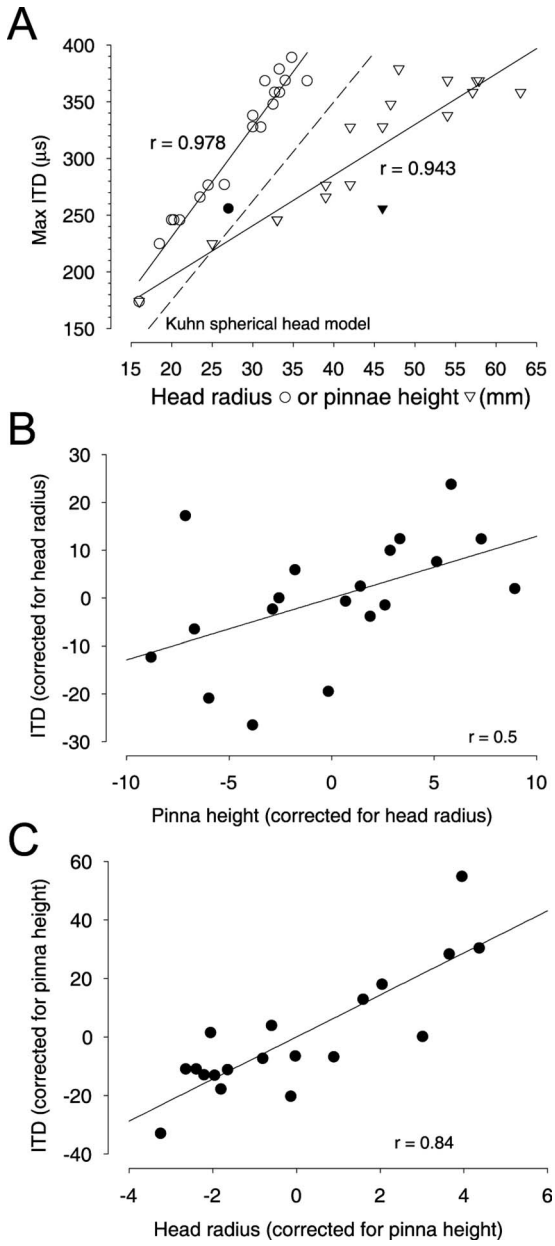


FIG. 8. (A) Maximum ITD as a function of head radius (circles) and outside pinnae height (triangles). Shaded data points were outliers that were removed from subsequent analyses. Solid lines indicate the linear regression through the remaining respective data. The associated correlation coefficients are indicated next to the fitted function. Spherical head model for low-frequency ITDs of Kuhn (1977) based on head radius (dashed). (B) Partial correlation of maximum ITD and pinnae height with the contribution of head radius held constant. (C) Partial correlation of maximum ITD and head radius with the contribution of pinnae height held constant. In (B) and (C) the solid line indicates the linear regression and the associated correlation coefficient is shown.

and the pinnae height (both corrected for their respective correlations with head diameter) also reached significance ($r=0.501$ and $p=0.034$). The partial correlation of the maximum ITD and pinna height adjusted for head radius is simply the correlation between the residuals from regressing maximum ITD on head radius and the residuals from regressing pinnae height on head radius [Fig. 8(B)]. Thus, pinnae height by itself accounted for ~25% of the variance in the increase in maximum ITD during development. Thus, when the two factors that were most likely to contribute to the increasing magnitude of the ITD cue during development were examined, both head diameter (or radius) and pinnae height emerged as determinants. Although maximum ITD was also highly correlated with animal weight ($r=0.916$, $p<0.0001$, and $n=19$) and weight was highly correlated with head radius ($r=0.944$, $p<0.0001$, and $n=19$) and pinnae height ($r=0.913$, $p<0.0001$, and $n=19$), the respective partial correlation analyses revealed no significant contribution to maximum ITD due to increases in animal weight ($r<0.4$ and $p>0.1$). In this latter case, the contributions of the head radius and pinnae height to the maximum ITD remained significant and comparable to that found above.

IV. DISCUSSION

A. Development of the linear dimensions of the head and pinnae

We studied the development of head and pinnae dimensions in cats beginning at 1.3 weeks after birth. Details are found in Tollin and Koka, 2009. These results are discussed here in the context of the development of the binaural cues to location. The head and the pinnae increased in size during development. Based on the growth curves, head diameter [Fig. 1(A)] increased by a factor of 2.17, from 28.6 mm at birth to 62 mm in adults and reached 90% of adult value by 23 weeks. These compare favorably with values of 29 and 63.02 mm measured in 35 newborn (Latimer, 1931) and 54 adult female cats (Latimer, 1936), respectively. Head diameter in adult males averages 5.5% larger (Latimer, 1936). The growth of the pinnae dimensions was more rapid. Outside pinnae height 1-2 [Fig. 1(B)] increased by a factor of 3.31 from 14.6 to 48.4 mm, reaching 90% at 16.4 weeks. Bodyweight (not shown) increased from 0.2 ± 0.06 kg at ~1.5 weeks and asymptoted at 2.6 ± 1.1 kg by ~18 weeks. Newborn weights in kittens average 0.15 ± 0.03 kg (n=35, Latimer, 1967).

Because the behavioral onset of hearing in cats is ~1.5 weeks (Ehret and Romond, 1981; Villablanca and Olmstead, 1979), the acoustical and behavioral consequences of the increasing size of the head and pinnae are functionally relevant only for ~1.5 weeks and older. Beginning at 1.5 weeks instead of birth the average head diameter increased by a factor of 2 and the outer pinnae dimension 1-2 increased by a factor of 2.3.

The rate of growth of the pinnae (given by parameter b in the equations in Fig. 1) was 30%–70% greater than that of the head. As such, the pinnae reach adult dimensions sooner than the head diameter. The major dimension of the pinnae, the outside height, reached 90% of adult size by 16 weeks while head diameter took 23 weeks to reach adult size. One implication of the rapid development of the pinnae is that the monaural acoustic transformations that are heavily determined by the pinnae, such as the spectral notches, the acoustic gain, and the acoustic axis, become adult-like before the acoustical transformations that depend on head diameter, such as the binaural cues to sound location discussed in this paper. The development of the monaural cues to sound source location in the cat were shown by Tollin and Koka (2009) to be mature by 16 weeks, consistent with the developmental growth of the pinnae.

B. Development of the pinnae dimensions contributes to the development of the ILD cues to location

To explore the hypothesis that the development of the pinnae dimensions determined the development of the ILD cue to location, Fig. 5 plotted one of the ILD metrics, the frequency at which the ILD reached 10 dB for a source at 90° azimuth (see Fig. 4). Although this metric is somewhat arbitrary, our purpose for this was to quantitatively capture the shift in the frequency range of ILD production during development. Although not shown here, comparable shifts were observed at locations between the midline and 90°. Analysis of the shift of the ILD functions along the frequency axis is similar to the frequency axis scaling studies of Middlebrooks (1999), Schnupp *et al.* (2003), and Maki and Furukawa (2005). Moore and Irvine (1979) demonstrated the shift in the frequency ranges of the ILDs cues from high to low during development of the cat for all azimuths tested, but the shifts were largest at 90°. Our results confirm their observation and this was one of our rationales for using 90° to make these measurements. The partial correlation analyses indicated that the growth of the pinnae accounted for ~31% of the variance in the development of this particular ILD metric, while the increasing head size accounted for virtually none. The results of this analysis along with the empirical observations that the various ILD metrics reached adult-like status by ~16 weeks fit with the empirical measurements of pinnae growth showing adult-like size and morphology by 16 weeks. 16 weeks also matches that necessary for the monaural acoustical cues, spectral notches, acoustical gain, acoustic axis, and ear canal resonance gain and frequency, to reach maturity (Tollin and Koka, 2009).

A spherical head model (Duda and Martens, 1998) was used here to isolate the frequency ranges and magnitudes of the ILD cue as the head diameter increased during development from the component of ILD cue development due to pinnae growth. Over the frequency range of interest in this paper (<32 kHz), a sphere of adult dimensions (32 mm radius) cannot generate the large ILDs of the magnitude measured here, particularly in the frequency ranges around ~10 kHz [see Fig. 3(D)] where the very largest ILDs were observed. We propose that the pinnae produced these large ILDs jointly through increases in acoustic gain by the pinnae ipsilateral to the source (re the gain due to just the sphere) and additional decreases (re simple spherical head shadow) due to the deep spectral notches produced by the pinnae contralateral to the source. Tollin and Koka (2009) showed that

the monaural acoustical gain increased systematically with the age of the cats, as well as pinnae size, and also moved from higher frequencies (~15 kHz for a 10 dB gain at 1.3 weeks age) toward lower frequencies (~5 kHz for a 10 dB gain in adults) similar to that observed here for ILD in Fig. 3. That this increased gain was due to the pinnae was shown by Tollin and Koka (2009) in one animal by comparing the monaural gains both before and after removal of the pinnae. The pinnae themselves produced an increase in gain of ~10 dB. The additional ipsilateral increase and contralateral decrease in gain contribute to the much larger ILDs than would otherwise be generated by a spherical head at these frequencies. One implication of large pinnae on small-headed mammals, then, may be to effectively shift the frequency range of substantial ILDs (>10 dB, see Fig. 3) toward much lower frequencies than would otherwise be the case without pinnae.

C. Development of the head and pinnae dimensions contributes to the development of the ITD cues to location

The maximum ITDs increased systematically from ~170 to nearly 400 μs, a factor of 2.35, as the head radius increased from 16 to 38 mm, a factor of 2.38. As might be expected, head size was directly correlated with the magnitude of the ITD cue to source location [Fig. 6(C)]. In general, the development of the ITD cue took longer than the development of the ILD cue. For example, the ITD-azimuth functions shown in Fig. 6(B) indicate that ITDs were not in the adult range until sometime around 22 weeks. Also, a grouping of the maximum ITDs and associated head radii in adults [shaded area, Fig. 6(C)] indicates that the head radius of ~29–30 mm is required to reach adult-sized maximum ITDs of 310 μs or greater. Head radii of 29–30 mm are only achieved by ~22–23 weeks of age, on average [Fig. 1(A)]. Finally, the ITD-azimuth slope first falls within the adult range [dashed line, Fig. 7(A)] at 23 weeks and with a head radius of ~32 mm. The development of the ITD cue appeared to be influenced most by the development of the head diameter (or radius) and the ITD cue was mature by ~22 weeks.

While the hypothesis that head diameter dictates ITDs is clear, the hypothesis that the pinnae might determine a portion of the ITD has not received much attention. The role of the pinnae in ITD production has been investigated by Roth *et al.* (1980) in the cat and Koka *et al.* (2008) in the rat. The pinnae in many mammals, such as cats, rabbits, rats, and bats to name just a few, are quite large, with dimensions on the same scale as that of the diameter of the head. In support of this hypothesis, recent reports from our laboratory have demonstrated that the pinnae in the rat contribute substantially to the overall magnitude of the ITDs—removal of the pinnae reduced the ITDs in rats by ~32%–36% (Koka *et al.*, 2008). Moreover, in the rabbit, systematic movements of the pinnae can produce quite substantial changes in the ITD cue (Bishop *et al.*, 2009).

Development of the maximum ITD was highly and significantly correlated with both head radius and pinnae height. Moreover, the growths of the head and the pinnae were highly correlated during development, as expected. This latter correlation was controlled for in the examination of the growth factor that determined the maximum ITD by performing partial correlations analysis with maximum ITD, head radius, and pinnae height. Increasing head diameter by itself was found to explain ~71% of the variance in the developmental increase in maximum ITD. Moreover, increasing pinnae height by itself was found to explain ~25% of the variance in the developmental increase in maximum ITD. Thus, when the two factors that were most likely to contribute to the increasing magnitude of the ITD cue during development were examined, the increase in head diameter emerged as the primary determinant, but there was also a sizable contribution from the pinnae height. One explanation for the role of the pinnae in ITD production as suggested by Koka *et al.* (2008) is that a considerable portion of the pinnae and the distal parts of the auditory meatus in mammals such as the cat and rat that have ear canals formed from cartilage remains in the path of the sound. Thus, the pinnae may function to make the acoustically effective diameter of the head larger, thus increasing the magnitude of the ITD. Consistent with this hypothesis, the predicted maximum ITD based on the empirically measured head radii and the spherical head model of Kuhn (1977) for low-frequency ongoing ITDs was ~23% less than the empirically measured ITDs [see Fig. 8(A)]. This discrepancy may result from the increased ITD due to the pinnae.

D. Implications for physiological and behavioral development

These data have implications for the concomitant development of the acoustical cues to sound source location, the neural encoding of these cues, and their ultimate use by the animal for the perception of source location. Behaviorally, adult cats localize sounds quickly and accurately with performance nearing that of humans (Moore *et al.*, 2008; Tollin *et al.*, 2004; May and Huang, 1996; Populin and Yin, 1998). And even kittens can approach sounds by around 24 days of age, although with much less precision (Clements and Kelly, 1978; Olmstead and Villablanca, 1980; Villablanca and Olmstead, 1979). The ability of kittens to make overt orienting responses to sounds suggests that the basic organization of the binaural auditory system may be established early in development. But physiological (Pujol and Hilding, 1973) and simple behavioral (Ehret and Romond, 1981) responses to sound are seen much earlier, a few days after birth. The rough circuitry of the mammalian binaural system appears to be in place and largely functional even while the peripheral system is still developing. The apparent delay in behavioral directional responding might be related to a slower rate of development of the binaural hearing mechanism, the specific cues for location, or simply motor control.

An attractive hypothesis for the prolonged development of spatial hearing is that auditory experience early in life calibrates the neural circuits that process sound location to the exact acoustical properties of the head and ears of each individual (Moore and King, 2004). In the most compelling example of this, Knudsen *et al.* (1984a, 1984b) revealed a sensitive period early in development of the barn owl where

normal acoustical input to the two ears, and thus normal cues to source location, must be present for normal sound localization behavior to develop. The duration of this sensitive period was shown to be correlated with the time course over which the head and facial ruff (like pinnae) dimensions reach maturity, ~8 weeks (Knudsen *et al.*, 1984a). These studies revealed that owls reared with altered acoustical cues (e.g., ear plug) prior to 8 weeks were able to adapt and regain normal behavioral sound localization abilities despite the altered cues; however, when the cues were altered in owls after ~8 weeks no adaptation was found. Thus, for a period of ~8 weeks, the internal mapping of the ensemble of acoustical cues to location and spatial location itself remains plastic. For the most part, the underlying anatomical and neural mechanisms of binaural hearing in the barn owl exhibited a similar time course of plasticity (Knudsen, 1983, 1985).

To the extent to which a sensitive period for the development of sound localization in cats exists, our present data detail the developmental constraints on when the peripheral acoustical transformations reach maturity. Here, the binaural spectral (ILDs) and temporal (ITDs) transformations were found to reach maturity by 16 and 22 weeks, respectively. The developmental growth of the pinnae was found to determine the development of the ILD cues, which agrees with the developmental time course of the monaural cues (spectral notches, acoustical gain, acoustic axis, etc.) demonstrated by Tollin and Koka (2009). And the developmental growth of not only the head diameter, but also the pinnae height, was found to determine the development of the ITD cues. The corresponding time periods for binaural ILD and ITD cue maturities of 16 and 22 weeks are in line with the development of the dimensions of the pinnae [Fig. 1(B)] and head [Fig. 1(A)], respectively. We hypothesize that a sensitive period for the consolidation of sound localization in the cat for the monaural cues and the ILD cues to location will occur within 16 weeks. Because the head dimensions and the associated ITD cues to location do not reach maturity until ~22 weeks, the sensitive period may be somewhat prolonged for ITD sensitivity.

ACKNOWLEDGMENTS

We thank Heath Jones and Jennifer Thornton for comments on the manuscript and Janet Ruhland and Mike Wells for assistance in some of the experiments. This work was supported by National Institutes of Deafness and Other Communicative Disorders Grant No. DC-006865 to D.J.T. We appreciate the assistance of Dr. Michael Hall for constructing custom experimental equipment (with support by NIH Grant No. P30 NS041854-05). Finally, we would like to thank Dr. George Kuhn for donating equipment to our laboratory.

Bishop, B., Kim, D. O., Sterbing-D'Angelo, S., and Kuwada, S. (**2009**). "Acoustic cues for sound localization measured in a rabbit and a tennis ball and computed using a rigid spherical head model," Assoc. Res. Otolaryngol. Abstr. **32**, 127.

Bugayevskiy, L. M., and Snyder, J. P. (**1995**). *Map Projections: A Reference Manual* (Taylor & Francis, London).

Calford, M. B., and Pettigrew, J. D. (**1984**). "Frequency dependence of directional amplification at the cat's pinna," Hear. Res. **14**, 13–19.

Casseday, J. H., and Neff, W. E. (**1973**). "Localization of pure tones," J. Acoust. Soc. Am. **54**, 365–372.

Clements, M., and Kelly, J. B. (**1978**). "Directional responses by kittens to an auditory stimulus," Dev. Psychobiol. **11**, 505–511.

Duda, R. O., and Martens, W. L. (**1998**). "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.

Ehret, G., and Romond, R. (**1981**). "Postnatal development of absolute auditory thresholds in kittens," J. Comp. Physiol. Psychol. **95**, 304–311.

Hays, W. L. (**1988**). *Statistics* (Holt, Rinehart, and Winston, Fort Worth, TX).

Huang, A. Y., and May, B. J. (**1996**). "Spectral cues for sound localization in cats: Effects of frequency domain on minimum audible angles in the median and horizontal planes," J. Acoust. Soc. Am. **100**, 2341–2348.

Irvine, D. R. F. (**1986**). *The Auditory Brainstem* (Springer-Verlag, Berlin).

Irvine, D. R. F. (**1987**). "Interaural intensity differences in the cat: Changes in sound pressure level at the two ears associated with azimuthal displacements in the frontal plane," Hear. Res. **26**, 267–286.

Kitzes, L. M. (**1990**). "Development of auditory system physiology," in *Development of Sensory Systems in Mammals*, edited by J. R. Coleman (Wiley, New York), pp. 249–288.

Knudsen, E. I. (**1983**). "Early auditory experience aligns the auditory map of space in the optic tectum of the barn owl," Science **222**, 939–942.

Knudsen, E. I. (**1985**). "Experience alters the spatial tuning of auditory units in the optic tectum during a sensitive period in the barn owl," J. Neurosci. **5**, 3094–3109.

Knudsen, E. I., Esterly, S. D., and Knudsen, P. F. (**1984a**). "Monaural occlusion alters sound localization during a sensitive period in the barn owl," J. Neurosci. **4**, 1001–1011.

Knudsen, E. I., Knudsen, P. F., and Esterly, S. D. (**1984b**). "A critical period for the recovery of sound localization accuracy following monaural occlusion in the barn owl," J. Neurosci. **4**, 1012–1020.

Koka, K., Read, H. L., and Tollin, D. J. (**2008**). "The acoustical cues to sound location in the rat: Measurements of directional transfer functions," J. Acoust. Soc. Am. **123**, 4297–4309.

Kuhn, G. F. (**1977**). "Model for the interaural time differences in the horizontal plane," J. Acoust. Soc. Am. **62**, 157–167.

Kuhn, G. F. (**1987**). "Physical acoustics and measurements pertaining to directional hearing," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York), pp. 3–25.

Latimer, H. B. (**1931**). "The prenatal growth of the cat. II. The growth of the dimensions of the head and trunk," Anat. Rec. **50**, 311–332.

Latimer, H. B. (**1936**). "Weights and linear measurements of the adult cat," Am. J. Anat. **58**, 329–347.

Latimer, H. B. (**1967**). "Variability in body and organ weights in the newborn dog and cat compared with that in the adult," Anat. Rec. **157**, 449–456.

Maki, K., and Furukawa, S. (**2005**). "Reducing individual differences in the external-ear transfer functions of the Mongolian gerbil," J. Acoust. Soc. Am. **118**, 2392–2404.

Martin, R. L., and Webster, W. R. (**1987**). "The auditory spatial acuity of the domestic cat in the interaural horizontal and median vertical planes," Hear. Res. **30**, 239–252.

Martin, R. L., and Webster, W. R. (**1989**). "Interaural sound pressure level differences associated with sound-source locations in the frontal hemisphere of the domestic cat," Hear. Res. **38**, 289–302.

May, B. J., and Huang, A. Y. (**1996**). "Sound orientation behavior in cats. I. Localization of broadband noise," J. Acoust. Soc. Am. **100**, 1059–1069.

Middlebrooks, J. C. (**1999**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C., and Green, D. M. (**1990**). "Directional dependence of interaural envelope delays," J. Acoust. Soc. Am. **87**, 2149–2162.

Middlebrooks, J. C., and Pettigrew, J. D. (**1981**). "Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location," J. Neurosci. **1**, 107–120.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Moore, D. R., and Irvine, D. R. F. (**1979**). "A developmental study of the sound pressure transformation by the head of the cat," Acta Oto-Laryngol. **87**, 434–440.

Moore, D. R., and King, A. J. (**2004**). "Plasticity of binaural systems," in *Development of the Auditory System*, Springer Handbook of Auditory Research, edited by T. N. Parks, E. W. Rubel, R. R. Fay, and A. N. Popper (Springer-Verlag, New York, NY), pp. 96–172.

Moore, J. M., Tollin, D. J., and Yin, T. C. T. (**2008**). "Can measures of sound

localization acuity be related to the precision of absolute localization estimates?," Hear. Res. **238**, 94–109.
Musicant, A. D., Chan, J. C., and Hind, J. E. (**1990**). "Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons," J. Acoust. Soc. Am. **87**, 757–781.
Olmstead, C. E., and Villablanca, J. R. (**1980**). "Development of behavioral audition in the kitten," Physiol. Behav. **24**, 705–712.
Phillips, D. P., Calford, M. B., Pettigrew, J. D., Aitkin, L. M., and Semple, M. N. (**1982**). "Directionality of sound pressure transformation at the cat's pinna," Hear. Res. **8**, 13–28.
Populin, L. C., and Yin, T. C. T. (**1998**). "Behavioral studies of sound localization in the cat," J. Neurosci. **18**, 2147–2160.
Pujol, R., and Hilding, D. (**1973**). "Anatomy and physiology of the onset of auditory function," Acta Oto-Laryngol. **76**, 1–10.
Rice, J. J., May, B. J., Spirou, G. A., and Young, E. D. (**1992**). "Pinna-based spectral cues for sound localization in cat," Hear. Res. **58**, 132–152.
Roth, G. L., Kochhar, R. K., and Hind, J. E. (**1980**). "Interaural time differences: Implications regarding the neurophysiology of sound localization," J. Acoust. Soc. Am. **68**, 1643–1651.
Schnupp, J. W. H., Booth, J., and King, A. J. (**2003**). "Modeling individual differences in ferret external ear transfer functions," J. Acoust. Soc. Am. **113**, 2021–2030.
Shaw, E. A. G. (**1974**). "The external ear," in *Handbook of Sensory Physiology: Vol. V/1: Auditory System*, edited by W. D. Keidel and W. D. Neff (Springer, New York), pp. 455–490.
Tollin, D. J. (**2008**). "Encoding of interaural level differences for sound localization," in *The Senses: A Comprehensive Reference*, edited by A. I. Basbaum, A. Kaneko, G. M. Shepherd, G. Westheimer, P. Dallos, and D. Oertel (Academic, San Diego, CA), Vol. **3**, pp. 631–654.
Tollin, D. J., and Koka, K. (**2009**). "Postnatal development of sound pressure transformations by the head and pinnae of the cat: Monaural characteristics," J. Acoust. Soc. Am. **125**, 980–994.
Tollin, D. J., Populin, L. C., Moore, J. M., Ruhland, J. L., and Yin, T. C. T. (**2004**). "Sound-localization performance in the cat: The effect of restraining the head," J. Neurophysiol. **93**, 1223–1234.
Villablanca, J. R., and Olmstead, C. E. (**1979**). "Neurological development of kittens," Dev. Psychobiol. **12**, 101–127.
Walsh, E. J., and McGee, J. (**1986**). "The development of function in the auditory periphery of cats," in *Neurobiology of the Cochlea*, edited by R. A. Altschuler, R. P. Bobbin, and D. W. Hoffman (Raven, New York), pp. 247–269.
Wiener, F. M., Pfeiffer, R. R., and Backus, A. S. N. (**1966**). "On the sound pressure transformation by the head and auditory meatus of the cat," Acta Oto-Laryngol. **61**, 255–269.
Xu, L., and Middlebrooks, J. C. (**2000**). "Individual differences in external-ear transfer functions of cats," J. Acoust. Soc. Am. **107**, 1451–1459.
Young, E. D., Rice, J. J., and Tong, S. C. (**1996**). "Effects of pinna position on head-related transfer functions in the cat," J. Acoust. Soc. Am. **99**, 3064–3076.

Use of the matching pursuit algorithm with a dictionary of asymmetric waveforms in the analysis of transient evoked otoacoustic emissions

W. Wiktor Jedrzejczak
Institute of Physiology and Pathology of Hearing, ul. Zgrupowania AK "Kampinos" 1, 01-943 Warszawa, Poland

Konrad Kwaskiewicz and Katarzyna J. Blinowska
Department of Biomedical Physics, Institute of Experimental Physics, Warsaw University, ul. Hoza 69, 00-681 Warszawa, Poland

Krzysztof Kochanek and Henryk Skarzynski
Institute of Physiology and Pathology of Hearing, ul. Zgrupowania AK "Kampinos" 1, 01-943 Warszawa, Poland

(Received 30 June 2009; revised 10 September 2009; accepted 12 September 2009)

Transiently evoked otoacoustic emissions (TEOAEs) are normally modeled as the sum of asymmetric waveforms. However, some previous studies of TEOAEs used time-frequency (TF) methods to decompose the signals into symmetric waveforms. This approach was justified mainly as a means to reduce the complexity of the calculations. The present study extended the dictionary of numeric functions to incorporate asymmetric waveforms into the analysis. The necessary calculations were carried out using an adaptive approximation algorithm based on the matching pursuit (MP) numerical technique. The classic MP dictionary uses Gabor functions and consists of waveforms described by five parameters, namely, frequency, latency, time span, amplitude, and phase. In the present investigation, a sixth parameter, the degree of asymmetry, was added in order to enhance the flexibility of this approach. The effects of expanding the available functions were evaluated by means of both simulations using synthetic signals and authentic TEOAEs. The resulting analyses showed that the contributions of asymmetric components in the OAE signal are appreciable. In short, the expanded analysis method brought about important improvements in identifying TEOAE components including the correct detection of components with long decays, which are often related to spontaneous OAE activity, the elimination of a "dark energy" effect in TF distributions, and more reliable estimates of latency-frequency relationships. The latter feature is especially important for correct estimation of latency-frequency data, which is a crucial factor in investigations of OAE-generation mechanisms.
 [DOI: 10.1121/1.3243294]

PACS number(s): 43.64.Jb, 43.60.Hj [BLM] Pages: 3137–3146

I. INTRODUCTION

Otoacoustic emissions (OAEs) are low-level sounds that are generated either spontaneously by the inner ear or in response to external stimulation (Probst *et al.*, 1991). Responses that are evoked by short-lasting stimuli and then measured are referred to as transiently evoked otoacoustic emissions (TEOAEs). One of the most common ways of measuring TEOAEs is to apply click stimuli and record the signals that follow. The response is normally described by spectra that consist of multiple peaks, distributed between 0.5 and 5 kHz. The other characteristic feature is the shorter latency of the high-frequency components and the longer latency of the low-frequency components. Because of the complex structure of TEOAEs, the use of time-frequency (TF) methods has been proposed to study the principal components of these signals (e.g., Wit *et al.*, 1994; Tognola *et al.*, 1997). Several methods have been devised to estimate the TF distributions of OAEs including the short-time Fourier transform (Hatzopoulos *et al.*, 2000) methods based on the Wigner–Ville transform (Cheng, 1995) or Choi–Williams transform (Ozdamar *et al.*, 1997), wavelet transform (WT) (Wit *et al.*, 1994), minimum variance spectral estimation (Zhang *et al.*, 2008), and adaptive approximation (Jedrzejczak *et al.*, 2004).

The most popular TF method in TEOAE analysis is that of the WT. To date, this method has been used in various OAE studies with its main focus being the latency-frequency dependence in adult (e.g., Tognola *et al.*, 1997) and child ears (e.g., Tognola *et al.*, 2001), and a power law was proposed as being the best fit for such experimental data. Sisto and Moleti (2002) extended these studies using the analysis of OAEs from hearing-impaired subjects, and showed that for ears with high-frequency hearing loss (HL), the latencies of the OAEs were greater. The results of TF analysis of OAEs were also applied to cochlear mechanics (e.g., Moleti and Sisto, 2003). Some attempts to incorporate the advantages of TF analysis for clinical applications have been made as well. For example, Marozas *et al.* (2006) proposed the use

of wavelets as a tool for achieving a more reliable detection of TEAOEs than that obtained using standard detection methods such as reproducibility (i.e., the correlation between two buffers of subaverages).

The adaptive approximation method, based on a matching pursuit (MP) algorithm, has also been shown to be very effective in its application to OAEs (e.g., Jedrzejczak *et al.*, 2004). This method may be considered as a generalized version of the wavelet-based approach with the main difference being that WT analysis is constrained within predefined frequency bands, while the MP algorithm may be fitted to waveforms without any limitations on the time and frequency parameters used. The other advantage is that the MP technique can detect components of different durations. This means that it can help to detect the presence of spontaneous components in TEOAEs (Jedrzejczak *et al.*, 2007, 2008a; Notaro *et al.*, 2007). To date, the MP algorithm has been used to study some of the subtle phenomena associated with the TEOAEs of subjects exposed to noise (Jedrzejczak *et al.*, 2005), of premature neonates (Jedrzejczak *et al.*, 2007), of patients with partial deafness (Jedrzejczak *et al.*, 2009), and in investigations of the two-tone suppression of TEOAEs (Jedrzejczak *et al.*, 2008b).

One important characteristic of TF methods is the nature of the basic waveform used in the analysis. In most applications that use wavelets or the MP approach, symmetric waveforms have generally been used. This is justified mainly in terms of the reduction in the complexity of the calculations and the types of waveform that give the best TF resolution (Gabor, 1946). The two most popular approaches are the Morlet wavelets used in WT (Goupillaud *et al.*, 1984) and the Gabor atom in the case of MP (Mallat and Zhang, 1993). In order to construct the synthetic signals that most closely resemble TEOAEs, however, the use of gammatones has frequently been considered (e.g., Wit *et al.*, 1994; Tognola *et al.*, 1997; Sisto and Moleti, 2002). Such waveforms are asymmetric, and are characterized by a sharper rise and a more gradual decay. Assuming that the cochlea behaves as a bank of gammatone filters (Patterson *et al.*, 1992) distributed according to the Greenwood frequency mapping (Greenwood, 1990), TEOAEs could be modeled as the sum of the impulse responses of these individual filters.

In general, TEOAEs consist of short bursts of activity of different frequencies, but are also characterized by some distinct components that decay slowly, or even last throughout the entire acquisition window. These components are related to spontaneous otoacoustic emissions (SOAEs) that are generated without stimulation. They are characterized by long-time spans and narrow frequency bands. To ease their detection, their occurrence can also be stimulated by sound. This kind of emission is called a synchronized spontaneous otoacoustic emission (SSOAE). In fact, SSOAEs are long-lasting OAEs comprised of two main kinds of activity, namely, components of (1) measurable equilibrium magnitude, and (2) characteristic time of exponential decay (e.g., Sisto *et al.*, 2001). The latter components are characterized by a steep rise and a slowly decaying tail. In order to account for this effect, it seems important to decompose the signal into asymmetric waveforms.

The primary aim of the study described herein was to extend the MP analysis of TEOAEs by incorporating the possibility of their decomposition into asymmetric waveforms. The performance of the new approach was initially assessed using synthetic signals. Next, the MP algorithm was applied to real TEOAEs using a dictionary of symmetric and asymmetric waveforms. Finally, the latency-frequency dependence obtained was compared for the existing and new MP approaches.

II. EXPERIMENTAL PROCEDURE

TEOAEs from both ears of 29 subjects (26 females and 3 males, aged 21–35 years) were measured under low ambient noise conditions using an ILO-96 apparatus (v5, Otodynamics Ltd., London, United Kingdom). These subjects had healthy larynxes and did not have any otoscopic ear abnormalities. Impedance audiometry tests revealed normal type A tympanograms and normal acoustic reflexes. According to pure-tone audiometry, hearing thresholds were better than 20 dB HL for all the frequencies tested (0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz). Standard click stimuli with average levels of 80 ± 3 peak dB sound pressure level (SPL) representing the default nonlinear averaging protocol were used to evoke a total of 520 TEOAE responses. The initial part of the emitted response was windowed automatically by the system, thus, minimizing the influence of stimulus artifacts on the output signal. The onset of the window was set at 2.5 ms. All recordings were performed using the standard acquisition window, which ended at 20 ms.

SSOAE spectra were also measured using the default ILO protocol. In accordance with this protocol, the TEOAEs evoked by click stimuli of 80 dB SPL were recorded in an 80-ms window. The first 20 ms obtained from each averaged response was discarded and the responses from the following 60 ms were considered to be SSOAEs.

III. METHOD

A. MP

In this study, the signals were analyzed using the MP algorithm introduced by Mallat and Zhang (1993). This method is widely used in the processing of various physiological signals (e.g., Wang *et al.*, 2001; Malinowska *et al.*, 2007). 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 a dictionary). The waveforms were fitted using an iterative procedure, starting with the waveform accounting for the largest part of the signal energy. Following the initial analysis, the next waveforms were fitted to the residues. This procedure was continued until either the desired percentage of the original signal energy was accounted for, or for a predetermined number of atoms. A detailed description of the MP procedure and its application to OAEs may be found in Jedrzejczak *et al.*, 2004. In previous MP studies of OAEs, only dictionaries composed of symmetric waveforms were used (e.g., Jedrzejczak *et al.*, 2004; Notaro *et al.*, 2007). Goodwin and Vetterli (1999) proposed damped sinusoids as the basic atoms for the decomposition of musical sounds, and for a simi-

lar purpose the atoms of a fixed asymmetric envelope were used by Gribonval and Bacry (2003). These concepts could also be applied to OAEs. In the present study, the dictionary was therefore composed of symmetric and asymmetric waveforms, which are described in more detail in Sec. III B.

B. Enriched dictionary with symmetric and asymmetric waveforms

Conventionally, adaptive decomposition may be performed using a dictionary that consists of symmetric Gabor atoms of the following form:

$$G(t;\mu,\sigma,\omega,\varphi) = Ne^{-(t-\mu)^2/2\sigma^2}\cos(\omega t+\varphi), \quad (1)$$

where N is the normalization constant, σ is the scale, μ is the position in time, ω is the angular frequency, t is the time, and φ is the phase. This type of waveform has been used in other applications of the MP algorithm to OAE decomposition (e.g., Jedrzejczak *et al.*, 2004, 2005). In this way, each single-frequency waveform obtained using the MP algorithm is described using physically meaningful parameters as follows: (i) The magnitude is determined as the maximum of the envelope of the waveform at a certain frequency, (ii) the latency is measured as the time taken from the 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 may be interpreted as representing the duration of the waveform. In order to improve the representative ability of the technique and account for the presence of asymmetric waveforms in OAEs, a dictionary composed of two-sided functions is herein proposed. These functions are composed of two parts, the first being based on a Gabor function and the second on an exponential function. This waveform could have different rise and fall times for the same frequency. Such a waveform could be described by the formula

$$\Lambda(t;\mu,\sigma,\omega,\varphi,T_f) = N\begin{pmatrix} e^{-(t-\mu)^2/2\sigma^2} & t\le T_f \\ e^{-\alpha(t-\tau)} & t>T_f \end{pmatrix}\cos(\omega t+\varphi),$$

where

$$\alpha = \frac{T_f-\mu}{\sigma^2}, \quad \tau = \frac{T_f+\mu}{2}. \quad (2)$$

The additional parameter $T_f > \mu$ determines the asymmetry of the atom. T_f describes the point at which the Gaussian envelope changes into an exponential decay function. The function thus obtained is continuous up to the first-order derivative. Examples of the functions based on this concept are shown in Fig. 1, and include symmetric Gabor functions as well as those atoms having different steepnesses of their rising compared to their decaying parts. In comparison with the size of the standard Gabor dictionary of the order of 10^6 atoms, the enriched dictionary consisted of about 3×10^7 waveforms described by the following parameters: frequency, latency, time span, amplitude, phase, and asymmetry. To avoid statistical bias caused by the structure of the dictionary, a stochastic dictionary, proposed by Durka *et al.* (2001a, 2001b), was used. The TF evolutions of the signal

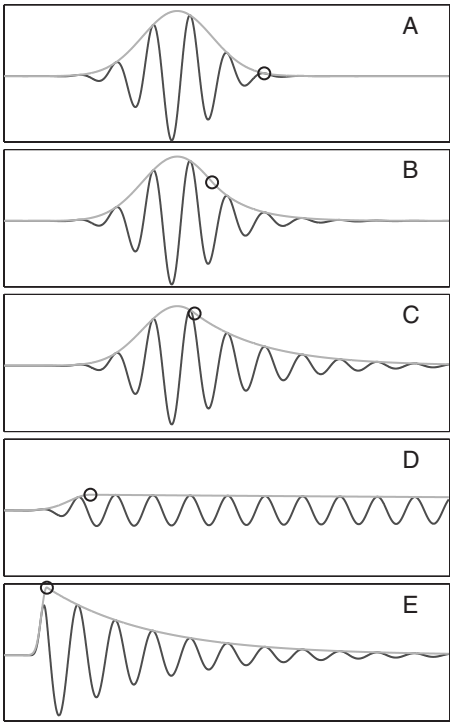


FIG. 1. Examples of structures with different asymmetries used in enriched dictionary—the Gabor dictionary used only symmetric fits: (A) symmetric, (B) with low asymmetry, (C) highly asymmetric, (D) with rise and constant amplitude, and (E) with steep rise and decaying amplitude. Circles mark the position where the Gaussian envelope changes into an exponent (parameter T_f).

are herein presented in the form of magnitude distributions, given by the expression

$$A(f,t) = Z^T(f)P(t), \quad (3)$$

where $\mathbf{Z}(f)$ is a vector that represents the absolute values of the Fourier transform of the atom and $\mathbf{P}(t)$ is a vector that describes the envelope of the atom. The symbol T indicates the transposition of the vector. The TF distribution of the decomposed signal is the sum of the distributions given by Eq. (3).

IV. RESULTS

A. Simulated signals

In order to compare the performances of the symmetric Gabor and enriched dictionaries, they were tested using simulated signals. In previous work (e.g., Wit *et al.*, 1994; Sisto and Moleti, 2002; Jedrzejczak *et al.*, 2004), gammatones were used as a basis for making synthetic TEOAE signals. Gammatones may be represented by the formula

$$\gamma(t) = \gamma_0 t^3 e^{-2\pi\Gamma t}\sin(2\pi f t), \quad (4)$$

where γ_0 is the level and is Γ the decay constant (expressed in hertz).

Examples of MP decompositions obtained for a long-decay gammatone (of a decay constant of 100 Hz) using the two dictionaries are given in Fig. 2(A). The decompositions were halted after they accounted for 98% of the energy. Using the Gabor dictionary, the MP algorithm tends to split this type of activity into several components and, in this example,

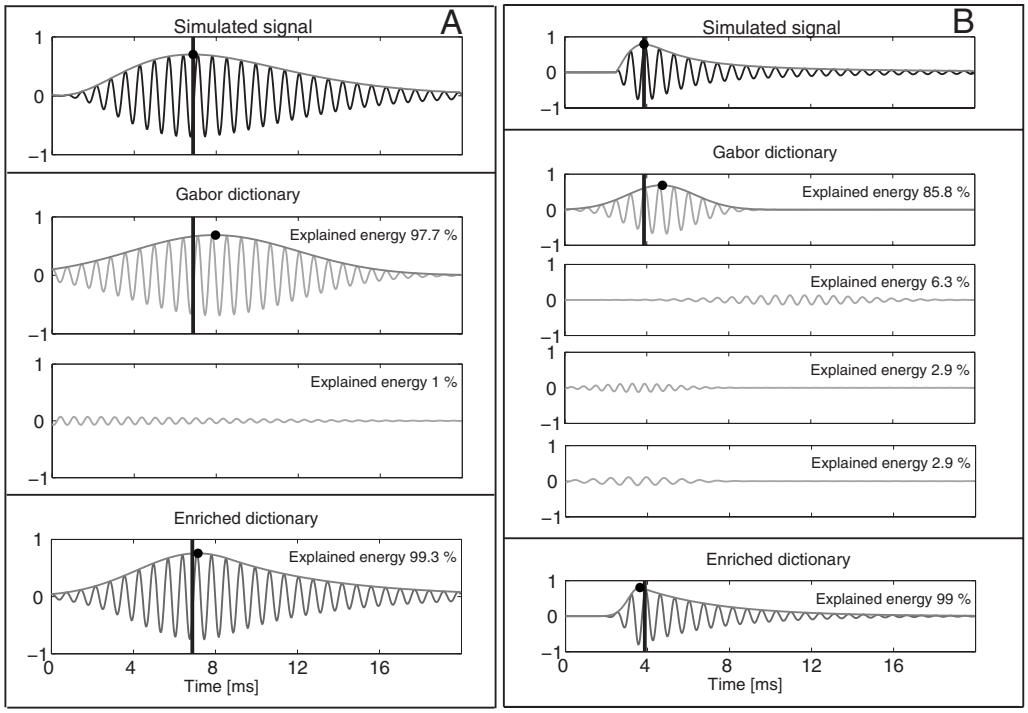


FIG. 2. (A) Example of the decomposition of a gammatone (top) using MP with the Gabor dictionary (below) and with the enriched dictionary (bottom). (B) Example of the decomposition of an asymmetric signal [waveform envelope given by $t(10^{-7}+t^{2.5})^{-1}$]. The latency positions of the original signal are marked with bars, and the latencies of the fitted waveforms marked with dots.

into two waveforms. The first explains the largest part of the energy and the second accounts for the initial part of it. In case of the Gabor function dictionary, two problems are apparent. The first concerns the fact that one waveform is split into several atoms. The second is that the Gabor atom shows the presence of some energy before the beginning of the actual signal. This has been referred to as the dark energy effect (e.g., Sturm *et al.*, 2008). The approach proposed herein solves both of these problems [see Fig. 2(A), bottom]. While both methods estimated the frequency with an adequate degree of precision, use of the Gabor dictionary shifted the latency of the dominant component by 1.08 ms, while using the enriched dictionary shifted it by only 0.28 ms compared with the simulated signal. As stated above, there is a specific type of long-lasting component that is often present in TEOAEs, which starts with a short burst of high magnitude and then slowly decays. This component has a higher asymmetry than the gammatone. An example of this phenomenon is depicted in Fig. 2(B). Again, using the enriched dictionary, the MP algorithm was able to fit a single waveform to the signal. The examples shown in Fig. 2 indicate that the use of the enriched dictionary is universally successful, as it is able to approximate gammatones and other asymmetric signals very well.

In order to compare the two dictionaries in a more quantitative manner, a set of 1000 signals was created that resembled TEOAEs. In most studies, TEOAEs are analyzed using five half-octave frequency bands centered on 1000, 1414, 2000, 2828, and 4000 Hz. For this reason, the synthesized signals were created using five gammatones, one from each different half-octave frequency band. These were drawn randomly from unbiased distributions within each frequency band. The latency depended on frequency according to the formula $t=10f^{-0.8}$ (t in ms and f in kHz), which approximated the experimental latency-frequency dependence estimated by Jedrzejczak *et al.* (2004). The decay constant Γ was randomly chosen from the six values 100, 200, 300, 400, 500, and 600. In this way, the signal had components of different durations, from short bursts of energy to longer-lasting ones that spanned most of the data-acquisition window. The lower the decay constant Γ, the longer was the gammatone. For values of Γ lower than 100 Hz, the duration of the gammatone exceeded 20 ms, which is the standard time window used for most TEOAE measurements (Kemp *et al.*, 1986). These values were therefore not used. Figure 3(A) shows one of the simulated signals and the five gammatones of which it is composed. Figure 3(B) shows the TF map of the original signal, and Figs. 3(C) and 3(D) illustrate the TF maps obtained using MP decomposition with the enriched and Gabor dictionaries, respectively. The differences between the distributions shown in Fig. 3 can clearly be seen, with the most obvious being apparent between Figs. 3(B) and 3(C). The Gabor dictionary comprised only of symmetrical waveforms does not reproduce the original distribution well. The distribution shown in Fig. 3(D) obtained by means of the enriched dictionary is very similar to the original one shown in Fig. 3(B), with the degree of asymmetry being the same. The minor differences at the 0.6% energy level are caused by differences between the waveforms from the enriched dictionary and the gammatones.

The results of comparison between the parameters of the simulated and fitted waveforms in case of both dictionaries for all 1000 simulated signals are presented in Fig. 4. They are shown in the form of their average deviations of the

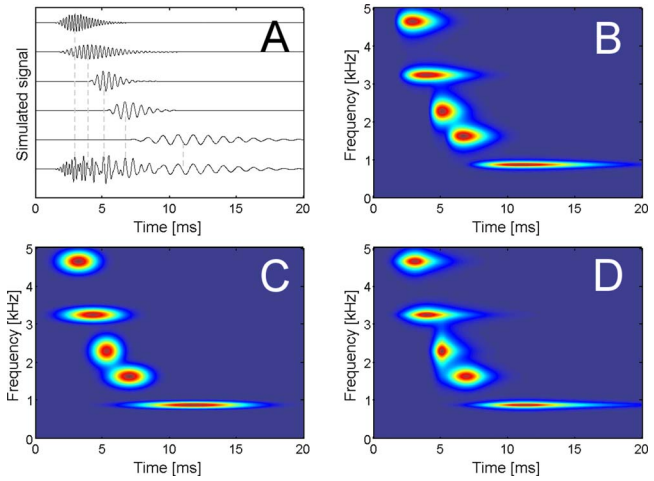


FIG. 3. (Color online) (A) Simulated signal constructed using five gammatones of different frequencies, latency, and decay constant. (B) TF amplitude distribution of the simulated signal. (C) TF amplitude distribution obtained from the decomposition of the simulated signal using MP with the Gabor dictionary. (D) TF amplitude distribution obtained from the decomposition of the simulated signal using MP with the enriched dictionary.

estimated frequency and latency from the original values as a function of the decay constant of the gammatone. It can be seen in Fig. 4(A) that the deviations of the frequency values are small (<10 Hz) and similar for both the Gabor and the enriched dictionaries. In terms of the latency shown in Fig. 4(B), the use of asymmetric waveforms resulted in similar deviations for different values of Γ, while the Gabor atoms tended to deviate more from the original values at the lower

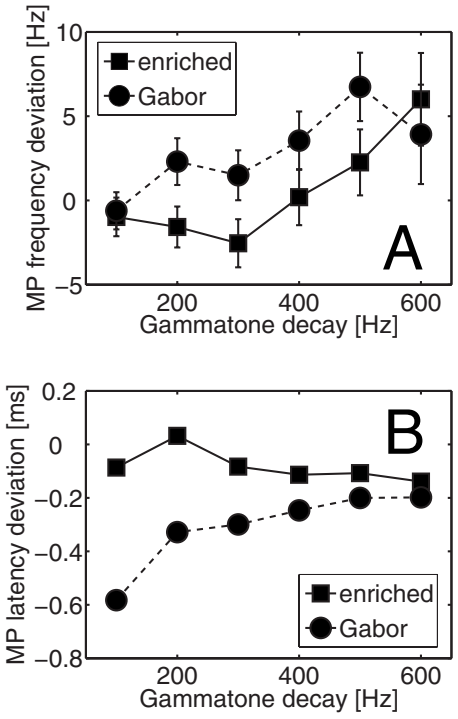


FIG. 4. Average deviation of the estimated frequency (A) and latency (B) using the original values of gammatone for both dictionaries, shown in relation to the decay constant of gammatone (horizontal axis). The higher the decay constant, the shorter is the waveform. Squares represent the enriched dictionary. Circles represent the dictionary of Gabor functions.

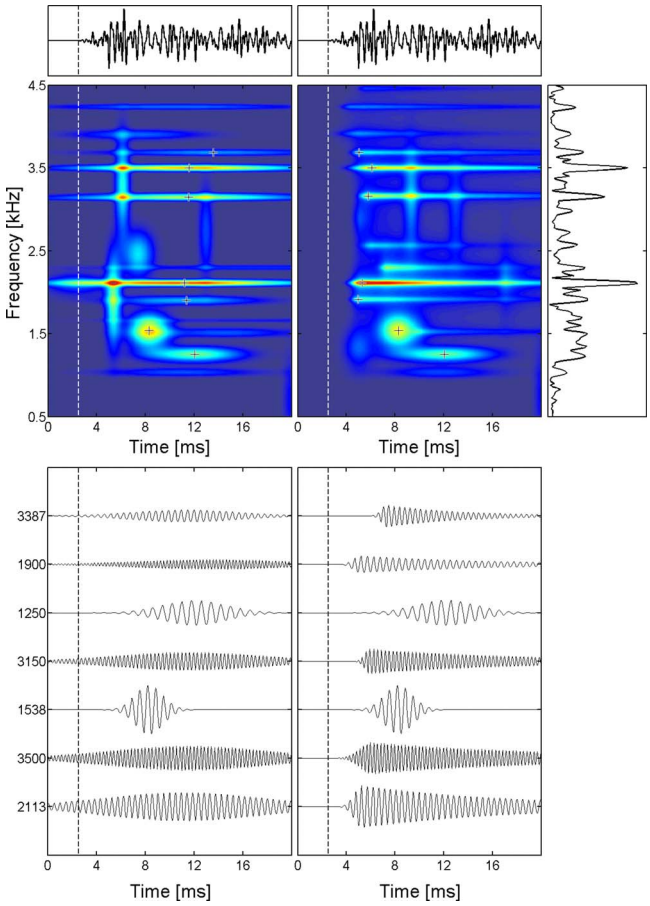


FIG. 5. (Color online) Example of TF amplitude distributions of TEOAEs for subject with SSOAEs. From left to right: map of the decomposition using MP with the Gabor dictionary, map of decomposition using MP with the enriched dictionary, and Fourier spectra. The crosses mark the positions of the maximum level in the TF plane. The time evolution of the signal is plotted above each TF map. Below each map are shown seven atoms of the decompositions. At the left of each waveform, frequency is also given.

values of Γ. For this simulation, the maximum deviation from the original values introduced by the Gabor dictionary was around 0.6 ms.

These simulations provide only a general assessment of the performance of the methods used. The simulated tests were designed only to give information about TEOAE-like signals. The estimates of performance and resolution should therefore be treated with caution, especially in relation to the analysis of the different signals used. A more general discussion of the properties of the MP algorithm may be found elsewhere (e.g., Delprat *et al.*, 1992; Gribonval *et al.*, 1996; Durka *et al.*, 2001a).

B. Results obtained using experimental data

The MP algorithm was used with the two different dictionaries to decompose actual TEOAE data. In Fig. 5, the analyzed signal is shown in the upper part of each figure, while the corresponding basic waveform is shown beneath it with the TF magnitude distribution at the bottom. The picture on the left of Fig. 5 corresponds to the standard Gabor dictionary (as used in previous reports, e.g., Jedrzejczak *et al.*, 2004, 2006). The illustration on the right shows the distribution of the results using the MP algorithm with both symmet-

ric and asymmetric waveforms. The spectrum of the signal is presented at the right, and rotated to match the frequency axis of the TF maps. On the TF maps shown in Fig. 5, it is clear that there were two main types of activity, namely, short (i.e., durations <5 ms, shown as "blobs") and long (i.e., durations >10 ms, shown as narrow traces). These narrow components are connected to long-lasting and sometimes spontaneous activity (Jedrzejczak *et al.*, 2008a). The main difference between the Gabor and enriched dictionaries is that the long-lasting components may be split into two or more waveforms by the Gabor dictionary (similar to the example shown above in Fig. 2). This characteristic is connected with the exponential decay of these components, which cannot be reproduced by the single symmetric Gabor atom. The enriched dictionary deals with components of this type in a more appropriate manner. It may also be noted that the effect of dark energy (i.e., the occurrence of activity that takes place before the beginning of the signal), which is visible in the map obtained using the Gabor dictionary, is eliminated when the enriched dictionary is used. In Fig. 5, atoms corresponding to transient components are visible as well as two structures below 1.5 kHz. In both dictionaries, they are represented by symmetric waveforms (i.e., Gabor atoms). This is the case also for the signal shown in the example in Fig. 6(A). In this plot, another interesting feature may be observed in the form of the atom at a frequency around 1.4 Hz, which is not a SOAE, although it reveals some degree of asymmetry. The effect of dark energy elimination by the enriched dictionary is visible as well in Fig. 6(A) for a low-frequency component. In Fig. 6, one can see clearly the differences between both dictionaries, namely, longer components are often described in the Gabor dictionary as separate atoms, which are especially pronounced in Fig. 6(B). This is the only way in which the Gabor dictionary can account for the decaying characteristics of the component.

The effect of the application of different dictionaries may be viewed in terms of the difference between the percentages of energy of the signal explained by the two dictionaries using the same number of atoms as shown in Fig. 7. This data set was divided into ears with ($n=33$) and without ($n=25$) detected SOAEs. The difference between the enriched and Gabor dictionaries is particularly apparent when they are applied to ears with SOAEs. This is due mainly to the fact that SOAE ears have more long-lasting components, as seen in Fig. 5. As stated above, the Gabor dictionary needs at least two waveforms to enable a fit to a long-lasting decaying component, while the enriched dictionary needs only one. A significant difference between the dictionaries was observed when the number of waveforms taken into account was between 1 and 11 (at the level $p<0.05$). The Gabor dictionary approaches the level of energy explained by the enriched dictionary after a greater number of iterations. The difference between the methods vanishes after a large number of iterations, since by then the main part of the energy has already been explained. The residual energy is usually dominated by the weaker fluctuations of the signal and the noise. This result is consistent with the previous studies where, based on the careful analysis of residual noise, the number of basic waveforms describing the TEOAE signals was estimated on average as 10 (Jedrzejczak *et al.*, 2006).

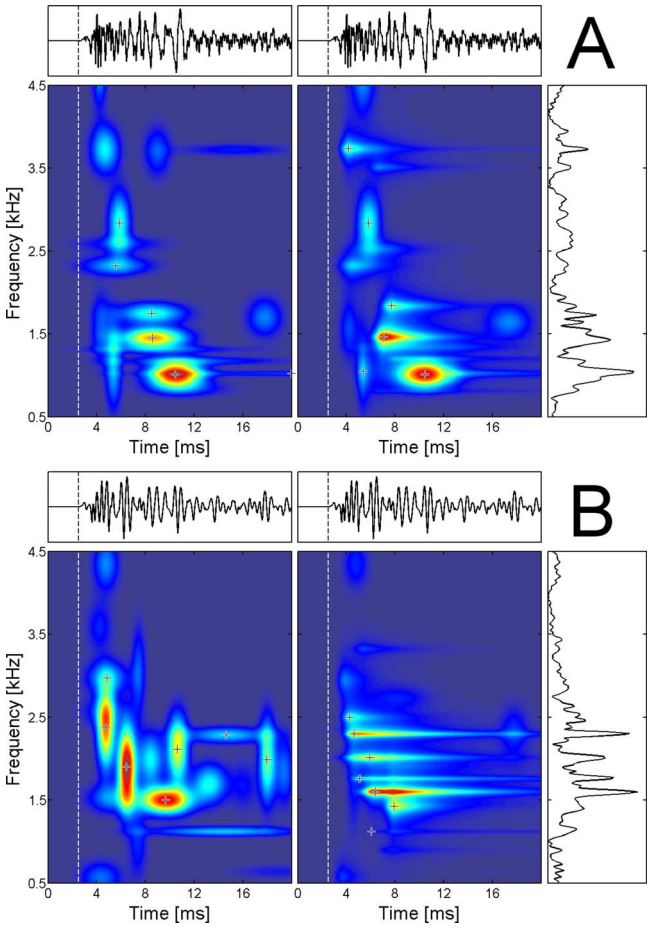


FIG. 6. (Color online) Examples of TF amplitude distributions of TEOAEs for two subjects without SOAEs. From left to right: map of the decomposition using MP with the Gabor dictionary, map of decomposition using MP with the enriched dictionary, and Fourier spectra. The crosses mark the positions of the maximum level in the TF plane. The time evolution of the signal is plotted above each TF map.

The implementation of both symmetric and asymmetric waveforms in the enriched dictionary provided an opportunity to check the prevalence of different types of components. The atom was considered asymmetric when the point

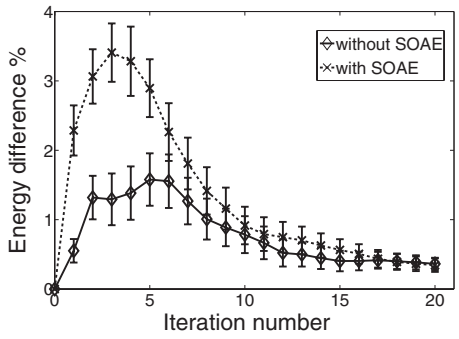


FIG. 7. The difference between the average percentages of signal energy accounted for by the given number of waveforms for the enriched and Gabor dictionaries. The data set was divided into ears with detected SOAEs (cross) and ears without SOAEs (diamond). A significant difference between ears with and without SOAEs was found when the number of waveforms taken into account was between 1 and 11 (at level $p<0.05$).

at which the Gaussian envelope changed into an exponential decay function was placed at half σ [Eq. (1)] from the maximum as in the example in Fig. 1(B). In all the investigated ears, of the 10 with the highest energy atoms, 63% were asymmetric, compared to 67% in ears with SOAEs and 58% in ears without SOAEs. Usually for the ears without SOAEs, the degree of asymmetry was small.

Finally, the latency-frequency dependence was estimated by means of both dictionaries for the whole group and for ears with and without SOAEs separately as shown in Fig. 8. These data were divided into nine quarter-octave frequency bands centered at 1000, 1189, 1414, 1682, 2000, 2378, 2828, 3364, and 4000 Hz to facilitate the comparison between them. Power-law relationships were then fitted to the data obtained using both methods (Table I). For ears without SOAEs in Fig. 8(A), the differences between latency-frequency estimations by both dictionaries were small, which could have been expected, since the degree of asymmetry of atoms was small. In case of the entire population of Fig. 8(C), and especially for ears with SOAEs as shown in Fig. 8(B), the fit for the data calculated by the Gabor dictionary was shifted toward longer latencies. This effect is especially pronounced at higher frequencies. It is also remarkable that in case of the Gabor dictionary the experimental points had a large spread and the goodness of fit to the exponential function was poorer (Table I). These effects were caused by an incorrect estimation of the latency of long-lasting components by the Gabor dictionary. The origin of this result can be seen directly by comparison of the latency of waveforms from different dictionaries in Fig. 5 (i.e., shifts of maxima toward longer latencies introduced by the Gabor dictionary). This effect was also evident in the examples with synthetic signals shown above in Fig. 2. One can see that time positions (marked by crosses) of long-lasting components are significantly shifted by the Gabor dictionary. On the other hand, for the enriched dictionary, the latency-frequency dependences from Fig. 8 were nearly the same in all three situations. The equation of the best fit line obtained for the enriched dictionary was $t=9.4f^{-0.54}$ for the whole data set [Fig. 8(C)].

V. DISCUSSION

The decomposition of a time series using the MP algorithm with an enriched dictionary encompassing both symmetric and asymmetric functions has herein been introduced for the analysis of TEOAEs. This set of numerical functions is broader than that previously used elsewhere (e.g., Jedrzejczak *et al.*, 2004, 2008a; Notaro *et al.*, 2007). The basic waveforms were constructed using a combination of two functions, namely, the Gabor and the exponential. This approach was dictated by the flexibility of the decomposition and the simplicity of the calculations. In this way, many different types of signal can be approximated, from highly asymmetric damped sinusoids to symmetrical signals with a Gaussian envelope. The performance of this new approach was evaluated in the present study using both simulations and the analysis of real TEOAEs. The simulations showed that the use of the MP algorithm with an enriched dictionary

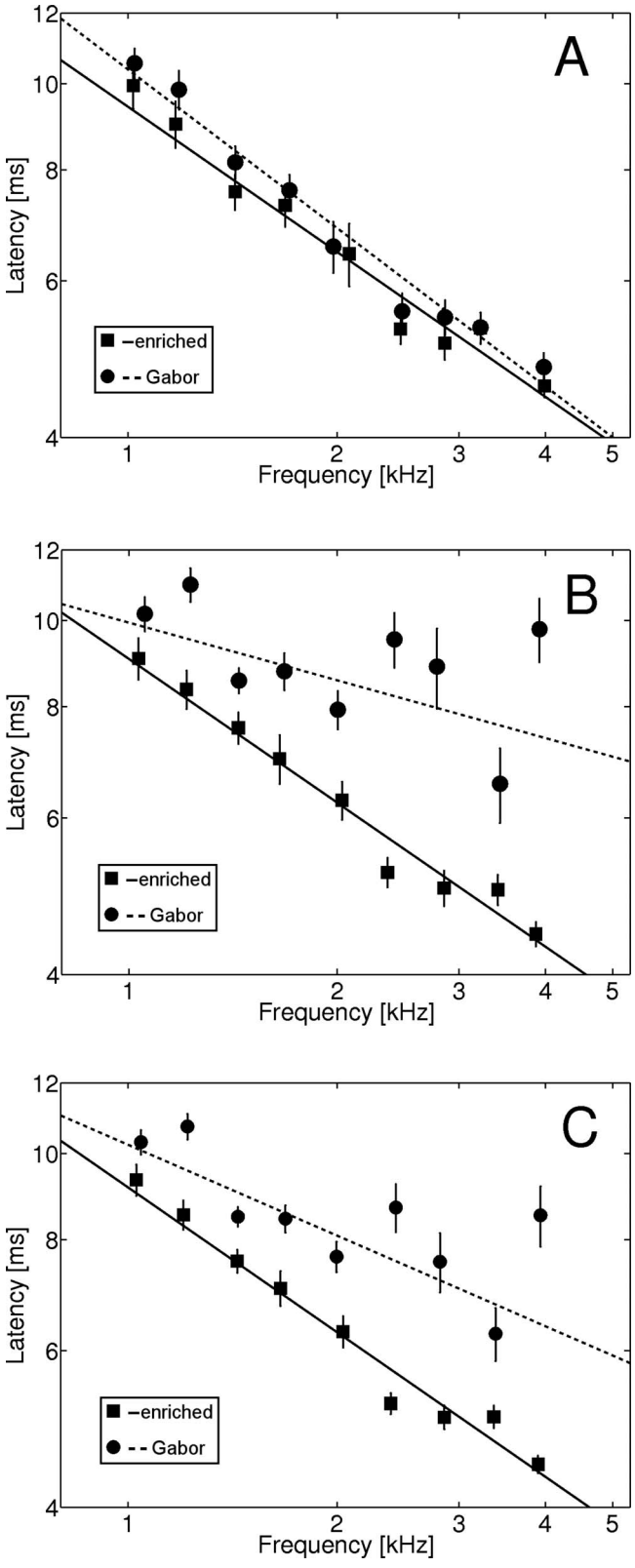


FIG. 8. Average latency in relation to frequency for TEOAE data calculated using MP with the enriched (square) and Gabor (circle) dictionaries for subjects without SOAEs (A), with SOAEs (B), and all the subjects together (C). The equations of the best fit lines obtained for different data sets were as follows: (A) $t=9.4f^{-0.54}$ for the enriched dictionary (solid line) and $t=10.3f^{-0.6}$ for the Gabor dictionary (dashed line). (B) $t=9.0f^{-0.54}$ for the enriched dictionary (solid line) and $t=9.9f^{-0.22}$ for the Gabor dictionary (dashed line). (C) $t=9.2f^{-0.54}$ for the enriched dictionary (solid line) and $t=10.2f^{-0.34}$ for the Gabor dictionary (dashed line).

TABLE I. The coefficients of power law fits ($t=af^b$) for TEOAE data calculated using MP analysis with the enriched and Gabor dictionaries for subjects without SOAEs, with SOAEs, and all the subjects together (see Fig. 8). The goodness of fit of the curves was estimated by means of R^2.

	Dictionary	a	b	R^2
Ears without SOAEs	Gabor	10.3	−0.60	0.98
	Enriched	9.4	−0.54	0.96
Ears with SOAEs	Gabor	9.9	−0.22	0.09
	Enriched	9.0	−0.54	0.98
All ears	Gabor	10.2	−0.34	0.46
	Enriched	9.2	−0.54	0.98

offers similar frequency resolution and better time resolution than the classic approach using a symmetric Gabor dictionary.

Synthetic signals resembling TEOAEs were created using gammatones. These artificial waveforms were introduced by Aertsen and Johannesma (1980) and have been used in other studies as a model of basilar membrane movement (e.g., Patterson *et al.*, 1992). In the current study, gammatones were selected in order to preserve some consistency with previous studies that assessed TF methods in relation to TEOAE analysis (e.g., Wit *et al.*, 1994; Tognola *et al.*, 1997; Sisto and Moleti, 2002; Jedrzejczak *et al.*, 2004). The gammatone is only one possible model of the TEOAE, and was therefore not chosen as a basis for the new dictionary. It is apparent that the enriched dictionary introduced here is more flexible and is capable of approximating not only gammatones, but also many other waveforms. The TEOAE simulations performed in this study showed that the frequency resolution of the MP algorithm was, on average, better than 20 Hz for complex signals resembling TEOAEs. By taking into account asymmetric waveforms, the temporal accuracy was improved in relation to the standard *Gabor dictionary.* When a synthetic signal created from gammatones was decomposed by the MP algorithm using symmetric waveforms, the latency values were higher than the simulated ones. The enriched dictionary performed equally well for different degrees of asymmetry, and its time accuracy was, on average, 0.1 ms, regardless of the degree of asymmetry or duration of the component. This high time resolution is especially important for studying subtle changes in latency. For example Tognola *et al.* (2005) reported shifts in latency on the level below 0.5 ms in subsequent TEOAE measurements in preterm neonates some 8 weeks after birth.

However, these estimates must be treated with caution. Despite the fact that the simulations were repeated 1000 times with parameters varying within certain ranges, simulations performed in a different way might have produced smaller or greater deviations. The MP algorithm gives a perfect frequency match when the simulated signal has only one frequency component as in the example shown above in Fig. 2. However, the main idea presented herein was to create a simple set of synthetic signals that mimicked the multicomponent characteristics of TEOAEs. In a case in which such signals consist of several components, the interference of waveforms that are close together in TF space may impose a limitation on the accuracy of the results. The basic property binding the TF resolution for discrete signals, namely, the Heisenberg–Gabor uncertainty principle (Gabor, 1946), must also be taken into account, as should the influence of noise when using experimental signals.

In common with other MP techniques, the proposed dictionary contains Gabor atoms. However, it also accounts for those waveforms that do not comply with a symmetric form. For example, a dictionary constructed only from gammatones would not be able to account for the presence of symmetric components; and neither could it approximate the stable long-lasting components. The type of asymmetric waveform proposed here is capable of successfully approximating gammatones in addition to many other functions.

The accuracy of the TF representations obtained by means of such different methods as the short-time Fourier transform, WT, the Wigner–Ville distribution, or the MP algorithm using a Gabor dictionary was considered earlier (Jedrzejczak *et al.*, 2004). The conclusions highlighted the advantages of the MP algorithm over other methods, which may be briefly summarized as (i) the objective parametrization of the signal with the only limitation being the size and complexity of the dictionary, (ii) the lack of additional assumptions (e.g., the arbitrary choice of the appropriate window in the case of a spectrogram), and (iii) a TF representation free of cross terms. The robustness of the MP algorithm with respect to noise has also been demonstrated (Jedrzejczak *et al.*, 2006). The introduction of the enriched dictionary further improves the accuracy of the TF representation.

The main advantage of using the MP algorithm with the enriched dictionary manifests itself in a better account of the long-lasting signal components, which are associated with SOAEs and in which some ears can form a substantial part of the TEOAE (e.g., Kulawiec and Orlando, 1995). According to Zhang and Penner (1998), SOAEs occur in 75% of population, and they are more common in females than in males. A similar or even higher prevalence was observed for long-lasting OAE components, which are characterized by an approximately exponential decay (Sisto *et al.*, 2001; Jedrzejczak *et al.*, 2008a). As noted by Notaro *et al.* (2007), the standard MP algorithm tends to split such activity into two (or even more) components, i.e., one with a high level and short duration related to the response onset and the other spanning the entire duration of the signal and associated with the tail of the decaying emitted activity. Use of the MP algorithm with the enriched dictionary can fit one waveform to such a component and correctly determines its maximum (i.e., its latency). In the case of ears with SOAEs, the enriched dictionary needs fewer waveforms to explain the energy of the main components, as shown in Fig. 7. The representation by the enriched dictionary is thus more efficient, which is a desirable property for any formalism. The improvement is clearly visible for first 11 atoms. It has been shown by Jedrzejczak *et al.* (2006) that the main features of the signal are explained, on average, by ten waveforms. Therefore, differences found for both dictionaries in this range of components point out to the significant progress introduced by application of the enriched dictionary.

Use of the enriched dictionary allows the correction of any artifacts produced by the symmetric dictionary. In particular, the use of the MP algorithm with the Gabor dictionary can lead to the dark energy effect, i.e., the occurrence of energy before the start of the actual signal (e.g., Gribonval *et al.*, 1996). This is particularly the case for acoustic signals, which often consist of a short rising component followed by a sustained or slowly decaying tail. The components of such shapes are visible in TF maps of TEOAEs with long-lasting components. The use of asymmetric waveforms in the dictionary eliminates the effect described above. The only drawback associated with the application of the enriched dictionary is that because of the additional parameter (asymmetry), the algorithm is more complex and the calculations take longer to perform.

Latency-frequency properties determined using the enriched dictionary slightly favor more lower-latency values in relation to results reported in previous studies from other laboratories. For example, 4- and 1-kHz latencies are 4 and 9 ms, respectively, while Sisto and Moleti (2002) using wavelets achieved functions shifted with respect to these values by 1 ms toward longer values, on average. Despite the fact that the wavelet analysis represents a different approach, the basic waveform applied by these authors is symmetrical, which may explain the shift in latency toward longer values.

The latency-frequency dependence is strongly influenced by the presence of long-lasting components. This problem was noted already in a study of TEOAEs using WTs (Sisto and Moleti, 2002). Later, it was shown by Jedrzejczak *et al.* (2007) that long-lasting components detected by MP with the Gabor dictionary do not reveal any dependence of latency on frequency. In fact, for symmetric dictionaries, component latencies are ill defined. The introduction of the asymmetric dictionaries eliminates this difficulty.

Usually, latency-frequency curves are constructed with the help of symmetric functions and without taking into account the presence of long-lasting components connected with SOAEs. This could be a source of significant errors. Notaro *et al.* (2007) determined several latency-frequency relationships using MP and WT analyses. In a case when they applied a procedure that arbitrarily removed atoms, which were connected to spontaneous and distortion OAE activities, they obtained a latency-frequency function that was closest to the one presented here in Fig. 8 for the enriched dictionary.

The present group of subjects consisted mainly of females, which exhibit more SOAEs than males. Therefore, the differences between functions obtained by means of enriched and symmetric dictionaries were especially pronounced. Nevertheless, the results of present study point to the importance of identifying the long-lasting components and the correct estimation of their latencies. This is especially important in the context that latency-frequency data provide basic evidence concerning the validity of different modes of OAE generation (e.g., Moleti and Sisto, 2008).

Introduction of asymmetric atoms improved the estimation of the latencies of principal components. The latency property is becoming an important parameter to measure, and in the future may be helpful in clinical diagnosis (e.g., Jedrzejczak *et al.*, 2005; Moleti *et al.*, 2005). For example, Jedrzejczak *et al.* (2007) demonstrated in a study of neonates that the presence of long-lasting components paradoxically increased the estimation of OAE latencies. The contribution of such an effect needs to be taken into account and the latency corrected accordingly in order to come to a correct diagnosis. Thus, using the enriched dictionary makes possible a better identification of long-lasting components and provides a correct estimation of their latencies. Asymmetric waveforms account well for longer-lasting activity, being able to better estimate their decaying nature and eliminate the effect of energy that appears before the start of the signal in the case of symmetric dictionaries.

VI. CONCLUSIONS

The aim of the present study was to extend the TF analysis of OAEs by adding the asymmetry parameter to those basic waveforms that have traditionally been fitted to biological and artificial signals. This resulted in a more general and flexible procedure for analyzing the principal components contained in a signal. The new approach, using an enriched dictionary consisting of both symmetric and asymmetric waveforms, was compared with the standard procedure based on using a symmetric Gabor atom dictionary. The tests made using synthetic signals showed that both dictionaries possess similar frequency-resolution capabilities, but that the enriched dictionary provides better time resolution. As may be expected, the enriched dictionary performed better when the signal contained asymmetric elements. Furthermore, the new approach negates the dark energy effect, i.e., the presence of false energy traces on TF maps created from MP based on the Gabor dictionary. Analyses of actual TEOAE signals demonstrated that the enriched dictionary provided a correct estimation of latencies for components having either short- or long-decay times. Long OAE components are often related to SOAEs, which are present in the majority of ears. Moreover, the correct estimation of latency is crucial for constructing accurate latency-frequency functions, especially since unbiased latency assessments may be important in diagnosing certain hearing impairments. Finally, the generation of OAEs is not yet fully understood, and the application of a more general, less restricted dictionary of functions in decomposing their principal components presents an opportunity to better understand and interpret the sources of emission generation.

ACKNOWLEDGMENTS

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. NN518 0924 33). We thank E. A. Bartnik, P. J. Durka, and the anonymous reviewers for their helpful comments. We are grateful to A. Matysiak for providing parts of the code for the implementation of the MP algorithm. We would also like to thank E. Pilka for assistance in data collection.

Aertsen, A. M., and Johannesma, P. I. (**1980**). "Spectro-temporal receptive fields of auditory neurons in the grassfrog. I. Characterisation of tonal and natural stimuli," Biol. Cybern. **38**, 223–234.

Cheng, J. (**1995**). "Time-frequency analysis of transient evoked otoacoustic emissions via smoothed pseudo Wigner-Ville distribution," Scand. Audiol. **24**, 91–96.
Delprat, N., Escudié, B., Guillemain, P., Kronland-Martinet, R., Tchamitchian, P., and Torrésani, B. (**1992**). "Asymptotic wavelet and Gabor analysis: Extraction of instantaneous frequencies," IEEE Trans. Inf. Theory **38**, 644–664.
Durka, P. J., Ircha, D., and Blinowska, K. J. (**2001a**). "Stochastic time-frequency dictionaries for matching pursuit," IEEE Trans. Signal Process. **49**, 507–510.
Durka, P. J., Ircha, D., Neuper, C., and Pfurtscheller, G. (**2001b**). "Time-frequency microstructure of event-related electro-encephalogram desynchronisation and synchronisation," Med. Biol. Eng. Comput. **39**, 315–321.
Gabor, D. (**1946**). "Theory of communication," J. Inst. Electr. Eng. **93**, 429–457.
Goodwin, M., and Vetterli, M. (**1999**). "Matching pursuit and atomic signal models based on recursive filter banks," IEEE Trans. Signal Process. **47**, 1890–1902.
Goupillaud, P., Grossman, A., and Morlet, J. (**1984**). "Cycle-octave and related transforms in seismic signal analysis," Geoexploration **23**, 85–102.
Greenwood, D. D. (**1990**). "A cochlear frequency position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Gribonval, R., and Bacry, E. (**2003**). "Harmonic decomposition of audio signals with matching pursuit," IEEE Trans. Signal Process. **51**, 101–111.
Gribonval, R., Depalle, P., Rodet, X., Bacry, E., and Mallat, S. (**1996**). "Sound signals decomposition using a high resolution matching pursuit," Proceedings of the International Computer Music Conference, August, pp. 293–296.
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.
Jedrzejczak, W. W., Blinowska, K. J., Kochanek, K., and Skarzynski, H. (**2008a**). "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., Lorens, A., Piotrowska, A., Kochanek, K., and Skarzynski, H. (**2009**). "Otoacoustic emissions evoked by 0.5 kHz tone bursts," J. Acoust. Soc. Am. **125**, 3158–3165.
Jedrzejczak, W. W., Smurzynski, J., and Blinowska, K. J. (**2008b**). "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.
Kulawiec, J. T., and Orlando, M. S. (**1995**). "The contribution of spontaneous otoacoustic emissions to the click evoked otoacoustic emissions," Ear Hear. **16**, 515–520.
Malinowska, U., Durka, P. J., Zygierewicz, J., Szelenberger, W., and Wakarow, A. (**2007**). "Explicit parameterization of sleep EEG transients," Comput. Biol. Med. **37**, 534–541.
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.
Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.
Moleti, A., and Sisto, R. (**2008**). "Comparison between otoacoustic and auditory brainstem response latencies supports slow backward propagation of otoacoustic emissions," J. Acoust. Soc. Am. **123**, 1495–1503.
Moleti, A., Sisto, R., Tognola, G., Parazzini, M., Ravazzani, P., and Grandori, F. (**2005**). "Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates," J. Acoust. Soc. Am. **118**, 1576–1584.
Notaro, G., Al-Maamury, A. M., Moleti, A., and Sisto, R. (**2007**). "Wavelet and matching pursuit estimates of the transient-evoked otoacoustic emission latency," J. Acoust. Soc. Am. **122**, 3576–3585.
Ozdamar, O., Zhang, J., Kalayci, T., and Ulgen, Y. (**1997**). "Time-frequency distribution of evoked otoacoustic emissions," Br. J. Audiol. **31**, 461–471.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," *Auditory Physiology and Perception*, *Proceedings of the 9th International Symposium on Hearing*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
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., Moleti, A., and Lucertini, M. (**2001**). "Spontaneous otoacoustic emissions and relaxation dynamics of long decay time OAEs in audiometrically normal and impaired subjects," J. Acoust. Soc. Am. **109**, 638–647.
Sturm, B. L., Shynk, J. J., and Daudet, L. (**2008**). "Dark energy in sparse atomic estimations," IEEE Trans. Audio, Speech, Lang. Process. **16**, 671–676.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
Tognola, G., Grandori, F., and Ravazzani, P. (**2001**). "Time-frequency analysis of neonatal click-evoked otoacoustic emissions," Scand. Audiol. Suppl. **52**, 135–137.
Tognola, G., Parazzini, M., de Jager, P., Brienesse, P., Ravazzani, P., and Grandori, F. (**2005**). "Cochlear maturation and otoacoustic emissions in preterm infants: A time-frequency approach," Hear. Res. **199**, 71–80.
Wang, W., Guo, Z., Yang, J., Zhang, Y., Durand, L. G., and Loew, M. (**2001**). "Analysis of the first heart sound using the matching pursuit method," Med. Biol. Eng. Comput. **39**, 644–648.
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.
Zhang, T., and Penner, M. J. (**1998**). "A new method for the automated detection of spontaneous otoacoustic emissions embedded in noisy data," Hear. Res. **117**, 107–113.
Zhang, Z. G., Zhang, V. W., Chan, S. C., McPherson, B., and Hu, Y. (**2008**). "Time-frequency analysis of click-evoked otoacoustic emissions by means of a minimum variance spectral estimation-based method," Hear. Res. **243**, 18–27.

Informational factors in identifying environmental sounds in natural auditory scenes

Robert Leech
Division of Neuroscience and Mental Health, Imperial College, London W12 0NN, United Kingdom

Brian Gygi
East Bay Institute for Research and Education, Martinez, California 94553

Jennifer Aydelott and Frederic Dick
School of Psychology, Birkbeck, University of London, London WC1E 7HX, United Kingdom

(Received 15 September 2008; revised 2 August 2009; accepted 24 August 2009)

In a non-linguistic analog of the "cocktail-party" scenario, informational and contextual factors were found to affect the recognition of everyday environmental sounds embedded in naturalistic auditory scenes. Short environmental sound targets were presented in a dichotic background scene composed of either a single stereo background scene or a composite background scene created by playing different background scenes to the different ears. The side of presentation, time of onset, and number of target sounds were varied across trials to increase the uncertainty for the participant. Half the sounds were contextually congruent with the background sound (i.e., consistent with the meaningful real-world sound environment represented in the auditory scene) and half were incongruent. The presence of a single competing background scene decreased identification accuracy, suggesting an informational masking effect. In tandem, there was a contextual pop-out effect, with contextually incongruent sounds identified more accurately. However, when targets were incongruent with the real-world context of the background scene, informational masking was reduced. Acoustic analyses suggested that this contextual pop-out effect was driven by a mixture of perceptual differences between the target and background, as well as by higher-level cognitive factors. These findings indicate that identification of environmental sounds in naturalistic backgrounds is an active process that requires integrating perceptual, attentional, and cognitive resources. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238160]

PACS number(s): 43.66.Dc, 43.66.Rq [RYL] Pages: 3147–3155

I. INTRODUCTION

In research on language comprehension, it has long been recognized that we can isolate an individual voice among many competing talkers, e.g., the famous "cocktail-party scenario" (Broadbent, 1958; Cherry, 1953). Listeners can attend to an individual voice in a multitalker environment by focusing on its perceptual characteristics (i.e., the distinctive acoustic properties associated with differences in vocal quality, pitch, and vocal tract length; Brungart, 2001; Darwin and Hukin, 2000; Bregman, 1990; Brokx and Nootebohm, 1982) and its spatial location (i.e., differences in the timing and intensity of the acoustic signal as it reaches the different ears; Cherry, 1953; Darwin and Hukin, 2000; Drennan *et al.*, 2003; Freyman *et al.*, 2001; Hawley *et al.*, 2004). While listeners are able to use these different cues to attend to a given speaker, the presence of competing speech makes language comprehension more difficult by masking and/or diverting attention from the target signal (e.g., Brungart and Simpson, 2002), and by introducing conflicting meaningful content (e.g., Moll *et al.*, 2001; Leech *et al.*, 2007).

Perhaps surprisingly, research on listeners' comprehension of similarly complex everyday non-linguistic auditory scenes (such as identifying the sound of a car start on a busy street) has not been the focus of such intensive research (but cf. Bregman, 1990; Ballas and Mullins, 1991; Gygi and Shafiro, 2007). This is despite the fact that environmental sounds are important and highly frequent auditory events encountered in everyday life (Ballas, 1993). Most studies into environmental sound processing have assumed ideal listening conditions, that is, identifying a single sound without accompanying distracting noise (e.g., Gygi *et al.*, 2004, 2007; Saygin *et al.*, 2005). To our knowledge, only a few studies (Ballas and Mullins, 1991; Oh and Lutfi, 1999; Kidd *et al.*, 2007; Gygi and Shafiro, 2007) have investigated environmental sound processing in the presence of noise or other distractors. This is despite the fact that ecologically important environmental sounds that need attending to (such as a cell phone ring or far-off police siren) can often occur in the midst of multiple auditory scenes that can also be spatially segregated to greater or lesser degrees. For instance, when a pedestrian walks along a city street lined with shops, traffic and construction sounds may predominate in the ear directly exposed to the street, whereas a blend of sounds (music, conversation, and air conditioning) may be more dominant in the ear facing away from the street.

The effects of such competing signals on sound processing have been addressed in speech perception research on so-called energetic and informational masking of auditory targets. Such studies have shown that interference from additional sources of acoustic information is greater than would be predicted by filter models of the auditory periphery (see

Durlach *et al.*, 2003). Informational masking has often been investigated by measuring the masking effect on speech identification of having a simultaneously presented competing speaker in the unattended ear (Brungart and Simpson, 2002; Brungart *et al.*, 2005; Wightman and Kistler, 2005). In these studies, the presence of competing speech in both the contralateral and ipsilateral ears makes it more challenging to detect a target word than purely ipsilateral masking, particularly when the target word is presented at low signal-to-noise levels (Brungart and Simpson, 2002). Such informational masking effects are also observed for non-linguistic stimuli, for instance, spatially segregated, simultaneously presented multitone masking of pure tones (Kidd *et al.*, 2003; Wightman *et al.*, 2003). Interestingly, such masking effects can also be affected by the relative familiarity of the masker. In a study of pure-tone detection, Oh and Lutfi (1999) observed that when spectrally sampled environmental sounds were used as maskers for pure tones, those environmental sounds rated as familiar evoked considerably reduced masking of the tone target relative to unfamiliar sounds.

Target sound detection is affected not only by other distracting sounds presented simultaneously, but by the characteristics of the *ongoing* context that precedes and accompanies the target. Even with simple artificial sounds such as pure tones, frequency-swept, or frequency modulated tones, the degree of *acoustic* overlap between a target sound and sequentially (as well as simultaneously) presented background sounds can affect detection, such that decreasing similarity leads to decreased detection thresholds and increased detection accuracy (Durlach *et al.*, 2003; Cusack and Carlyon, 2003). Similarity of non-target stimuli can also degrade performance in some linguistic tasks. For example, in phoneme monitoring, target phonemes are more salient and easier to detect if they are perceptually incongruent with phonemes that were contained in the *preceding* word in the speech stream (Stemberger *et al.*, 1985). (It is worth noting that phonemes are more like meaningful sounds than a non-speech tone in that their accurate perception requires a many-to-one mapping from acoustics to phonemic category.) Such results suggest that listeners are building up perceptual expectations based on what they have heard previously.

In addition to such perceptually driven "lead-up" effects on sound detection, contextual effects linked to higher-level cognitive processes such as associative memory and semantic processing have also been shown to have effects on sound target detection. Ballas and Mullins (1991) found that the meaningful, real-world contextual information in a sequence of environmental sounds could bias the subsequent identification of an acoustically ambiguous target sound that was embedded in that sequence, such that the preceding sound context could lead listeners to incorrectly identify the ambiguous sound as one that fits with the context (e.g., the sound of a burning fuse misidentified as food frying when preceded by sounds associated with food preparation). Further, Gygi and Shafiro (2007) reported a preliminary study showing that unambiguous environmental sound targets are easier to detect and identify when they occur in an ongoing auditory scene where they are contextually incongruent (e.g., a doorbell presented in the context of a barn scene). Thus, a pop-out effect emerges for targets that are incompatible with the meaningful background in which they appear.

THE PRESENT STUDY

The current experiment builds on this initial work (in particular, Gygi and Shafiro, 2007) and introduces a novel experimental paradigm designed to investigate how detection and identification of short, naturalistic environmental sound targets presented in one ear or the other are affected by the auditory background scene in which the target is heard. Here, participants listened to a series of background auditory scenes, such as a seashore, a restaurant, and so forth. Some trials featured only a single stereophonic background scene, whereas others featured dual monaural scenes, e.g., a seashore scene was played to one ear, and a restaurant scene was simultaneously played to the other. Before listening to each scene, participants were shown one to three photographs of objects, and were played the short sounds associated with those objects (e.g., a glass breaking or a cow mooing). These short sounds—presented at different signal intensities—were embedded in one channel of the background scene that was presented next. The participant's task was to listen to the scene, and as soon as she/he heard one of the target short sounds, to push the button under the picture corresponding to that sound.

This paradigm requires the participant to perform a number of tasks that vary in their degree of difficulty depending on the characteristics of the background and target stimuli. In all conditions, the listener must simultaneously attend to two continuous auditory streams while maintaining one or more target sounds in memory. The listener must then isolate the target(s) from the background sounds (*detection*), and match each individual target to its corresponding picture (*identification*). Target detection depends on the successful perceptual separation of the target and background signals, which is influenced by the level of energetic and informational masking imposed on the target by the background. In multiple competing backgrounds, the total number of elements in the auditory scene is substantially increased, making detection more difficult by introducing a greater number of potential targets. Further, as noted above, previous studies of speech identification (Brungart and Simpson, 2002) and pure-tone detection (Kidd *et al.*, 2003; Wightman *et al.*, 2003) indicate that the presence of multiple competing backgrounds increases informational masking effects relative to a single background, resulting in decreased accuracy, particularly at low signal-to-noise ratios (SNRs). It was predicted that similar results would emerge for naturalistic environmental sounds.

In addition, the findings of Gygi and Shafiro (2007) suggest that the detection and identification of environmental sounds can be influenced by their compatibility with the meaningful context represented by the background auditory scene. Multiple factors may contribute to the contextual pop-out effect observed for incongruent targets. One possibility is that sounds from the same real-world auditory environment tend to share perceptual characteristics, such that targets may be acoustically more similar to compatible backgrounds than

incompatible backgrounds. This account predicts that the more accurate detection of incongruent targets can be accounted for in terms of the acoustic properties of the stimuli, specifically the degree of overlap between targets and backgrounds.

Another possibility is that the meaningful background activates associated real-world knowledge in semantic memory, allowing the listener to generate expectancies about the sound events that are likely to occur within a given auditory scene. Incongruent targets would violate these expectancies, which might serve to enhance their salience relative to the competing background sounds, thereby contributing to the pop-out effect. In evaluating this interpretation, however, it is important to consider the separate tasks of detection and identification, and how these may be affected by context-generated expectancies. More easily detected target stimuli may also be more readily identified, in which case both detection and identification accuracy should be greater for incongruent targets. Alternatively, stimuli that violate expectancies may be easier to detect, but their incompatibility with active memory representations may interfere with identification.

In both of the above accounts, the presence of multiple competing backgrounds would be expected to reduce or eliminate the pop-out effect, by increasing the complexity of the acoustic signal so that incongruent targets no longer stand out in terms of their perceptual characteristics, and/or by disrupting the activation of a set of memory representations that are consistent with a single, unified auditory scene.

These possibilities were investigated in the present study by manipulating various aspects of the experimental paradigm.

(1) The effect of relative signal intensity on the detection and identification of short, naturally occurring sound targets by other natural background scenes was assessed by manipulating target/background SNR at four different levels: −6, −3, 0, and +3 dB.
(2) Informational masking of sound targets was assessed by comparing performance on sound targets presented in one channel of a stereophonic single background (e.g., the sound of a dog barking embedded in the left channel of a stereo recording of a beach scene) versus a target presented in one channel of dual monaural scenes (e.g., the dog bark embedded in the beach scene, which is played to the left ear, with a traffic scene played to the right ear).
(3) As observed above, previous studies (e.g., Brungart and Simpson, 2002) have shown that informational masking may be enhanced when targets are presented at lower SNRs. This possibility is investigated here by testing for an interaction between SNR and number of background scenes, where increased informational masking should be observed when targets are presented at lower SNR, but not at higher SNR.
(4) Following on from Gygi and Shafiro (2007), contextual influences were assessed by comparing performance on sound targets that were embedded in background scenes that were contextually congruent or incongruent with the target (e.g., a cow moo in a farm versus factory background). In order to tease apart more expectancy-driven contextual effects from those due to acoustic similarities between target and background sounds, both contextual and acoustical predictors were entered into the regression model used to analyze the data.
(5) Finally, contextual "pop-out" effects might be lessened or extinguished by additional informational masking. This can be shown by testing for the interaction between the number of background scenes and contextual congruence of the target and background.

II. METHOD

A. Participants

48 adults participated in the experiment (25 females, mean age 35.5 years, range, 19–46 years). 26 of these participants were recruited specifically for the experiment and a further 22 took part in the study as part of a larger battery of auditory and language skills. All participants reported normal hearing and were native British English speakers.

B. Stimuli

Stimuli were 23 single environmental sounds, which were used as targets and 10 natural auditory scenes, used as backgrounds. Each target sound had a color image (250 × 250 pixels) associated with it, which corresponded to the real-world object being represented. All stimuli (backgrounds and targets) were meaningful, naturally occurring sounds, e.g., a "barn sound" background or a "dog bark" target sound. The target and background sounds were selected because they were reliably identified with appropriate verbal labels in previous studies: The background sounds were taken from Gygi and Shafiro (2007) and the target sounds from Cummings *et al.* (2006). The sounds were chosen to represent a wide range of sound categories (e.g., animals, vocal sounds, machines, actions, nature, etc). Target and background sounds were taken as edited in the previous studies and resampled to 22 050 Hz. Background auditory scenes varied in length from 8.00 to 14.91 s, and were scaled to have the same average intensity using PRAAT (Boersma and Weenink, 2009).

III. EXPERIMENTAL DESIGN

This was a 3-within-subjects factorial design. Within-subjects factors were background type [single background/dual background], target/background SNR [low (−6 dB, −3 dB) and high (0 dB, +3 dB)], and target/background contextual congruence [congruent/incongruent].

There were 40 trials in the experiment. In each trial, either a single stereophonic background sound was played in the left and right channels, or two distinct mono background sounds were presented in separate channels, e.g., a casino scene in the left channel, and a beach scene in the right channel. In the case of two different background scenes, the longer scene's duration was edited to match the duration of the shorter scene. One, two, or three short mono target sounds (e.g., "car horn," "door knock," and "flute") were

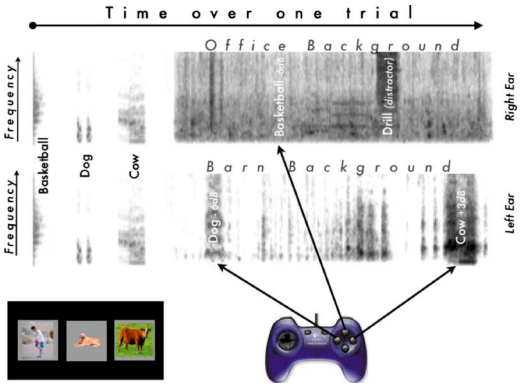


FIG. 1. (Color online) Schematic of the experiment design and presentation.

mixed into one channel of the background sound (see Fig. 1). An example of a congruent sound would be the car horn target sound mixed with the left channel of the "racing cars'" background sound. Multiple target sounds could appear in the same or different channels; e.g., two targets could appear in the left channel, and one in the right. Each target sound was either contextually congruent or incongruent with the background—for instance, the "sheep baaing" target sound was considered congruent with the "barn" background sound, but incongruent with the "office" background. Binary contextual congruence was assessed subjectively by the authors; a full list of target sounds and background scenes along with can be viewed in the supplementary materials (see EPAPS). In addition, acoustic measures of the similarity between each target and its ipsilateral background were calculated (see Sec. III B). Each target occurred twice in a congruent background context and twice in an incongruent background.

The 23 sound targets each occurred at four SNRs (−6, −3, 0, and 3 dB), with SNR calculated relative to the background sound (using rms, scaled in PRAAT) in the same auditory channel over the time window the target was being presented. (Choice of SNR levels was based on results from pilot work and previous studies; SNRs of −6 and −3 dB were classed as "low SNR," and 0 and +3 dB as "high SNR.") There were 92 target sounds presented over the 40 trials of the experiment, with 8 trials containing one target, 12 containing two targets, and 20 containing three targets. In addition, each target sound was also presented once as a contextually incongruent distractor, where the sound was played at 0 dB SNR without any associated picture. There was at most one incongruent distractor sound in any trial.

Target sound assignment was counterbalanced across background and contextual congruence conditions, with each target sound appearing once in each of the dual/congruent, dual/incongruent, single/congruent, and single/incongruent conditions. For each participant, the same number of targets was presented to each ear in each background and target contextual congruence.

Target-to-background signal-to-noise was counterbalanced within individual participants, and an additional level of counterbalancing in SNR levels *across* participants was employed to eliminate potential item-specific SNR effects. Participants were assigned to 1 of 4 different stimulus sets (2 sets were presented to 13 participants each, and 2 sets to 11 participants each), with each set varying in terms of specific target/background SNRs such that across participants, each target sound in each possible combination of contextual and background conditions appeared in every SNR. Similarly, for each trial, the location of the target pictures on the screen (left, middle, or right) was counterbalanced across participants. Finally, individual pairings of target side, background condition, and target contextual congruence were counterbalanced by reversing headphone side (left ear→right headphone) in half of the participants.

A. Procedure

The experiment was presented on an Apple G4 14-in. iBook laptop using the Psychophysics toolbox (Brainard, 1997) in MATLAB. Sounds were presented through Sennheiser HD 25-1 headphones, and participant responses were collected using a Logitech Precision gamepad. Participants held the gamepad while sitting in front of the laptop placed on a desk. Before starting the experiment, participants watched a 2-min-long QuickTime movie, which explained the task and provided an initial two practice trials. Before the main test session participants had three practice trials, which provided visual feedback. After these practice trials, participants received no further feedback beyond general encouragement.

Each trial had two parts: a target/picture familiarization phase and the main target detection phase. During the familiarization phase, three gray boxes appeared on the left, middle, and right of the screen. Depending on the number of target sounds, one to three target-associated pictures would appear in the boxes; pictures remained on the screen until the end of the trial (see Fig. 1). After a 500 ms silence, the target sound corresponding to the left-most picture was presented. After another 500 ms gap, the target sound corresponding to the next picture to the right was presented in sequence. There was 1.0 s of silence between the offset of the last target and the beginning of the main detection phase.

Participants heard the stereo background sound with between one and three mono target sounds inserted onto either the right or left channel. An incongruent distractor sound was present on half of all trials. Participants responded with one of three buttons on the gamepad, which corresponded to one of the three target boxes. For instance, if the participant heard the target represented by the picture on the left, the participant pressed the left button as soon as they heard the sound. Participants were not told whether the target sounds would appear in the left or right channel. Responses were recorded throughout the entire trial, and counted as "correct" only if participants pushed the button corresponding to the target within a 2-s time window starting 300 ms after target onset.

B. Acoustic measures

As noted above, congruence between the ipsilateral background and the target sound could arise from either acoustic similarity between the sounds and/or via real-world associative or semantic knowledge. To establish whether

contextual congruence between background and target was driven by acoustical properties of the sounds alone, a range of spectral, envelope, and periodicity similarity measures was computed. The acoustic measures were the same as those used in Gygi *et al.*, 2007 to characterize variation in environmental sounds. All measures were calculated using MATLAB (Mathworks, Natick, MA).

The predictive value of the acoustic measures for detection and identification of a sound target and for the degree of masking with ipsilateral background sounds was assessed by comparing each acoustic measure for each sound with mean identification accuracy. This was conducted in two ways: first with univariate models with each acoustic variable predicting response accuracy, and second with a stepwise multiple regression model that also included the binary experimenter-defined contextual congruence and the number of distinct background sounds (dual or single).

In addition, for each of the 38 acoustic variables, a measure of the similarity between each occurrence of target and ipsilateral background sounds was calculated as follows: (1) For each target sound, its time-matched segment of ipsilateral background sound was extracted; and (2) for each acoustic measurement, the absolute difference between the measurements for the ipsilateral background and target sounds was calculated. This resulted in an approximate similarity distance for each target/background pairing, which was subsequently entered into stepwise multiple regression models.

The acoustic measures were as follows.

(a) *Envelope measures.* (1) Long term rms/pause-corrected rms (an index of the amount of silence), (2) number of peaks (peak is defined as a point in a vector that is greater in amplitude than the preceding point by at last 80% of the range of amplitudes in the vector), (3) number of bursts [amplitude increases of at least 4 dB sustained for at least 20 ms, based on an algorithm developed by Ballas (1993), (4) total duration, and (5) burst duration/total duration (a measure of the "roughness" of the envelope).

(b) *Autocorrelation statistics.* (1) Number of peaks, (2) maximum, (3) mean peak, and (4) standard deviation (SD) of the peaks. Peaks (as defined above) in the autocorrelation function reveal periodicities in the waveform, and the statistics of these peaks measure different features of these periodicities, such as the strength of a periodicity and the distribution of periodicities across different frequencies.

(c) *Correlogram-based pitch measures (from* Slaney, 1995*).* (1) Mean pitch, (2) median pitch, (3) SD of pitch,(4) maximum pitch, (5) mean pitch salience, and (6) maximum pitch salience. The correlogram measures the pitch and pitch salience by autocorrelating in sliding 16 ms time windows. This captures spectral information and provides measures of the distribution of that information over time.

(d) *Moments of the spectrum.* (1) Mean (centroid), (2) SD, (3) skew, and (4) kurtosis.

(e) rms energy in octave-wide frequency bands from 63 to 16 000 Hz.

(f) *Spectral shift in time measures.* (1) Centroid mean, (2) centroid SD,(3) mean centroid velocity,(4) SD centroid velocity, and (5) maximum centroid velocity. The centroid mean and SD are based on consecutive 50-ms time windows throughout the waveform. The spectral centroid velocity was calculated by measuring the change in spectral centroid across sliding 50-ms rectangular time windows.

(g) *Cross-channel correlation.* This is calculated by correlating the envelopes in octave-wide frequency bands (or channels) ranging from 150 to 9600 Hz. It measures the consistency of the envelope across channels.

(h) *Modulation spectrum statistics.* The modulation spectrum, first suggested by Houtgast and Steeneken (1985), reveals periodic temporal fluctuations in the envelope of a sound. The algorithm used here divides the signal into frequency bands approximately a critical band wide, extracts the envelope in each band, filters the envelope with low-frequency bandpass filters (upper f_c ranging from 1 to 32 Hz), and determines the power at that frequency. The result is a plot of the depth of modulation by modulation frequency. The statistics measured were (1) the height and (2) frequency of the maximum point in the modulation spectrum, as well as the (3) number, (4) mean, and (5) variance of bursts in modulation spectrum (using the burst algorithm described above).

(i) *Spectral flux statistics.* Spectral flux is another measure of the change in the spectrum over time. It is the running correlation of spectra in short (50-ms) time windows. The (1) mean, (2) SD, and (3) maximum value of the spectral flux were used in this analysis.

IV. RESULTS

Background condition, contextual congruence, and SNR were entered into a full-factorial analysis of variance (ANOVA) model using SPSS V16. In order to make model interpretation more transparent, all factors were coded as binary variables by collapsing target/background SNR into lower (−6 dB, −3 dB) and higher (−0 dB, +3 dB) SNR conditions. Preliminary analyses on sublevels of SNR demonstrated no significant or marginal effects.

Accuracy was used as the dependent measure, as opposed to signal-detection statistics such as A' or D', because the probability of misses and false alarms was not stationary over multi-target trials. In order to ensure that participants were not simply responding haphazardly, the responses to the incongruent distractor targets were first analyzed. Participants incorrectly responded to 6.5% of the distractors, but the response rate to targets was considerably higher (87.3% and 86.2%).

To simplify the reporting of results, all significant main effects and interactions are listed in Table I, and are described below and displayed in Figs. 2 and 3.

SNR. As expected, participants were more accurate in identifying targets when they were presented at higher SNR (0 and +3 dB: 88.7% correct, s.e. 1.6%) relative to lower SNR [−3 and −6 dB: 78.8% correct, standard error (s.e.) 1.6%].

TABLE I. Significant effects from ANOVA analysis with accuracy.

Background type (single versus dual)	$F(1,47)=29.556$	$p=<0.001$	$\eta^2=0.386$
Contextual congruence	$F(1,47)=34.059$	$p=<0.001$	$\eta^2=0.42$
Target/background SNR	$F(1,47)=77.116$	$p=<0.001$	$\eta^2=0.621$
Background type × congruence	$F(1,47)=7.233$	$p=<0.01$	$\eta^2=0.133$
Background type × SNR	$F(1,47)=12.728$	$p=p<0.001$	$\eta^2=0.213$

Background type (dual versus single background). Targets heard in the single background condition were more accurately identified (86.8% correct, s.e. 1.3%) than in the dual background condition (80.8% correct, s.e. 1.8%).

Target/background contextual congruence. Participants were also more accurate in identifying targets that were contextually incongruent with the background scene (86.5%, s.e. 1.5%) than congruent targets (81%, s.e. 1.5%). There was no significant interaction between contextual congruence and SNR.

SNR × background type (Fig. 2). When targets were presented at an intensity equal to or greater than that of the ongoing background scene, participants' performance differed little over single and dual backgrounds. However, when target/background SNR was lower, participants were significantly more accurate at identifying targets in the single rather than the dual background condition.

Target/background contextual congruency × background type (Fig. 3). In a single (stereophonic) background, participants were significantly more accurate at identifying targets when they were contextually incongruent with the background. However, this benefit of contextual incongruency (the pop-out effect) became statistically insignificant when participants had to attend to two different background scenes.

A. Acoustic analyses of target and background sounds

Mean performance for each ipsilateral background sound segment and for each target sound (averaged across participants) was first compared with the acoustic measurements using univariate regression. This allowed an assessment of the general acoustic aspects of the masking background sound or the target sound that were implicated in identification accuracy. For the target sounds, there were significant positive relationships between identification and the following acoustic measures: number of peaks and the range of the autocorrelation peaks (a measure of the spread of periodic information across the frequency spectrum). There were also negative relationships with the skewness of the spectrum and the frequency of the maximum burst in the modulation spectrum.

For the background sounds, analyses revealed significant ($p<0.05$) positive relationships between (target) identification accuracy with spectral skew, spectral kurtosis and mean spectral flux, and the mean and SD of the peaks of the autocorrelation (measures of the amount of periodicity and its spread across frequency ranges). There were negative relationships with measures of spectral centroid (i.e., both windowed spectral mean centroid and the first moment of the spectrum).

Finally, to account for overlapping variance between variables, a mixed stepwise multiple regression (with a liberal $p<0.1$ threshold) was calculated starting with all ipsilateral background and target acoustic measures, as well as the binary contextual congruence and dual/single background conditions. The resulting model (all $p<0.05$) found positive relationships between target identification accuracy and the maximum spectral velocity of the background, the maximum burst of the modulation spectrum, the number of peaks in the target, and the range of peaks of the autocorrelation of the target sounds. Negative relationships were

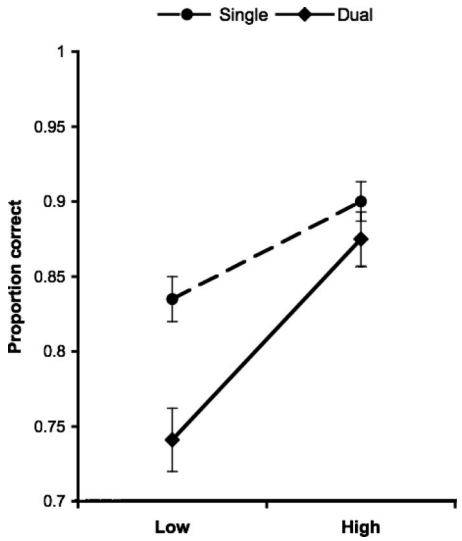


FIG. 2. SNR × background type.

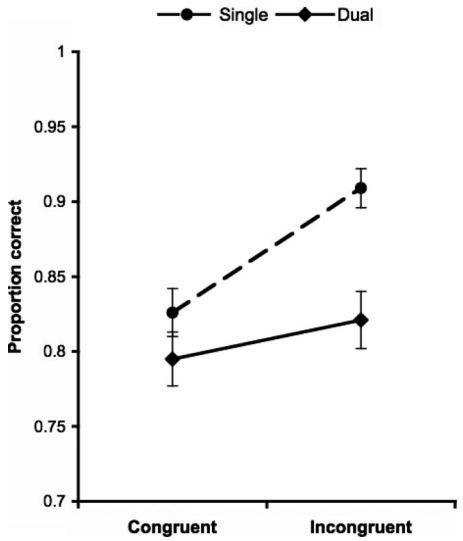


FIG. 3. Target/background contextual congruency × background type.

found between background centroid, SD of the spectral velocity of the background, and the frequency of the maximum burst in the modulation spectrum of the targets. In addition, experimenter-defined contextual congruence and dual/single background conditions were significant predictors of identification accuracy alongside the acoustic variables in the final model.

B. Acoustic similarity and contextual congruence

It is possible that the subjective assessment of contextual congruency may serve as a proxy measure for underlying acoustical similarity between target sound and background. Thus, to determine whether the experimenter-defined binary contextual congruency measures between target and background might be related to acoustic similarity or other factors, logistic and stepwise regression models were constructed using the battery of 38 acoustic similarity measures as predictor variables and contextual congruence as the binary dependent variable for each of the 92 different target/background pairings. The logistic regression model was nonsignificant [chi-square $(36)=25.00$, $p=0.916$]. In the stepwise regression model, no acoustic similarity variable was entered at a liberal $p<0.1$ level. It was also possible that latent acoustic variables underlying the 38 acoustic measures might relate to the contextual congruence measure. To test this, all 38 principal components of the acoustic variables were entered into a stepwise model with contextual congruence as the dependent variable. Again, the stepwise model entered no predictor variable into the model ($p>0.1$). These results suggested that subjective contextual congruence between target and background was not transparently related to acoustical factors.

Turning to the experimental results, a mixed stepwise regression assessed whether the accuracy difference between contextually congruent and incongruent targets could be accounted for by metrics of acoustic similarity alone, or whether they could be accounted for by a combination of acoustic similarity and the binary contextual variable. Here, the experimenter-defined congruency factor (congruent or incongruent), 38 acoustic similarity measures, and single/dual background condition were passed to the model as regressor terms ($p<0.1$). The stepwise procedure converged on a model that included the contextual congruence term ($p<0.05$) and the dual versus single background condition term ($p<0.05$). In addition, the final model included three acoustic measures significant at the $p<0.05$ level: the SD and maximum value of the pitch, and the range of peaks of the autocorrelation statistics. Two further acoustic measures were marginally significant predictors: the maximum value of the spectral velocity, and the mean spectral flux. These results suggest that the contextual congruence and acoustic similarity of target and background account for different variance in participants' accuracy scores.

V. SUMMARY AND DISCUSSION

In summary, the experimental results were as follows: (1) Overall, reducing the SNR of natural sound targets decreased participants' accuracy in detecting and identifying those targets; (2) participants were less accurate in detecting and identifying sounds when they had to attend to two different auditory scenes (presented to two ears) as opposed to a single stereophonic scene; (3) this greater accuracy in one versus two scenes emerged only at low SNRs; (4) participants were more accurate in detecting and identifying targets that were contextually incongruent with the background; (5) acoustic similarity and congruency with the real-world auditory scene accounted for different aspects of the variance in accuracy; and (6) accuracy was only affected by target/background congruence for single stereophonic scenes.

These findings support the general predictions outlined in the Introduction. As expected, based on previous findings, increased contralateral masking in the multiple background condition disrupted target detection and identification overall and reduced detection accuracy at low SNRs. In addition, a contextual pop-out effect was observed for incongruent targets in single backgrounds in the detection accuracy measure. This effect was eliminated when the target stimuli were presented in multiple backgrounds.

The present study explored a number of possible explanations for the contextual pop-out effect. One account is that participants are better able to detect and identify contextually incongruent targets simply because they are more acoustically dissimilar to the background scene, relative to contextually congruent targets. To evaluate this claim, detailed acoustic analyses of the target and background sounds were conducted, to determine which acoustic properties of the stimuli predicted performance in the task, and to establish whether the pattern of findings could be accounted for in terms of the acoustic overlap between backgrounds and targets. The results of the acoustic analyses indicate that accuracy depended on a mixture of spectral and temporal qualities of both the target and background sounds. For the background sounds, increased spectral (mean spectral flux and maximum spectral velocity) and temporal fluctuations (maximum modulation spectrum) led to improved performance. Similarly for target sounds, accuracy increased with envelope measures of fluctuation (the number of amplitude peaks) These findings suggest that temporal peaks in the target sound envelope, as well as spectral and temporal dips in the masking background sounds, served to aid target performance. This is consistent with findings from speech recognition (e.g., Moore *et al.*, 1999) in which listeners have lower speech reception thresholds for masking sounds with spectral and temporal dips. In the current study, detection and identification were also more accurate for periodic background sounds and for backgrounds and target sounds with periodic information spread over a broader range of frequencies, which is roughly consistent with the reduced masking for harmonic background sounds observed in previous speech recognition studies (Treurniet and Boucher, 2001), Gygi *et al.* (2004) also found that greater periodicity in environmental sounds predicted higher accuracy in identifying these sounds when they were vocoded. Greater periodicity was also associated with faster response times to identifying environmental sounds (Ballas, 1993). Finally, general spectral properties (the first and third moments of the spectrum) of the background and target sounds were predictive of success-

ful target detection and identification. Background sounds with higher spectral centroids were harder to detect and identify, while highly skewed background sounds were easier, perhaps because the spectral energy associated with these sounds is not evenly distributed across high and low frequencies, providing less consistent masking across the frequency range. Mirroring this finding, target sounds with less skewed spectra were easier to detect and identify. (Spectral skewness also modulated reaction times for environmental sound identification in Ballas, 1993.)

Measures of target/background similarity revealed that the contextual pop-out effect is only partially explained by acoustic factors. Acoustic variables that contributed significantly to the pop-out effect included correlogram-based pitch measures and peak autocorrelation statistics reflecting the degree of overlap in the spectro-temporal characteristics of the stimuli and the distribution of periodic elements across different frequencies. However, acoustic similarity was not the sole predictor of the congruence of a target stimulus with a background scene, or the emergence of the behavioral pop-out effect. This finding suggests that higher-level cognitive factors may play an important role in the detection and identification of meaningful sounds in complex auditory scenes.

One possibility explored in the present study is that the meaningful sound environments suggested by real-world auditory scenes activate associated memory representations, resulting in the generation of expectancies, such that unexpected sounds are more salient and therefore better detected. A subjective measure of the semantic compatibility of the target sounds with the meaningful context represented by the background proved a significant predictor of detection performance, providing support for this interpretation. The conflict between the expectancies generated by the context and the presentation of an incongruent target might have been expected to have a detrimental effect on identification performance, as has been observed in other domains. In spoken language processing, for example, a biasing sentence context facilitates recognition of words that are compatible with the contextual meaning, and inhibits recognition of incongruent words (Stanovich and West, 1983; Aydelott and Bates, 2004). However, in the present study, incongruent target sounds were *more* accurately identified than congruent targets, suggesting that it is ease of detection, rather than compatibility with the context, that facilitates target identification in this paradigm.

The specific demands of the detection task may account for this finding, as monitoring a continuous auditory scene for a single event requires the listener to suppress responses to potential false alarms. Thus, congruent sounds are more likely to be interpreted as elements of the background scene, and therefore to escape detection. A paradigm in which the background scene served to predict a specific event (e.g., the sound of tires screeching followed by the sound of a car crash) might better show the effects of contextual buildup on target identification (cf. Ballas and Mullins, 1991). In addition, the present study only relies on a limited set of target and background sounds, and thus the generalizability of the results would be enhanced by using a more comprehensive set of sounds that sample the full range of naturally occurring auditory events.

As noted above, the detection of tone, noise, and speech targets is vulnerable to informational masking—and this study extends this finding to the detection and identification of environmental sounds in more naturalistic settings. As in previous studies (e.g., Brungart and Simpson, 2002), this informational masking effect is only apparent when targets are presented at low SNRs within the ipsilateral background. Potentially more noteworthy, however, is the finding that the introduction of different contralateral acoustic information disproportionately reduces identification accuracy for contextually incongruent targets. Indeed, the main effect of background scene is primarily driven by the drop in accuracy for incongruent targets in the presence of a dual background. This supports the claim that the congruency effect is being driven by more attentional or cognitive factors, in that additional information from the second background scene is completely spatially segregated from the target and its incongruent background. Thus, it is informational rather than energetic masking that disrupts contextual pop-out.

As mentioned above, the contralateral masking result is very similar to that reported in several cocktail-party-inspired studies of speech processing, such as the coordinate response measure paradigm (Kidd *et al.*, 2007; Brungart and Simpson, 2002). Brungart and Simpson (2002) and Brungart *et al.* (2005) observed that contralateral informational masking of speech is greater the more speech-like the masking signal, indicating that the masking effect occurs when the listener must segregate simultaneously presented auditory stimuli with similar perceptual characteristics. These results suggest a set of follow-ups to the present study: If indeed there are similar informational masking effects occurring in these non-linguistic scenes, then one should see that reversed or combined environmental sound scenes presented contralaterally to the target should induce significantly more informational masking than artificial noise distractors, such as white noise convolved with the amplitude envelope of an environmental sound scene.

Another possibility is that the contralateral masking effect in the present study is at least partially driven by the increased attentional demand imposed by the presentation of an additional background scene. Having to monitor two distinct, complex, and changing auditory streams may interfere with listeners' ability to generate expectancies (cf. Aydelott and Bates, 2004), thereby contributing to the elimination of the pop-out effect. A similar result has been reported in the phoneme monitoring literature. The phoneme monitoring task resembles the paradigm used in the present study, in that it requires listeners to respond whenever they hear a particular phoneme occurring within a speech stream. Of particular interest is the finding that listeners' detection accuracy is poorer when they are required to perform a secondary task (Martin, 1977; Treisman and Squire, 1974), suggesting that the generation of expectancies based on the information provided by the auditory context is disrupted by increased attentional load, even in the absence of a competing acoustic signal. By manipulating the predictability of the target in the

present paradigm (e.g., by holding constant the temporal or spatial location of the target and the number of targets), it may be possible to disambiguate whether the contralateral masking is driven by the difficulty of segregating and monitoring distinct auditory streams in the face of conflicting information, or other processes of distraction such as stimulus uncertainty or overwhelming attentional load.

In conclusion, the present study demonstrates that real-world contextual information from the auditory scene influences the accurate detection and identification of natural environmental sounds. The acoustic properties of the target and background sounds offer only a partial account of the context effect; the extent to which the target sound is predictable within the meaningful auditory environment represented by the background also plays a significant role. The emergence of context-driven effects on target identification depends on the perception of a single, unified auditory environment, and is therefore highly vulnerable to contralateral masking by a competing background scene. The respective contributions of expectancy, auditory segregation, and attentional demand to the observed pattern of results remain topics for future research.

Aydelott, J., and Bates, E. (**2004**). "Effects of acoustic distortion and semantic context on lexical access," Lang. Cognit. Processes **19**, 29–56.

Ballas, J. A. (**1993**). "Common factors in the identification of an assortment of brief everyday sounds," J. Exp. Psychol. Hum. Percept. Perform. **19**, 250–267.

Ballas, J. A., and Mullins, T. (**1991**). "Effects of context on the identification of everyday sounds," Hum. Perform. **4**, 199–219.

Boersma, P., and Weenink, D. (**2009**). "Praat: Doing phonetics by computer," Version.5.1.18 (Computer program), http://www.praat.org (Last accessed Oct. 15, 2009).

Brainard, D. H. (**1997**). "The psychophysics toolbox," Spatial Vis. **10**, 433–436.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Broadbent, D. E. (**1958**). *Perception and Communication* (Pergamon, Oxford).

Brokx, J. P. L., and Nootebohm, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Brungart, D. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D., and Simpson, B. (**2002**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

Brungart, D., Simpson, B., Darwin, C., Arbogast, T. L., and Kidd, G. (**2005**). "Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**, 292–304.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Cummings, A., Ceponiene, R., Koyama, A., Saygin, A. P., Townsend, J., and Dick, F. (**2006**). "Auditory semantic networks for words and natural sounds," Brain Res. **1115**, 92–107.

Cusack, R., and Carlyon, R. P. (**2003**). "Perceptual asymmetries in audition," J. Exp. Psychol. Hum. Percept. Perform. **29**, 713–725.

Darwin, C., and Hukin, R. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**, 970–977.

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**). "Perceptual segregation of competing speech sounds: The role of spatial location," J. Acoust. Soc. Am. **114**, 2178–2189.

Durlach, N., Mason, C., Shinn-Cunningham, B., Arbogast, T., Colburn, H., and Kidd, G. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

EPAPS Document No. E-JASMAN-126-044911 for a list of the background and target sounds heard by participants and the different conditions these sounds were heard in (e.g., signal to noise ratios). For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.

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.

Gygi, B., Kidd, G. R., and Watson, C. S. (**2004**). "Spectral-temporal factors in the identification of environmental sounds," J. Acoust. Soc. Am. **115**, 1252–65.

Gygi, B., Kidd, G. R., and Watson, C. S. (**2007**). "Similarity and categorization of environmental sounds," Percept. Psychophys. **69**, 839–55.

Gygi, B., and Shafiro, V. (**2007**). "Effect of auditory context on the identification of environmental sounds," in Proceedings of the 19th International Congress of Acoustics, Madrid, Spain.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Kidd, G., Jr., Mason, C. R., and Richards, V. M. (**2003**). "Multiple bursts, multiple looks, and stream coherence in the release from informational masking," J. Acoust. Soc. Am. **114**, 2835–2845.

Kidd, G. R., Watson, C. S., and Gygi, B. (**2007**). "Individual differences in auditory abilities," J. Acoust. Soc. Am. **122**, 418–435.

Leech, R., Aydelott, J., Symons, G., Carnevale, J., and Dick, F. (**2007**). "The effect of semantic and attentional distractors on syntactic processing in typical development and adulthood," Dev. Sci. **10**, 794–813.

Martin, M. (**1977**). "Reading while listening: A linear model of selective attention," J. Verbal Learn. Verbal Behav. **16**, 453–463.

Moll, K., Cardillo, E., and Aydelott Utman, J. (**2001**). "Effects of competing speech on sentence-word priming: Semantic, perceptual, and attentional factors," in *Cognitive Science*, edited by J. D. Moore and K. Stenning (Lawrence Erlbaum Associates, Edinburgh), pp. 651–656.

Moore, B. C. J., Peters, R. W., and Stone, M. A. (**1999**). "Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips," J. Acoust. Soc. Am. **105**, 400–411.

Oh, E. L., and Lutfi, R. A. (**1999**). "Informational masking by everyday sounds," J. Acoust. Soc. Am. **106**, 3521–3528.

Saygin, A. P., Dick, F., and Bates, E. (**2005**). "An on-line task for contrasting auditory processing in the verbal and nonverbal domains and norms for younger and older adults," Behavior Research Methods **37**, 99–110.

Slaney, M. (**1994**). "Auditory toolbox: A Matlab toolbox for auditory modeling work," Apple Computer Technical Report No. 45, Apple Computer Inc., Cupertino, CA.

Stanovich, K. E., and West, R. F. (**1983**). "On priming by a sentence context," J. Exp. Psychol. Gen. **112**, 1–36.

Stemberger, J. P., Elman, J. L., and Haden, P. (**1985**). "Interference between phonemes during phoneme monitoring: Evidence for an interactive activation model of speech perception," J. Exp. Psychol. Hum. Percept. Perform. **11**, 475–489.

Treisman, A., and Squire, R. (**1974**) "Listening to speech at two levels at once," Q. J. Exp. Psychol. **26**, 82–97.

Treurniet, W. C., and Boucher, D. R. (**2001**). "A masking level difference due to harmonicity," J. Acoust. Soc. Am. **109**, 306–320

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.

Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.

Perceptual and emotional categorization of sound

Penny Bergman, Anders Sköld, and Daniel Västfjäll
Department of Civil and Environmental Engineering, Division of Applied Acoustics, Chalmers University of Technology, SE-41296 Gothenburg, Sweden

Niklas Fransson
Department of Energy and Environment, Division of Building Services Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden

(Received 28 October 2008; revised 27 July 2009; accepted 13 September 2009)

This paper investigates how different types of data from psychoacoustical experiments may be combined to render further knowledge about the mechanisms underlying sound perception. Two studies were conducted with auditory alerts of short duration. First, an experiment where participants rated the dissimilarity among the auditory alerts was performed. This resulted in a two-dimensional multi-dimensional scaling solution. Second, an experiment where participants evaluated the stimuli with semantic descriptors and rated their emotional reactions to the sounds was performed. The output of this experiment was a reduced set of underlying perceptual and emotional dimensions. The results of the two experiments were then integrated by the use of multi-dimensional perceptual unfolding and a set mediation analyses. The integrative analyses showed that part of the cognitive categorization of the semantic descriptors was mediated by the emotional reactions to the sounds. The results are discussed in relation to theories of auditory perception and emotional response categorization. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3243297]

PACS number(s): 43.66.Lj, 43.66.Ba, 43.50.Qp, 43.50.Rq [BSF] Pages: 3156–3167

I. INTRODUCTION

In traditional psychoacoustics the emphasis has for a long time been to establish the perceptual determinants for the physical parameters in sounds. For example, Grey (1977) studied the dimensions of how timbre in music instruments are perceived and noted that, e.g., the distribution of spectral energy, spectral fluctuation, and presence of low-amplitude high-frequency energy in the initial attack play a role in how we categorize different aspects of timbre. Other studies have examined how changes in fundamental frequency, pitch, harmonics, etc., impact identification (Pollack and Ficks, 1954; Graham, 1999), the perceived level of urgency (Edworthy *et al.*, 1991; Burt *et al.*, 1995; Haas and Casali, 1995), the learnability of its meaning (Ballas, 1993; Keller and Stevens, 2004), the perception (Repp and Knoblich, 2007; Warren and Verbrugge, 1984), and categorization (Gygi *et al.*, 2007; Dubois, 2000).

In perception and categorization of sound, an important role is played by the content features of the sound (i.e., psychological associations to the sound producing source). Gygi *et al.* (2007) examined categorization of environmental sounds and showed that there were physical aspects that played a part in categorization but also that sounds from similar sources were categorized together. This has also been shown by Dubois (2000) where the participants more often categorized the sounds based on the sound source rather than the physical attributes. Genell (2008) meaning-neutralized environmental sounds by means of frequency smearing using different window lengths and frequency bands. This procedure thus retained psychoacoustical features of the sound, but made the source unidentifiable. The result of the study showed that overall, content, rather than form, appeared to have the biggest impact on emotional reactions. This appears to be the same for visual stimuli where categorization is done rather by function than by surface features (Guastavino, 2007).

The basis for why people group objects together is based on either structural (perceptual) or functional similarity. Structural similarity points to the underlying theory of the object whereas the functional similarity refers to how the objects are used (Halberstadt and Niedenthal, 1997). Another proposed definition of two aspects divides categorization into similarity and theory. In this approach, the similarity is the apprehended correlation between the physical aspects of the objects, whereas the theory based approach includes or supersedes the underlying idea or deeper structure of the objects (Niedenthal *et al.*, 1999). In addition to the other grounds for categorization, a third aspect proposed is the affective or emotional response categorization. This was proposed by Bruner *et al.* (1956) who termed it "affective response categorization" (see Halberstadt and Niedenthal, 1997). They referred to the "evocation of a defining affective response" (p. 1017, Halberstadt and Niedenthal, 1997) as the basis for categorization. Objects that elicit the same emotion are grouped together.

We argue that emotional categorization is an important aspect in how individuals categorize sounds, both environmental sounds and auditory alerts that are more of caricatures of naturally occurring sounds. Niedenthal *et al.* (1999) pointed toward that any emotional response can form the ground for a category; however, to test this it is beneficial to limit the amount of dimensions. A common approach is to use the two dimensions of valence and activation (or arousal) (e.g., Västfjäll, *et al.*, 2002; Niedenthal *et al.*, 1999; Russell, 2003; Feldman Barrett, 1998). Valence is a basic dimension

of all emotional responses and may account for between 50% and 60% of the variance in emotional responses. It ranges from negative over neutral to positive (Niedenthal *et al.*, 1999; Västfjäll *et al.*, 2003). Activation is a second orthogonal dimension of experience that relate to how active versus passive the experience is. Activation often account for half as much variance as valence and ranges between activated to calm (Västfjäll *et al.*, 2002; Niedenthal *et al.*, 1999). A two-dimensional model gives a more complete and parsimonious description of the basic building blocks of our emotional experiences. The actual experience (e.g., sadness) is a distinct feeling that can be described as a combination of these two dimensions (negative and low activation). Valence and arousal also have the advantage of being pan-cultural (Feldman Barrett, 1998).

Emotional response categorization has primarily been studied outside the auditory domain. In research on visual face perception, Niedenthal *et al.* (1999) showed that affective categorization was even more pronounced when being in an emotional state. A similar result was shown by Halberstadt and Niedenthal (1997) where the participants rated different emotional faces. Later studies of Nygaard and Queen (2008) showed that the emotional tone of a voice acts as an additional information source when categorizing and occurs relatively early in the processing of the spoken language.

Much of the previous research on auditory categorization has relied on one type of data (e.g., dissimilarity ratings) derived using one psychophysical method (e.g., pairwise comparisons). In this article, we will combine different methods of evaluating sounds and different types of data (perceptual ratings, pair comparison, reactions, and physical descriptions). By doing this, we will hopefully shed further light on mechanisms underlying auditory perception and the basis for categorization of sounds.

The present paper will use auditory alerts as stimuli. Auditory alerts can be described as a "caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes" (Gaver, 1986) and have an intuitive link to the event they describe (Graham, 1999). This makes them highly efficient in learning and retention processes. The use of auditory alerts in products and environments are increasing. There are good reasons for this; in visually complex environments, auditory information may be very beneficial (Burt *et al.*, 1995). In general, auditory warnings also have a higher compliance level than the visual counterpart (Edworthy *et al.*, 1995). Hearing is also omnidirectional and therefore cannot easily be shut off or ignored (Haas and Edworthy, 2006). But the increase in auditory alerts also increases the demands of the alerts. A review of auditory alerts in aircraft in 1990 showed that there were as many as 15 different auditory warnings, with no internal structure to provide assistance in learning as well as retention of the warnings (Patterson, 1990).

II. OVERVIEW OF PRESENT STUDY

The present paper aims at investigating some of the processes underlying auditory perception. To achieve this, two integrated analyses were conducted: a multi-dimensional perceptual unfolding (MDU) and a set of mediation analyses. For this purpose, inputs from two different experiments were used: (1) dissimilarity ratings and (2) semantic rating of sound stimuli. We integrate the outcomes of these experiments in an overall analysis with the hope to show that this additional analysis step will provide further understanding of the underlying perceptual processes. A MDU combines semantic evaluations with a multi-dimensional scaling (MDS) of dissimilarity ratings. The semantic evaluation then serves as a tool to the dimensions of the MDS.

The stimuli used in the two experiments originate from a development process for new information and warning sounds intended for use in a vehicle. The auditory alerts had different levels of importance and urgency, from "information" to "severe warning." The sounds were approximately 3 s long (s.d. 1 s) and loudness equalized (Zwicker ISO, 532B).

The first experiment uses dissimilarity ratings in order to obtain a multi-dimensional scaling of the stimuli, free of the restrictions imposed by predetermined scales or response criteria. When rating dissimilarities between complex sounds, listeners are only able to focus on a limited number of parameters (Miller and Carterette, 1975). MDS is therefore an efficient tool to distinguish the prevalent or dominant perceptual features of stimuli. A preference mapping was also conducted. Preference mappings may sometimes facilitate the interpretation of the dissimilarity ratings (Clark *et al.*, 2001). Dissimilarity ratings were analyzed using the individual difference scaling (INDSCAL) model. The INDSCAL model assumes that all participants share the same psychological space but attends differently to the underlying psychological dimensions (Ashby *et al.*, 1994). An advantage with the INDSCAL model is that it provides a unique configuration solution that requires no further rotation of the model (Martens and Zacharov, 2000). The analysis resulted in a two-dimensional solution without clear interpretation of the dimensions solely based on knowledge of the stimuli.

In the second experiment, the same auditory alerts were mapped with perceptual ratings of various semantic attributes. The attributes included both physical properties and psychological properties of the stimulus. In addition to evaluating the sound attribute ratings, participants were asked to rate their affective reactions to the sounds (measurement procedure and rationale described later). The results were first analyzed separately with regard to if the participant could differentiate the sounds. 14 of the 15 items and the affective ratings met this request. A factor analysis of the remaining 14 items resulted in five different factors; these were interpreted as describing emotional reactions, complexity of the sound, size of the sound source, and the ecological validity.

The results from the two experiments were then combined in the MDU and mediation analyses where the dimensions from the MDS solution were related to the dimensions from the factor analysis and ratings of emotional reactions.

III. EXPERIMENT 1

The aim of the first experiment was to perform a multi-dimensional scaling of a set of 13 auditory alerts. Multi-

dimensional scaling is a commonly used tool to reduce the number of parameters and to distinguish the more salient ones. It also has an advantage in not requiring previous assumption regarding the possible underlying dimensions. Several techniques to do this have been developed (Caclin *et al.*, 2005). In this experiment, the participants performed pairwise ratings of dissimilarities between the different auditory alerts using a sliding scale. They also performed a preference mapping. In total, 78 pairs of sounds were presented in a half matrix design to 11 participants. This was then analyzed with the INDSCAL method. An advantage with the INDSCAL model is that it provides a unique configuration solution that requires no further rotation of the model (Martens and Zacharov, 2000).

The participants were not aware of the sounds being auditory alerts when rating them and were only asked to rate differences and preference. The preference of the sounds shall therefore not be interpreted as preferred sound for the warning level they represent but a general preference (or dis-preference). It can thus be expected that the lower warning levels should be more preferred than the higher levels of warnings.

A. Participants

11 students participated in the test, 7 male and 4 female. All tested for normal hearing. The average age was 25.9 years.

B. Design

1. Stimuli

The sounds were 13 auditory alerts, sounds 1–13 in Table I. The sounds were 3 s long (s.d. 1 s) and all adjusted to equal loudness (Zwicker ISO, 532B). The stimuli varied in five different levels of urgency and information. The first level is called "information;" there is new information somewhere, no immediate reaction needed. The second level is "advice;" the listener is recommended to do something, i.e., put on the seat belt. In the third through fifth level, something has happened but the required action differs. The third level is called "caution;" when possible, due to traffic, action is required. The fourth level "warning" requires some kind of action as soon as possible. The fifth level "severe warning" requires immediate reaction, i.e., brake the car. The design of the different auditory alerts was done by three different designers (each assigned a concept number, see Table I) in a development process of new icons to be used in vehicles. Within a concept, some of the sounds have similarities in regard to, e.g., timbre and pitch of tonal components. The sounds were created to be used as mono-sounds and in the experiment the participants listened to the same sound in both ears.

The auditory alerts were presented in a half matrix design, $n\text{pairs}=n\text{sounds}(n\text{sounds}-1)/2$, thus resulting in 78 pairs of sound. The limiting factor of the design was the test length.

2. Set up

The dissimilarity ratings and the preference mapping were performed by the use of a computer interface (Lise) in MATLAB [developed by Rioux (2001)]. The stimuli were presented through electrostatic headphones (STAX Lambda), fed through a digital soundcard (RME) and a professional D/A converter (Lucid) to a STAX amplifier.

3. Procedure

The participants were individually tested in a sound-attenuated room with no visual distractions. The participants were first welcomed and given a set of demographic questions and were tested for normal hearing. Before the listening test, they were instructed in how to perform the test. The participants were asked to listen to the sounds and rate the level of difference between them. The ratings were conducted on a slider ranging from different to similar. In addition to the rating, they had to decide which sound of the pair they preferred. The preference ratings were done on radio buttons. The participant had full control over the playback rate and, if needed, they could repeat the sounds within the present pair but not go back to the previous pair.

The results of the dissimilarity ratings were processed and analyzed according to the INDSCAL model for multi-dimensional scaling.

One of the participants did not comply with the instructions and was therefore removed from the analysis. Data were visually inspected before analyses and it was determined that all remaining participants discriminated between sounds or groups of sounds.

The analysis resulted in a multi-dimensional scaling with two dimensions. The result had a fair goodness of fit (normalized raw stress of 0.093) according to Kruskal (1964) (see Fig. 1).

In the first dimension of the MDS solution, tendencies to discriminate based on sharpness were evident. Variation in the second dimension suggests that discrimination is based on inter-onset intervals, repetition, as well as roughness. Taken together, however, the interpretation of the two dimensions from the available data appeared ambiguous. Additional types of data may, however, help to better understand why the sounds were rated as similar/dissimilar.

The result of the preference mapping was analyzed by the method of Bradley–Terry–Luce (BTL) (See Wickelmaier and Schmid, 2004). The BTL-method results in a ratio-scale assuming a linear preference. See Fig. 2 for a summary of the result.

C. Result

The overall results of the first study are that the different stimuli were perceptually discriminated. The preference ratings also indicate that sounds of lower "urgency," the information and advice alerts are more preferred than the warning alerts. This may be expected since these sounds lower in urgency were intended to be less annoying and more informative. Participants rated the sounds without knowledge of

TABLE I. Description of the stimuli used.

Stimuli number	Concept	Warning level	Sound description	Sound spectrum	Temporal description	Rate (IOI in s)	Repetitions	Sharpness (acum)	Roughness (asper)
1	I	Information	Dong-ding	2 harmonics at 293, 392, Hz	⋯	⋯	1	1.4	1.92
2	I	Caution	Beeps	>10 harmonics at 737, 1463, 2225, 2938, 4418 Hz	2 bursts (L=31 ms, T=94 ms) followed by 4 echoes	0.8	2	5.1	6.28
3	I	Warning	Low frequent buzz	>10 harmonics, 84.6, 127, 170, 211, 296 Hz	Burst, L=800 ms	1.4	3	2.7	5.14
4	I	Severe warning	Buzz	>10 harmonics, 94.6, 142.6, 237 Hz	Burst, L=270 ms	0.5	4	2.9	5.53
5	II	Information	Slow on and offset beeps	2 Harmonics, 312, 941 Hz	Burst, L=400 ms, ons.=40 ms, offs.=100 ms	0.4	3	1.2	0.85
6	II	Advice	Artificial bell	2 Harmonics, 277, 830 Hz	L=600 ms	1.7	2	1	1.125
7	II	Caution	Artificial bell	2 harmonics, 310, 910 Hz	L=750 ms, ons.=28 ms, offs.=500 ms	1.5	2	1.2	0.62
8	II	Warning	Double bell	2 harmonics, 278, 834 Hz	L=610 ms, delay of 2nd harmonic 200 ms	1.1	3	1.2	0.62
9	II	Severe warning	Beeps	>10 harmonics, 207, 620, 829, 1037, 1452 Hz	4 beeps in a group L=95 ms, T=100 ms	0.5	4	2.8	2.75
10	III	Information	Short knocks	Peak at 1760 Hz, −3 dB 100 Hz	2 knocks in a group, L=10 ms, T=320 ms	2.1	2	2.7	18.4
11	III	Caution	Beeps	2 harmonics, 985, 2955 Hz	2 beeps, L=110 ms, T=320 ms	2.2	2	3	12.5
12	III	Warning	(Dissonant) beeps	3 prominent harmonics, 99, 156, 215 Hz	Ons=20 ms, offs=20 ms, L=300 ms	0.9	4	2.6	6
13	III	Severe warning	Beeps	2 harmonics, 2497, 2649 Hz	4 grouped beeps, ons.=17 ms, offs.=17 ms, L=120 ms, T=130 ms	0.6	4	4	2.25
14	IV	Information	Very short beep, tick character	>10 harmonics, 9 prominent, 300, 900, 2097, 2995, 7201 Hz	L=30 ms	0.1	9	6.5	4.35
15	IV	Advice	Ding+belt icon	5 harmonics, 261, 328, 391, 523, 1047 Hz	2 dings, L=450 ms	0.5	2	4.2	12.5
16	IV	Serious advice	Ding+belt icon	>10 harmonics, 5 prominent, 261, 328, 391, 523, 1047 Hz	3 dings, L=350 ms, T=350 ms	1.4	2	3.3	5.1
17	IV	Caution	Bells+icon of malfunction engine	5 harmonics, 261, 328, 391, 523, 1047 Hz	Bell, L=450 ms, T=450 ms	0.5	2	2.9	2
18	IV	Warning	Bells+icon of malfunction engine	5 harmonics, 261, 328, 391, 523, 1047 Hz	3 bells, L=350 ms, T=350 ms	1.4	2	2.8	1.87
19	IV	Severe warning	Repeated horn sound +car screech	>10 harmonics, 335, 435, 474, 671, 869 Hz	Car horn, L=340 ms, T=350 ms	0.4	6	4	1.75

Note: IOI=inter-onset interval, ons=onset, and offs=offset

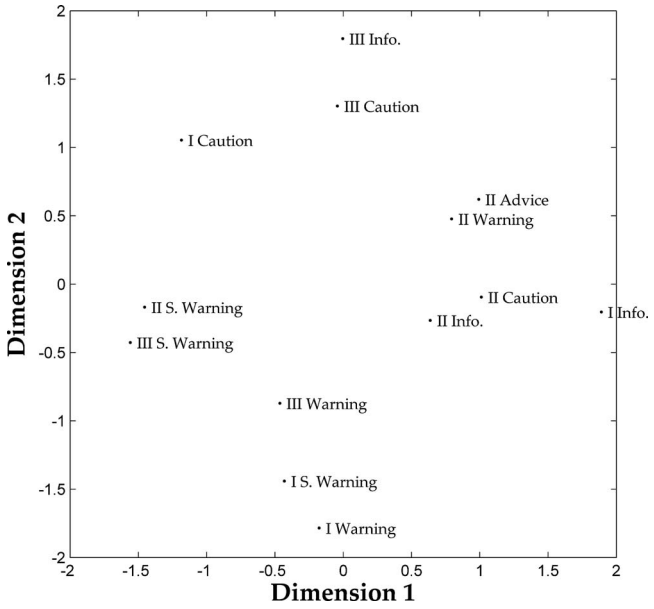


FIG. 1. The result of the multi-dimensional scaling, the different stimuli positioned along the two dimensions. The roman number denotes the concept of the stimuli.

the intended usage as in-vehicle alerts. It may therefore be concluded that the design scheme applied was relatively successful.

However, the dissimilarity ratings were only partly based on physical sound properties. Clusters can be seen for the different "concepts" where the sounds partly share same physical properties, e.g., concept II. It is, however, from the MDS solution alone, difficult to interpret the results and find the most salient parameters.

IV. EXPERIMENT 2

Experiment 2 aims to further investigate the perception of the auditory alerts. In this experiment, we used semantic characteristics that included both physical properties of the sounds and psychological properties that possibly could be linked to the sounds (see Table II).

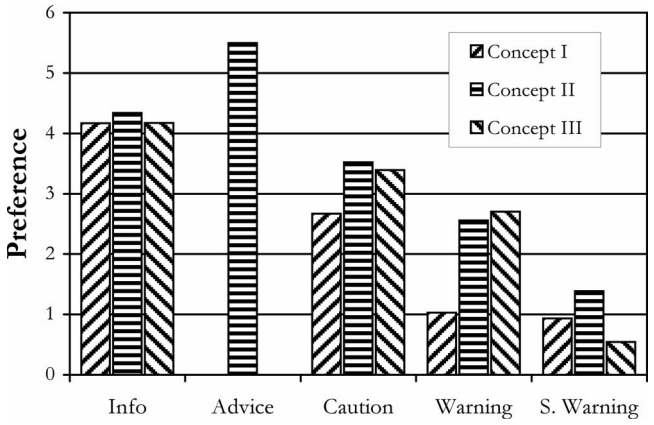


FIG. 2. The result of the preference mapping from the pair-wise comparison. Higher values equals more often preferred.

The items were formulated by a panel of five expert listeners (from the fields of psychoacoustics and psychology) that reviewed the sounds. Based on their experience and a previous pilot MDS, they formed 15 items.

Individuals may, however, use other aspects than auditory sensations to categorize, label, or interpret their experiences. As previously argued, we believe that the emotional responses may be an important part of auditory perception and a mechanism underlying categorization of sound (Niedenthal *et al.*, 1999). Auditory stimuli tend to provoke strong emotional reactions (Bradley and Lang, 2000) and it is therefore likely that participants may use their emotions as decision criterion when rating the similarity among auditory alerts. For this reason, we assessed the emotional reactions to the sounds. Participants thus also rated their level of experienced valence and activation when listening to the sounds. The stimuli were 19 auditory alerts, the 13 previously used in experiment 1 and another 6 that formed a fourth concept of auditory alerts, concept IV. Twenty-six participants participated. Each rating was first analyzed separately to see if the participants discriminated between the sounds and then a factor analysis was conducted.

A. Participants

26 participants, 12 men and 14 women, participated in the experiment. All participants were tested for normal hearing. The average age was 25.5 years old, ranging between 21 and 33 years.

B. Design

A panel of five expert listeners (including two of the present authors) in psychoacoustic and psychology listed the most prominent features in the auditory alerts described in Table I. They listened to the alerts, and to their assistance they used a pilot MDS obtained in a previous experiment with the same sounds. They listed both physical properties of the sounds and psychological aspects that could be linked to the sound. These topics include tempo, level of information, and experienced sound size; see Table II for a full description of the items. Affective ratings were also conducted in the test using the paper and pencil version of the nine-point self-assessment manikin scale (SAM-scale) (Fig. 3) (Bradley and Lang, 1994; see Västfjäll and Gärling, 2007 for a recent validation using a similar sample as the present one). The SAM-scale measures the dimensions of valence and activation.

1. Stimuli

The sounds used in the experiment were 19 auditory alerts (see Table I); apart from the 13 used in experiment, another concept (from a fourth sound designer) was added to the stimuli list, concept IV. This concept includes more ecological alerts than the first three concepts and was therefore deemed not suitable for the dissimilarity ratings of experiment 1. The sounds were 3 s long (s.d. 1 s) and all adjusted to equal loudness (Zwicker ISO, 532B). The stimuli originated from a development process of new icons to be used in vehicles. The stimuli varied in six different levels of urgency and information. The first level is called information; there is

TABLE II. The items used for the semantic evaluation (experiment 2).

Number	Item	Range 0–8
1	How slow do you think the sound decay?	Very quick–very slow
2	Do you think the sound is dissonant (false, impure) or harmonic?	Harmonic–dissonant
3	Do you think the sound is shrill or soft?	Soft–shrill
4	What tempo do you think the sound has?	Slow–quick
5	How do you experience the sound level?	Faint–loud
6	Do you think the sound is dynamic/varied or monotonic/stable?	Monotonic/stable–dynamic/varied
7	To what extent do you consider the sound to be attention demanding?	Not at all–a lot
8	Do you think it is a sound with a musical structure?	No, not at all–yes, a lot
9	Do you consider it to be an informative sound? (i.e., does it give you any information)	No, not at all–yes, a lot
10	Does the sound remind you of something you've heard in real life?	Never heard it–seldom–yes, I hear it every day
11	Do you perceive the sound source to be close to you or far away?	Close–far away
12	Do you consider the sound to be bright or dull?	Bright–dull
13	Do you consider the sound to be complex or simple?	Simple–complex
14	Does the sound begin abruptly?	No, not at all–yes, really
15	How big does the sound feel?	Small – big

new information somewhere, and no immediate reaction needed. The second level is advice; the listener is recommended to do something, i.e., put on the seat belt. The third level "serious advice" has a similar meaning as advice but with a higher level of urgency. In the fourth through sixth level, something has happened but the required action differs. The fourth level is called caution; when possible, due to traffic, action is required. The fifth level warning requires some kind of action as soon as possible. The sixth level severe warning requires immediate reaction, i.e., brake. Within a concept, some of the sounds have similarities with regard to, e.g., timbre and pitch of tonal components. The alerts were created to be used as mono-sounds and in the experiment the participants listened to the same sound in both ears.

2. Apparatus

The listening test took place in a room with low background noise (<20 dBA) without any identifiable sound sources. Sound-absorbing black screens were placed around the listening position, forming a booth. All apparatus were placed outside the room. The stimuli were presented through electrostatic headphones (STAX Ear speakers, SR lambda Signature), with a Lucid DA9624 D/A converter. The software for the presentation of the sounds was Microsoft PowerPoint 2003, controlled by the participant.

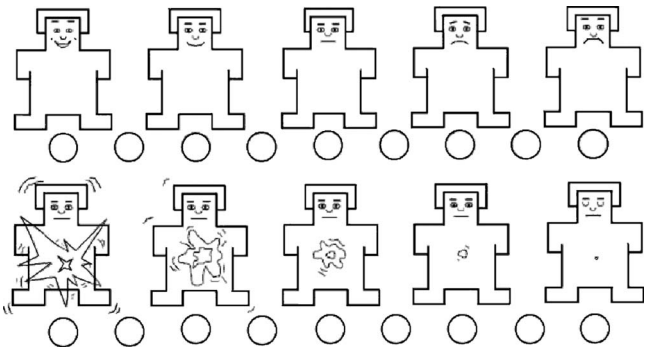


FIG. 3. The SAM-scale for valence (upper) and activation (lower) (Bradley and Lang, 1994). Both are coded 0–8 where 0 stand for negative and passive and 8 for positive and activated.

3. Procedure

The participants were individually tested in a sound-attenuated room. The participants were first welcomed and given a set of demographic questions and tested for normal hearing. After that, they were instructed in how to perform the test. The two affective and descriptive ratings were done in separate blocks where half of the participants begun with the affective ratings and the other half with the descriptive ratings. The participants were allowed to listen to the sounds several times and in their own pace. They were asked to go with the first feeling of the sound and not peruse it.

C. Result

First, separate analyses of variance (ANOVAs) were conducted on each semantic attribute and on the affective ratings to test if the participants discriminated between the sounds. Greenhouse–Geissers' corrected F-value was used for all analyses to correct for possible unequal variance (violation of sphericity). For item 11 (Do you perceive the sound source to be close or far away from you?), the participants did not separate the sounds from each other and the item was therefore removed from further analysis (see Table III).

The remaining 14 items were analyzed in a factor analysis resulting in four different factors, (see Table IV, including the previously omitted item that did not substantially change

TABLE III. The results from the ANOVA in experiment 2.

No.	Item	df	F	p-value
Q1	How slow do you think the sound decay?	7.917	19.138	<0.01
Q2	Do you think the sound is dissonant (false, impure) or harmonic?	8.832	29.156	<0.01
Q3	Do you think the sound is shrill or soft?	7.907	29.995	<0.01
Q4	What tempo do you think the sound has?	9.154	31.818	<0.01
Q5	How do you experience the sound level?	8.473	10.438	<0.01
Q6	Do you think the sound is dynamic/varied or monotonic/stable?	8.171	17.244	<0.01
Q7	To what extent do you consider the sound to be attention demanding?	8.641	11.833	<0.01
Q8	Do you think it is a sound with a musical structure?	7.28	27.784	<0.01
Q9	Do you consider it to be an informative sound? (i.e., does it give you any information)	8.444	8.298	<0.01
Q10	Does the sound remind you of something you've heard in real life?	8.643	5.915	<0.01
Q11	Do you perceive the sound source to be close or far away from you?	7.919	1.597	0.128
Q12	Do you consider the sound to be bright or dull?	8.331	16.814	<0.01
Q13	Do you consider the sound to be complex or simple?	8.009	12.533	<0.01
Q14	Does the sound begin abruptly?	7.332	14.116	<0.01
Q15	How big does the sound feel?	9.153	21.473	<0.01
	Valence	4.615	12.479	<0.01
	Arousal	7.935	23.665	<0.01

the factor solution). The number of components was determined by the Kaiser criterion: the individual factors should account for at least the variance of a single variable, i.e., a restriction of eigenvalue greater than 1 (Kline, 1994). The components explained 66% of the variance present.

Interpretations of the principal component attached to each of the four factors are perceived pleasantness, complexity, size of the sound source, and the ecological validity of the sound.

The *first component* includes shrillness, dissonance, sound level, musicality, sound attack and level of attention demand and could be summarized into the feeling of pleasantness when hearing the sound. Since it concerns auditory alerts, one may also interpret it to be the urgency or potency of the sounds. Item 3, the shrillness of the sound, has a natural link to pleasantness following the psychoacoustic definition of sensory pleasantness of Zwicker and Fastl (1990). The softer the sound is perceived to be, the more pleasant it is rated. The inclusion of perceived sound level (item 5) follows the same reasoning, as sound level also is a

TABLE IV. The result of the principal component analyses, experiment 2.

		Components			
No.	Items	1	2	3	4
Q3	Do you think the sound is shrill or soft?	0.807		−0.305	
Q2	Do you think the sound is dissonant (false, impure) or harmonic?	0.798			
Q5	How do you experience the sound level?	0.697			
Q8	Do you think it's a sound with a musical structure?	−0.697	0.438		
Q14	Does the sound begin abruptly?	0.652			
Q7	To what extent do you consider the sound to be attention demanding?	0.571	0.443		
Q6	Do you think the sound is dynamic/varied or monotonic/stable?		0.820		
Q13	Do you consider the sound to be complex or simple?		0.817		
Q12	Do you consider the sound to be bright or dull?			0.728	
Q15	How big does the sound feel?	0.305		0.679	
Q4	What tempo do you think the sound has?	0.559		−0.569	
Q1	How slow do you think the sound decay?	−0.441	0.404	0.452	
Q10	Does the sound remind you of something you've heard in real life?				0.872
Q9	Do you consider it to be an informative sound? (i.e., does it give you any information)				0.850

part of the psychoacoustic definition of sensory pleasantness. In this context, it is, however, more likely that sounds perceived as pleasant are also perceived as less loud as the sounds are loudness equalized. The first component also includes item of dissonance and musicality (items 2 and 8), the more dissonant and less musical a sound is, the less pleasant. It should be noted that this is the participants' perception of dissonance, not a musical analysis or whether partials are present within the same critical band (Plomp, 1967). The sound attack (item 14) could be included in the perceived pleasantness of the sound as an abrupt beginning of the stimuli may be considered as a threat or cause an unpleasant startle response. As the factor analysis showed that also the sound decay has a link to the pleasant component, we argue that it is better linked to the perceived size of the sound (component 3). The final item included in component 1 concerns the perceived attentional capture of a sound. Another benefit of using emotional responses as a guideline for sound design is the cognitive association (imagery or episodic memory) to the sounds may show substantial inter-individual variation (Guillaume and Roberts, 2007). It is possible that by creating the emotional response the different possibilities in what sound could mean to the listener could be narrowed down. An attention demanding sound is more intrusive and therefore considered as less pleasant than a non-attention demanding sound.

The *second component* deals with variations within the sound or the presence of different parts in a sound. Item 6 regarding the dynamics of the sound represents the variation, whereas item 13 concerns the complexity that represents how many different aspects are present within the sound.

The *third component* relates to the perceived size of the sound source and includes the brightness, the size, the tempo, and the sound decay. A presence of higher frequencies (a bright sound) is more often associated with smaller objects than larger objects, e.g., a squeak of a mouse vis-à-vis the sound of an elephant. Item 15 was a direct question about the perceived size of the sound. The tempo of the sound (item 4) is also often linked to object size, e.g., a pendulum. Finally, the decay of a sound is prolonged in bigger objects (item 1).

The *fourth component* deals with the ecological validity of the sound. Ecological sounds contain information of the sound source in itself or information given to a certain sound, like speech. Ecological sounds are well known to have a strong communicative power (Gaver, 1993), and it is therefore a logical link between the reminiscence of the sound (item 10) and the level of information in the sound (item 9).

The affective ratings showed distinct differences between the different levels of urgency (see Fig. 4): valence—$F=57.999$, $p<0.001$; activation—$F=71.696$, $p<0.001$. Pair-wise comparison showed that the higher urgency levels differed from the lower levels for both valence and activation. No significant differences between information and advice levels were, however, found for neither valence nor activation.

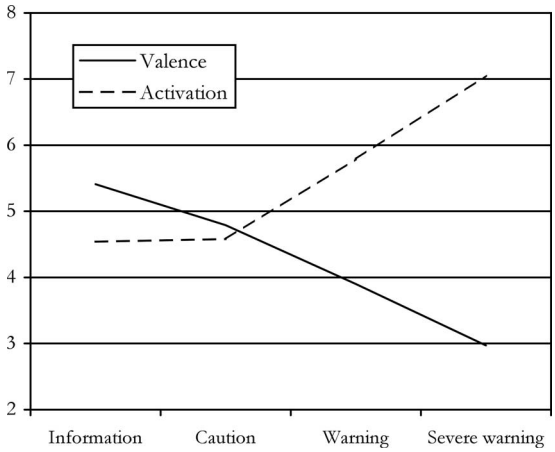


FIG. 4. The affective ratings over level of auditory alerts. The levels of advice and serious advice have been removed for sake of comparison.

V. INTEGRATED ANALYSES

The results from experiments 1 and 2 both rendered results of interest for design of auditory alerts. By an additional analysis combining the results from the experiments, we hope to further understand the underlying perceptual processes behind the responses in experiments 1 and 2. This will be done by MDU and a set mediation analyses.

In a MDU, the semantic evaluation from experiment 2 is combined with the MDS and preference mapping from experiment 1. The underlying semantic dimensions serves as a basis for interpreting the MDS dimensions. The semantic scale evaluation and the preference mapping are integrated with the MDS solution by means of regression analysis. In this way, the MDU will help highlight salient parameters used for the dissimilarity rating. The MDU helps in interpreting the dimensions of MDS and to understand what in the stimuli is relevant for the perception. However, to investigate underlying mechanisms, mediation analyses are beneficial.

The advantage of mediation is that it provides a more detailed understanding of the relations among different variables and explains how and why two variables are related. MacKinnon and Fairchild (2009) formulated it: "The promise of mediation analysis is that it can identify fundamental processes underlying human behavior that are relevant across behaviors and contexts" (p. 16). Mediation occurs when there is a causal relationship between an input variable and the outcome, and this relationship could be explained by an intervening variable. The requirements for a variable to function as a mediator is thus that the variations in the input variable significantly account for the variations in the mediator and that the variations in the mediator significantly account for the variations in the output (Baron and Kenny, 1986) The semantic descriptors used in experiment 2 may account for the variations in the multi-dimensional scaling provided in experiment 1 through a mediator. We argue that an often neglected mechanism in perception and categorization of sound is the emotional reaction to the sound. In the mediation analysis, we will therefore analyze whether part of the cognitive responses to the semantic descriptors are me-

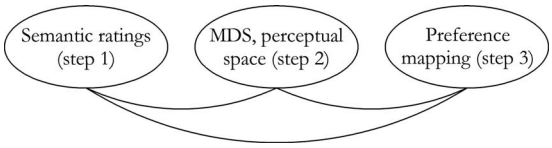


FIG. 5. The connection of the three parts in multi-dimensional perceptual unfolding.

diated by emotion. The emotional reaction to the sound was measured by account of valence and activation in experiment 2.

A. Multi-dimensional perceptual unfolding

Multi-dimensional scaling is a beneficial method to evaluate and to locate the salient parameters of sounds. It reduces the number of dimensions to otherwise complex data. It also has an advantage in not requiring previous assumption regarding the possible dimensions (Caclin *et al.*, 2005). However, it has a clear limitation in the interpretability of the dimensions created. The multi-dimensional perceptual unfolding solves this by the use of attribute ratings for each stimulus, i.e., humans are used as an analysis system by their responses (Sköld, 2005). Attribute ratings are integrated with the multi-dimensional scaling and serves as an interpreter of the dimensions. Within the field of psychoacoustics, this method has primarily been used with attribute ratings such as preference of sounds. The preference of sounds is of assistance when evaluating the perceived sound quality (e.g., Martens and Zacharov, 2000; Västfjäll *et al.*, 2004; Sköld *et al.*, 2004). In present case, the stimuli of this study are auditory alerts designed with different levels of urgency, which makes the level of preference of less interest. The auditory alerts of higher urgency levels will have a lower level of preference than the auditory alerts of lower urgency. This MDU will therefore focus on the semantic scale evaluation of the sound, rated by naive listeners in the second experiment.

The process of a multi-dimensional unfolding follows four steps. The first three steps contain the semantic scale evaluation done in experiment 2, the mapped perceptual space done by the multi-dimensional scaling in experiment 1, and the mapped preference done in experiment 1 (See Fig. 5 for a graphic description). The fourth step synthesizes the results of the semantic scale evaluation and the preference mapping [together with traditional instrumental psychoacoustic parameters if they provide relevant data (Sköld, 2005) to the results of the multi-dimensional scaling by the use of regression analysis].

Each of the four factors from experiment 2 is thus used as regressands in a multiple regression and the two dimensions from the MDS-analysis are the regressors. The regression coefficients for each dimension serve as coordinates in the multi-dimensional space. The p-values for the different factors are as follows: F1: <0.05, F2: 0.950, F3: <0.01, and F4: 0.175. The result of the synthesis of the previous described data resulted in the following MDU (Fig. 6).

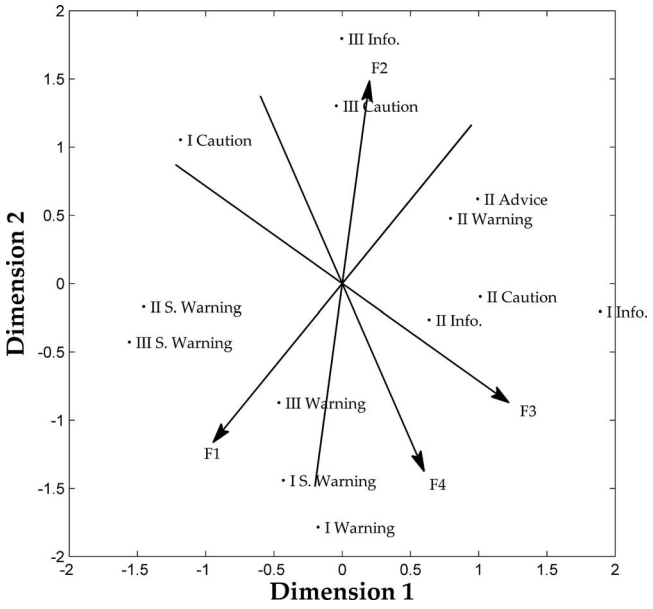


FIG. 6. The result of the MDU. The arrows depict the directions of each factor in the two dimensions from the MDS. The directions are given by the regression coefficients for each dimension. The lengths are here made equal for simplicity of viewing.

B. Mediation

Mediation occurs when there is a causal relationship between an input variable and an outcome, and this may be explained by an intervening variable. That is, the requirements for a variable to function as a mediator are that the variations in the input variable significantly account for the variations in the mediator and that the variations in the mediator significantly account for the variations in the output (Baron and Kenny, 1986). In line with our hypothesis, this would imply that there is a causal relationship between a physical descriptor and the perceptual space of the sound (the MDS) and this may be explained by the emotional reaction to the same sound.

A mediation model may be depicted as three causal chains: (1) The direct impact of the input variable on the output, (2) the impact of the input variable on the mediator, and 3) the impact of the mediator on the output (see Fig. 7). A direct measured mediation occurs when there is a significant connection between the input variable on the output and, when the mediator is added to the regression, the significance of the input variable on the output variable disappears. If there is an *a priori* belief of a small or suppressed effect size of the input variable, the variations in the input variable do not have to significantly account for the variations in the output (Shrout and Bolger, 2002). In those cases, one does an indirect analyze for mediation. The basic re-

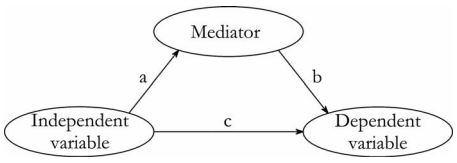


FIG. 7. Mediation model. (a)–(c) denote the connection between the independent variable, the dependent variable, and the mediator.

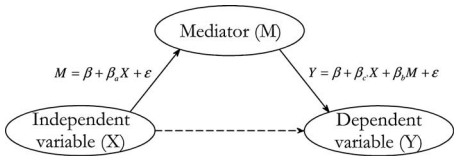


FIG. 8. Indirect mediation model. The regression show the dependence of X on M, and the dependence of X and M on Y.

quirements for an indirect mediation to occur are as follows. (1) The variations in the input variable significantly account for the variations in the mediator. (2) The variations in the mediator significantly account for the variations in the output. A figure of an indirect mediation is shown in Fig. 8.

With the two above regressions, we may assess the reduced association between X and Y when adding M. If the reduction is significant, there is a significant indirect mediation. The significance of the reduction is measured by a z-test where the reduction is divided by its standard error. The reduction is the product of the regression coefficient $\beta_a * \beta_b$ (Muller *et al.*, 2005). The standard error can be calculated in numerous ways. In this article, we will use one of the more strict methods developed by Aroian (1947), also known as the Sobel method (popularized in Baron and Kenny, 1986):

$$Z_{\text{Sobel}} = \frac{\beta_a \beta_b}{\sqrt{\beta_b^2 \beta_a^2 + \beta_a^2 s_b^2 + s_a^2 s_b^2}}. \quad (1)$$

The multi-dimensional scaling showed that there exist several parameters that affect the appearance of the perceptual space. Auditory alerts are complex sounds including several physical descriptors. It is therefore reason to believe that the effect size is small or suppressed for each descriptor. Due to this, we investigate whether there exists an indirect mediation of emotion from the perceptual ratings done in experiment 2 to the multi-dimensional scaling conducted in experiment 1. The 14 items and the two different mediators each form a mediation model for the two dimensions. In each dimension, 28 different mediation models will be thus tested for mediation. After steps 1 and 2 in the requirements for indirect mediation was conducted, 14 potential mediation models remained (see Table V). Of the 14 potential indirect mediations, 9 were found to be significant.

In dimension 1, one of the perceptual ratings was indirectly mediated by activation, the perceived sound level. In dimension 2, a total of eight mediation models were significant. The items indirectly mediated by valence were sound decay, dissonance/harmonic, shrill/soft, musical structure, and sound attack. The items indirectly mediated by activation in the second dimension were dissonance/harmonic, shrill/soft, musical structure, and sound attack. For Z-values see Table V.

TABLE V. The result of the mediation analyses

Item	Mediator	Z_{Sobel}	p-value
	Dimension 1		
Sound decay	Activation	1.1323	Non-sig
Sound level	Activation	−2.3286	<0.05
Dynamic/monotonic	Valence	1.915	Non-sig
Reminding of something	Valence	0.4462	Non-sig
	Dimension 2		
Sound decay	Valence	2.175	<0.05
	Activation	1.216	Non-sig
Dissonance/harmonic	Activation	−2.1876	<0.05
Shrill/soft	Valence	−2.4675	<0.05
	Activation	−2.9075	<0.01
Musical structure	Valence	2.7833	<0.01
	Activation	2.318	<0.05
Reminding of something	Valence	0.4478	Non-sig
Sound attack	Valence	−2.184	<0.05
	Activation	−3.1594	<0.01

VI. CONCLUSIONS

This research set out to shed further light on the underlying mechanisms in the perception of sound. Previous research has mainly focused on one type of data (e.g., dissimilarity ratings) derived using one psychophysical method (e.g., pair-wise comparisons). We argue that it is necessary to integrate different data to reach a further knowledge in perception of sounds. In the present study, we have conducted two kinds of experiments: (1)dissimilarity ratings in combination with preference mapping and (2) semantic ratings and emotional reactions.

The two experiments were first analyzed separately and then combined with each other. The combination included a multi-dimensional perceptual unfolding and a mediation analysis.

The results of the dissimilarity ratings show that there are some clusters based on the temporal and semantic ratings (of the psychoacoustic parameters listed in Table I). The participants also seemed to discriminate sounds based on sharpness (dimension 1) and on the inter-onset intervals, repetition, and roughness (dimension 2)

In the second experiment, the participant rated the semantic characteristics and their emotional reactions to the sounds. A factor analysis was conducted on the semantic ratings and the parameters were reduced to four main components. These were interpreted to be (1) perceived pleasantness or, urgency/potency of the sound; (2) the complexity of the sounds; (3) the perceived size of the sounds; and (4) the ecological validity of the sound.

The second part of this paper integrated the results of these experiments and made a combined analysis. The first analysis was a MDU. The result of the MDU (see Fig. 6) both provides assistance in what the parameters for the perceptual categorization behind the dissimilarity ratings were and also show which parameters that do *not* interfere with each other. The latter is of great importance in the design of sounds.

Two main concerns when designing auditory alerts are to make the listener understand *what* to do and *when* to do it. To design sounds with unambiguous meaning is perhaps the most important task in auditory warning design but still the cognitive response linked to sounds is not clarified. Edworthy and Hellier (2006) suggested that abstract sounds can be interpreted very differently depending on the many possible meanings that can be linked to them, and in large depending

on the surrounding environment and the listener. Several studies, however, embrace an ecological approach to the design of auditory icons and by that increasing the informational content of the sound (Gaver, 1993; Keller and Stevens, 2004; Ballas, 1993). The second concern in design of auditory icons is when to react, the urgency of the sound. Urgency is both a cognitive and emotional sensation with the function of motivating behavior (Scherer, 2001).

Much of the previous research has focused on this as a measure of the effectiveness of the design (Haas and Edworthy, 2006). Haas and Casali (1995) showed that warning sounds with higher perceived urgency produce faster response times, a measure relevant for many real-life actions in working environments (i.e., braking or pushing the correct button). In the multi-dimensional perceptual unfolding figure (Fig. 6), we can see that the urgency or potency (F1) and the ecological validity (F4) are almost perpendicular which implies that they are not perceptually categorized together. This implies that when designing auditory alerts, it is possible to increase the urgency level as well as the ecological level independent of each other.

It should be noted that two of the principal components of the semantic evaluation (F2, complexity, and F4 ecological validity) had not a significant regression in the MDU. This may be due to the small variation in the stimuli used in the study, but it may also be that they are of important for categorization of the sounds used in the present study. Future research should further address this issue.

The second integrated analysis of the results of the two experiments was a mediation analysis. Where the multidimensional perceptual unfolding provided assistance in interpretation of the salient parameters, the mediation goes deeper and finds the underlying mechanisms. Mediation has been widely used as a way to find explanations of how and why two variables are related in fields such as psychology (MacKinnon and Fairchild, 2009). We argue that one of the underlying mechanisms in perception of sound is the emotional response (see also Niedenthal *et al.*, 1999). The mediation model was thus designed to test if the semantic descriptors of experiment 2 were mediated by the emotional responses to the same sounds on the perceptual dimensions provided from experiment 1. The results of the analysis provided partial support for this hypothesis.

For the first dimension of the multi-dimensional scaling, the experienced sound levels were mediated by activation. This is in line with the primary interpretation of experiment 2; as the sounds were loudness equalized, participants recall their experienced feelings when judging the physical property (in this case loudness; for conceptually similar results see Neuhoff, 1998; Tajadura *et al.*, in press).

For the second dimension, five different semantic items were mediated by valence and/or activation: the sound decay, the level of dissonance, the shrillness, the musicality, and the attack of the sound. In the factor analysis, these were clustered in the same component (1) interpreted as perceived pleasantness or urgency or potency (Item 1 was sorted in component 3 but had a strong connection with component 1). In line with these results, a previous study on auditory alerts and auditory icons has shown a high correlation between urgency and activation as well as valence, suggesting that urgency is a combination of high activation and unpleasantness (Sköld, 2008). Also Cabanac (1992) suggested that the emotional reaction is a common currency with which the urgency, behavioral significance, and needed action will be evaluated against.

The aim of the present study was to show how integrated analyses could provide further knowledge in perceptual categorization and emotional responses but the results also provide direct implication in theories regarding design of warning sounds. The MDU showed that the ecological validity (what the sound means) and the urgency level (when to react) are unrelated in terms of perceptual categorization. Further, the mediation analyses showed that urgency is mediated by the emotional responses to the sounds, measured with valence and activation. This suggests that it is possible that some of the perceptual or cognitive responses typically obtained in psychoacoustic experiments are, in fact, based on emotional responses. Future research should further address this possibility for other contexts.

In the present context, these results suggest that design improvement of warning sounds should include considerations of the induced emotional response as well as the perceived ecological validity of the sound. A possible benefit of using emotional responses as a guideline for sound design is the cognitive association (imagery or episodic memory) with the sounds that may show substantial inter-individual variation (Guillaume and Roberts, 2007). It is possible that by creating the auditory-induced emotions, the different possibilities in what a sound could mean to a listener could be limited.

Aroian, L. A. (**1947**). "The probability function of the product of two normally distributed variables," Ann. Math. Stat. **18**, 265–271.

Ashby, G. F., Maddox, W. T., and Lee, W. W. (**1994**). "On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model," Psychol. Sci. **5**, 144–151.

Ballas, J. A. (**1993**). "Common factors in the identification of an assortment of brief everyday sounds," J. Exp. Psychol. [Hum Learn] **19**, 250–267.

Baron, R. M., and Kenny, D. A. (**1986**). "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations," J. Pers. Soc. Psychol. **51**, 1173–1182.

Bradley, M. M., and Lang, P. J. (**1994**). "Measuring emotion: The self-assessment manikin and the semantic differential," J. Behav. Ther. Exp. Psychiatry **25**, 49–59.

Bradley, M. M., and Lang, P. J. (**2000**). "Affective reactions to acoustic stimuli," Psychophysiology **37**, 204–215.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. (**1956**). *A Study of Thinking* (Wiley, New York).

Burt, J. L., Bartolome, D. S., Burdette, D. W., and Comstock, J. R., Jr. (**1995**). "A psychophysiological evaluation of the perceived urgency of auditory warning levels," Ergonomics **38**, 2327–2340.

Cabanac, M. (**1992**). "Pleasure: The common currency," J. Theor. Biol. **155**, 173–200.

Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (**2005**). "Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones," J. Acoust. Soc. Am. **118**, 471–482.

Clark, W. C., Janal, M. N., Hoben, E. K., and Carroll, J. D. (**2001**). "How separate are the sensory, emotional, and motivational dimensions of pain? A multidimensional scaling analysis," Somatosens Mot Res. **18**, 31–39.

Dubois, D. (**2000**). "Categories as acts of meaning: The case of categories in olfaction and audition," Cognit. Sci. Q. **1**, 35–68.

Edworthy, J., and Hellier, E. (**2006**). "Complex nonverbal auditory signals and speech warnings," in *Handbook of Warnings*, edited by M. S. Wogalter (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 199–220.

Edworthy, J., Loxley, S., and Dennis, I. (**1991**). "Improving auditory warn-

ing design: Relationship between warning sound parameters and perceived urgency," Hum. Factors **33**, 205–231.
Edworthy, J., Stanton, N., and Hellier, E. (**1995**). "Warnings in research and practice: Editorial," Ergonomics **38**, 2145–2154.
Feldman Barrett, L. (**1998**). "Discrete emotions or dimensions? The role of valence focus and arousal focus," Cognit. Emotion **12**, 579–599.
Gaver, W. W. (**1986**). "Auditory icons: Using sound in computer interfaces," Hum. Comput. Interact. **2**, 167–177.
Gaver, W. W. (**1993**). "What in the world do we hear?: An ecological approach to auditory event perception," Ecological Psychol. **5**, 1–29.
Genell, A. (**2008**). "Perception of sound and vibration in heavy trucks," Ph.D. thesis, Chalmers University of Technology, Gothenburg.
Graham, R. (**1999**). "Use of auditory icons as emergency warnings: Evaluation within a vehicle collision avoidance application," Ergonomics **42**, 1233–1248.
Grey, J. M. (**1977**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270–1277.
Guastavino, C. (**2007**). "Categorization of environmental sounds," Can. J. Exp. Psychol. **61**, 54–63.
Guillaume, T., and Roberts, M. (**2007**). "Event-related potential study of attention capture by affective sounds," NeuroReport **18**, 245–248.
Gygi, B., Kidd, G. R., and Watson, C. S. (**2007**). "Similarity and categorization of environmental sounds," Percept. Psychophys. **69**, 839–855.
Haas, E., and Casali, J. (**1995**). "Perceived urgency and response time to multi-tone and frequency-modulated warning signals in broadband noise," Ergonomics **38**, 2313–2326.
Haas, E., and Edworthy, J. (**2006**). "An introduction to auditory warnings and alarms," in *Handbook of Warnings*, edited by M. S. Wogalter (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 189–198.
Halberstadt, J. B., and Niedenthal, P. M. (**1997**). "Emotional state and the use of stimulus dimensions in judgement," J. Pers. Soc. Psychol. **72**, 1017–1033.
Keller, P., and Stevens, C. (**2004**). "Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons," J. Exp. Psychol., Appl. **10**, 3–12.
Kline, P. (**1994**). *An Easy Guide to Factor Analysis* (Routledge, New York).
Kruskal, J. B. (**1964**). "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika **29**, 1–27.
MacKinnon, D. P., and Fairchild, A. J. (**2009**). "Current directions in mediation analysis," Curr. Dir. Psychol. Sci. **18**, 16–20.
Martens, W. L., and Zacharov, N. (**2000**). "Multidimensional perceptual unfolding of spatially processed speech I: Deriving stimulus space using INDSCAL," in Proceedings AES 109th Convention, Los Angeles, CA, Preprint No. 5224.
Miller, J. R., and Carterette, E. C. (**1975**). "Perceptual space for musical structures," J. Acoust. Soc. Am. **58**, 711–720.
Muller, D., Judd, C. M., and Yzerbyt, V. Y. (**2005**). "When moderation is mediated and mediation is moderated," J. Pers. Soc. Psychol. **89**, 852–863.
Neuhoff, J. G. (**1998**). "Perceptual bias for rising tones," Nature (London) **395**, 123–124.
Niedenthal, P. M., Halberstadt, J. B., and Innes-Ker, Å. (**1999**). "Emotional response categorization," Psychol. Rev. **106**, 337–361.
Nygaard, L. C., and Queen, J. S. (**2008**). "Communicating emotion: Linking affective prosody and word meaning," J. Exp. Psychol. Hum. Percept. Perform. **34**, 1017–1030.
Patterson, R. D. (**1990**). "Auditory warning sounds in the work environment," Philos. Trans. R. Soc. London, Ser. B **327**, 485–492.
Plomp, R. (**1967**). "Beats of mistuned consonances," J. Acoust. Soc. Am. **42**, 462–474.
Pollack, I., and Ficks, L. (**1954**). "Information of elementary multidimensional auditory displays," J. Acoust. Soc. Am. **26**, 155–158.
Repp, B. H., and Knoblich, G. (**2007**). "Action can affect auditory perception," Psychol. Sci. **18**, 6–7.
Rioux, V. (**2001**). "Sound quality of flue organ pipes—An interdisciplinary study on the art of voicing," Ph.D. thesis, Chalmers University of Technology, Gothenburg.
Russell, J. A. (**2003**). "Core affect and the psychological construction of emotion," Psychol. Rev. **110**, 145–172.
Scherer, K. R. (**2001**). "Appraisal considered as a process of multilevel sequential checking," in *Appraisal Processes In Emotion: Theory, Methods, Research*, edited by K. R. Scherer, A. Schorr, and T. Johnstone (Oxford University Press, New York), pp. 92–120.
Shrout, P. E., and Bolger, N. (**2002**). "Mediation in experimental and nonexperimental studies: New procedures and recommendations," Psychol. Methods **7**, 422–445.
Sköld, A. (**2005**). "Perceived character and quality of vehicle sound," Licentiate Report No. Lic 2005:8, Chalmers University of Technology, Gothenburg.
Sköld, A. (**2008**) "Integrative analyses of perception and reaction to information and warning sounds in vehicles," Ph.D. thesis, Chalmers University of Technology, Gothenburg.
Sköld, A., Västfjäll, D., and Kleiner, M. (**2004**). "Predicting consumer preference from expert sensory ratings of sounds," in Proceedings of the ICA 18th International Congress, Kyoto, Japan, pp. 1825–1828.
Tajadura, A., Väljamäe, A., Västfjäll, D., and Asutay, E. "Embodied affect in auditory perception: The impact of approaching and receding sounds on emotional responses," Emotion (in press).
Västfjäll, D., Friman, M., Gärling, T., and Kleiner, M. (**2002**). "The measurement of core affect: A Swedish self-report measure," Scand. J. Psychol. **43**, 19–31.
Västfjäll, D., and Gärling, T. (**2007**). "Validation of a Swedish short self-report measure of core affect," Scand. J. Psychol. **48**, 233–248.
Västfjäll, D., Kleiner, M., and Gärling, T. (**2003**). "Affective reactions to interior aircraft sound quality," Acta Acust. **89**, 693–701.
Västfjäll, D., Sköld, A., and Kleiner, M. (**2004**). "On the use of psychoacoustical and psychophysical methods to predict consumer preference for interior car sounds," in Proceedings of the CFA/DAGA, Strasbourg, pp. 24–28.
Warren, W. H., Jr., and Verbrugge, R. R. (**1984**). "Auditory perception of breaking and bouncing events: A case study in ecological acoustics," J. Exp. Psychol. [Hum Learn] **10**, 704–712.
Wickelmaier, F., and Schmid, C. (**2004**). "A Matlab function to estimate choice-model parameters from paired-comparison data," Behav. Res. Methods Instrum. Comput. **36**, 29–40.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics: Facts and Models* (Springer-Verlag, Berlin).

A comparison of the temporal weighting of annoyance and loudness

Kerstin Dittrich[a] and Daniel Oberfeld
Department of Psychology, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany

(Received 20 August 2008; revised 16 July 2009; accepted 26 August 2009)

The influence of single temporal portions of a sound on global annoyance and loudness judgments was measured using perceptual weight analysis. The stimuli were 900-ms noise samples randomly changing in level every 100 ms. For loudness judgments, Pedersen and Ellermeier [J. Acoust. Soc. Am. **123**, 963–972 (2008)] found that listeners attach greater weight to the beginning and ending than to the middle of a stimulus. Qualitatively similar weights were expected for annoyance. Annoyance and loudness judgments were obtained from 12 listeners in a two-interval forced-choice task. The results demonstrated a primacy effect for the temporal weighting of both annoyance and loudness. However, a significant recency effect was observed only for annoyance. Potential explanations of these weighting patterns are discussed. Goodness-of-fit analysis showed that the prediction of annoyance and loudness can be improved by allowing a non-uniform weighting of single temporal portions of the signal, rather than assuming a uniform weighting as in measures like the energy-equivalent level (L_{eq}). A second experiment confirmed that the listeners were capable of separating annoyance and loudness of the stimuli. Noises with the same L_{eq} but different amplitude modulation depths were judged to differ in annoyance but not in loudness.
 [DOI: 10.1121/1.3238233]

PACS number(s): 43.66.Mk, 43.66.Cb, 43.66.Ba, 43.50.Ba [BCM] Pages: 3168–3178

I. INTRODUCTION

A. Annoyance and noise exposure

The perceptual dimension annoyance has received considerable interest over the last few decades (for recent reviews see Kryter, 2007; Marquis-Favre *et al.*, 2005a, 2005b). Parameters influencing annoyance can be divided into acoustical parameters (cf. Zwicker, 1991), such as the presence of tonal components (e.g., Hellman, 1984, 1985) or frequency (e.g., Leventhall, 2004), and non-acoustical variables, such as individual noise-sensitivity (e.g., Zimmer and Ellermeier, 1996). The non-acoustical variables could explain why listeners' evaluations of the annoyance of sounds differ widely. Loud rock music, for example, can be a pleasant event as well as an annoying disturbance.

An important focus of research on noise is to examine the consequences of noise exposure, which is commonly associated with annoyance reactions. Noise can have negative effects on the auditory system, for example, inner ear damage. Non-auditory effects also occur, for example, sleep disturbance, impairment of work performance, or interference with daily activities (e.g., Michaud *et al.*, 2008; see Marquis-Favre *et al.*, 2005a for a recent review). Given the fact that noise exposure has a lot of negative consequences, engineering standards and laws have been developed to protect people against these negative consequences (for an international example see, e.g., Guidelines for Community Noise, WHO, 1999). These regulations mostly use technical measures like the energy-equivalent level (L_{eq} or L_{Aeq}) in order to assess the annoyance or loudness of noises.[1] The measures used to quantify noise can be divided into different categories (cf. Marquis-Favre *et al.*, 2005a), for example, those related to the sound pressure level (e.g., L_A), energy-based indices (e.g., L_{eq}), or statistical indices (e.g., N_5).[2] Given the fact that sound intensity is an important factor for both annoyance and loudness (e.g., Zwicker, 1991; Hellman, 1982), the same measures are frequently used for the two dimensions (cf. Marquis-Favre *et al.*, 2005a; Schomer *et al.*, 2001). Nevertheless, annoyance and loudness depend in a different manner on the characteristics of sounds. For example, Zwicker (1991) proposed that besides loudness, amplitude modulation depth and sharpness should be taken into account in annoyance calculations.

For assessing the annoyance of longer sounds that fluctuate in level, several alternative measures have been proposed (e.g., N_5, L_{eq}, or L_A; cf. Zwicker and Fastl, 1999). Most countries use some variant of the A-weighted energy-equivalent level (cf. Schomer *et al.*, 2001). The validity of L_{eq} and L_{Aeq} for estimating the annoyance of real-world noises was partially confirmed in some studies (e.g., Hiramatsu *et al.*, 1983; Kuwano and Namba, 2000). These measures take into account acoustical parameters such as sound pressure level and frequency spectrum. However, the correlations between these measures and annoyance judgments are frequently found to be rather weak (see Marquis-Favre *et al.*, 2005a, 2005b, for recent reviews). The weak correlation can be ascribed to at least two different causes. First, these measures do not take into account non-acoustical factors such as individual sound sensitivity. Second, relevant acoustical parameters might not be considered in these measures.

[a]Author to whom correspondence should be addressed. Present address: Department of Psychology, Albert-Ludwigs-Universität Freiburg, 79085 Freiburg, Germany. Electronic mail: dittrich@psychologie.uni-freiburg.de

B. Temporal aspects of annoyance

One acoustical parameter which has not received much consideration until now is the *temporal aspect* of annoyance (notable exceptions are Hiramatsu *et al.*, 1983; Dornic and Laaksonen, 1989; Namba and Kuwano, 1979, 1980). This study is concerned with the question of whether and how the influence of single temporal portions of a longer stimulus on annoyance varies as a function of the temporal position within the sound. In a two-interval forced-choice task, two noises consisting of nine contiguous 100-ms segments were presented. The task was to select the more annoying noise. On each trial, the sound pressure levels of the nine segments were drawn independently from a normal distribution for each of the two noises, with a 1 dB difference in mean level between the two intervals. In such a setting, the *perceptual weight* is defined as the relative influence that the level of a given temporal segment had on the decision of the listener. These weights can be estimated from the trial-by-trial data using *molecular* analyses (e.g., Ahumada and Lovell, 1971; Berg, 1989; Richards and Zhu, 1994).

If listeners are asked to judge the overall loudness of the described type of sounds, the initial and final portions of the stimulus receive greater weight than its temporal center (e.g., Ellermeier and Schrödl, 2000; Oberfeld, 2008a, 2008b; Pedersen and Ellermeier, 2008). In other words, primacy and recency effects are observed. Pedersen and Ellermeier (2008) suggested that an interaction of perceptual and cognitive processes leads to the observed primacy/recency weighting pattern. This assumption seems to be plausible given the fact that primacy and recency effects are not specific to this type of loudness judgment, but are ubiquitous in cognitive psychology. In studies of learning and memory of serially sorted information, the serial position curve frequently shows both a primacy and a recency effect (e.g., Postman and Phillips, 1965; Anderson *et al.*, 1998; Jones *et al.*, 2004). Similar memory effects have been found for the recall of nonverbal acoustical stimuli (McFarland and Cacace, 1992; Surprenant, 2001). In these studies, serial position effects were examined for tonal sequences with an overall duration up to 4 s. For loudness, one can assume that the levels of the single segments of a noise are processed as serially sorted information in a system exhibiting similar characteristics to short-term memory (Oberfeld, 2008b). The beginning and the ending can be assumed to be more distinct than the middle of the noise, and therefore have a stronger influence on a decision, as, for example, a loudness judgment (see Neath *et al.*, 2006 for a detailed discussion).

The present study compared the temporal weighting of loudness and annoyance. Primacy and recency effects were expected to show for both perceptual dimensions. One reason for this expectation was the close relation between loudness and annoyance (e.g., Zwicker, 1966; Hellman, 1984, 1985). Additionally, if the processing of the segments as serially sorted information caused the non-uniform temporal weighting, this effect should show for annoyance as well as for loudness.

Insight into the temporal weighting of annoyance is especially relevant for technical measures used in noise quantification. Conventional measures like L_{eq} assume that listeners weight the information provided by each temporal portion of a sound uniformly. The present study examined whether this approach is compatible with the perception of annoyance or whether temporal aspects should be considered in the estimation of annoyance.

In Experiment 1, listeners evaluated the relative annoyance and the relative loudness of two 900-ms samples of noise. The sound pressure level of the noise was changed randomly every 100 ms by drawing the level repeatedly and independently from a normal distribution. The influence of single temporal segments of this level-fluctuating noise on annoyance and loudness judgments was estimated using perceptual weight analysis (cf. Berg, 1989). Goodness-of-fit analysis was used to test whether the prediction of annoyance and loudness can be improved by allowing for a non-uniform weighting of single temporal portions of the signal.

A potential problem for the within-subjects comparison of temporal weights for loudness and annoyance is that listeners may not be capable of separating loudness and annoyance when repeatedly judging the same type of stimuli. If Experiment 1 actually showed the expected differences between the temporal weighting patterns for annoyance and loudness, then this would demonstrate that loudness and annoyance represented separate dimensions. If, on the other hand, an identical pattern of weights was found for annoyance and loudness, it might have been the case that subjects always evaluated the noises according to their loudness, even when they were asked to judge the stimuli according to their annoyance, or vice versa.

We used two methods to assess these possibilities. First, in Experiment 1, two groups of listeners were assigned to different task orders. One group made only annoyance judgments in the first part of the experiment, and only loudness judgments in the second part. For the other group, the order of the tasks was reversed. If the "true" weighting patterns for loudness and annoyance differed in any respect, then an effect of task order on the patterns of weights would indicate a failure of the listeners to switch between loudness and annoyance judgments. Second and more important, the listeners from Experiment 1 participated in an additional experiment (Experiment 2b) designed to more directly test whether they were capable of judging the level-fluctuating stimuli independently according to their loudness and their annoyance. To this end, we presented noises differing in modulation depth, that is, in the variability of the nine segment levels. Modulation depth has been reported to produce a dissociation between loudness and annoyance (Widmann, 1994). In Experiment 2, the two noises presented on each trial had the same L_{eq} while differing in modulation depth. We expected the listeners to perceive these sounds as similar in loudness, but to perceive the sound with the higher modulation depth as more annoying (Zwicker, 1991), in line with our assumption that listeners were able to separate the dimensions loudness and annoyance for our stimuli.

II. EXPERIMENT 1: COMPARISON OF THE TEMPORAL WEIGHTING OF ANNOYANCE AND LOUDNESS

A. Method

1. Listeners

Twelve listeners (8 women, 5 men, age 20–31 years) participated. Most were psychology students at the Johannes Gutenberg-Universität Mainz, participated for course credit, and had no experience in comparable psychoacoustic tasks. All listeners reported normal hearing. Detection thresholds in the right ear, as measured by a two-interval forced-choice, adaptive procedure with a three-down, one-up rule (Levitt, 1971), were better than 15 dB HL (hearing level; relative to the reference levels provided by Han and Poulsen, 1998) at all octave frequencies between 250 and 4000 Hz.

The individual noise-sensitivity of the listeners was assessed using the noise-sensitivity questionnaire of Zimmer and Ellermeier (1996). Noise-sensitivity is viewed as a trait reflecting individual differences in the tolerance of environmental noise. The questionnaire assesses the perceptual, cognitive, affective, and behavioral reactions to noise in different contexts. The mean noise-sensitivity was $M=85$, standard deviation $SD=8$, with a range from 71 to 100. Since the scale ranges from 0 to 186, none of the listeners could be considered as being extremely high or low in his or her noise-sensitivity. However, the listeners differed widely concerning their self-rated noise-sensitivity. On an 11 point rating scale ranging from 0 (not at all noise sensitive) to 10 (very noise sensitive), the range of self-rated noise-sensitivity was 2–9 ($M=5.7$, $SD=2.6$).

2. Apparatus

The stimuli were generated digitally, played back via two channels of an RME ADI/S digital-to-analog converter ($f_S=44.1$-kHz, 24-bit resolution), attenuated (two TDT PA5s), buffered (TDT HB7), and presented diotically via Sennheiser HDA 200 headphones calibrated according to IEC 318 (1970). The experiment was conducted in a single-walled sound-insulated chamber. Listeners were tested individually.

3. Stimuli and experimental procedure

The stimuli were presented in a two-interval procedure. Figure 1(a) shows a schematic depiction of a trial. On each trial, two level-fluctuating noises were presented. The stimuli were Gaussian wide-band noises consisting of nine contiguous temporal segments. The duration of each segment was 100 ms. On each trial and for each interval, the sound pressure levels of the nine temporal segments were drawn independently from a normal distribution. In the interval containing the less intense noise, the mean of the distribution was $\mu_L=64.5$-dB sound pressure level (SPL) and the SD was 2.5 dB. In the interval containing the more intense noise, the mean was $\mu_H=65.5$-dB SPL, also with $SD=2.5$ dB. Although the estimation of perceptual weights would be possible without a difference in mean level between the two intervals, we introduced this difference in level mainly to make the task easier for the subjects and also to be compatible with previous experiments (e.g., Berg, 1989; Ellermeier and Schrödl, 2000). The more intense noise was presented in interval 1 or interval 2 with identical *a priori* probability. To avoid overly loud sounds, the range of levels was restricted to $\mu \pm 2.5$SD. Therefore, the maximal level difference between the most intense and the least intense segment within a given noise was 12.5 dB. The standard deviation of the nine segment levels (SD_{levels}) is a measure of the modulation depth. Across all trials, the mean modulation depth was 2.3 dB ($SD=0.55$ dB, range of 0.48–4.46 dB).

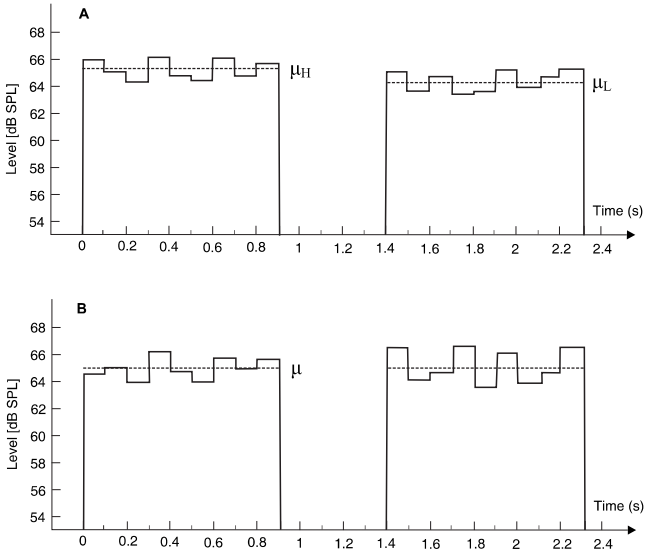


FIG. 1. Trial configurations used in Experiments 1 and 2. Two broad-band-noises consisting of nine contiguous segments were presented. On each trial, the level of each segment was drawn independently from one of two normal distributions differing in their means (Experiment 1) or standard deviations (Experiment 2). Task: First or second noise more annoying/louder? The dashed line represents the mean level. Panel (A): Experiment 1. Means of the normal distribution: $\mu_L=64.5$-dB SPL in one interval; $\mu_H=65.5$-dB SPL in the other interval. The SD was 2.5 dB in both intervals. Panel (B): Experiment 2. Small modulation depth (normal distribution with $SD=2$ dB) in one interval; large modulation depth (normal distribution with $SD=4$ dB) in the other interval. For the noise with the small modulation depth, the mean of the normal distribution was $\mu=65$ dB SPL. The two noises had the same L_{eq} or N_5.

The two noises were presented with a silent interstimulus interval of 500 ms. Depending on the task, the listeners selected the interval containing the more annoying or louder sound. No feedback was provided. The next trial followed the response after an inter-trial interval of 2 s.

In the last part of the experiment, magnitude estimates of the annoyance of the stimuli were obtained using a procedure without reference (e.g., Hellman and Zwislocki, 1961) and essentially the same instructions as in Hellman and Meiselman (1988). A single noise, fluctuating in level, was presented on each trial, with the segment levels drawn from a normal distribution with mean $\mu=65$ dB SPL and $SD=2$ dB.[3] Each listener judged 15 noises four times in randomized order. The listeners were asked to choose any positive number which seemed adequate to describe the annoyance of the presented noise. The geometric mean of the 60 numerical judgments was taken as the individual annoyance estimate. Across listeners, the mean magnitude estimate of annoyance was $M=0.75$ ($SD=0.13$, range of 0.55–1.00).

The listeners were randomly assigned to two experimental groups. Group 1 made only annoyance judgments in the first part of the experiment, and only loudness judgments in the second part. For Group 2, the order of tasks was reversed. The experiment was arranged in blocks of 50 trials. Each session comprised ten blocks and lasted approximately 60 min. Each listener completed six sessions. At the beginning of a session, the listeners received 50 practice trials. In the first session, the listeners from both groups completed the detection threshold measurements. In the second part of session 1, Group 1 received 300 trials of Experiment 2b (see description below) and judged the noises according to their annoyance. In sessions 2 and 3, the listeners of Group 1 received 1000 trials in the annoyance task of Experiment 1. In the second part of the experiment (sessions 4–6), Group 1 first made loudness judgments for 300 trials of Experiment 2b, and then completed 1000 trials in the loudness task of Experiment 1. For Group 2 the procedure was analogous.

At the end of session 6, the listeners provided magnitude estimates of the annoyance of the level-fluctuating noises (see above) and filled in the noise-sensitivity questionnaire (Zimmer and Ellermeier, 1996).

4. Estimation of temporal weights

Multiple binary logistic regression (PROC LOGISTIC, SAS 8.01) was used to estimate the weights from the trial-by-trial data.[4] For each trial and each segment ($i=1,\ldots,9$), the difference between the level of segment i in interval 2 and the level of segment i in interval 1 was computed. The binary responses served as the dependent variable and the nine within-trial segment level differences served as predictors. Due to the difference in mean level between the two intervals, the within-trial segment level differences were correlated. Therefore, separate logistic regression analyses were conducted for the trials in which the noise with the higher mean level (μ_H) occurred in interval 1, and for the trials in which the position of the noise with mean level μ_H was interval 2. Thus, a logistic regression was conducted for each combination of subject, task (annoyance/loudness), and position μ_H. Because modulation depth, that is, the variability of the levels within a sound, has an influence on annoyance (e.g., Widmann, 1994), the within-trial difference between the standard deviations of the nine segment levels (SD_{levels}) in interval 2 and the nine segment levels in interval 1 was included as a predictor. A comparison of the goodness-of-fit for models containing or not containing the within-trial difference in SD_{levels} as a predictor will be presented in Sec. II B.

The regression coefficients for the nine segment level differences were taken as weight estimates. The weights were normalized such that the sum of the absolute values was unity (see Kortekaas *et al.*, 2003), resulting in a set of relative temporal weights for each listener, task (annoyance/loudness), and position μ_H.

The unweighted residual sum-of-squares test (Copas, 1989) was used for assessing global goodness-of-fit. This test has been shown to perform favorably compared to some alternative tests (Hosmer *et al.*, 1997; Kuss, 2002). An SAS macro (GOFLOGIT; Kuss, 2001) was used to compute the test statistics. In global goodness-of-fit tests, the hypothesis is tested that the saturated (full) model containing as many parameters as observations does not provide a better description of the data than the fitted (restricted) model (cf. Agresti, 2002). Small p-values indicate lack-of-fit of the restricted model. It is usual to take p-values of less than 0.2 as an indication that the model did not fit adequately (cf. Agresti, 2002). For the 48 (Listener × Task × Position μ_H) fitted multiple logistic regression models the test produced a p-value below 0.2 in only five cases.

A summary measure of the predictive power of a logistic regression model is the area under the receiver operating characteristic (ROC) curve (cf. Agresti, 2002, Swets, 1986). This measure provides information about the degree to which the predicted probabilities are concordant with the observed outcome (see Hosmer and Lemeshow, 2000, for a critical discussion). The logistic regression model predicts the probability of a response as a function of the values of the predictors (e.g., the nine within-trial differences in segment level). For example, let $y=0$ denote the observed response "Louder noise in interval 1" and $y=1$ denote the response "Louder noise in interval 2." The logistic regression equation models the probability $\hat{\pi}$ of y being equal to 1. The predicted response is $\hat{y}=1$ when $\hat{\pi}>n_0$, and $\hat{y}=0$ when $\hat{\pi}\leq\pi_0$, for some cutoff π_0. The *sensitivity* of the model is $P(\hat{y}=1|y=1)$ and the *specificity* is $P(\hat{y}=0|y=0)$. In signal detection theory terms, the sensitivity corresponds to the proportion of hits, and the specificity corresponds to one minus the proportion of false alarms. The sensitivity and the specificity depend on the arbitrary cutoff π_0. For example, a value of π_0 close to 0 maximizes the sensitivity but minimizes the specificity. The area under the ROC curve, which is a plot of sensitivity as a function of (1—specificity), overcomes this limitation because it summarizes the predictive power for all possible cutoffs. In practice, the area under the ROC curve (AUC) is often computed via a Mann–Whitney U type of statistic for all pairs of $y=0$ and $y=1$ trials (Bamber, 1975). Areas of 0.5 and 1.0 correspond to chance performance and perfect performance, respectively. Across the 48 fitted logistic regression models, AUC ranged between 0.55 and 0.89 ($M=0.74$, $SD=0.097$), indicating reasonably good predictive power (Hosmer and Lemeshow, 2000). A repeated-measures analysis of variance (ANOVA) on AUC, with the within-subjects factors (annoyance and loudness) and position μ_H (first interval and second interval) and the between-subjects factor order of tasks (annoyance judgments first and loudness judgments first), showed no significant effects (all $p>0.1$).

B. Results and discussion

1. Temporal weights

The individual temporal weighting patterns are displayed in Fig. 2 separately for each task and each position of the noise with higher mean level. Clear primacy and recency effects for both tasks were observed for listeners 2, 4, 5, 8, and 9, while other listeners showed a more equal weighting pattern, e.g., 7 and 10. The position of the noise with higher mean level did not usually produce strong differences in the weighting patterns, an exception being for listener 1.

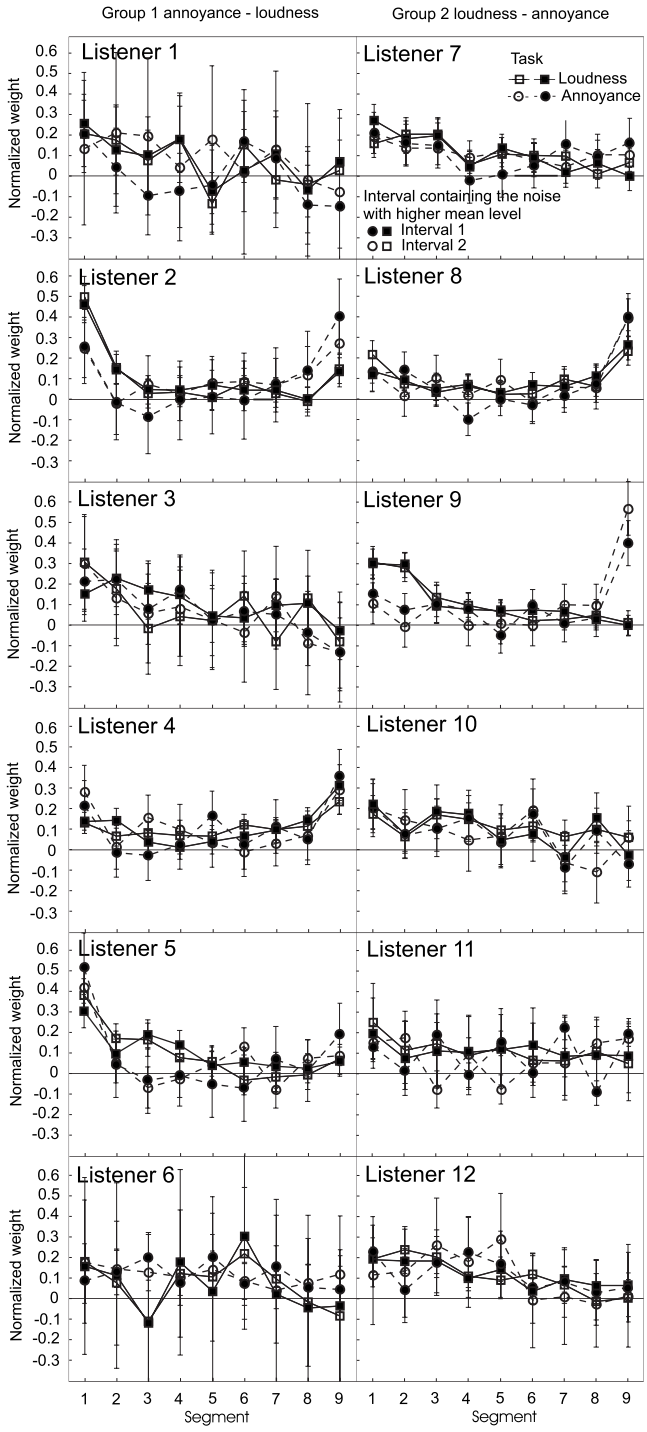


FIG. 2. Experiment 1. Individual relative normalized temporal weights are plotted as a function of segment number. Panels represent listeners. Left column: Group 1 (task order: annoyance-loudness). Right column: Group 2 (task order: loudness-annoyance). Squares and continuous lines: Loudness judgments. Circles and dashed lines: Annoyance judgments. Filled symbols: Interval 1 contained the noise with the higher mean level. Open symbols: Interval 2 contained the noise with the higher mean level. Error bars show the 95%-confidence intervals.

The normalized weights were analyzed via a repeated-measures ANOVA using a univariate approach. The Huynh–Feldt correction for the degrees-of-freedom was used where applicable (Huynh and Feldt, 1976), and the value of the df correction factor $\tilde{\varepsilon}$ is reported. The three within-subjects factors were segment (1–9), task (annoyance and loudness), and position μ_H (first interval and second interval). The order of tasks (annoyance judgments first and loudness judgments first) was included as a between-subjects factor. The results are displayed in Table I. There was a significant effect of segment. The Segment × Task interaction was significant, possibly because annoyance and loudness differed in their recency effects (see below). The effect of order of tasks was not significant. The Segment × Order of Tasks interaction and the Task × Order of Tasks interaction were also not significant. Thus, task order had no significant effect on the pattern of weights, indicating that the first task performed in the experiment did not strongly influence listeners' behavior in the second task. Due to the normalization of the weights, the main effect of task was also not significant. Because neither the main effect of the position of the noise with the higher mean level nor any interactions with this factor were significant, the weights were averaged across the two positions for further analysis. Figure 3 displays the mean temporal weights.

Primacy effects for the two conditions were compared in a three-factor repeated-measures ANOVA with the within-subjects factor section (weight assigned to Segment 1 versus average weight assigned to Segments 2–8) and task, and order of tasks as between-subjects factor. The analysis revealed a significant main effect of section $[F(1,10)=37.04,\ p<0.001]$, confirming that there was a primacy effect in both tasks. The effect of task was not significant ($p=0.09$). The Section × Task interaction was also not significant ($p>0.1$), failing to confirm our hypothesis that primacy effects are stronger for annoyance than for loudness. Neither the main effect of order of tasks nor the Section × Order of Tasks and Task × Order of Tasks interactions were significant (all $p>0.1$). Recency effects for the two conditions were compared in an ANOVA with the within-subjects factor section (weight assigned to Segment 9 versus average weight assigned to Segments 2–8) and task, and order of tasks as a between-subjects factor. The Section × Task interaction was significant $[F(1,10)=5.41,\ p=0.046]$, confirming the observation of a stronger recency effect for annoyance than for loudness. A *post-hoc* pairwise comparison between the weights assigned to Segment 9 showed a significant difference between the two tasks, $t(11)=3.71$, $p=0.003$ (two-tailed). In the ANOVA, no further main effects and interactions were significant (all $p>0.1$).

Additionally, it was tested whether the temporal weighting patterns for loudness and annoyance differed in uniformity. Therefore, we examined whether the variance of the nine temporal weights differed between loudness and annoyance. For this purpose, the coefficient of variation ($CV=SD/M$) of the nine temporal weights was calculated for each listener and each task. For annoyance, the mean CV was 1.21 ($SD=0.57$). For loudness, the mean CV was 0.97 ($SD=0.45$). A repeated-measures ANOVA with the within-subjects factor task and order of tasks as between-subjects factor showed no significant difference between the CVs of the weights in the two tasks $[F(1,10)=1.55,\ p=0.241]$. Neither the main effect order of task nor the Task × Order of Tasks interaction was significant (both $p>0.3$).

TABLE I. Results of the ANOVA conducted for the normalized weights from Experiment 1. Within-subjects factors: segment (S), task (T), and position μ_H (P). Between-subjects factor: order of tasks (O). Values in parentheses represent mean square errors. $\underline{S}$=subjects. Partial η^2: Variance due to the effect of interest expressed as a proportion of the sum of the error variance and the effect-of-interest variance.

Source	df	F	p	Partial η^2	$\tilde{\epsilon}$
		Between subjects			
Order of tasks (O)	1	3.43	0.094	0.255	
$\underline{S}$ within-group error	10	(0.007)			
		Within subjects			
Segment number (S)	8	7.759[a]	0.001	0.430	0.394
Task (T)	1	2.844	0.123	0.202	
Position (P)	1	0.264	0.618	0.011	
S × T	8	2.276[b]	0.045	0.181	0.788
S × P	8	0.450	0.887	0.052	1.0
S × O	8	1.095	0.357	0.107	
T × P	1	2.423	0.151	0.166	
T × O	1	0.122	0.734	0.013	
P × O	1	0.026	0.875	0.012	
S × T × P	8	0.505	0.771	0.043	0.626
S × T × O	8	0.715	0.572	0.064	
S × P × O	8	0.749	0.648	0.075	
T × P × O	1	2.362	0.155	0.191	
S × T × P × O	8	0.639	0.671	0.049	
S×$\underline{S}$	80	(0.047)			
T×$\underline{S}$	10	(0.005)			
P×$\underline{S}$	10	(0.002)			
S×T×$\underline{S}$	80	(0.013)			
S×P×$\underline{S}$	80	(0.004)			
T×P×$\underline{S}$	10	(0.002)			
S×T×P×$\underline{S}$	80	(0.006)			

[a]$p<0.01$.
[b]$p<0.05$.

2. Model comparisons

The weights analyzed above were estimated via a multiple logistic regression model including the within-trial difference in modulation depth as a predictor. For comparison, a simpler regression model containing as predictors only the nine within-trial differences in segment level was fitted. The two models are nested so that a likelihood-ratio test (Agresti, 2002) can be used for model comparison. This test uses the statistic $-2(LL_{\text{restricted}}-LL_{\text{full}})$, where $LL_{\text{restricted}}$ and LL_{full} are the log likelihood of the model containing only the nine differences in segment level and the model additionally containing the difference in modulation depth, respectively. The test statistic is asymptotically distributed as χ^2_1 because the full model contains one additional free parameter. Of the 48

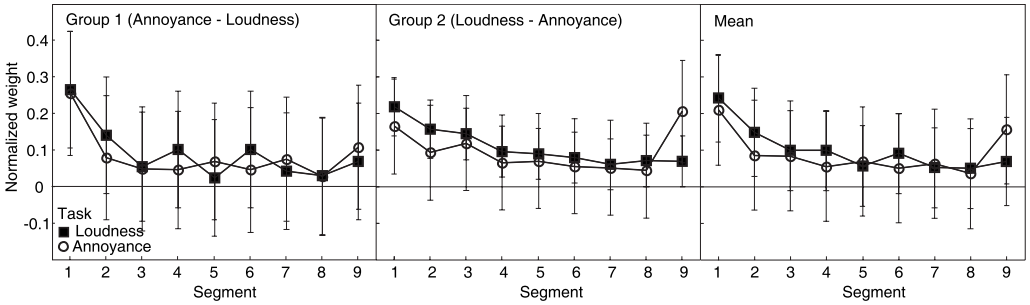


FIG. 3. Experiment 1. Mean relative normalized temporal weights plotted as a function of segment number. Squares: Loudness judgments. Circles: Annoyance judgments. Left panel: Group 1 (task order: annoyance-loudness). Middle panel: Group 2 (task order: loudness-annoyance). Right panel: All listeners (Groups 1 and 2 aggregated). Error bars show the 95%-confidence intervals.

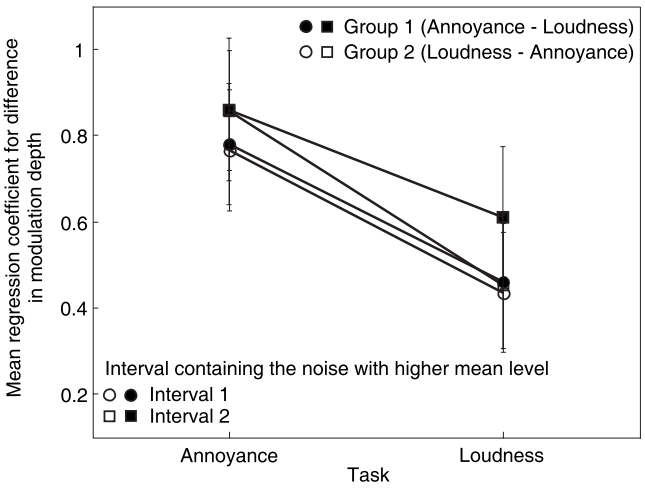


FIG. 4. Multiple logistic regression model for Experiment 1, containing as predictors the within-trial differences in segment level, and the difference in modulation depth. Shown are mean regression coefficients for the difference in modulation depth as a function of task and order of tasks. Circles: Interval 1 contained the noise with the higher mean level. Squares: Interval 2 contained the noise with the higher mean level. Filled symbols: Group 1 (task order: annoyance-loudness). Open symbols: Group 2 (task order: loudness-annoyance). Error bars show the 95%-confidence intervals.

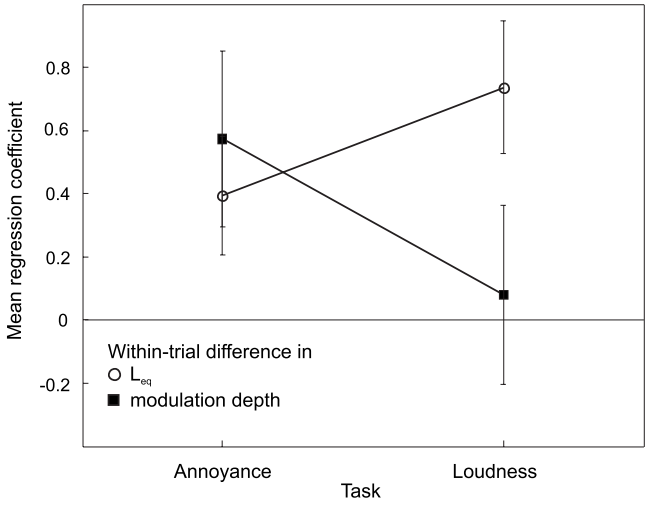


FIG. 5. Mean regression coefficients for a multiple logistic regression model for Experiment 1, containing as predictors the within-trial difference in L_{eq} (circles), and the within-trial difference in modulation depth (squares). Error bars show the 95%-confidence intervals.

(Listener × Task × Position μ_H) fitted models including both the differences in segment level and the difference in modulation depth, all but 14 fitted the data significantly better than the model not containing the difference in modulation depth ($p<0.05$). Notably, of these 14 models, only 2 were from the annoyance task, compatible with the expected stronger influence of modulation depth on annoyance than on loudness judgments. This pattern was also evident in the regression coefficients for the difference in modulation depth obtained for the full model, which are shown in Fig. 4. These regression coefficients were analyzed via a repeated-measures ANOVA with the within-subjects factors task and position μ_H and the between-subjects factor order of tasks. A significant effect of task confirmed the observation that the difference in modulation depth had a stronger influence for the annoyance than for the loudness task [$F(1,10)=7.40$, $p=0.022$]. The remaining effects were not significant ($p>0.15$).

As discussed in the Introduction, the traditional view is that the loudness and the annoyance of a sound can be predicted from L_{eq} and, at least in the case of annoyance, the modulation depth (e.g., Zwicker, 1991). To test whether a model allowing for a non-uniform temporal weighting of the nine segment levels provides a better account of loudness and annoyance, the fit of two different multiple logistic regression models was compared. The restricted model was compatible with the traditional approach and contained as predictors the difference between L_{eq} in interval 2 and L_{eq} in interval 1 (ΔL_{eq}), and the difference between the modulation depths (i.e., the standard deviation of the nine segment levels) in interval 2 and in interval 1 (ΔSD_{levels}). The full model also contained the predictors ΔL_{eq} and ΔSD_{levels}, but additionally the nine within-trial differences in segment level.

For the restricted model, the regression coefficient for the within-trial difference in L_{eq} was generally positive. It was not significantly different from 0 ($p>0.05$, two-tailed) in only five cases. Figure 5 shows the mean data. Across the 48 fitted models, the regression coefficient for the within-trial difference in modulation depth was mostly greater than 0, and in only one case significantly smaller than 0. In 26 of the 48 fitted models, it was significantly different from 0 ($p<0.05$, two-tailed). Figure 5 shows the mean data. As expected, the influence of modulation depth was stronger for annoyance than for loudness.

The full model containing as predictors the nine segment levels as well as L_{eq} and modulation depth, and the restricted model containing only L_{eq} and modulation depth are nested. The full model has nine additional free parameters so that the test statistic is distributed as χ^2_9. In 30 of the 48 cases, the fit of the full model was significantly better than the fit of the restricted model ($p<0.05$), compatible with the hypothesis that the prediction of loudness and annoyance can be improved by allowing for a non-uniform weighting of the sound pressure level of single temporal portions of the signal. This notion was corroborated by an analysis of the predictive power of the two alternative models in terms of the area under the ROC curve. Figure 6 shows the mean outcome. A repeated-measures ANOVA with the within-subjects factors model (full versus restricted), task, and position μ_H, and the between-subjects factor order of tasks showed a significant effect of model [$F(1,10)=45.92$, $p<0.001$], confirming the higher predictive power of the model with a non-uniform weighting of the segment levels. The remaining effects were not significant ($p>0.1$). Two *post-hoc* ANOVAs were conducted to examine the AUC differences for the loudness task and the annoyance task separately. There was a significant effect of model for both tasks [$F(1,10)=30.99$, $p<0.001$ and $F(1,10)=31.71$, $p<0.001$, respectively], demonstrating that including temporal weights can improve the prediction of both perceptual dimensions.

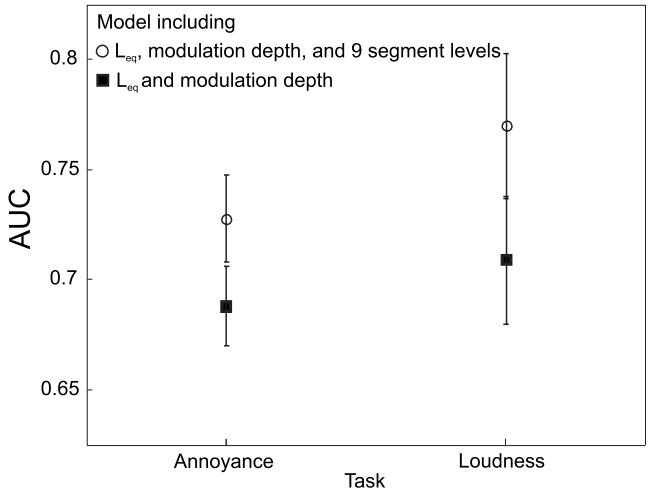


FIG. 6. Mean predictive power of two models for the results of Experiment 1, in terms of the AUC, as a function of task. Squares: Model containing as predictors the within-trial differences in L_{eq} and in modulation depth. Circles: Model containing as predictors the within-trial differences in L_{eq} and in modulation depth, and the nine within-trial differences in segment level. Error bars show the 95%-confidence intervals.

III. EXPERIMENT 2: CAN THE STIMULI USED IN EXPERIMENT 1 BE JUDGED SEPARATELY ACCORDING TO THEIR ANNOYANCE AND LOUDNESS?

In the Introduction, we discussed the potential problem of listeners failing to separate loudness and annoyance. The significant differences between the patterns of temporal weights for loudness and annoyance found in Experiment 1 and the only very weak effect of task order indicated that this problem was negligible in the present study. In Experiment 2, we used a different and more direct approach to examine whether the listeners were capable of independently judging the type of stimuli used in Experiment 1 according to their loudness or their annoyance. The rationale of this experiment was to introduce a parameter previously reported to produce a dissociation between loudness and annoyance. To this end, we presented noises differing in amplitude modulation depth. Level-fluctuating sounds have been suggested to be similar in loudness to steady sounds with the same L_{eq} or N_5 (Berglund *et al.*, 1976; Zwicker and Fastl, 1999, Chap. 16), although some studies reported amplitude-modulated sounds to be slightly louder than steady sounds with the same root-mean-square level (e.g., Zhang and Zeng, 1997; Grimm *et al.*, 2002), while Moore *et al.* (1999) found a small effect in the opposite direction for sinusoidal rather than noise carriers. For annoyance, on the other hand, it has been proposed that sounds differing in modulation depth may differ considerably in their annoyance even if the L_{eq} or the N_5 is constant (Zwicker, 1991; Widmann, 1994; but see Hiramatsu *et al.*, 1983). The method for Experiment 2b was similar to Experiment 1, but we constructed the two noises presented within a trial so that they were similar in loudness while differing in modulation depth. We expected the listeners to perceive these sounds as similar in loudness, but to perceive the sound with the higher modulation depth as more annoying, in line with our assumption that listeners were able to separate the dimensions loudness and annoyance. Additionally, we varied the task order (loudness judgments followed by annoyance judgments or vice versa) just as in Experiment 1 so that a tendency of the listeners to adhere to the type of judgment they had given during the first several hundreds of trials could be detected.

A. Experiment 2a: Comparison of L_{eq} and N_5

We are not aware of studies reporting psychophysical loudness measurements for exactly the same type of stimuli used in our experiments. Therefore, we conducted Experiment 2a as a pretest in order to decide whether N_5 or L_{eq} is more suitable for constructing nine segment noises differing in modulation depth but similar in loudness.

On each trial, two noises fluctuating in level were presented in a two-interval forced-choice task. One of the intervals contained a noise with higher modulation depth. The listeners decided which of two noises presented in a given trial was louder. Within a trial, the stimuli were constructed so that they had either the same L_{eq} or the same N_5.

1. Method

Thirteen listeners (8 women, 5 men, age 22–51 years) participated. None of them had taken part in Experiment 1. Most were psychology students and had no experience in comparable psychoacoustic tasks. All listeners reported normal hearing.

The same apparatus as in Experiment 1 was used. On each trial, two level-fluctuating noises with different modulation depth (small or large) but the same L_{eq} or N_5 were presented. The stimuli were identical to Experiment 1 except for the following differences. On each trial, the level-fluctuating noise with small modulation depth was generated by independently drawing the sound pressure levels of the nine temporal segments from a normal distribution with mean μ=65 dB SPL and SD=2 dB. The noise with large modulation depth was generated by independently drawing the nine segment levels from a normal distribution with μ=65 dB SPL and SD=4 dB. Subsequently, the mean level of the noise with large modulation depth was adjusted so that either L_{eq} or N_5 (depending on the experimental condition) was identical to that of the noise with small modulation depth. For identical L_{eq}, the level of each segment of the noise with large modulation depth was adjusted by an identical amount (e.g., +1.1 dB). For identical N_5, the level of each segment of the noise with large modulation depth was adjusted by an identical amount so that the highest segment level was identical to the highest segment level for the noise with small modulation depth.[5]

The procedure was the same as before except for the following differences. Each participant received 300 trials with the same L_{eq} and 300 trials with the same N_5. Trials with identical L_{eq} and N_5 were randomly interleaved in each block. The noise with large modulation depth was presented in interval 1 or interval 2 with identical *a priori* probability. Each block consisted of 50 trials. The listeners' task was to decide whether the first or the second noise was louder. The experiment started with a practice block consisting of 40

trials. See Fig. 1(b) for a schematic depiction of a trial. The duration of the experiment was approximately 35 min.

2. Results and discussion

The proportion of trials in which the noise with the large standard deviation was chosen as louder was analyzed. A mean proportion of 0 would indicate that the noise with the large standard deviation was never chosen as louder; a mean proportion of 1 would indicate that the noise with the large standard deviation was always chosen as louder. For noises identical in L_{eq}, the mean proportion was $M=0.50$, $SD=0.09$, indicating that noises with large and small modulation depths were perceived as equally loud. A one-sample t-test showed that this mean proportion was not significantly different from 0.5 [$t(12)=0.13$, $p=0.899$ (two-tailed)]. For noises identical in N_5, the mean proportion was $M=0.28$, $SD=0.08$, indicating that the noise with the smaller modulation depth was perceived as louder. The mean proportion differed significantly from 0.5 [$t(12)=7.64$, $p=0.001$ (two-tailed)]. A repeated-measures ANOVA with the within-subjects factor loudness measure (L_{eq},N_5) showed a significant effect [$F(1,12)=105.45$, $p<0.01$]. The results demonstrate a clear advantage for L_{eq} compared to N_5 for the purpose of constructing noises with different modulation depth but identical loudness. Therefore, L_{eq} was used in Experiment 2b. The results are consistent with the report of Moore *et al.* (1999) that the loudness of sounds with the same rms level is only weakly influenced by amplitude modulation, but see Zwicker and Fastl (1999) for a different claim.

B. Experiment 2b

1. Method

The same listeners as in Experiment 1 participated in this experiment. The division of the listeners into Groups 1 and 2 was the same as before. The same apparatus and essentially the same stimuli and experimental procedure as in Experiment 2a were used. In a two-interval task, listeners selected the louder or more annoying sound. On each trial, two noises with the same L_{eq} (and therefore approximately the same loudness) but with different modulation depths were presented. Noises with small and large modulation depths were generated by independently drawing the sound pressure levels of the nine temporal segments from a normal distribution with mean $\mu=65$ dB SPL and $SD=2$ or 4 dB. See Fig. 1(b) for a schematic depiction of a trial. The level of each segment of the noise with the large modulation depth was displaced by the same amount so that L_{eq} was equal to that of the noise with small modulation depth. Fifteen trials with the large SD in the first interval and the small SD in the second interval and 15 trials with the reverse order of modulation depths were generated and stored before the experiment started. Thus, the listeners evaluated exactly the same set of stimuli according to both their loudness and their annoyance. The same set of 30 trials was used for all listeners.

As already reported, Experiments 1 and 2b were interleaved. Experiment 2b consisted of practice blocks and experimental blocks. In session 1, Group 1 received each of the 30 stored trials ten times in random order and decided which interval contained the more annoying sound. In session 4, Group 1 again received 300 trials, but this time they made loudness judgments. The stimuli were presented in blocks of 50 trials. For Group 2, which started with the loudness judgments, the procedure was analogous.

2. Results and discussion

The proportion of trials in which the noise with the large standard deviation was chosen as louder or more annoying was analyzed. For loudness, the mean proportion was $M=0.52$ ($SD=0.09$), compatible with the hypothesis that noises with the same L_{eq} are perceived as equally loud. A one-sample t-test showed that this mean proportion was not significantly different from 0.5 [$t(11)=0.63$, $p=0.54$]. For annoyance, the mean proportion was $M=0.66$ ($SD=0.08$), indicating that, as expected, the noise with the larger modulation depth was perceived as more annoying. The mean proportion differed significantly from 0.5 [$t(11)=6.79$, $p<0.001$]. A repeated-measures ANOVA with the within-subjects factor task and the between-subjects factor order of tasks showed a significant effect of task [$F(1,10)=45.75$, $p<0.001$, $\tilde{\varepsilon}=1.0$]. Neither the effect of order of tasks [$F(1,10)=2.47$] nor the Task $\times$ Order of Tasks interaction [$F(1,10)=0.25$] was significant.

The results of Experiment 2b showed that noises with the same energy-equivalent level but different modulation depths were judged to differ in annoyance but not in loudness. These findings indicate that listeners were capable of separating loudness and their annoyance for the type of stimuli presented in Experiment 1. Thus, Experiment 2 provided further evidence for the appropriateness of the within-subjects design used in Experiment 1.

IV. GENERAL DISCUSSION

The present study examined the temporal weighting of annoyance for broadband noises fluctuating in level. The pattern of temporal weights was compared to the temporal weighting of loudness for the same stimuli and the same listeners. The results of Experiment 2 and the non-significant effect of task order in Experiment 1 showed that the listeners were able to separate loudness and annoyance. Additional evidence for the separability of the perceptual dimensions annoyance and loudness was provided by the goodness-of-fit analyses which showed that in Experiment 1 the variance of the intensity fluctuations (modulation depth) had a clear effect on annoyance but not on loudness.

Consistent with previous studies (e.g., Ellermeier and Schrödl, 2000; Plank, 2005; Oberfeld, 2008b; Pedersen and Ellermeier, 2008), a primacy effect was found for loudness and for annoyance. The listeners assigned higher weight to the level of the beginning of a sound than to its middle portion. The size of the primacy effect did not differ between annoyance and loudness. A significant recency effect (i.e., higher weight assigned to the end than to the temporal center of the sound) was observed for annoyance only. In the study of Pedersen and Ellermeier (2008) a recency effect for loudness was evident in the mean data. However, only three of

the five listeners in their experiment showed a recency effect, while all listeners showed a primacy effect. Thus, just as in Experiment 1 of the present study, the primacy effect was stronger than the recency effect. Note that Plank (2005) also found no recency effect in a loudness judgment task where the noise segments were separated by pauses.

As discussed in the Introduction, the processing of the nine segments as serially sorted information represents a potential explanation of the primacy and the recency effects, for example, within the framework of the distinctiveness concept (e.g., Neath *et al.*, 2006). Supporters of this account hypothesize that the difficulty in correctly recalling an item depends on the degree to which it is distinct (or "stands out"). In serially sorted information sets, for example, word lists or the stimuli used in the present study, middle items have two neighboring items. However, end and beginning items only have one neighboring item and therefore are more distinct. It remains for future work to assess whether the temporal weights in loudness and annoyance can be explained in this way.

How can the stronger recency effect for the temporal weighting of annoyance be explained? One potential framework is the "peak-end rule" (Fredrickson and Kahneman, 1993) observed for retrospective evaluations of negative experiences such as painful medical treatments (Redelmeier and Kahneman, 1996) or exposure to aversive sounds (Schreiber and Kahneman, 2000).[6] Kahneman and coworkers found that such judgments are strongly influenced by the worst and the final part of the episode. Thus, a tentative explanation for our observation that the recency effect was significant for annoyance judgments only would be that judging annoyance implies negative emotions while judging the loudness of moderately loud sounds is a "neutral" task and therefore does not elicit a "peak-end" type of retrospective evaluation.

The level profile of stimuli presented in this study was flat because all segment levels within a noise were drawn from the same distribution. For loudness judgments, a gradual increase in level over the first few segments results in a delayed primacy effect: the weights assigned to the attenuated fade-in part are close to zero, and the maximum weight is assigned to the first segment presented at the full level (Oberfeld and Plank, 2005; Oberfeld, 2008a, 2008b). It remains to be shown whether this pattern is paralleled in annoyance judgments.

An important implication of the results of the present study is that technical measures of annoyance and loudness used in noise quantification should consider temporal aspects. Particularly the beginning and ending of a noise should be taken into account more strongly. The goodness-of-fit analyses conducted for Experiment 1 demonstrated that the prediction of both loudness and annoyance can be improved significantly by allowing for a non-uniform weighting of single temporal portions of the signal, rather than assuming that each temporal portion of a sound contributes equally to annoyance and loudness, which is the concept underlying measures like L_{eq} or N_5.

With respect to noise quantification, a limitation of the present study is that the stimuli were shorter than environmental noises, for example, aircraft noise. However, this does not preclude practical applications of our findings. For instance, car alarms frequently use short repeating patterns.[7] Our results indicate that sound designers trying to either increase or decrease the annoyance or loudness of such warning sounds (cf. Suied *et al.*, 2008) should focus on the beginning of the repetitive patterns. Nevertheless, additional research is necessary to clarify whether primacy and recency effects in the temporal weighting pattern of annoyance can be found for longer stimuli.

Finally, it should be noted that Experiment 2a demonstrated that L_{eq} is a better estimate of the loudness of the type of noises used in our experiments than N_5, which is favored by some authors (e.g., Zwicker and Fastl, 1999).

ACKNOWLEDGMENTS

We thank Brian Moore and two anonymous reviewers for helpful comments on earlier versions of this article.

[1] L_{eq}: Energy-equivalent sound pressure level. The sound pressure level of a steady sound that has the same total acoustic energy as a fluctuating sound with the same duration. The equivalent sound pressure level with an A-weighting is referred to as L_{Aeq}.

[2] L_A: A-weighted sound pressure level. This rating is based on the 40-phone equal loudness contour and is expressed in dB(A). N_5 is the maximal loudness (frequently estimated via a loudness model) that is reached or exceeded in 5% of the measurement time (i.e., the 95th percentile of the loudness distribution; Zwicker and Fastl, 1999).

[3] In the magnitude estimation procedure we presented noises with the same modulation depth as for Experiment 2b (levels drawn from a normal distribution with $SD=2$ dB). Therefore, the modulation depth was smaller than for the stimuli of the first part of Experiment 1, where the *SD* was 2.5 dB. Thus, due to the effect of modulation depth on annoyance, the stimuli in Experiment 1 were likely slightly more annoying than the stimuli for which the magnitude estimates were obtained.

[4] Besides multiple binary logistic regression, there exist other techniques for weight estimation (e.g., Ahumada and Lovell, 1971; Berg, 1989; Richards and Zhu, 1994), which are all based on a similar decision model (cf. Lutfi and Jesteadt, 2006) and produce similar estimates (Plank, 2005; Tang *et al.*, 2005). See Pedersen and Ellermeier (2008) for a discussion of the advantages of multiple logistic regression.

[5] The stimuli consisted of nine 100-ms segments. For these stimuli, N_5 corresponds to the loudness of the segment with the highest sound pressure level because this loudness is reached during one-ninth of the stimulus duration.

[6] We are grateful to Pieter Jan Stallen for suggesting this explanation.

[7] We thank an anonymous reviewer for suggesting the potential applications discussed in this paragraph.

Agresti, A. (**2002**). *Categorical Data Analysis* (Wiley, New York).
Ahumada, A., and Lovell, J. (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.
Anderson, J. R., Bothell, D., Lebiere, C., and Matessa, M. (**1998**). "An integrated theory of list memory," J. Mem. Lang. **38**, 341–380.
Bamber, D. (**1975**). "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," J. Math. Psychol. **12**, 387–415.
Berg, B. G. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.
Berglund, B., Berglund, U., and Lindvall, T. (**1976**). "Scaling loudness, noisiness, and annoyance of community noises," J. Acoust. Soc. Am. **60**, 1119–1125.
Copas, J. (**1989**). "Unweighted sum of squares test for proportions," Appl. Stat. **38**, 71–80.
Dornic, S., and Laaksonen, T. (**1989**). "Continuous noise, intermittent noise and annoyance," Percept. Mot. Skills **68**, 11–18.
Ellermeier, W., and Schrödl, S. (**2000**). "Temporal weights in loudness summation," in *Fechner Day 2000. Proceedings of the 16th Annual Meeting of*

the International Society for Psychophysics, edited by C. Bonnet (Université Louis Pasteur, Strasbourg, France), pp. 169–173.
Fredrickson, B. L., and Kahneman, D. (**1993**). "Duration neglect in retrospective evaluations of affective episodes," J. Pers. Soc. Psychol. **65**, 45–55.
Grimm, G., Hohmann, V., and Verhey, J. L. (**2002**). "Loudness of fluctuating sounds," Acta. Acust. Acust. **88**, 359–368.
Han, L. A., and Poulsen, T. (**1998**). "Equivalent threshold sound pressure levels for Sennheiser HDA 200 earphone and Etymotic Research ER-2 insert earphone in the frequency range 125 Hz to 16 kHz," Scand. Audiol. **27**, 105–112.
Hellman, R. P. (**1982**). "Loudness annoyance, and noisiness produced by single-tone-noise complexes," J. Acoust. Soc. Am. **72**, 62–73.
Hellman, R. P. (**1984**). "Growth rate of loudness, annoyance, and noisiness as a function of tone location within the noise spectrum," J. Acoust. Soc. Am. **75**, 209–218.
Hellman, R. P. (**1985**). "Perceived magnitude of two-tone-noise complexes: Loudness, annoyance, and noisiness," J. Acoust. Soc. Am. **77**, 1497–1504.
Hellman, R. P., and Meiselman, C. H. (**1988**). "Prediction of individual loudness exponents from cross-modality matching," J. Speech Hear. Res. **31**, 605–615.
Hellman, R. P., and Zwislocki, J. (**1961**). "Some factors affecting the estimation of loudness," J. Acoust. Soc. Am. **33**, 687–694.
Hiramatsu, K., Takagi, K., and Yamamoto, T. (**1983**). "Experimental investigation on the effect of some temporal factors of nonsteady noise on annoyance," J. Acoust. Soc. Am. **74**, 1782–1793.
Hosmer, D. W., Hosmer, T., leCessie, S., and Lemeshow, S. (**1997**). "A comparison of goodness-of-fit tests for the logistic regression model," Stat. Med. **16**, 965–980.
Hosmer, D. W., and Lemeshow, S. (**2000**). *Applied Logistic Regression*, 2nd ed. (Wiley, New York).
Huynh, H., and Feldt, L. S. (**1976**). "Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs," J. Educ. Stat. **1**, 69–82.
IEC 318 (**1970**). "An IEC artificial ear, of the wide band type, for the calibration of earphones used in audiometry" (International Electrotechnical Commission, Geneva).
Jones, D. M., Macken, W. J., and Nicholls, A. P. (**2004**). "The phonological store of working memory: Is it phonological and is it a store?," J. Exp. Psychol. Learn. Mem. Cogn. **30**, 656–674.
Kortekaas, R., Buus, S., and Florentine, M. (**2003**). "Perceptual weights in auditory level discrimination," J. Acoust. Soc. Am. **113**, 3306–3322.
Kryter, K. D. (**2007**). "Acoustical sensory, and psychological research data and procedures for their use in predicting effects of environmental noises," J. Acoust. Soc. Am. **122**, 2601–2614.
Kuss, O. (**2001**). "A SAS/IML Macro for goodness-of-fit testing in logistic regression models with sparse data," in Proceedings of the 26th Annual SAS Users Group International Conference, Paper No. 265–226.
Kuss, O. (**2002**). "Global goodness-of-fit tests in logistic regression with sparse data," Stat. Med. **21**, 3789–3801.
Kuwano, S., and Namba, S. (**2000**). "Psychological evaluation of temporally varying sounds with L_{Aeq} and noise criteria in Japan," J. Acoust. Soc. Jpn. (E) **21**, 319–322.
Leventhall, H. G. (**2004**). "Low frequency noise and annoyance," Noise Health **6**, 59–72.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lutfi, R. A., and Jesteadt, W. (**2006**). "Molecular analysis of the effect of relative tone level on multitone pattern discrimination," J. Acoust. Soc. Am. **120**, 3853–3860.
Marquis-Favre, C., Premat, E., and Aubrée, D. (**2005a**). "Noise and its effects—A review on qualitative aspects of sound. Part II: Noise and annoyance," Acta. Acust. Acust. **91**, 626–642.
Marquis-Favre, C., Premat, E., Aubrée, D., and Vallet, M. (**2005b**). "Noise and its effects—A review on qualitative aspects of sound. Part I: Notions and acoustic rating," Acta. Acust. Acust. **91**, 613–625.
McFarland, D. J., and Cacace, A. T. (**1992**). "Aspects of short-term acoustic recognition memory: Modality and serial position effects," Audiology **31**, 342–352.
Michaud, D. S., Keith, S. E., and McMurchy, D. (**2008**). "Annoyance and disturbance of daily activities from road traffic noise in Canada," J. Acoust. Soc. Am. **123**, 784–792.
Moore, B. C. J., Vickers, D. A., Baer, T., and Launer, S. (**1999**). "Factors affecting the loudness of modulated sounds," J. Acoust. Soc. Am. **105**, 2757–2772.
Namba, S., and Kuwano, S. (**1979**). "An experimental study on the relation between long-term annoyance and instantaneous judgment of level-fluctuation sound," in *Proceedings Inter-noise '79*, edited by S. Czarnecki (Institute of Fundamental Technological Research Polish Academy of Sciences, Warsaw, Poland), Vol. **II**, pp. 837–842.
Namba, S., and Kuwano, S. (**1980**). "The relation between overall noisiness and instantaneous judgment of noise and the effect of background noise level on noisiness," J. Acoust. Soc. Jpn. (E) **1**, 99–106.
Neath, I., Brown, G. D. A., McCormack, T., Chater, N., and Freeman, R. (**2006**). "Distinctiveness models of memory and absolute identification: Evidence for local, not global, effects," Q. J. Exp. Psychol. **59**, 121–135.
Oberfeld, D. (**2008a**). "Does a rhythmic context have an effect on perceptual weights in auditory intensity processing?," Can. J. Exp. Psychol. **62**, 24–32.
Oberfeld, D. (**2008b**). "Temporal weighting in loudness judgments of time-varying sounds containing a gradual change in level," J. Acoust. Soc. Am. **123**, 3307.
Oberfeld, D., and Plank, T. (**2005**). "Temporal weighting of loudness: Effects of a fade in," in *Fortschritte der Akustik (Advances in Acoustics)—DAGA '05*, edited by Deutsche Gesellschaft für Akustik (DEGA, Berlin), pp. 227–228.
Pedersen, B., and Ellermeier, W. (**2008**). "Temporal weights in the level discrimination of time-varying sounds," J. Acoust. Soc. Am. **123**, 963–972.
Plank, T. (**2005**). "Auditive Unterscheidung von zeitlichen Lautheitsprofilen (Auditory discrimination of temporal loudness profiles),"Ph.D. thesis, Universität Regensburg, Regensburg, Germany.
Postman, L., and Phillips, L. W. (**1965**). "Short-term temporal changes in free-recall," Q. J. Exp. Psychol. **17**, 132–138.
Redelmeier, D. A., and Kahneman, D. (**1996**). "Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures," Pain **66**, 3–8.
Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.
Schomer, P. D., Suzuki, Y., and Saito, F. (**2001**). "Evaluation of loudness-level weightings for assessing the annoyance of environmental noise," J. Acoust. Soc. Am. **110**, 2390–2397.
Schreiber, C. A., and Kahneman, D. (**2000**). "Determinants of the remembered utility of aversive sounds," J. Exp. Psychol. Gen. **129**, 27–42.
Suied, C., Susini, P., and McAdams, S. (**2008**). "Evaluating warning sound urgency with reaction time," J. Exp. Psychol., Appl. **14**, 201–212.
Surprenant, A. M. (**2001**). "Distinctiveness and serial position effects in tonal sequences," Percept. Psychophys. **63**, 737–745.
Swets, J. A. (**1986**). "Indices of discrimination or diagnostic accuracy: Their ROCs and implied models," Psychol. Bull. **99**, 100–117.
Tang, Z., Richards, V. M., and Shih, A. (**2005**). "Comparing linear regression models applied to psychophysical data," J. Acoust. Soc. Am. **117**, 2597.
WHO (**1999**). "Guidelines for community noise," edited by B. Berglund, T. Lindvall, and D. H. Schwela, World Health Organization, Geneva, http://www.who.int/docstore/peh/noise/giudelines2.html (Last viewed 8/20/08).
Widmann, U. (**1994**). "Zur Lästigkeit von amplitudenmoduliertem Breitbandrauschen (The annoyance of amplitude modulated broadband noises)," in *Fortschritte der Akustik (Advances in Acoustics)—DAGA '94, Aachen*, edited by Deutsche Gesellschaft für Akustik (DEGA, Berlin), pp. 1121–1124.
Zhang, C., and Zeng, F.-G. (**1997**). "Loudness of dynamic stimuli in acoustic and electric hearing," J. Acoust. Soc. Am. **102**, 2925–2934.
Zimmer, K., and Ellermeier, W. (**1996**). "Construction and evaluation of a noise-sensitivity questionnaire," in *Recent Trends in Hearing Research: Festschrift for Seiichiro Namba*, edited by H. Fastl (BIS, Bibliotheks- und Informationssystem der Universität Oldenburg), pp. 163–170.
Zwicker, E. (**1966**). "Ein Beitrag zur unterscheidung von Lautstärke und Lästigkeit (A contribution to the distinction between loudness and annoyance)," Acustica **17**, 22–25.
Zwicker, E. (**1991**). "Ein Vorschlag zur Definition und zur Berechnung der unbeeinflussten Lästigkeit (A proposal for defining and calculating the unbiased annoyance)," Zeitschrift für Lärmbekämpfung **38**, 91–97.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics—Facts and Models* (Springer, Berlin).

What makes a melody: The perceptual singularity of pitch sequences

Marion Cousineau[a)]
Laboratoire de Psychologie de la Perception (UMR CNRS 8158), Université Paris-Descartes and Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, F-75230 Paris Cedex 05, France

Laurent Demany
Laboratoire Mouvement, Adaptation, Cognition (UMR CNRS 5227) BP 63, Université de Bordeaux, 146 Rue Leo Saignat, F-33076 Bordeaux, France

Daniel Pressnitzer
Laboratoire de Psychologie de la Perception (UMR CNRS 8158), Université Paris-Descartes and Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, F-75230 Paris Cedex 05, France

(Received 5 January 2009; revised 25 September 2009; accepted 28 September 2009)

This study investigated the ability of normal-hearing listeners to process random sequences of tones varying in either pitch or loudness. Same/different judgments were collected for pairs of sequences with a variable length (up to eight elements) and built from only two different elements, which were 200-ms harmonic complex tones. The two possible elements of all sequences had a fixed level of discriminability, corresponding to a d' value of about 2, irrespective of the auditory dimension (pitch or loudness) along which they differed. This made it possible to assess sequence processing per se, independent of the accuracy of sound encoding. Pitch sequences were found to be processed more effectively than loudness sequences. However, that was the case only when the sequence elements included low-rank harmonics, which could be at least partially resolved in the auditory periphery. The effect of roving and transposition was also investigated. These manipulations reduced overall performance, especially transposition, but an advantage for pitch sequences was still observed. These results suggest that automatic frequency-shift detectors, available for pitch sequences but not loudness sequences, participate in the effective encoding of melodies.
 [DOI: 10.1121/1.3257206]

PACS number(s): 43.66.Mk, 43.66.Hg [BCM] Pages: 3179–3187

I. INTRODUCTION

To understand speech or appreciate music, one has to process sequences of sounds. Among the diverse auditory features that may vary between sounds of a sequence, pitch seems to have a peculiar importance for human listeners. Pitch sequences constitute melodies and play a major role in music (Dowling and Harwood, 1986). Melodies have also been the example of choice used by Gestalt psychologists to illustrate the emergence of form out of the combination of discrete elements (Wertheimer, 1924).

Do sequences of pitch truly have special perceptual properties, compared to sequences of other auditory features such as loudness or timbre? Sequences of pitch have the ability to convey contour and relative interval information (Dowling and Fujitani, 1971), a property extensively used in music. Moore and Rosen (1979) suggested that familiar melodies cannot be recognized when the pitch changes are replaced by equivalent loudness changes. However, this suggestion is at odds with the results of a more recent study by McDermott *et al.* (2008). McDermott *et al.* (2008) argued that contour perception is not specific to pitch. They found that contours in the dimension of either pitch, loudness, or timbre can be recognized cross dimensionally. They contended that pitch is particularly useful to construct melodies merely because this perceptual feature is encoded very accurately over a wide range.

McFarland and Cacace (1992) reached an opposite conclusion. They employed pure-tone sequences built from only two possible values of a given feature, separated by a fixed number of just-noticeable differences. This was intended to equate the discriminability of individual elements of the sequences, in order to reveal mechanisms of sequence processing that otherwise might have been obscured by the peculiarities of feature encoding for single sounds. An advantage for pitch sequences was found: listeners could recognize longer pitch sequences than loudness or duration sequences. There are, however, potential concerns about the experimental procedure. First, it was assumed that a fixed number of just-noticeable differences would produce equal discriminability. This hypothesis is supported by introspective data (Terhardt, 1968), but has not been confirmed by performance data. Second, a large number of elements had to be memorized on each trial, so that the task presumably recruited high-level cognitive strategies (Cowan, 2001). The important question of a possible sensory advantage for pitch-sequence processing thus remains unresolved.

In the present study, we reassessed the processing of pitch sequences and loudness sequences. Binary sequences

a)Author to whom correspondence should be addressed. Electronic mail: marion.cousineau@ens.fr

of tones were constructed and the discriminability between the two possible component tones was equated in terms of the d' sensitivity index of signal detection theory (Green and Swets, 1966). Each sequence contained at most eight elements, which were complex tones. The tones consisted of harmonics with variable ranks, so that in some sequences the spectral components of the tones were completely resolvable in the auditory periphery, whereas in other sequences the spectral components were completely unresolvable. These two kinds of complex tones convey musical pitch (Moore and Rosen, 1979), but with a different accuracy (Kaernbach and Bering, 2001) and perhaps by means of different mechanisms (de Cheveigné, 2005). We wished to determine whether the advantage of pitch over loudness suggested by McFarland and Cacace (1992) for pure tones would generalize to complex tones of either type when discriminability is equalized.

II. EXPERIMENT 1: SEQUENCES OF PITCH AND LOUDNESS

In this experiment, sequence processing was evaluated using tones differing in (1) loudness (condition L), (2) pitch produced by resolved harmonics (condition P-R), and (3) pitch produced by unresolved harmonics (condition P-U). In order to dissociate single-sound encoding from sequence processing per se, we equalized the discriminability index d' between elements of the sequences in these three conditions.

A. Method

The sequence elements (see Fig. 1) were complex tones with a fundamental frequency (F0) close to 125 Hz. The tones were bandpass-filtered click trains (eighth order Butterworth filters). The pass-band of the filter was set between 125 and 625 Hz in the P-R and L conditions, and between 3900 and 5400 Hz in the P-U condition. Identical tones had been used by Shackleton and Carlyon (1994). Based on frequency difference limens and phase sensitivity measures, Shackleton and Carlyon (1994) argued that spectral components would be resolved in the P-R condition and unresolved in the P-U condition. In order to mask distortion products that could affect the internal spectrum of the tones in the P-U condition (Pressnitzer and Patterson, 2001), the tones were mixed with pink noise in all conditions. The pink noise was generated in the spectral domain with components between 62.5Hz and half the sampling rate. For each stimulus, the overall level of the noise was set at 6 dB below the overall level of the tone; then, the sound pressure level (SPL) of the tone plus noise compound was set close to 65 dB. Each stimulus had a duration of 200 ms and was gated on and off with 25-ms raised-cosine ramps.

In a preliminary adjustment phase, listeners had to perform a same/different task on two tones with variable differences in F0 or SPL. The reference tone had an F0 of 125 Hz and a SPL of 65 dB, while the other tone was higher in either F0 or SPL. In a given block of trials, only one tone was used in addition to the reference tone; each of the two tones to be compared on a given trial was randomly chosen to be the reference or the other tone. For each condition (P-R, P-U, or L) and listener, several adjustment blocks were run to estimate the stimulus difference Δ yielding a d' value of 2. The Δ values tested as well as the number of adjustment blocks were determined heuristically by the experimenter. Note, however, that this adjustment was independently verified in the main part of the experiment.

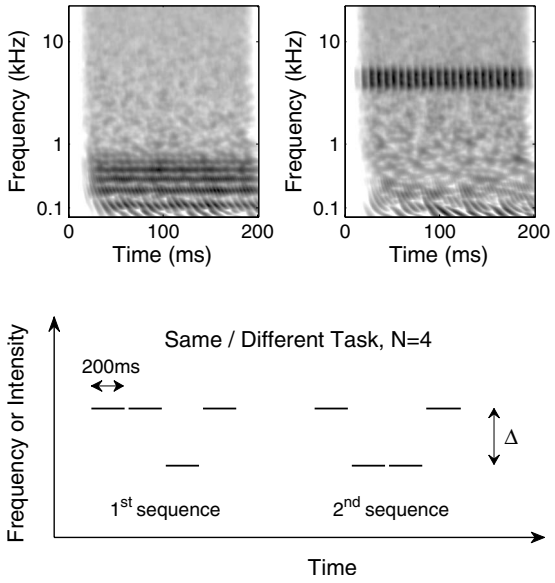


FIG. 1. Top panels: Stimuli of Experiment 1. A simulation of peripheral auditory processing (Patterson *et al.*, 1995) illustrates the predicted activation of different frequency channels after cochlear encoding. (Top-left) When the tones were filtered in a low frequency region (125–625 Hz), their spectral components were processed in independent channels and thus "resolved" (conditions P-R and L). (Top-right) When the tones were filtered in a high frequency region (3900–5400 Hz), their spectral components were unresolved (condition P-U). (Bottom panel) Experimental task. Listeners made same/different judgments on binary sequences. The elements of the sequences could take only two values of either F0 (conditions P-R and P-U) or SPL (condition L). These two values differed by Δ. The number of elements, N, could be equal to 1, 2, 4, or 8. On a given trial, only one element (chosen at random) could change. In this example of a "different" trial, N is equal to 4 and the second element changes.

In the subsequent and main part of the experiment, the previously determined values of Δ were used in binary sequences of $N=1$, 2, 4, or 8 tones. On each trial, two sequences separated by a 400-ms silence were presented (see Fig. 1, bottom panel). In the first sequence, each tone was, at random, either a reference stimulus, A, with a 125-Hz F0 and a 65-dB level, or another stimulus, B, differing positively from A by Δ. The second sequence was equiprobably identical to the first sequence or different from it. In the latter case, a single, randomly chosen tone was changed from A to B or vice versa. The listener had to make a same/different judgment. For each listener, condition, and N value, four blocks of 50 trials were run. The ordering of conditions and N values was randomized. No training was provided apart from the series of adjustment blocks, which corresponds to sequences with $N=1$.

Thirteen listeners with no self-reported hearing disorder participated in the experiment (mean age=24.0, SD=6.1, six female). A questionnaire was used to evaluate musical training. The measure used was the number of years participants had practiced a musical instrument (mean=6.9, SD=7.2). For displaying individual data, participants were sorted in

TABLE I. Means and standard deviations of the Δ values used in Experiment 1. These values are expressed as percentages of the reference frequency or sound pressure. The values in parentheses are conversions to semitones (P-R or PU conditions) or dB (L condition).

Experiment 1			
Condition	P-R	P-U	L
Reference	125 Hz	125 Hz	65 dB
Mean Δ	1.9 (0.3)	38.8 (5.7)	49.6 (3.5)
s.d.	0.9 (0.2)	29.6 (4.5)	37.3 (2.8)

three groups: no musical training (4), less than 10 years (5), and more than 10 years (4). Stimuli were generated with an RME Fireface audio sound card and digital to analog converter, with 16-bit coding accuracy and 44.1 kHz sampling rate. They were delivered diotically by means of Sennheiser HD 250 linear II headphones. Listeners were seated in a double-walled sound insulated booth (IAC). Responses were given by means of button press. No feedback was provided.

B. Results

Table I displays the mean results of the adjustment phase and shows that Δ was much higher in the P-U condition than in the P-R condition. This is consistent with many previous data showing that the pitch percepts evoked by unresolved harmonics are less precise than those evoked by resolved harmonics (see, e.g., Shackleton and Carlyon, 1994; Houtsma and Smurzynski, 1990). In the L condition, the average Δ value was 3.5 dB. If relative thresholds are compared across attributes, in terms of percent of the reference, Δ values for loudness were larger than Δ values in the two pitch conditions. A significant correlation was observed between musical expertise (years of musical training) and Δ values for the P-R condition ($r=-0.80$, $t=-4.44$, $P=0.001$). Consistent with previous reports (Micheyl *et al.*, 2006), musicians displayed smaller thresholds. Similar correlations were observed between musical expertise and Δ values in the P-U condition ($r=-0.70$, $t=-3.29$, $P=0.007$) and Δ values in the L condition ($r=-0.66$, $t=-2.90$, $P=0.014$).

The top-left panel of Fig. 2 shows the effect of N and condition on d'. For $N=1$, d' was similar in all three conditions, indicating that the preliminary adjustment phase had been successful. Performance was higher than 2, suggesting equivalent improvement between adjustment and test phase in all conditions. When N was greater, however, a large difference was observed between conditions. For conditions L and P-U, almost identical trends were obtained: performance steadily decreased as soon as N exceeded 1. For P-R, in contrast, performance was approximately constant up to $N=4$ and decreased only when N reached its highest value, 8.

A repeated-measures analysis of variance (ANOVA) ($N\times$condition) confirmed the existence of a significant interaction between the two experimental factors [$F(6,72)=4.95$, $P=0.0002$], in addition to main effects of N [$F(3,36)=12.96$, $P<0.0001$] and condition [$F(2,24)=6.45$, $P=0.0057$]. *Post-hoc* tests (Tukey's HSD) confirmed that, for P-R, performance decreased only when N varied from 4 to 8.

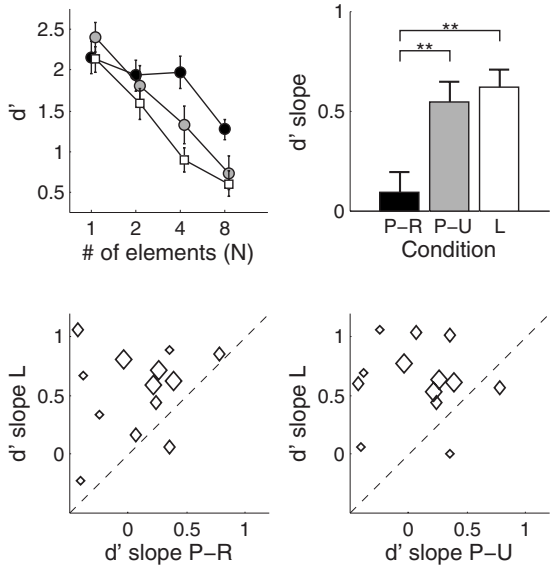


FIG. 2. Results of Experiment 1. (Top-left) Mean value of d' as a function of N in the three experimental conditions (P-R: black circles; P-U: gray circles; L: white squares). Error bars are standard errors of the means. (Top-right) Mean and standard error of the d' slope, summarizing the decrease in performance from $N=1$ to $N=4$. (Bottom-left) Individual data for the d' slope in the P-R and P-U conditions. Each diamond represents one listener. The size of the diamonds represents the musical experience of listeners: small for no experience, medium for less than 10 years, and large for more than 10 years. (Bottom-right) Same, but for the P-U and L conditions.

Figure 2 and the *post-hoc* tests indicated that, between $N=1$ and $N=4$, d' remained constant in the P-R condition but decreased regularly in the P-U and L conditions. We performed an additional analysis to summarize these patterns: Straight lines were fitted to the individual data obtained for $N=1$, 2 and 4, using a log-scale for N. The slope of the fitted lines characterizes the initial effect of N on performance, while normalizing for possible individual differences in performance at $N=1$. The top-right panel of Fig. 2 displays the results of this analysis. The slope found for P-R was small and differed significantly from those found for both P-U and L ($P<0.01$), whereas slopes for P-U and L did not differ reliably from each other ($P>0.10$). The bottom panels of Fig. 2 display individual results for the d' slope. All but one listener displayed a shallower slope for P-R than P-U, that is, an advantage for P-R sequences. Some listeners even displayed a negative slope for P-R, indicating that their performance actually increased between sequences of $N=1$ and $N=4$ elements. In contrast, results seem evenly distributed when P-U and P-L are compared.

There was no obvious effect of musical training on the magnitude of the pitch advantage, as indicated by the individual data plots. The correlation between musical expertise and the magnitude of the pitch advantage (d' slope for loudness minus d' slope for pitch) was not significant ($r=-0.20$, $t=-0.69$, $P=0.51$).

In summary, Experiment 1 demonstrated superior processing of sequences for pitch compared to loudness, but only when resolved harmonics were present.

III. IDEAL OBSERVER SIMULATION

Within the framework of signal detection theory, it is possible to simulate an ideal observer required to make

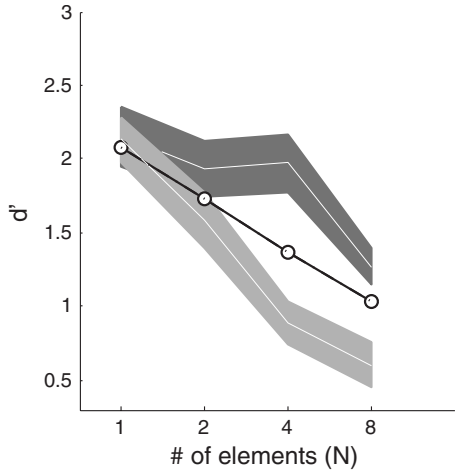


FIG. 3. Ideal observer simulation. The solid curve shows the prediction derived from an ideal observer simulation based on independent processing of single elements. The experimental data for conditions P-R (dark gray) and L (light gray), Experiment 1, are replotted as shaded areas encompassing two standard errors of the mean.

same/different judgments on two elements (Green and Swets, 1966; Dai *et al.*, 1996). Observation variables are associated with each element, taking into account internal noise. If the two observations are independent, the optimal strategy is to compare each of them with a fixed criterion to determine which distribution they belong to and then take the same/different decision (Dai *et al.*, 1996).

Our experimental task was not a comparison between two elements, but rather a comparison between sequences comprising several elements. If no specific sequence-processing mechanism is assumed, optimal performance on this task may be simulated by repeating the ideal observer's strategy on corresponding elements across sequences, under the assumption that the elements are encoded independently of each other and with perfect memory.

We implemented such a model to quantify performance for N independent comparisons. Binary sequences of A and B elements were generated as in Experiment 1. Each element was transformed into its observation variable by combining its true value, 0 or 1, with a Gaussian noise of zero mean and standard deviation σ. No bias was assumed, so the criterion was set to 0.5 (note that changing the criterion would not affect the obtained d'). For each element, the observation variable was compared to the criterion and an A or B decision was taken. All corresponding pairs of observed elements, matched across the two sequences, were then compared and a "same" decision was taken if all pairs were identical. The magnitude of the noise, σ, was the only free parameter of the model. It was fitted to the experimental data for $N=1$.

For each value of N, we computed d' from 100 000 simulated trials. The simulation results are displayed in Fig. 3, together with the data of Experiment 1 (for simplicity, we omit condition P-U). As N increased, the predicted d' decreased less rapidly than in condition L but more rapidly than in condition P-R. Further simulations showed that changing σ would shift overall performance but would not change the predicted slope of the function relating d' to N. We compared, by means of student t tests, the d' slopes of the experimental data (Fig. 2) to the predicted slope. It appeared that the model significantly outperformed listeners for condition L but was significantly outperformed by listeners for condition P-R ($P<0.025$, one-tailed tests).

That the ideal observer model outperformed listeners in condition L can be easily accounted for, by assuming for instance that listeners were unable to achieve optimal performance because their memory was not perfect. In condition P-R, however, listeners were more efficient than the ideal observer. This seemingly paradoxical finding shows that the most important assumption of the model does not hold: the high performance of listeners in the P-R condition cannot be based on independent processing of the elements of the sequences. Instead, listeners had to use sequence-specific mechanisms in this condition.

TABLE II. Means and standard deviations of the Δ values used in Experiment 2. These values are expressed as percentages of the reference frequency. The values in parentheses are conversions to semitones.

Experiment 2			
Condition	P-F0high	P-F0mid	P-F0low
Reference	250 Hz	125 Hz	62.5 Hz
Mean Δ	5.8 (1.0)	9.6 (1.6)	32.3 (4.9)
s.d.	9.2 (1.5)	9.6 (1.6)	22 (3.4)

IV. EXPERIMENT 2: EFFECT OF FREQUENCY REGION AND HARMONIC RESOLVABILITY

In Experiment 1, complex tones were filtered in different frequency regions to produce resolved or unresolved harmonics. Resolvability was therefore confounded with frequency region, which could have influenced pitch processing by itself (Meddis and O'Mard, 1997; Pressnitzer *et al.*, 2001). Experiment 2 controlled for this possibility by manipulating resolvability within a fixed frequency region.

A. Method

As in Experiment 1, band-pass filtered click trains were used. The pass-band was fixed here between 1375 and 1875 Hz. There were three pitch conditions, with reference F0s of 250, 125, and 62.5 Hz. They were, respectively, termed P-F0high, P-F0mid, and P-F0low. Again, our choice of the stimulus parameters was motivated by observations of Shackleton and Carlyon (1994). They found low frequency difference limens for P-F0high, suggesting that spectral components would be resolved in this case. In contrast, difference limens were poor for P-F0low, suggesting that spectral components would be unresolved in this case. Results for P-F0mid were intermediate, with some variability across subjects.

The tones were mixed with pink noise and all other experimental details were as in Experiment 1, except that here only two values of N were used: 1 and 4. The Δ values are displayed in Table II. Five listeners with no self-reported hearing disorder participated (mean age=24.4, SD=3.4, three female). Musical training was evaluated as in Experiment 1 and quantified by the number of years of musical

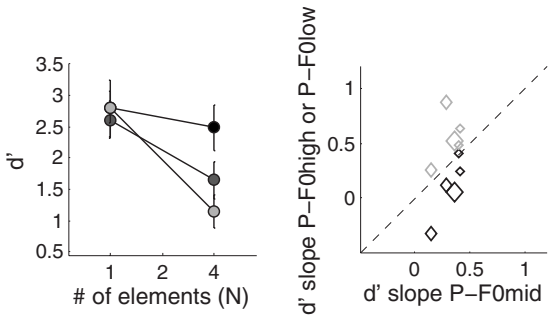


FIG. 4. Results of Experiment 2. (Left) Mean and standard error of d' for each F0 condition as a function of N (P-F0high: black circles; P-F0mid: dark gray circles; P-F0low: light gray circles). (Right) Individual data for the d' slope in the P-F0mid condition, plotted against d' slope for P-F0high (black diamonds) or d' slope for P-F0low (light gray diamonds). The size of the markers indicates musical experience as in Fig. 2.

practice (mean=4.8, SD=6.2; two participants with no musical training, two with less than 10 years, and one with more than 10 years).

B. Results and discussion

The left panel of Fig. 4 displays the mean results. A repeated-measures ANOVA revealed a significant interaction between N and F0 [$F(2,8)=10.77$, $P=0.0053$], in addition to main effects of N [$F(1,4)=17.84$, $P=0.013$] and F0 [$F(2,8)=6.19$, $P=0.023$]. The hypothesis tested by the experiment was that performance would be degraded in the case of unresolved harmonics. To test for this, *post-hoc* analyses (Fisher LSD) were performed. For $N=4$, performance was significantly better for P-F0high than for P-F0low ($P=0.00018$); performance for P-F0mid was intermediate, significantly poorer than for P-F0high ($P=0.0035$) and significantly better than for P-F0low ($P=0.041$). No significant differences were found across conditions for $N=1$. It can also be seen in Fig. 4 that when N varied from 1 to 4, performance was essentially unchanged for P-F0high but decayed in the other two conditions.

Individual d' slopes for this experiment are shown in the right panel of Fig. 4. For most listeners, the d' slopes were larger for P-F0mid than for P-F0high, and smaller for P-F0mid than for P-F0low. There was no obvious effect of musicianship on the pattern of results.

In Experiments 1 and 2, performance with pitch sequences appears to be determined by the rank of the lowest harmonic present in the pass-band of the stimulus (R_{lh}) rather than by the absolute frequency region. Poor performance was observed in Experiment 1 for a high frequency region (lower limit: 3.9 kHz) and a R_{lh} of 32. A similarly poor performance was observed in Experiment 2 for a medium frequency range (lower limit: 1.375 kHz) but a R_{lh} which was again very high (22). In Experiment 2, performance was good even when relatively high harmonics were used (P-F0high condition, $R_{lh}=6$). Estimates of resolution based on hearing out individual harmonics would fail with such stimuli (Plomp, 1964; Moore, 1973). However, a large number of studies based on frequency difference limens find a transition region from good to poor performance between the 10th and 13th harmonics (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003; de Cheveigné and Pressnitzer, 2006). It is therefore likely that at least some of the harmonics of the P-F0high condition were resolved. For the P-F0mid condition, R_{lh} was 11, which is within the transition region for estimates of resolvability based on difference limens (Shackleton and Carlyon, 1994). The small but significant pitch advantage that we observed in this condition could be due to partial resolvability of the lower harmonics of the stimuli.

Moore and Rosen (1979) showed that listeners are able to recognize familiar melodies produced by varying the pitch of unresolved complexes. The condition in which they observed best performance (2 kHz highpass filtering and F0 between 100 and 200 Hz) resembles the P-F0mid condition of the present experiment, for which we also found evidence of an advantage for pitch sequences. In another condition which resembles the P-U condition of Experiment 1 (4 kHz highpass filtering), Moore and Rosen (1979) found that recognition performance degraded significantly. The two sets of results are thus consistent if one considers the different possible definitions of resolvability: pitch sequences may be processed efficiently when composed of tones with harmonics that cannot be heard out individually (Plomp, 1964) but for which there is evidence of at least partial peripheral resolvability (Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003).

In summary, here and in Experiment 1, performance with pitch sequences was better when at least some of the harmonics of the stimuli were partially resolved than when all harmonics were clearly unresolved.

V. EXPERIMENT 3: EFFECT OF FREQUENCY AND AMPLITUDE ROVE

The previous experiments used the same two elements for all sequences, which were presented in blocks of 50 contiguous trials. It is conceivable that listeners could have memorized the absolute values of each attribute within the block, which could in turn interfere with sequence processing. We tested for this possibility by applying a random rove on the dimension tested within each block. From trial to trial, the reference value of F0 for pitch sequences and SPL for loudness sequences was randomly varied. The two sequences within a trial still had identical reference tones, as in Experiment 1.

A. Method

The sequences were similar to those of Experiment 1. However, only conditions P-R and L were used and the maximum value of N was 4. We chose to adjust the Δ value for each condition and listener without any rove (the Δ values are displayed in Table III). Then, a first series of four blocks of 50 trials was run with the Δ value selected and $N=1$, for each listener and condition, in order to check for equal discriminability without rove. In subsequent blocks, roving was applied. For pitch sequences, the reference F0 on a given trial was randomly chosen between 125 Hz and 125 Hz+2Δ. For loudness sequences, the reference SPL was randomly chosen between 65 dB and 65 dB+2Δ. The

TABLE III. Means and standard deviations of the Δ values used in Experiments 3 and 4. These values are expressed as percentages of the reference frequency or sound pressure. The values in parentheses are conversions to semitones (P-R condition) or dB (L condition).

Experiments 3 and 4		
Condition	P-R	L
Reference	125 Hz	65 dB
Mean Δ	1.27 (0.2)	28.16 (2.2)
s.d.	0.38 (0.1)	6.74 (0.6)

amount of roving was thus aimed to be comparable across attributes and listeners. We measured discriminability for roved sequences with $N=1$, 2, and 4 elements in interleaved blocks of 50 trials (four blocks per condition and listener). Six listeners with no self-reported hearing impairment participated (mean age=24.0, SD=2.2, four female). Musical training was evaluated as in Experiment 1 and quantified by the number of years of musical practice (mean=9.7, SD =8.5; one participant with no musical training, three with less than 10 years, and two with more than 10 years). All listeners also participated in Experiment 4 (described below). Half of the subjects started by Experiment 3 and the other half by Experiment 4.

B. Results and discussion

The left panel of Fig. 5 displays the mean results. The preliminary adjustment phase was satisfactory: a paired sample t-test revealed no significant difference between P-R and L for $N=1$, no rove ($P=0.5$). For the data with rove, performance was globally lower than in Experiment 1, where no rove was applied. The pattern of the results, however, was highly similar, with an advantage of P-R over L as sequences increased in length. A repeated-measures ANOVA revealed a significant effect of N [$F(2,10)=6.86$, $P=0.003$] and a significant interaction between N and condition [$F(2,10)=7.29$, $P=0.01$]. The right panel of Fig. 5 displays the individual d' slopes. All but one listener displayed a pitch advantage. The largest pitch advantage was observed for a participant with no musical training, but large effects were also observed for musicians.

It is not surprising that roving should degrade overall performance. Discrimination tasks involving pitch (Harris, 1948; Demany and Semal, 2005; Ahissar *et al.*, 2006) or loudness (Oxenham and Buus, 2000) always result in poorer performance with roving than without. Many factors may explain this finding, including perceptual learning (Demany and Semal, 2002) or stimulus uncertainty (Watson *et al.*, 1976). Importantly, in our case, performance was equally degraded by roving for pitch and loudness. As a result, roving did not change the main features of the results and the advantage for pitch sequences was replicated. In Experiment 1, therefore, the possibility of memorizing the sequence elements across trials was not the cause of the observed advantage of pitch sequences over loudness sequences.

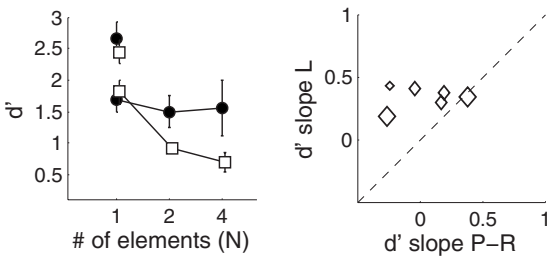


FIG. 5. Results of Experiment 3. (Left) Mean and standard error of d' in conditions P-R (black circles) and L (white squares) as a function of N, with rove. The disconnected symbols show mean and standard error of d' for N=1, with the same Δ values but without rove. (Right) Individual data for the d' slope in the P-R condition, plotted against d' slope for L. The size of the markers indicates musical experience as in Fig. 2.

VI. EXPERIMENT 4: EFFECT OF TRANSPOSITION

In all experiments reported so far, the two sequences presented on a given trial were built from the same pair of tone elements. Absolute cues for each attribute were thus potentially available within a trial. In musical situations, by contrast, melodies may be transposed; that is, the same sequence of pitch intervals may be presented with different starting frequencies (Dowling and Harwood, 1986). Listeners are able to recognize transposed melodies, but the task becomes increasingly difficult if the melody is unfamiliar and if contour cues are not affected (Dowling and Fujitani, 1971; Kidd and Watson, 1996). Recently, it has also been shown that listeners are able to recognize transposed loudness sequences (McDermott *et al.*, 2008). In this final experiment, we introduce transpositions in the pitch or the loudness domain, in order to investigate whether the pitch-sequence advantage remains with transposed material.

There is, however, a fundamental limitation with introducing transposition in our task: a same/different task with only one element and transposition is not possible. It is therefore impossible to control for performance at $N=1$, which was a strong prerequisite for each of the experiments reported above.

A. Method

The stimuli were similar to those of Experiments 1 and 3, but N had only two possible values: 2 and 4. Only conditions P-R and L were tested. As mentioned above, the subjects were the same as in Experiment 3. Δ also had the same values as in Experiment 3 (displayed in Table III). In the first sequence presented on each trial, the reference tone had an F0 of 125 Hz and a SPL of 65 dB. In the second sequence, the reference F0 was changed to 125 Hz+2Δ for condition P-R and the reference SPL was changed to 65 dB+2Δ for condition L. Again, this was aimed at introducing comparable amounts of transposition across listeners and conditions.

B. Results and discussion

Mean results are shown in the left panel of Fig. 6. Transposition clearly had a deleterious effect since performance was generally quite poor: d' was always close to 0.5, except when N was equal to 4 in condition P-R. A repeated-measures ANOVA revealed that the main effects of condition

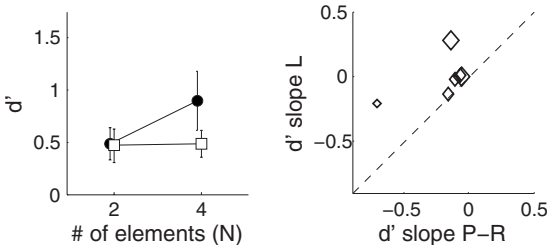


FIG. 6. Results of Experiment 4. (Left) Mean and standard error of d' in conditions P-R (black circles) and L (white squares) as a function of N. (Right) Individual data for the d' slope in the P-R condition, plotted against d' slope for L. The size of the markers indicates musical experience as in Fig. 2.

and N were not significant, but that the interaction between the two factors was marginally significant [$F(1,5)=5.5$, $P=0.06$]. *Post-hoc* tests (Fisher LSD) indicated that performance was significantly better in condition P-R than in condition L for $N=4$ ($P=0.017$), whereas there was no significant difference between the two conditions for $N=2$ ($P=0.87$).

Individual data are shown on the right panel of Fig. 6. Only two listeners out of six, a musician and a non-musician, showed a strong pitch advantage. For the other four listeners, the d' slopes were generally small. However, all six listeners consistently showed negative slopes in the pitch condition, which was not the case for loudness.

The fact that transposition made the comparison between pitch sequences more difficult is in line with previous data obtained with unfamiliar melodies (Dowling and Fujitani, 1971). Moreover, Kidd and Watson (1996) reported that the adverse effect of transposition was already substantial when the amount of transposition was modest and the transposition interval was constant from trial to trial, as in the present experiment. We also observed that longer sequences ($N=4$) showed improved performance compared to short sequences ($N=2$), corresponding to negative d' slopes. This is again consistent with results of Kidd and Watson (1996), who found better performance for changes embedded in sequences of five elements than for changes embedded in sequences of two elements. They interpreted this finding by remarking that a change in the pitch of a single tone modifies two consecutive pitch intervals if and only if the tone that changes is surrounded by two other tones. When only relative information is relevant (which is the case in the presence of a transposition), this predicts an increase in performance for sequences with more than two elements.

In our L condition, performance was close to the chance level for each value of N. This suggests that contour cues are not available for the discrimination of loudness sequences. McDermott *et al.* (2008) suggested the opposite, but in their study the loudness changes taking place from tone to tone were larger than in our study. It seems likely that human listeners are sensitive to *relative* loudness only for rather large loudness changes.

VII. GENERAL DISCUSSION

The following conclusions can be drawn from our data: (1) pitch sequences are processed more efficiently than loudness sequences, (2) the discriminability of pitch sequences is better than predicted from independent processing of their individual elements, (3) the perceptual advantage of pitch sequences over loudness sequences requires that resolved harmonics be available, (4) this advantage does not depend on the familiarity of the sequence elements, and (5) transposed sequences are less efficiently compared than non-transposed sequences, but transposition does not abolish the advantage of pitch sequences for sequences of more than two elements.

A first possible explanation for the advantage of pitch sequences would be that pitch has access to a specific and efficient short-term memory store. However, one key aspect of our data is not consistent with this interpretation. Here, resolved harmonics were required to support high performance, whereas short-term pitch memory seems to have the same temporal characteristics for resolved and unresolved harmonics when differences are adjusted in discriminability (Clément *et al.*, 1999). Also, listeners outperformed the ideal observer model which had perfect memory.

Another explanation is that superior performance for the pitch of resolved harmonics could be caused by a greater familiarity with such sequences compared to sequences of unresolved harmonics or loudness. In this case, musically experienced listeners who have received extensive training with pitch sequences should show a larger advantage. An effect of musicianship was observed in the initial discriminability adjustment step, but it was present for all dimensions tested. In the main experiment, when the discriminability of individual elements was factored out, musicians did not display any increased advantage for pitch-sequence processing.

A more likely explanation is that the advantage is based on the recruitment of an additional mechanism to encode pitch sequences. There are different strands of evidence for the existence of such a mechanism. Brain imaging studies showed that secondary auditory regions in the right hemisphere respond more strongly to melodies with pitch changes than to sequences of tones with a fixed pitch (Patterson *et al.*, 2002; Hyde *et al.*, 2008). Neuropsychological studies have shown that lesions lateralized in the right auditory cortex can impair the sensitivity to frequency-shift direction without impairing absolute frequency discrimination (Johnsrude *et al.*, 2000). Behavioral data indicate that listeners can consciously perceive an upward or downward pitch shift between two consecutive pure tones even when they did not hear out the first tone because it was fused in a complex chord (Demany and Ramos, 2005; Demany *et al.*, 2008). The latter result has been interpreted as evidence for automatic frequency-shift detectors (FSDs).

A mechanism based on FSDs can account for most of our findings. First, since FSDs detect shifts in frequency, they should not be activated by amplitude shifts, consistent with the results of Experiment 1. Second, the fact that FSDs identify by definition a *relation* between tonal elements is consistent with the refutation, by our ideal observer simulation, of the element-independence assumption for pitch sequences consisting of resolved harmonics. Third, behavioral evidence for the FSDs only exists for *spectral* shifts, that is,

shifts of resolved spectral components. The poorer performance we observed with stimuli consisting of unresolved components suggests that there are no equivalent periodicity-shift detectors (see also Demany and Semal, 2008).

The FSD hypothesis may provide a low-level basis for the long-standing observation that contour is an essential cue to melody recognition (White, 1960; Dowling and Fujitani, 1971). There are differences between our task and realistic musical situations, however. Because we aimed at equating strictly all aspects of the task that were not directly related to sequence processing, a single interval between elements of the sequences was used in all conditions and this interval was close to threshold. Within these constraints, we found evidence for a contour-extraction mechanism in pitch sequences only. Consistent with previous findings (Kidd and Watson, 1996), we also observed that even a modest amount of transposition had a strong deleterious effect on sequence discrimination. This may seem contradictory with a relative, contour-extraction mechanism, as well as with recent observations that pitch and loudness sequences can be recognized after transposition (McDermott *et al.*, 2008). McDermott *et al.* (2008) used acoustic changes not matched in discriminability and much larger than the ones used in our experiments. Using smaller changes, Moore and Rosen (1979) failed to find any evidence for contour recognition with loudness sequences. All of these observations may be reconciled if one assumes two distinct steps in any contour-matching task: (1) a sensory encoding stage, where absolute and relative cues may be pooled, followed by (2) a decision stage. If transposition introduces a fixed amount of noise in the decision stage, consistent with the observation of Kidd and Watson (1996) that transposition produces essentially the same impairment over a wide range (from 2 to 12 semitones), then this noise will swamp small differences at the sensory encoding stage, as in our experiment, but it will be overcome by large differences, as investigated by McDermott *et al.* (2008) and as used in realistic musical melodies. To test for this hypothesis, further experiments are required where the steps on each dimension are larger than in the present series of experiments, but still controlled for equal discriminability across dimensions.

Sensory encoding of pitch contour by means of FSDs may be one of the several mechanisms that "make" a melody, especially as it may occur without attention (Demany and Ramos, 2005). This hypothesis has important implications for hearing-impaired listeners. Cochlear implant users, for instance, hear through a device that directly stimulates the auditory nerve, but with a limited number of frequency channels that, to date, cannot transmit resolved harmonics. While speech intelligibility in quiet can be high, music listening is more challenging. In particular, melody recognition is poor (Kong *et al.*, 2004; Pressnitzer *et al.*, 2005; Cooper *et al.*, 2008). Providing the necessary cues for automatic contour encoding may help to improve music perception with such devices.

ACKNOWLEDGMENTS

We wish to thank Josh H. McDermott and Brian C. J. Moore for suggestions that led to the design of Experiments 3 and 4. Portions of these results were presented at the 155th Meeting of the Acoustical Society of America.

Ahissar, M., Lubin, Y., Putter-Katz, H., and Banai, K. (**2006**). "Dyslexia and the failure to form a perceptual anchor," Nat. Neurosci. **9**, 1558–1564.
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.
Clément, S., Demany, L., and Semal, C. (**1999**). "Memory for pitch versus memory for loudness," J. Acoust. Soc. Am. **106**, 2805–2811.
Cooper, W. B., Tobey, E., and Loizou, P. C. (**2008**). "Music perception by cochlear implant and normal hearing listeners as measured by the Montreal battery for evaluation of amusia," Ear Hear. **29**, 618–628.
Cowan, N. (**2001**). "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," Behav. Brain Sci. **24**, 87–114.
Dai, H., Versfeld, N. J., and Green, D. M. (**1996**). "The optimum decision rules in the same-different paradigm," Percept. Psychophys. **58**, 1–9.
de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch, Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, New York), pp. 169–233.
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.
Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: perceiving shifts in nonperceived pitches," J. Acoust. Soc. Am. **117**, 833–841.
Demany, L., and Semal, C. (**2002**). "Learning to perceive pitch differences," J. Acoust. Soc. Am. **111**, 1377–1388.
Demany L. and Semal C. (**2005**). "The slow formation of a pitch percept beyond the ending time of a short tone burst," Percept. Psychophys. **67**, 1376–1383.
Demany, L.. and Semal, C. (**2008**). "The role of memory in auditory perception," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 77–113.
Demany, L., Trost, W., Serman, M., and Semal, C. (**2008**). "Auditory change detection—Simple sounds are not memorized better than complex sounds," Psych. Sci. **19**, 85–91.
Dowling, W. J., and Fujitani, D. S. (**1971**). "Contour, interval, and pitch recognition in memory for melodies," J. Acoust. Soc. Am. **49**, 524–531.
Dowling, W. J., and Harwood, D. L. (**1986**). *Music Cognition* (Academic, Orlando, CA).
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Harris, J. (**1948**). "Discrimination of pitch: Suggestions toward method and procedure," Am. J. Psychol. **61**, 309–322.
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.
Hyde, K. L., Peretz, I., and Zatorre, R. J. (**2008**). "Evidence for the role of the right auditory cortex in fine pitch resolution," Neuropsychologia **46**, 632–639.
Johnsrude, I. S., Penhune, V. B., and Zatorre, R. J. (**2000**). "Functional specificity in the right human auditory cortex for perceiving pitch direction," Brain **123**, 155–163.
Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.
Kidd, G. R., and Watson, C. S. (**1996**). "Detection of frequency changes in transposed sequences of tones," J. Acoust. Soc. Am. **99**, 553–566.
Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.
McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2008**). "Is relative pitch specific to pitch?" Psych. Sci. **19**, 1263–1271.
McFarland, D. J., and Cacace, A. T. (**1992**). "Aspects of short-term acoustic recognition memory: Modality and serial position effects," Audiology **31**, 342–352.
Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.
Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training of pitch discrimination," Hear. Res. **219**, 36–47.
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., and Rosen, S. M. (**1979**). "Tune recognition with reduced pitch and interval information," Q. J. Exp. Psychol. **31**, 229–240.
Oxenham, A. J., and Buus, S. (**2000**). "Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions," J. Acoust. Soc. Am. **107**, 1605–1614.
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.
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.
Plomp, R. (**1964**). "The ear as a frequency analyser," J. Acoust. Soc. Am. **36**, 1628–1636.
Pressnitzer, D., Bestel, J., and Fraysse, B. (**2005**). "Music to electric ears: Pitch and timbre perception by cochlear implant patients," Ann. N. Y. Acad. Sci. **1060**, 343–345.
Pressnitzer, D., and Patterson, R. D. (**2001**). "Distortion products and the pitch of harmonic complex tones," in *Physiological and Psychophysical Bases of Auditory Function*, edited by A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht, The Netherlands), pp. 97–104.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
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.
Terhardt, E. (**1968**). "Über ein äquivalenzgesetz für intervalle akustischer empfindungsgrössen [On an equivalence rule for intervals between the magnitudes of acoustic sensations]," Kybernetik **5**, 127–133.
Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.
Wertheimer, M. (**1924**). *Gestalt Theory* (Hayes Barton, Raleigh, NC).
White, B. W. (**1960**). "Recognition of distorted melodies," Am. J. Psychol. **73**, 100–107.

The role of across-frequency processes in dichotic listening conditions[a)]

Marc Nitschmann,[b)] Jesko L. Verhey, and Birger Kollmeier
International Graduate School "Neurosensory Science and Systems," Carl-von-Ossietzky-Universität Oldenburg, 26111 Oldenburg, Germany

(Received 19 March 2008; revised 11 September 2009; accepted 14 September 2009)

In the bandwidening experiment with a diotic noise masker, an apparently wider critical bandwidth has often been reported when a dichotic signal ($S\pi$) is used instead of a diotic signal (So). Two competing across-channel processes were proposed to account for this apparently wider critical bandwidth: (i) A detrimental across-channel effect reducing the binaural masking-level difference (BMLD) for broadband maskers and (ii) a beneficial integration of information across channels for narrowband maskers. The two hypotheses result in different predictions of the BMLD in the notched-noise experiment: According to the first hypothesis, the change in BMLD with notch width is determined by the level-dependence of the BMLD for a narrowband masker centered at the signal frequency, whereas the second hypothesis predicts that it is determined by the level-dependence of the BMLD for a broadband masker. To test the hypotheses, masked thresholds of a diotic or dichotic 500-Hz signal were measured for a diotic notched-noise masker as a function of notch width. In addition, thresholds were measured for a diotic broadband and narrowband masker as a function of masker level. The data indicate that neither of the two hypotheses is able to predict the continuous decrease in the BMLD as the notch width increases. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3243307]

PACS number(s): 43.66.Pn, 43.66.Dc, 43.66.Ba, 43.66.Rq [RYL] Pages: 3188–3198

I. INTRODUCTION

A fundamental characteristic of the auditory system is its frequency selectivity which can be modeled as a bank of overlapping bandpass filters (e.g., Patterson and Nimmo-Smith, 1980; Moore, 2003). One way to characterize auditory frequency selectivity is the critical bandwidth which is determined in the bandwidening experiment (Fletcher, 1940). In the bandwidening experiment, the threshold of a tone is measured in the presence of a signal-centered noise masker as a function of masker bandwidth. Several authors have used such an experimental paradigm to get insights into the frequency selectivity of the binaural system (e.g., Bourbon and Jeffress, 1965; Wightman, 1971; Sever and Small, 1979; Hall *et al.*, 1983; Zurek and Durlach, 1987). They measured thresholds in a condition where the signal had an interaural phase difference of 180° in the presence of a diotic masker. This condition is commonly referred to as $NoS\pi$ condition. For comparison, they also estimated the monaural critical bandwidth by measuring thresholds in the NoSo condition, where both signal and masker were presented diotically. Their threshold data showed that $NoS\pi$ thresholds increased as the masker bandwidth increased even beyond the critical bandwidth that was derived from the NoSo thresholds. Thus, the difference between the thresholds in the diotic and the dichotic condition, referred to as the binaural masking-level difference (BMLD), decreased as the bandwidth increased. This result was initially interpreted as a poorer frequency selectivity of the binaural system compared to that of the monaural pathway (see also Sondhi and Guttman, 1966; Yama and Robinson, 1982). However, two later studies (Hall *et al.*, 1983; van de Par and Kohlrausch, 1999) argued that the frequency selectivity is the same for the monaural and binaural systems and that the apparently larger binaural critical bandwidth in the bandwidening experiment may reflect a process across critical bands (across-channel process). The nature of the across-channel process differs between the two studies being either beneficial (van de Par and Kohlrausch, 1999) or detrimental (Hall *et al.*, 1983). The present study investigates if these hypotheses can also account for the data of a notched-noise experiment which is another experiment commonly used to characterize the frequency selectivity of the auditory system (de Boer and Bos, 1962; Patterson, 1976; Patterson and Henning, 1977). In notched-noise experiments, thresholds of a tone are measured in the presence of a noise masker with a spectral notch around the signal frequency as a function of the notch width.

Hall *et al.* (1983) measured the BMLD for both types of experiments, i.e., the notched-noise experiment and the bandwidening experiment. To account for the wider binaural critical bandwidth in the bandwidening experiment they suggested that the interaural difference cues in the dichotic condition are biased by the information from remote auditory filters indicating no interaural difference. According to their line of arguments, the larger bandwidth in the dichotic bandwidening experiment is accounted for by an increase in the strength of the detrimental effect as the masker bandwidth is increased.

[a)] Parts of this paper have been presented at the 2006 Meeting of the Deutsche Gesellschaft für Akustik (Nitschmann and Verhey, 2006).

[b)] Author to whom correspondence should be addressed. Electronic mail: marc.nitschmann@uni-oldenburg.de

In contrast to the bandwidening experiment, Hall *et al.* (1983) derived a similar critical bandwidth for the monaural and binaural systems from the notched-noise data. They concluded that the two experiments measure different aspects of binaural processing. While the notched-noise experiment measures the binaural frequency selectivity, the bandwidening experiment reveals a process across frequency channels which depends on the presence of noise in off-frequency filters. Obviously this cannot mean that the detrimental across-channel process is not affecting the notched-noise data since the no-notch condition of the notched-noise experiment is equivalent to a broadband condition for the bandwidening experiment. Hall *et al.* (1983) did not discuss explicitly how this detrimental across-channel process affects the BMLD in the notched-noise experiment. However, their comparison of the BMLD for a narrowband masker to the notched-noise data provides indirect information about the size of the effect and how it varies with notch width. According to Hall *et al.* (1983), level effects of the BMLD should be assessed using narrowband maskers centered at the signal frequency since in this condition a negligible influence of the detrimental across-frequency effects could be assumed. Hall *et al.* (1983) reported a quantitative agreement between the threshold data in the notched-noise experiment and the 10-Hz bandwidth condition of the bandwidening experiment: Whenever the NoSo thresholds decreased by 20 dB, either due to a decrease in spectrum level of a narrowband masker or due to the introduction of a notch in a broadband noise reducing the excitation of the filter centered at the signal frequency, the corresponding NoSπ thresholds showed a decrease of about 13 dB. This comparison implicitly assumes that the magnitude of the BMLD reduction due to the detrimental across-channel process is determined by the difference between the BMLD for the narrowband masker and the broadband (no-notch) masker with the same spectrum level and does not vary with notch width.

Due to the relatively coarse sampling of level in the bandwidening experiment the comparison is limited to only a few data points. It is also somewhat unfortunate that Hall *et al.* (1983) showed only mean thresholds since the BMLDs for narrowband maskers tend to show large individual differences (Bernstein *et al.*, 1998; Buss *et al.*, 2007). These individual differences may also hamper the comparison to other publications (e.g., Hall and Harvey, 1984; Buss *et al.*, 2003). If the hypothesis by Hall *et al.* (1983) is correct, a good correspondence between the notched-noise results and the level-dependence of the BMLD for the narrowband masker should also be observed when compared individually. Such a comparison is not possible on the basis of the results of the previous studies.

In contrast to the detrimental across-channel process proposed by Hall *et al.* (1983), van de Par and Kohlrausch (1999) suggested a beneficial across-channel combination of information. They argued that when detecting a dichotic signal in the presence of a diotic narrowband masker, subjects might use information in filters adjacent to the peripheral filter centered at the signal frequency as an additional cue since the filters have signal-to-noise ratios (SNRs) similar to the one in the filter centered at the signal frequency. This beneficial across-channel combination of information cannot be used in a diotic condition because in this condition the fluctuations of the masker limit detection (Bos and de Boer, 1966) and these masker fluctuations are the same in the on- and off-frequency filters. For a broadband masker, the SNR is high in the filter centered at the signal frequency only. Thus, the auditory system cannot benefit from information in off-frequency filters neither in a diotic nor in a dichotic condition.

According to van de Par and Kohlrausch (1999), the apparently wider binaural bandwidth decreases with decreasing level because of the reduced number of peripheral filters that are excited by the stimulus. Van de Par and Kohlrausch (1999) concluded that in a bandwidening experiment at a low masker level as well as in all broadband masking paradigms the binaural processor has to rely on the on-frequency filter and that, in these cases, a single-filter binaural model is sufficient to predict the thresholds. Breebaart *et al.* (2001b) showed that a model based on the hypothesis of van de Par and Kohlrausch (1999) can predict the wider binaural bandwidth observed in the bandwidening experiment.

A notched-noise experiment can be considered as a broadband masking situation. This has an implication on the change in the BMLD with notch width similar to those drawn in Hall *et al.* (1983). Whenever the NoSo thresholds decreased either due to a decrease in spectrum level of a broadband masker or due to the introduction of a notch in the noise masker reducing the excitation of the filter centered at the signal frequency, the corresponding change in BMLD should be the same. Thus the only difference between the two hypotheses is which bandwidth should be used for the comparison of the notched-noise data and the dependence of the BMLD on the masker level for a bandpass noise centered at the signal frequency. The masker should be narrowband for the hypothesis of a detrimental across-channel process and broadband for the hypothesis of a beneficial across-channel process.

The present study investigates which of the hypotheses can account for the notched-noise data. The dependence of the BMLD on the masker level is measured for a broadband and a narrowband masker and compared to results with notched-noise threshold data within the same subjects. To show all data sets within one figure, the dichotic thresholds are plotted as a function of the respective diotic thresholds.[1]

According to van de Par and Kohlrausch (1999), the BMLD in all broadband masking conditions including the notched-noise masking condition is solely determined by the information within the auditory filter centered at the signal frequency. Based on this hypothesis, the BMLD for the notched-noise masker and the broadband-noise masker should be the same for the same diotic threshold. Thus, the same threshold curves should be obtained for the threshold data for the notched-noise masker and for the broadband-noise masker when plotted as dichotic thresholds as a function of diotic thresholds (Verhey and Zerbs, 2001).

To test the hypothesis of a detrimental across-frequency process and its implication for the notched-noise experiment, more subjects and a higher level resolution than in Hall *et al.* (1983) were used in the present study to increase the accu-

racy of the estimated effect of level on the magnitude of the BMLD for narrowband maskers. According to Hall *et al.* (1983), the decrease in BMLD for the narrowband masker as the masker level decreases should be the same as observed in the notched-noise data for a comparable decrease in the diotic threshold. Thus, for the data representation chosen in the present study, the slope of the threshold curves for the narrowband experiment and the notched-noise experiment should be the same.

II. METHODS

A. Apparatus and stimuli

Stimuli were generated digitally at a sampling rate of 44.1 kHz. A standard personal computer controlled stimulus generation and presentation and recorded results using a software package developed at the University of Oldenburg. Stimuli were D/A converted (RME ADI-8 DS, 32 bits), amplified (Tucker-Davis HB7), and presented via Sennheiser HD 580 headphones. Subjects were seated in a double-walled, sound-insulated booth.

The target signal was a sinusoid of 500 Hz that was either in phase (So) or antiphase ($S\pi$) between the two ears. Its duration was 300 ms including two 50-ms raised-cosine ramps at on- and offsets so that the steady state portion of the signal was 200 ms. The diotic masking noises (No) were generated in the frequency domain using a 2^{17}-point (0.34-Hz resolution) buffer. Their spectrum had a constant nonzero amplitude in the passband regions and a random phase. The noises were transformed to the time domain (inverse fast Fourier transform) and restricted to the desired length of 127 890 samples (i.e., a duration of 2.9 s). The noise masker was gated with 50-ms raised-cosine ramps at on- and offsets.

For the notched-noise experiment, a constant spectrum level of 50 dB/Hz and six notch widths of 0, 50, 200, 400, 600, and 800 Hz were used. All notches were arithmetically centered at 500 Hz. The lower frequency limit was set to 30 Hz, the higher to 1 kHz. Stimuli and procedure were similar to those used by Hall *et al.* (1983).

In the experiment with a broadband noise, the masker spectrum level was varied from 0 to 50 dB/Hz in steps of 10 dB/Hz. Like in the notched-noise experiment, the lower and upper cut-off frequencies of the masker were 30 and 1000 Hz, respectively. For the highest spectrum level the experimental condition was the same as in the notched-noise experiment with a notch width of 0 Hz. The thresholds for this spectrum level were taken from the notched-noise experiment.

In the experiment with a narrowband-noise masker, six different masker spectrum levels (10–60 dB/Hz in steps of 10 dB/Hz) were used. The masker was 10 Hz wide and arithmetically centered at the signal frequency.

B. Procedure

A three-interval forced-choice procedure with adaptive signal-level adjustment was used to determine detection thresholds in masking noise. Temporally centered in the masker, there were three intervals of 300-ms duration separated by 300-ms pauses. The intervals were indicated on the screen in front of the subject. One randomly chosen interval contained the signal. The subject's task was to indicate this interval. Responses were given by pressing the corresponding button 1, 2, or 3 on a computer keyboard. Trial-by-trial feedback was provided. On every trial a new sample of noise was generated.

Signal level was adjusted according to a one-up two-down rule tracking the 70.7% correct response level (Levitt, 1971). The initial step size of the signal level was 8 dB. The step size was halved after each second reversal of the level adjustment procedure until a step size of 1 dB was reached. After that the run continued for another six reversals. The mean value over these last six reversals was used as a threshold estimate.

Each subject did the experiments in the following order: Notched-noise experiment, broadband-noise experiment, and finally the narrowband-noise experiment. Runs with diotic and dichotic stimulus conditions were mixed in the notched-noise and broadband masking experiments. In the narrowband masking experiment, thresholds were first measured in the NoSo and then in the $NoS\pi$ condition because of the higher difficulty of this listening task compared to the other two experiments.[2] At least three threshold estimates were obtained and averaged for each parameter value and subject.

C. Subjects

Eight normal-hearing subjects (4 male, 4 female, aged 19–30 years, mean 25 years) participated in all three experiments. Three of the subjects including the first author were members of the research group, the others were paid volunteers. All subjects had normal audiograms with hearing thresholds lower than or equal to 10 dB HL (hearing level, i.e., dB *re* pressure at average normal-hearing absolute thresholds) at the standard audiometric frequencies from 125 Hz to 1 kHz, i.e., the audiometric frequencies in the relevant frequency range. With the exception of subject 4 this also holds for the audiometric frequencies from 1.5 to 8 kHz. Subject 4 had absolute thresholds up to 20-dB HL at 1.5 kHz and higher frequencies at the left ear and at 2 kHz and higher frequencies at the right ear.

III. RESULTS

Figure 2 shows the detection thresholds measured in the notched-noise, broadband, and narrowband experiments. $NoS\pi$ thresholds are plotted as a function of NoSo thresholds with triangles, squares, and diamonds, respectively. Using this data representation all threshold data can be plotted within one figure. The figure shows individual thresholds of eight subjects and the average thresholds over these subjects in the lower right panel. The dotted diagonal line denotes the same NoSo as $NoS\pi$ thresholds. The BMLD is the vertical distance between this line and a data point. The gray line is the narrowband threshold curve shifted such that its threshold at the masker spectrum level of 50 dB/Hz coincides with the threshold at the notch width of 0 Hz of the notched-noise experiment. The reason for plotting this shifted threshold curve is based on an analysis by Hall *et al.* (1983) (see Sec.

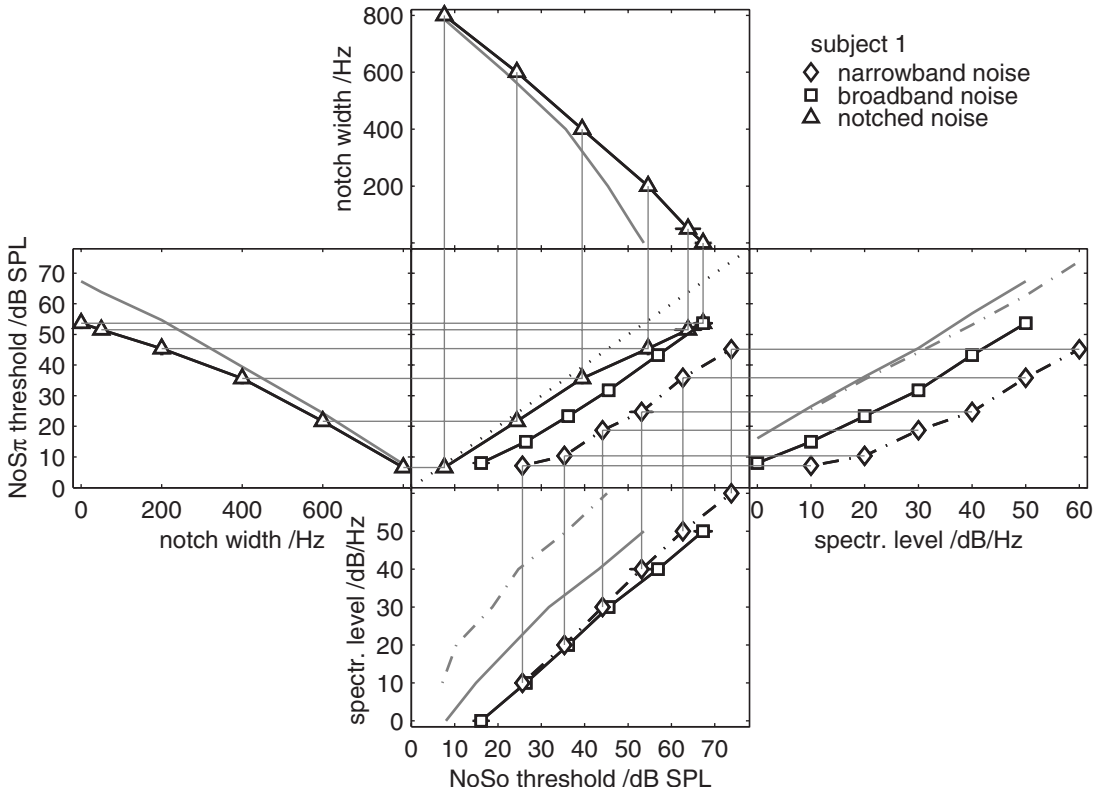


FIG. 1. Illustration of how the threshold data from the different experiments are plotted within one graph. This is the data representation used in Figs. 2 and 3. As an example, the figure shows threshold data for subject 1. Thresholds of the notched-noise experiment are plotted in the left panel and in the upper panel. Thresholds of the broadband experiment (solid lines) and of the narrowband experiment (dash-dotted lines) are plotted in the right panel and in the lower panel. The symbols denote NoSo thresholds in the upper and lower panels and NoSπ thresholds in the left and right panels. The other thresholds (NoSo or NoSπ, respectively) are indicated by gray lines. The vertical gray lines connecting the upper and lower panels with the center panel denote the projection of NoSo thresholds to the center panel. The horizontal gray lines connecting the left and right panels with the center panel denote the projection of NoSπ thresholds to the center panel. Thus, the center panel shows the threshold data of the notched-noise experiment (triangles), broadband experiment (squares), and narrowband experiment (diamonds) in a representation where NoSπ thresholds are plotted as a function of NoSo thresholds. The dotted line in the center panel denotes the same NoSo and NoSπ thresholds.

IV D). Error bars denote plus minus one intraindividual or interindividual standard deviation for the individual or for the mean thresholds, respectively.

A. Notched-noise

The highest threshold (no notch) of the notched-noise experiment is at the upper right end of each notched-noise threshold curve, the lowest (notch width of 800 Hz) at the lower left. For all subjects except for subject 6, the threshold data points of the notched-noise experiment exhibit the highest NoSπ thresholds for given NoSo thresholds.

For all subjects, NoSπ thresholds decrease with decreasing NoSo thresholds. This decrease is shallower for narrow than for large notch widths. Toward lower threshold the curve approaches the dotted diagonal line (i.e., the BMLD approaches 0 dB). The decrease of BMLD with increasing notch width differs interindividually: For the 600-Hz notch width, subjects 2 and 3 show almost no BMLD (smaller than 1 dB), whereas subjects 4–6 still measure a BMLD of about 4 dB.

For the mean data, NoSo thresholds decrease with increasing notch width from 68 dB in a broadband noise without a notch to about 8 dB at a notch width of 800 Hz. Standard deviations are around 1 dB for notch widths up to 50 Hz, 2 dB for the largest notch width, and between 2 and 3 dB for intermediate notch widths. The large standard deviations for intermediate notch widths are due to individual differences in filter shape. Mean NoSπ thresholds decrease with increasing notch width from 53 dB in a broadband noise without a notch to 7 dB for a notch width of 800 Hz. Standard deviations as a function of notch width are similar to those of the diotic thresholds, except for values around 2 dB for intermediate notch widths. The mean BMLD decreases gradually with increase in notch width. It is about 15 dB in a broadband noise without a notch and reduces to a value clearly below the standard deviations of corresponding thresholds for the two largest notch widths.

B. Broadband noise

The highest threshold (at the spectrum level of 50 dB/Hz) of the broadband experiment is at the upper right end of each broadband threshold curve. It is the same as the highest threshold of the notched-noise experiment. Thresholds for lower spectrum levels are below the notched-noise threshold curve for all subjects. The lowest threshold (at the spectrum level of 0 dB/Hz) is at the lower left end of each broadband threshold curve.

Individual differences are small for the broadband-noise threshold data. They are found mainly in the decrease in BMLD toward low spectrum levels. Mean NoSo thresholds

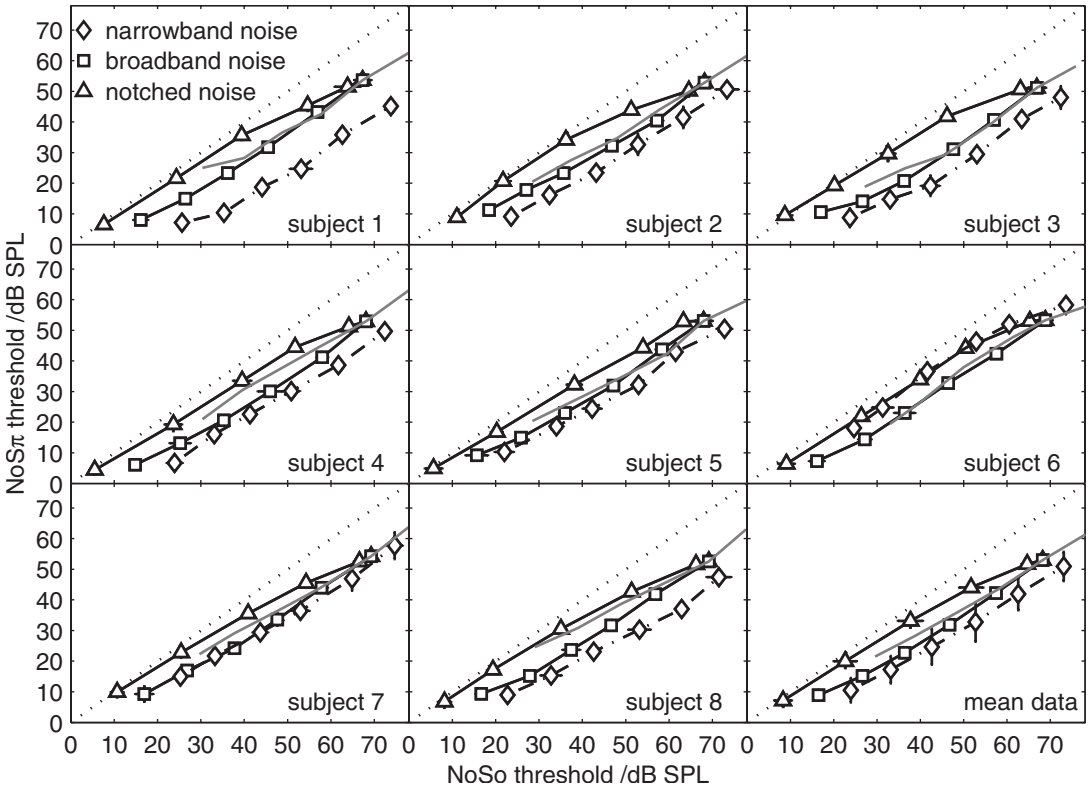


FIG. 2. Individual and mean (lower right panel) detection thresholds. The figure displays NoSπ thresholds as a function of NoSo thresholds, as illustrated in Fig. 1. Thresholds of the notched-noise, broadband, and narrowband experiments are denoted with triangles, squares, and diamonds, respectively. The gray line is the narrowband threshold curve shifted such that its 50-dB/Hz threshold coincides with the threshold at the notch width of 0 Hz of the notched-noise experiment. The dotted diagonal line denotes the same NoSo and NoSπ thresholds. Error bars show the intraindividual standard deviations for the individual thresholds and the interindividual standard deviations for the mean thresholds. Standard deviations are not shown if they are below the size of the symbol.

increase as a function of spectrum level from 16.5 dB at 0 dB/Hz to 68 dB at 50 dB/Hz. The standard deviation of the mean NoSo thresholds is mostly below 1 dB. The standard deviations of the mean NoSπ thresholds are below 1 dB at 50 dB/Hz as well, but increase toward lower spectrum levels to 1.7 dB at the spectrum level of 0 dB/Hz. Mean dichotic thresholds increase with increase in spectrum level from 9 dB at 0 dB/Hz to 53 dB at 50 dB/Hz. The mean BMLD increases with spectrum level from below 8 dB at 0 dB/Hz to 15 dB at 50 dB/Hz. The mean BMLD is clearly above at least twice the standard deviations of the corresponding thresholds even at the lowest spectrum level.

C. Narrowband noise

The highest threshold (at 60 dB/Hz) of the narrowband experiment is at the upper right end of each narrowband threshold curve, the lowest threshold (at 10 dB/Hz) is at the lower left end. The narrowband threshold curve is below the broadband threshold curve for most subjects (except for subjects 6 and 7).

The thresholds show large interindividual differences that also have been reported by Bernstein *et al.* (1998) and Buss *et al.* (2007). Individual diotic thresholds range from 22 to almost 26 dB at 10 dB/Hz and from 71.5 to almost 75 dB at 60 dB/Hz. Individual dichotic thresholds for the lowest spectrum level range from almost 7 to 18 dB, for the highest spectrum level they are in a range from 45 to 58 dB. Individual BMLDs range from almost 7 to 18.5 dB for the lowest and 15 to nearly 29 dB for the highest spectrum level. In particular, subject 6 apparently had problems with the dichotic stimulus condition and achieved, in contrast to all other subjects, a smaller BMLD than in the broadband masking experiment.

Mean NoSo thresholds in the narrowband experiment increase with increasing spectrum level from 24 dB at 10 dB/Hz to 73 dB at 60 dB/Hz. The interindividual standard deviations are around 1 dB for all spectrum levels for the NoSo thresholds, whereas they are in a range of 4 to above 6 dB for the NoSπ thresholds. Mean NoSπ thresholds are 10.5 and 51 dB for the spectrum levels of 10 and 60 dB/Hz, respectively. The BMLD of the mean thresholds increases with increasing spectrum level from above 13 dB at 10 dB/Hz to 22 dB at 60 dB/Hz.

IV. DISCUSSION

A. Comparison to previous studies

The bottom two panels of Fig. 3 show threshold data from previous studies related to the present study. In order to facilitate the comparison between the present data and the data in the literature, their results are plotted using the same data representation as in Fig. 2. The top left panel redraws the average thresholds of the present study, which were already shown in the bottom right panel of Fig. 2. The other panels show the threshold data of the present study with a gray curve without symbols. For comparison of the present

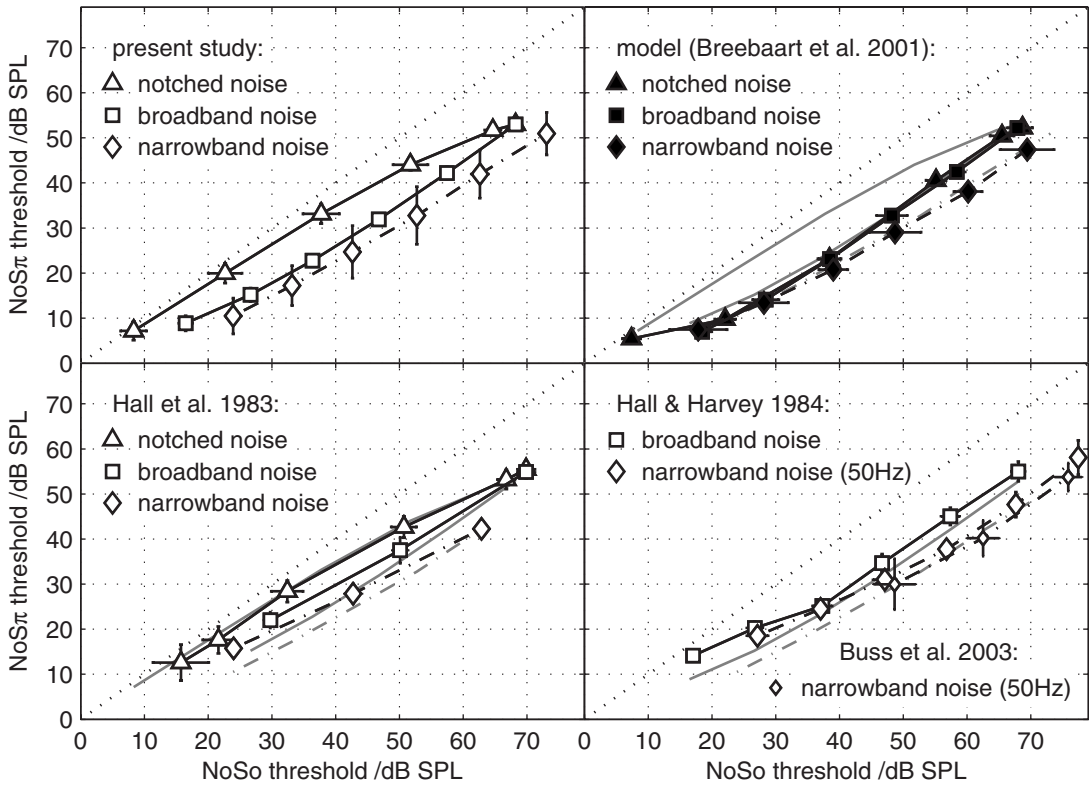


FIG. 3. Average detection thresholds of the present study (upper left panel and gray curves in the other panels) together with literature data (Hall *et al.*, 1983; Hall and Harvey, 1984; Buss *et al.*, 2003) and threshold predictions of the model by Breebaart *et al.* (2001a). The data representation is the same as in Fig. 2, as illustrated in Fig. 1.

data to previous studies, only a subset of the present data is shown which corresponds to the parameter range used in the respective study in the literature. To our knowledge, the study by Hall *et al.* (1983) was the only study measuring the BMLD for notched noise and the BMLD for a bandwidening experiment at different levels. Their threshold data are shown on the bottom left panel of Fig. 3.

Although the previous studies by Hall *et al.* (1983) track a different point on the psychometric functions than the present study, the data of the studies are similar. For the two smallest notch widths, i.e., the two highest NoSo thresholds, the thresholds of the present study and the thresholds of Hall *et al.* (1983) at the same spectrum level of 50 dB/Hz agree quantitatively.

For intermediate notch widths, NoSo thresholds tend to be lower in the study by Hall *et al.* (1983) than in the present study. This is presumably due to different average monaural auditory filter shape of the two groups of subjects. The best-fitting gammatone filter parameters for the mean diotic thresholds of the present study are a filter of sixth order and a width of 1.4 ERB (equivalent rectangular bandwidth, for the definition of ERB in hertz see, e.g., Glasberg and Moore, 1990; Kollmeier and Holube, 1992). The filter parameters were derived from the thresholds on the basis of a simple energy-detector model (Fletcher, 1940) with a gammatone filter using the filter implementation by Hohmann (2002). The diotic threshold data of Hall *et al.* (1983) for a spectrum level of 50 dB/Hz are best described by 1-ERB wide fourth-order gammatone filters (see also Zerbs, 2000). Thus, the filter parameters for the present data indicate a narrower filter with respect to filter tail and a broader filter with respect to the tip than those for the subjects participating in the experiment by Hall *et al.* (1983). Only for subject 3 of the present study a fourth-order gammatone filter with a bandwidth of 1 ERB produces the best fit to the diotic thresholds. The NoSo thresholds of the other subjects indicate higher order of the gammatone filter and wider bandwidth. The diotic and dichotic thresholds at the largest notch width are lower in the present study than in the study by Hall *et al.* (1983), indicating different absolute thresholds of the two groups of subjects.

Despite the differences in diotic thresholds, the threshold curves for the notched-noise experiments of the two studies almost overlie one another when plotted as dichotic thresholds as a function of diotic thresholds.

At the spectrum level of 50 dB/Hz, i.e., the highest diotic threshold, there is a quantitative agreement between the threshold data of Hall *et al.* (1983) and of the present study for the broadband masker and the narrowband masker. Toward lower NoSo thresholds the threshold curves for the two masker bandwidths are shallower in Hall *et al.* (1983) than in the present study. Thus, the decrease in BMLD as diotic thresholds decrease is faster than in the present data. For both maskers the difference is primarily due to higher NoSπ thresholds at the masker spectrum levels of 10 and 30 dB/Hz in the study by Hall *et al.* (1983). For the broadband masker, diotic and dichotic thresholds are both higher for these spectrum levels. The reason for this may be comparatively large individual differences at levels slightly above absolute threshold.

The shallower slope for the narrowband masker leads to the good correspondence of the change in BMLD between the notched-noise data and the data for the narrowband noise, which was already mentioned in the Introduction. A slope of the threshold curve for the narrowband masker similar to the one in Hall *et al.* (1983) was found in Hall and Harvey (1984) for a masker bandwidth of 50 Hz (bottom right panel of Fig. 3). For spectrum levels of 40 dB/Hz (third threshold point from the right) and below, the BMLD decreases at a higher rate in Hall and Harvey (1984) than in the present study as spectrum level decreases. Recent experimental results by Buss *et al.* (2003) with the same masker bandwidth as used in Hall and Harvey (1984) show a quantitative agreement with the data in the present study when plotted as dichotic thresholds as a function of diotic thresholds (small diamonds in the bottom right panel of Fig. 3).

For the broadband masker, Hall and Harvey (1984) measured an average threshold curve with a similar shape as the one of the present study. The BMLD was nearly constant for spectrum levels between 20 and 50 dB/Hz (four right threshold points). Below 20 dB/Hz, the BMLD decreased as spectrum level decreased. The decrease was slightly larger in their study than in the present study: In the present study the BMLD was nearly 14 dB at a spectrum level of 20 dB/Hz and decreased down to almost 8 dB at a spectrum level of 0 dB/Hz. For the same range of spectrum levels the BMLD decreased from 12 to 3 dB in Hall and Harvey (1984).

In general, the data in the literature show the same tendency as found in the present study: For the data representation used in Fig. 3, all studies show the lowest threshold curve for the narrowband masker and the highest threshold curve for the notched-noise data. The slope of the threshold curves is usually different for the different masker types.

B. Possible role of beneficial across-frequency processing

A beneficial across-channel process is the basis of one hypothesis to account for the apparently wider bandwidth for the binaural system compared to the monaural frequency selectivity (van de Par and Kohlrausch, 1999, see Sec. I). This beneficial across-channel process should increase the BMLD for narrowband maskers centered at the signal frequency. According to van de Par and Kohlrausch (1999) this beneficial across-channel process should have no effect in all broadband masking conditions including notched-noise masking since in these cases the SNR is high only in the auditory filter centered at the signal frequency and for this reason combination of information across filters is unhelpful. Verhey and Zerbs (2001) argued that the hypothesis by van de Par and Kohlrausch (1999) would predict the same threshold curves for the notched-noise experiment and for the broadband experiment when plotted as dichotic thresholds as a function of diotic thresholds. A possibility to test this prediction is to use the model by Breebaart *et al.* (2001a) which included an implementation of the beneficial across-channel process proposed by van de Par and Kohlrausch (1999).

Breebaart *et al.* (2001a) developed a model of binaural auditory perception based on the monaural model by Dau *et al.* (1996) combined with an equalization-cancellation (EC) stage (Durlach, 1963) as first described by Holube *et al.* (1995) (see also Zerbs, 2000 for a slightly different implementation of the model). For the prediction of the present threshold data a filterbank with eight filters was used. The parameters of the auditory filters were adjusted to predict the diotic notched-noise threshold data, i.e., 1.4-ERB wide sixth-order gammatone filters (see Sec. IV A). The spectral spacing was one filter per ERB. The level of the noise added after the filtering was set to 0.0 dB SPL (sound pressure level, i.e., dB *re* 20 micropascals) as used in Zerbs (2000) since the predicted thresholds for the largest notch widths were too high when using a level of 9.4 dB SPL as in Breebaart *et al.* (2001a). The rest of the model parameters were the same as in Breebaart *et al.* (2001a). Stimuli and procedure were the same as in the respective psychoacoustic experiments.

Simulated thresholds are shown in the top right panel of Fig. 3. The effect of the beneficial across-frequency processing implemented in the model by Breebaart *et al.* (2001a) is reflected by the lower threshold curve for the narrowband masker than for the other two masker types. If instead of a filterbank a single filter centered at the signal frequency is used, the curves of predicted thresholds of all three masker types overlie one another (not shown). The beneficial across-channel effect results in a decrease in dichotic thresholds in narrowband of about 6 dB at the highest masker spectrum levels (see also Breebaart *et al.*, 2001b). For the broadband and for the narrowband maskers, there is a quantitative agreement between measurement and simulation in threshold curve (difference smaller than 3 dB).

As argued in Verhey and Zerbs (2001), based on the hypothesis of van de Par and Kohlrausch (1999), the same threshold curve is predicted for the notched-noise experiment and the broadband experiment. The NoSπ thresholds of the notched-noise experiment, however, are only predicted quantitatively at the notch widths of 0 and 800 Hz. For intermediate notch widths, the predicted NoSπ thresholds are about 10 dB too low. Thus, the model cannot predict the difference in threshold curve between the broadband masker and the notched-noise masker which is observed in the average and individual threshold data shown in Fig. 2.

For the thresholds measured in the present study, the difference of the BMLD between the two masker types for the same diotic threshold amounts to a maximum of 9.4 dB and a mean of 7.1 dB for the average data. The largest difference is obtained for subject 3, who shows a maximum difference of 12.6 dB and a mean of 9.3 dB. Subjects 5–7 show the smallest differences with a maximum distance of 8.5 dB and an average distance of 6.5 dB. A similar (though slightly smaller) difference is found when comparing the data for the different masker types of Hall *et al.* (1983) shown in the bottom left panel of Fig. 3 (5.4-dB maximum and 4.0-dB mean).

The comparison of model prediction with the measured data indicates that the beneficial across-channel process accounts for the apparently wider bandwidth for the binaural bandwidening experiment. However, a model with a realistic monaural filter width including only a beneficial across-frequency process cannot predict the measured monotonic

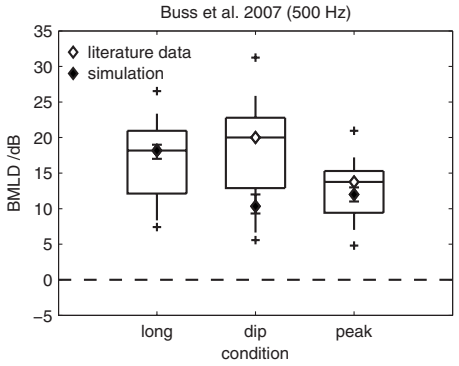


FIG. 4. Measured (open symbols) and simulated (filled symbols) BMLDs for an experiment by Buss *et al.* (2007) for three different conditions: long, where the signal was a 200-ms-long sinusoid masked by the noise, peak, where a 30-ms sinusoid was positioned in a peak of the masker envelope, and dip, where a 30-ms sinusoid was placed into a dip of the masker envelope. For the data measured by Buss *et al.* (2007), the median is indicated with an open symbol, the span between the 25th and 75th percentiles with a box, the 10th–90th percentile with bars, and the maximum and minimum with crosses. For the simulation data, the median is indicated with a filled symbol; the error bars denote the quartiles. The dashed horizontal line denotes a BMLD of 0 dB.

decrease in BMLD as the notch width increases. Thus, other processes are needed to account for this effect.

C. Role of envelope fluctuations

Recently, the possible role of masker-envelope fluctuations was discussed in relation to the BMLD for narrowband maskers (Buss *et al.*, 2003, 2007). They compared the BMLD for a long signal with the BMLD for a short signal that was either placed in the minimum or in the maximum of the masker envelope. Figure 4 shows, as an example, the data from Buss *et al.* (2007) for long (200-ms) and short (30-ms) sinusoids of 500 Hz. The noise spectrum level was 55 dB/Hz. The median is indicated with an open symbol, the span between the 25th and 75th percentiles with a box, the 10th–90th percentile with bars, and the maximum and minimum with crosses. In agreement with the previous study (Buss *et al.*, 2003), Buss *et al.* (2007) found a similar median BMLD for the long signal and the short signal that was positioned in the dip of the masker envelope. The BMLD was reduced by about 5 dB for a short signal positioned at a peak of the masker envelope. The similarity of the BMLD for the dip position of the short signal and the long signal was interpreted as evidence for the ability of the subjects to listen into the dips of the masker when detecting the long signals. This interpretation will be referred to as the dip-listening hypothesis in the following. In the light of the dip-listening hypothesis, the decrease in the BMLD observed in the bandwidening experiment with increasing masker bandwidth may be interpreted as a result of a reduced effective duration of the dips as the masker bandwidth increases.

Figure 4 shows, apart from the BMLDs measured by Buss *et al.* (2007), predictions (filled diamonds) of the model by Breebaart *et al.* (2001a) using the same parameters as described in Sec. IV B. The model predicts the BMLD quantitatively for the long signals. The predicted BMLDs for the peak condition are within the 25th and 75th percentiles of the measured BMLDs. For the dip condition, the predicted BMLD is similar to the one for the peak condition and is between the 10th and 25th percentiles of the BMLD measured by Buss *et al.* (2007). The model by Breebaart *et al.* (2001a) does not predict a similar median BMLD for short signals positioned in a masker-envelope dip and the long signal. This inability of the model may be interpreted as evidence for an inappropriate cue used by the model to predict the different BMLD for the narrowband and broadband masker quantitatively. Note, however, that large individual differences are observed in the data by Buss *et al.* (2003, 2007) and that some subjects at some levels show a similar pattern of results as predicted by the model.

One may argue that the model is not able to predict the dichotic notched-noise threshold data because it is not in agreement with the dip-listening hypothesis. This raises the question if the notched-noise data can be understood on the basis of this hypothesis. In order to answer this question the average envelope fluctuation rate for the notched-noise masker, here defined as the centroid of the envelope power spectrum, is calculated. For the calculations, the filtering properties of the outer, middle, and inner ear are accounted for by analyzing the Hilbert envelope after the masker has been preprocessed by an outer and middle ear filter as proposed in Breebaart *et al.* (2001a) and a 1.4-ERB wide sixth-order gammatone filter centered at the signal frequency. The average envelope fluctuation rate is about 44 Hz for a masker with no notch and increases as the notch width increases up to a maximum of 77 Hz for a notch width of 100 Hz. A further increase in notch width results in a decrease in the average envelope fluctuation rate. For the 10-Hz wide narrowband-noise masker used in the present study the average fluctuation rate is about 3 Hz. For the 50-Hz wide noise band used in Buss *et al.* (2007) the average envelope fluctuation rate is 15 Hz.[3] Thus the average envelope fluctuation rate for the notched noise is considerably higher (and thus the average dip duration shorter) than for the narrowband noise. The reduced dip duration implies that the dip-listening cues are largely reduced for the notched-noise masker. This may account for the difference in the average BMLD between the notched-noise experiment and narrowband-noise experiment. It is, however, unlikely that dip listening is able to account for the difference in slope of the decrease in BMLD between the two studies. Assuming a monotonic relation between dip duration and reduction in the magnitude of BMLD, the BMLD in the notched-noise experiment should be smallest for the notch width of 100 Hz (i.e., for the smallest average dip duration). In contrast, a continuous decrease in the BMLD is observed when the notch width is increased. This argues against an explanation of the difference between the notched-noise data and the narrowband-noise data solely on the basis of a dip-listening hypothesis. Another argument against a strong influence of dip-listening cues in the notched-noise data is the relatively small interindividual variability for the results of the notched-noise experiment. In contrast, the narrowband data for which the dip-listening hypothesis was hypothesized show a high inter-subject variability.

Although it is unlikely that cues based on the envelope

fluctuations play an important role in the notched-noise experiment of the present study, modulation cues may be used in other binaural masking experiments. For example, Zwicker and Henning (1984) showed in a masking pattern experiment that the BMLD decreases rapidly when the signal is outside the bandwidth of the masker. This may reflect a difference in the ability of the monaural and binaural systems to use modulation cues. Derleth and Dau (2000) showed that a model using a modulation filterbank to analyze the modulation predicts much steeper threshold curves than a simple energy-detector model. If the ability to use modulation cues is reduced in the binaural system, the threshold curve for the dichotic condition may be shallower than that for the monaural one, as found in the experimental data by Zwicker and Henning (1984). Further studies are needed to investigate the role of modulation cues in binaural detection tasks.

D. Possible role of detrimental across-frequency processing

For the results of Hall *et al.* (1983) shown in Fig. 3, the upper part of the notched-noise threshold data curve and the narrowband threshold data curve are parallel. This led Hall *et al.* (1983) to hypothesize that the BMLD in a notched noise is basically determined by the BMLD in a narrowband masking condition.

According to the hypothesis of Hall *et al.* (1983) the threshold curve for the notched-noise experiment should be an upward shifted version of the data for the narrowband-noise masker when plotted as dichotic thresholds as a function of diotic thresholds. Thus, the magnitude of the vertical shift reflects the constant reduction in the BMLD due to the detrimental across-frequency process (see Sec. I).[4]

In order to facilitate the comparison between the notched-noise data and the data for the narrowband noise in Fig. 2, a shifted version of the narrowband-noise data is shown with a gray line. Following the comparison strategy used by Hall *et al.* (1983), the data point for the narrowband noise at a spectrum level of 50 dB/Hz was aligned with the data point for the notched-noise experiment with no notch. For all subjects this alignment results in a shift toward higher NoSo and NoSπ thresholds, though there are large interindividual differences. On average, the threshold data shift on the abscissa is almost 6 dB and on the ordinate 11 dB. For the individual results the shifts on the abscissa range from above 3 to about 8 dB, shifts on the ordinate from about 1 (subject 6) to almost 18 dB (subject 1).

For two subjects (4 and 8) the shifted version of the narrowband-noise data is close to the notched-noise data. For all other subjects the data sets are still considerably different from one another (see Fig. 2). Note that for subject 6 the original narrowband threshold curve is close to the subject's notched-noise threshold curve, whereas the shifted version is not. Thus, only two out of eight subjects support the hypothesis of Hall *et al.* (1983) that the change in BMLD in the narrowband condition determines the change in BMLD in the notched-noise data. The reason for this discrepancy is unclear. It may be partly due to the very limited set of data points that were used in Hall *et al.* (1983) to compare the data for the two masker types. The relatively good correspondence between the two data sets for two of the eight subjects in the present study also indicates that it is partly due to a different set of subjects. The majority of the data of the present study shows that it is unlikely that a constant detrimental across-channel effect accounts for the BMLD in the notched-noise experiment. It is, however, possible to reconcile the hypothesis of a detrimental across-channel process with the notched-noise data if it is assumed that the strength of the detrimental effect increases as the ratio between the level of the off-frequency components and the level of the on-frequency components increases. This ratio increases as the notch width increases, i.e., according to this hypothesis the BMLD should decrease in agreement with the experimental results.

V. SUMMARY AND CONCLUSION

Diotic and dichotic detection thresholds of a sinusoidal signal were measured for a notched-noise masker with varying notch widths and compared to threshold data for broadband- and narrowband-noise maskers with varying spectrum levels. This comparison was motivated by the hypotheses of across-frequency processes to account for the apparently wider critical bandwidth in the bandwidening experiment in dichotic compared to monaural or diotic conditions. Both hypotheses predict that the level-dependence of the BMLD for a masker centered at the signal frequency determines the decrease in BMLD as the notch width increases. The data of the present study indicate that, for the same diotic threshold, the BMLD is different in the notched-noise and broadband-noise experiments. In the notched-noise experiment, the BMLD decreases gradually with increasing notch width, whereas in the broadband masking experiment the BMLD is constant for most spectrum levels. In contrast to this finding, a model assuming a beneficial across-channel process to account for the large BMLD for a narrowband masker (van de Par and Kohlrausch, 1999) predicts the same BMLD for the broadband masker and the notched-noise masker if compared at the same diotic threshold for the two masker types. Thus, the hypothesis of a beneficial across-channel integration of information in the dichotic condition is not sufficient to account for both broadband-noise and notched-noise data. The model is also unable to predict recent experimental results (Buss *et al.*, 2007) relating to the temporal position of a short signal with respect to the masker envelope, indicating a higher weight to masker-envelope minima. However, a strategy based on the listening into the valleys is unlikely to account for the BMLD in the notched-noise experiment since the change in BMLD as the notch width increases is different from the change in the average duration of the masker dips. The measured threshold data are also not consistent with the hypothesis of Hall *et al.* (1983) that the change in BMLD in the notched-noise experiment is determined by the level-dependence of the BMLD for a narrowband masker. According to this hypothesis, a decrease in the diotic threshold of the notched-noise experiment leads to the same decrease in BMLD as found for a narrowband masker if the masker level is decreased by the same amount. In contrast to this prediction the decrease in BMLD for the

notched-noise experiment as the notch width increases is considerably faster than observed in the narrowband masking condition. Thus, the data are inconsistent with the predictions of both hypotheses. The notched-noise data may, however, be reconciled with the hypothesis of a detrimental across-channel process if it is assumed that the strength of the detrimental effect is determined by the level of the off-frequency components and its relation to the level of the on-frequency components.

ACKNOWLEDGMENTS

This study was supported by the Deutsche Forschungsgemeinschaft (DFG GRK 591/3). We would like to thank Associate Editor Ruth Litovsky and three anonymous reviewers for their helpful comments on earlier versions of the manuscript.

[1]This comparison is motivated by the discussion in Hall *et al.* (1983) on the similarity of the narrowband-noise data and notched-noise data. It implicitly assumes that the NoSo thresholds provide an estimate of the energy in the filter centered at the signal frequency. Such an assumption was the basis for the derivation of the filter characteristics, i.e., the energy in the filter centered at the signal frequency determines threshold (see, e.g., Fletcher, 1940; Patterson, 1976). Thus the thresholds on the abscissa can be regarded, to a first approximation, as an estimate of the energy in the auditory filter. A limitation of this interpretation is that thresholds are also influenced by the masker variability and envelope statistics (e.g., Verhey *et al.*, 2007). For the bandwidening experiment, Bos and de Boer (1966) showed that the masker variability decreases with the square root of the bandwidth. As a consequence, thresholds for narrowband signals are higher than for broadband signals with the same energy in the passband of the auditory filter centered at the signal frequency. For the notched-noise experiment, however, Patterson and Henning (1977) argued that, assuming an exponential filter shape, changes in notch width do not lead to changes in the variability of the decision statistics. Since a broadband masker is a special case of a notched-noise masker (i.e., a masker with no notch), the above prediction of the same threshold curves for notched-noise experiments and the broadband masker is also true if the masker variability is taken into account.

[2]The sinusoidal signal and the narrowband-noise masker have both a clear pitch which is the same for the condition considered here, i.e., a narrowband masker centered at the signal frequency. In contrast, for the broadband maskers, only the signal has a clear pitch. Thus due to the similarity of the subjective quality of masker and signal it is more difficult to detect the signal in narrowband noise than in broadband noise. In addition, even for subcritical bandwidths, the threshold of a sinusoidal signal in the presence of a signal-centered bandpass noise decreases as the bandwidth increases. The decrease is similar to the decrease observed for the difference limen of bandpass noise (e.g., Bos and de Boer, 1966). This decrease is likely to be due to level fluctuations of the masker which depend on the bandwidth (e.g., de Boer, 1966; Buus, 1990; Verhey, 2002). Thus, at the same signal-to-noise ratio it is harder to detect the diotic signal in narrowband noise than in broadband noise.

[3]For small bandwidths, the average envelope fluctuation rate of bandpass noise is proportional to the bandwidth of the noise. The equation proposed in Rice (1954), which relates the average number of envelope maxima per second to the bandwidth of bandpass noise, shows the same linear relation between these two quantities. Eddins and Wright (1994) used this equation to estimate the average envelope fluctuation rate of narrowband-noise maskers. This estimate results in slightly higher rates since a calculation based on the maxima of the envelope favors high fluctuation rates. Numerical simulation showed that the average envelope fluctuation rate was similar to the centroid if it is defined as half of the average rate of zero crossings of the ac-coupled envelope. The average rate of zero crossings (rather than the average rate of maxima) was also used in Carlyon *et al.* (1989) to estimate the average modulation rate of lowpass-noise modulators. Due to the auditory preprocessing all three definitions of the average envelope fluctuation rate predict hardly any change in rate for large bandwidths.

[4]Note that the hypothesis of a detrimental across-channel process as proposed in Hall *et al.* (1983) does only assume a constant reduction for the notched-noise data at the same masker spectrum level. If the same reduction had been assumed for all masker spectrum levels, the same BMLD would be predicted for the same diotic thresholds. This is not in agreement with the data in Hall *et al.* (1983): For example, the BMLD is about 12 dB at a masker spectrum level of 30 dB/Hz and no notch, whereas it amounts to only about 8 dB at a masker spectrum level of 50 dB/Hz at a notch width of about 200 Hz although both result in the same diotic thresholds. Thus, the notched-noise data in Hall *et al.* (1983) for the different spectrum levels indicate that the reduction in the BMLD due to the hypothesized across-channel process decreases as the masker spectrum level decreases.

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.

Bos, C. E., and de Boer, E. (**1966**). "Masking and discrimination," J. Acoust. Soc. Am. **39**, 708–715.

Bourbon, W. T., and Jeffress, L. A. (**1965**). "Effect of bandwidth of masking noise on detection of homophasic and antiphasic tonal signals," J. Acoust. Soc. Am. **37**, 1180–1181.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters," J. Acoust. Soc. Am. **110**, 1089–1104.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2003**). "The masking level difference for signals placed in masker envelope minima and maxima," J. Acoust. Soc. Am. **114**, 1557–1564.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2007**). "Individual differences in the masking level difference with a narrowband masker at 500 or 2000 Hz," J. Acoust. Soc. Am. **121**, 411–419.

Carlyon, R. P., Buus, S., and Florentine, M. (**1989**). "Comodulation masking release for three types of modulator as a function of modulation rate," Hear. Res. **42**, 37–46.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

de Boer, E. (**1966**). "Intensity discrimination of fluctuating signals," J. Acoust. Soc. Am. **40**, 552–560.

de Boer, E., and Bos, C. E. (**1962**). "On the concept of the critical band," 4th International Congress on Acoustics, Copenhagen, Denmark, H13, pp. 1–4.

Derleth, R. P., and Dau, T. (**2000**). "On the role of envelope fluctuation processing in spectral masking," J. Acoust. Soc. Am. **108**, 285–296.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Eddins, D. A., and Wright, B. A. (**1994**). "Comodulation masking release for single and multiple rates of envelope fluctuation," J. Acoust. Soc. Am. **96**, 3432–3442.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Hall, J. W., and Harvey, A. D. G. (**1984**). "NoSo and NoSπ thresholds as a function of masker level for narrowband and wideband masking noise," J. Acoust. Soc. Am. **76**, 1699–1703.

Hall, J. W., Tyler, R. S., and Fernandes, M. A. (**1983**). "Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking," J. Acoust. Soc. Am. **73**, 894–898.

Hohmann, V. (**2002**). "Frequency analysis and synthesis using a gammatone filterbank," Acta. Acust. Acust. **88**, 433–442.

Holube, I., Colburn, H. S., van de Par, S., and Kohlrausch, A. (**1995**). "Model simulations of masked thresholds for tones in dichotic noise maskers," J. Acoust. Soc. Am. **97**, 3411–3412.

Kollmeier, B., and Holube, I. (**1992**). "Auditory filter bandwidths in binaural and monaural listening conditions," J. Acoust. Soc. Am. **92**, 1889–1901.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).
Nitschmann, M., and Verhey, J. L. (**2006**). "Spektrale Effekte beim binauralen Hören (Spectral effects in binaural hearing)," in *Fortschritte der Akustik, DAGA 2006* [Deutsche Gesellschaft für Akustik e.V. (DEGA), Berlin], pp. 79–80.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., and Henning, G. B. (**1977**). "Stimulus variability and auditory filter shape," J. Acoust. Soc. Am. **62**, 649–664.
Patterson, R. D., and Nimmo-Smith, I. (**1980**). "Off-frequency listening and auditory-filter asymmetry," J. Acoust. Soc. Am. **67**, 229–245.
Rice, S. O. (**1954**). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York), pp. 133–294.
Sever, J. C., Jr., and Small, A. M., Jr. (**1979**). "Binaural critical masking bands," J. Acoust. Soc. Am. **66**, 1343–1350.
Sondhi, M. M., and Guttman, N. (**1966**). "Width of the spectrum effective in the binaural release of masking," J. Acoust. Soc. Am. **40**, 600–606.
van de Par, S., and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," J. Acoust. Soc. Am. **106**, 1940–1947.
Verhey, J. L. (**2002**). "Modeling the influence of inherent envelope fluctuations in simultaneous masking experiments," J. Acoust. Soc. Am. **111**, 1018–1025.
Verhey, J. L., and Zerbs, C. (**2001**). "Spektrale Selektivität und frequenzgruppenübergreifende Verarbeitung beim binauralen Hören (Spectral selectivity and across-frequency processing in binaural hearing)," in *Fortschritte der Akustik, DAGA 2001* [Deutsche Gesellschaft für Akustik e.V. (DEGA), Oldenburg], pp. 490–491.
Verhey, J. L., Rennies, J., and Ernst, S. M. A. (**2007**). "Influence of envelope distributions on signal detection," Acta. Acust. Acust. **93**, 115–121.
Wightman, F. L. (**1971**). "Detection of binaural tones as a function of masker bandwidth," J. Acoust. Soc. Am. **50**, 623–636.
Yama, M. F., and Robinson, D. E. (**1982**). "Comparison of frequency selectivity for the monaural and binaural hearing systems: Evidence from a probe-frequency procedure," J. Acoust. Soc. Am. **71**, 694–700.
Zerbs, C. (**2000**). "Modelling the effective binaural signal processing in the auditory system," Ph.D. thesis, physics department, Carl-von-Ossietzky-Universität, Oldenburg, Germany.
Zurek, P. M., and Durlach, N. I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.
Zwicker, E., and Henning, G. B. (**1984**). "Binaural masking-level differences with tones masked by noises of various bandwidths and levels," Hear. Res. **14**, 179–183.

Effects of bandwidth on auditory localization with a noise masker

Douglas S. Brungart[a] and Brian D. Simpson
Air Force Research Laboratory, 2610 Seventh Street, WPAFB, Ohio 45433

(Received 24 September 2008; revised 9 September 2009; accepted 14 September 2009)

Although high-frequency content is known to be critically important for the accurate location of isolated sounds, relatively little is known about the importance of high-frequency spectral content for the localization of sounds in the presence of a masker. In this experiment, listeners were asked to identify the location of a pulsed-noise target in the presence of a randomly located continuous noise masker. Both the target and masker were low-pass filtered at one of eight cutoff frequencies ranging from 1 to 16 kHz, and the signal-to-noise ratio was varied from −12 to +12 dB. The results confirm the importance of high frequencies for the localization of isolated sounds, and show that high-frequency content remains critical in cases where the target sound is masked by a spatially separated masker. In fact, when two sources of the same level are randomly located in space, these results show that a decrease in stimulus bandwidth from 16 to 12 kHz might result in a 30% increase in overall localization error. [DOI: 10.1121/1.3243309]

PACS number(s): 43.66.Qp [RYL] Pages: 3199–3208

I. INTRODUCTION

One of the most established principles in the study of spatial hearing is the important role that high-frequency auditory content plays in sound localization (Middlebrooks and Green, 1990). Although low-frequency content (below 3 kHz) is generally sufficient to accurately determine the left-right location of a sound source, accurate distinction between sounds in the front and rear hemispheres requires a bandwidth of at least 6–9 kHz, and optimal localization in elevation requires a stimulus bandwidth of as much as 13 kHz (King and Oldfield, 1997).

Despite the clear importance of high-frequency content for the accurate localization of isolated sounds, little is known about the importance of high-frequency content for the localization of sound sources in the presence of other interfering sounds. However, there is at least some reason to believe that high-frequency content might be substantially less important in multisource environments than in single-source environments because of the inherent fragility of the high-frequency cues listeners use for determining the elevations of sounds. These cues are derived from complex patterns of spectral "peaks" and "notches" that are introduced into the head-related transfer function (HRTF) by interactions of the acoustic stimulus with the complex geometry of the human pinna. When multiple sounds are present in the environment, it is possible that these peaks and notches could become filled in or distorted, thus reducing or eliminating the potential spatial information listeners might be able to extract from the high-frequency regions of these stimuli. Indeed, Good and Gilkey (1996) already showed that localization in the up-down and front-back dimensions degrades more rapidly with decreases in signal-to-noise ratio (SNR) than localization in the left-right dimension when a noise masker is introduced into the environment. At the same time, there is also some possibility that high-frequency sound might contribute to the disambiguation of the locations of multiple sound sources, and the larger head-shadowing effects that exist at high frequencies can, in some cases, lead to better detection performance of spatially separated high-frequency stimuli than for spatially separated low-frequency stimuli (Good *et al.*, 1997).

To this point, very little data are available to provide a detailed analysis of the impact of bandwidth on the localization of a target sound in a noise masker. In this paper, we present the results of an experiment that required listeners to localize a pulsed-noise target in the presence of a randomly located noise masker. The results are examined both in terms of the SNR of the stimulus and the low-pass filter cutoff frequency applied to both the target and masker sounds.

II. METHODS

A. Participants

A total of eight paid volunteer listeners participated in the experiment, including four males and four females. All had normal audiometric hearing, and their ages ranged from 19 to 23 (mean age 21.9 years).

B. Facility

The localization experiments were conducted by listeners standing in the center of the Auditory Localization Facility (ALF) at Wright-Patterson Air Force Base (Fig. 1). The ALF consists of a large geodesic sphere (4.3 m in diameter) suspended in an anechoic chamber to reduce reverberation and ambient noise levels. There are 277 Bose 11-cm full-range loudspeakers mounted at each of the structural vertices of the sphere, with an angular spacing of approximately 15° between each two loudspeaker locations. However, for the purposes of this experiment, only the 224 loudspeakers at

[a]Author to whom correspondence should be addressed. Present address: Walter Reed Army Medical Center, Washington, DC. Electronic mail: douglas.brungart@us.army.mil

FIG. 1. The Auditory Localization Facility where the localization experiments were conducted.

locations greater than −45° in elevation were used to generate target or masker stimuli. Each loudspeaker in the ALF facility is also fitted with a visual display consisting of a cluster of four independently addressable light emitting diodes (LEDs).

C. Stimuli

The target stimulus always consisted of a sequence of five brief pulses of white Gaussian noise. Each pulse was rectangularly gated to a length of 25 ms, with a 20-ms period of silence between each pulse.

The masker always consisted of a 250-ms burst of continuous white Gaussian noise that was statistically independent of the target and was gated on and off with a 25-ms $\cos^2$ ramp.

Within each stimulus presentation, the masker always started first, with the onset of the first target burst occurring 25 ms after the onset of the masker. The offset of the last target burst occurred 45 ms prior to the end of the target burst. This sequence is illustrated in Fig. 2.

The pulsed-noise target signal was selected because it was found to provide the best balance of high localization accuracy when the stimulus was presented in isolation and relatively easy segregation from the masker when the stimulus was presented in the presence of a 250-ms noise burst. However, it should be noted that the pulses did provide some brief glimpses where the masking sound could be heard without interference from the target, which may have provided the listener with more information about the location of the masking sound than would have been available in a stimulus where the target and masker sounds were gated on and off simultaneously.

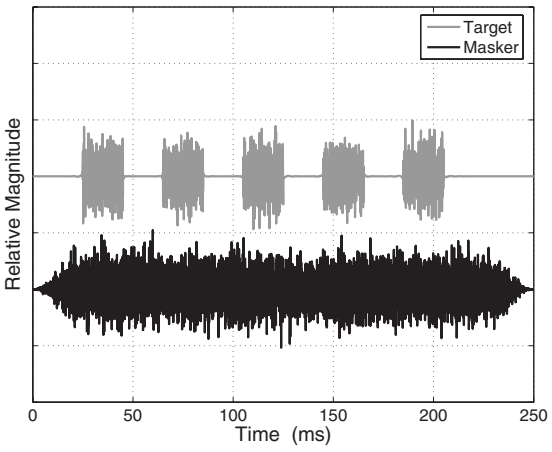


FIG. 2. Target and masker waveforms typical of those used in the experiment.

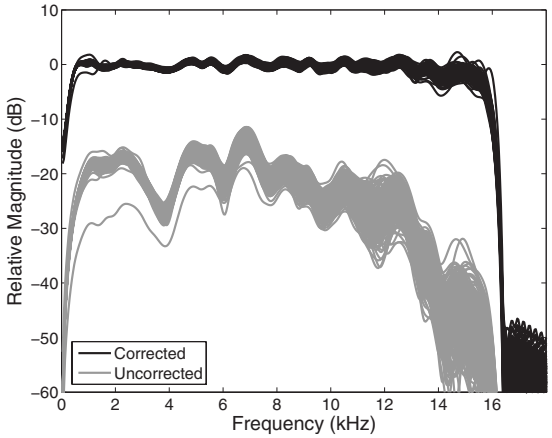


FIG. 3. Frequency responses of the 224 loudspeakers used in the experiment, as measured by a probe microphone (Etymotics ER-7) located exactly in the center of the sphere. The gray lines show the frequency responses of individual speakers prior to the inverse filtering, and the black lines show the frequency responses of the individual speakers after the inverse filtering.

For the purposes of this experiment, both the wideband target and masker noises were bandpass filtered by a fifth-order Butterworth filter with a low-frequency cutoff of 200 Hz and one of eight different high-frequency cutoff values: 1, 2, 4, 6, 8, 10, 12, and 16 kHz.

Prior to presentation in the ALF facility, the target and masking stimuli were inverse-filtered with two calibration filters. One was a 512-point finite impulse response (FIR) filter designed to match the frequency response of the individual loudspeaker to the response of the median loudspeaker in the 277-speaker ALF facility. The second was a 1024-point FIR filter designed to flatten the response of the median loudspeaker between the frequencies of 500 Hz and 15 kHz. This resulted in a loudspeaker response that varied by no more than ±1 dB at frequencies from 500 Hz to 11 kHz, and by no more than ±2.5 dB at frequencies up to 15 kHz (Fig. 3).

The presentation level of the masker was fixed at 65 dB sound pressure level (SPL) (as measured at the center of the sphere), and the SNR of the stimulus was adjusted by scaling the level of the target signal (as measured at the center of the sphere) to one of five levels relative to the masker (−12, −6, 0, +6, or +12 dB). A target-only control condition was also tested where the masker was eliminated and the target was set to a level of 65 dB SPL.

D. Procedure

The experimental data were collected with the listener standing on a platform in the center of the ALF facility while wearing a headband equipped with a six-degree-of-freedom

head-tracking device (Intersense IS-900). Prior to beginning a block of trials, the listener was asked to face a loudspeaker located directly in front of the sphere (0° azimuth, 0° elevation) and adjust the headband until the headtracker registered a location within ±5° of the loudspeaker location in both azimuth and elevation (this was indicated by the activation of a LED cursor on the front loudspeaker). Then a block of 20 trials was collected with the following procedure.

(1) The location of the listener's head was repeatedly measured at 200-ms intervals by the headtracker until two consecutive measurements were made that were separated by no more than 3° in great circle angle. This procedure ensured that the head was not moving faster than 15°/s at the stimulus onset.
(2) The target stimulus was presented from a location that was randomly selected from the 224 possible loudspeaker locations in the sphere and, when present, the masker was presented from a second location, different than that of the target, that was also randomly selected from the 224 possible loudspeaker locations.
(3) The listeners responded by moving a LED cursor to the perceived loudspeaker location of the target sound and pressing the trigger on a position-tracked response wand. This LED cursor was created by monitoring the orientation of the response wand and turning on the LED cluster located on the loudspeaker most closely matching the direction of the wand. Thus, the listeners could essentially make their response by pointing the wand in the direction of the perceived target sound and pressing the response button.
(4) The listener was given feedback about the actual location of the target sound by lighting a LED at the target location and presenting a 250-ms burst of noise from the target location.
(5) The listener acknowledged the feedback signal by holding his or her head steady and pressing the trigger on the response wand to hear the next stimulus presentation.

Although listeners were required to keep their heads still prior to each stimulus presentation, they were not asked to reorient to a fixed point prior to each trial. Instead, the head orientation at the time of each stimulus presentation was measured and, in each case, the locations of the target and masking sounds were calculated relative to the orientation of the listener's head. These "head-relative" coordinates were then used in all subsequent analyses of the localization data.

Within each block of 20 trials, the bandwidth of the stimulus was held constant, but the SNR value and the target and masker locations were selected randomly in each trial. In all, a total of 48 blocks of trials were collected for each listener, with a total of 20 trials collected at each possible combination of six SNR values [−12, −6, 0, +6, +12, and no masker (NM)] and eight high-pass cut-off conditions (1, 2, 4, 6, 8, 10, 12, and 16 kHz). Thus, a total of 7680 trials were collected across the eight listeners in the experiment.

III. RESULTS

The results of the experiment were analyzed using the interaural-polar coordinate system illustrated in Fig. 4

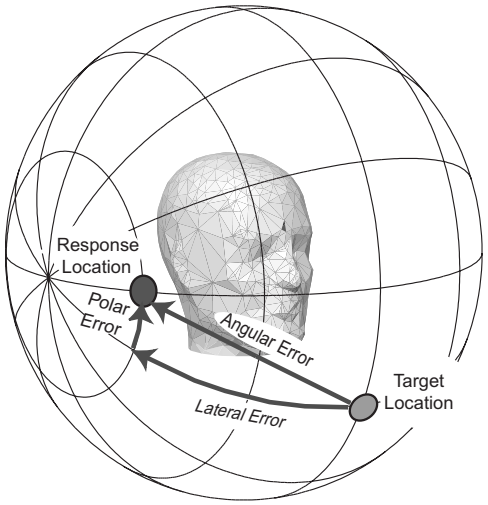


FIG. 4. Interaural-polar coordinate system used to measure angular, vertical, and polar errors in the experiment. Figure adapted from Macpherson and Middlebrooks (2003).

(Macpherson and Middlebrooks, 2003). Three different types of localization error metrics were calculated. The first was the *angular error*, defined as the great circle error between the location of the stimulus and the location of the response. This is an overall measure of localization accuracy that incorporates errors in both azimuth and elevation. The second was the *lateral error*, defined as the difference between the angle of the target location relative to the listener's median sagittal plane and the angle of the response location relative to the listener's median sagittal plane. This error metric primarily reflects localization errors in azimuth related to interaural time and intensity differences. The third was the *polar error*, defined as the difference between the angles of the target and masker when both locations are projected into the listener's median sagittal plane. This error metric primarily reflects front-back confusions and localization errors in elevation that are generally related to spectral filtering effects by the outer ear or pinnae. However, the polar angle error is not meaningful for responses near ±90° in azimuth, where a very small change in location can produce a very large polar angle change. Thus, in this experiment, trials where the absolute value of the lateral angle exceeded 75° for either the stimulus location or the response location were excluded from the analysis of polar error. The percentage of front-back confusions was also calculated in each condition. Front-back confusions were defined as trials where the actual target location was in the front hemisphere and the response location was in the rear hemisphere (or vice versa) and the absolute polar angle error was greater than 30° (to avoid counting trials where the target and response were located near the boundary between the front and rear hemispheres).

As would be expected, the results of the experiment show that all four of these error metrics varied systematically with both the SNR of the target and the low-pass cutoff frequency (LPF) of the stimulus. This is illustrated in Table I, which shows the results of two-factor, within-subjects analyses of variance (ANOVAs) conducted on the individual scores of the eight subjects for each type of error. As indicated in the table, these ANOVAs showed that the main ef-

TABLE I. Within-subjects analysis of variance on the factors of SNR and LPF for each error metric tested in the experiment.

Error metric	SNR	LPF	SNR*LPF
Angular error	$F_{(5,35)}=223.98$, $p<0.001$	$F_{(7,49)}=54.972$, $p<0.001$	$F_{(35,245)}=2.472$, $p<0.001$
Lateral error	$F_{(5,35)}=94.075$, $p<0.001$	$F_{(7,49)}=24.686$, $p<0.001$	$F_{(35,245)}=1.857$, $p<0.004$
Polar error	$F_{(5,35)}=140.64$, $p<0.001$	$F_{(7,49)}=50.10$, $p<0.001$	$F_{(35,245)}=2.82$, $p<0.001$
Front-back confusions	$F_{(5,35)}=83.15$, $p<0.001$	$F_{(7,49)}=19.14$, $p<0.001$	$F_{(35,245)}=1.87$, $p<0.003$

fects of both SNR and LPF were statistically significant at the $p<0.001$ level in all cases, and that the interaction between SNR and LPF was significant at the $p<0.005$ level.

Because of the large number of conditions tested in the experiment, it is helpful to discuss the different aspects of the results one at a time. First, consider the effects of bandwidth on performance in the control condition with no masking signal (Fig. 5). When a full 16-kHz bandwidth signal was presented in isolation, overall angular accuracy was quite good, with an average error on the order of 10°. However, the localization errors systematically increased as the bandwidth of the signal decreased, reaching a maximum of 50° when the cutoff frequency was 1 kHz. While undeniably poor, this level of localization accuracy is well above chance performance in the task (roughly 90° in angular error). The reason for this is clearly seen in the second panel of Fig. 5, which shows that bandwidth limitations had only a very small impact on the listener's ability to determine the left-right position of a signal. Even in the 1 kHz condition, listeners were able to determine the left-right position of the target stimulus within 10°–12°. This is consistent with earlier findings (King and Oldfield, 1997; Wightman and Kistler, 1992) that have shown that localization accuracy in the left-right dimension depends primarily on low-frequency interaural time delays that are largely unaffected by the low-pass filtering of the stimulus.

Low-pass filtering does, however, have a profound effect on localization accuracy in the vertical dimension. The third panel of Fig. 5 shows the mean polar angle error as a function of stimulus bandwidth. The results in the figure indicate that there is a significant increase in polar localization error with decreasing bandwidth as soon as the bandwidth is reduced below 16 kHz (one-tailed, Bonferroni-corrected t-test, $p<0.05$).[1] However, most of the increase in polar angle error occurs when the bandwidth is reduced below 10 kHz. The reason for this is clearly seen in the last panel of Fig. 5, which shows that the number of front-back confusions did not increase significantly from the 16-kHz level until the bandwidth of the stimulus decreased to 8 kHz (one-tailed Bonferroni-corrected t-test, $p<0.05$). Again, this is consistent with the results of King and Oldfield (1997), which showed that most of the information for resolving front-back confusions was contained in the 6–9-kHz frequency range. The most likely explanation for this is that the front-back confusions are resolved by effects related to the acoustic shadow of the pinna flange, which has mean dimensions of 6.4×2.9 cm^2 (Algazi *et al.*, 2001) and thus should start to have a significant acoustic impact on the HRTF for sounds in the 5–10-kHz range (where its dimensions are on the order of a single wavelength of sound).

Next consider the effect of SNR on localization accuracy in each bandwidth condition of the experiment. Each panel of Fig. 6 shows localization performance as a function of SNR. The lines within the panels each show performance for a different low-pass filter cutoff frequency. In order to make the figures more legible, the eight low-pass conditions have been binned into five frequency ranges, as indicated by the legends in the figure.

The top panel shows overall angular error as a function of SNR. As would be expected, there is a systematic decrease in angular error with increasing SNR in all of the bandwidth conditions tested. It is notable that, at stimulus bandwidth values greater than 4 kHz, there was a significant decrease in performance between the +12 dB SNR condition and the "no masker" condition (as indicated by the non-overlapping error bars in Fig. 6), suggesting that the presence of a masker has a detrimental effect on localization accuracy even at relatively high SNR values.

The overall angular error results also show that localization accuracy systematically decreased with the bandwidth of

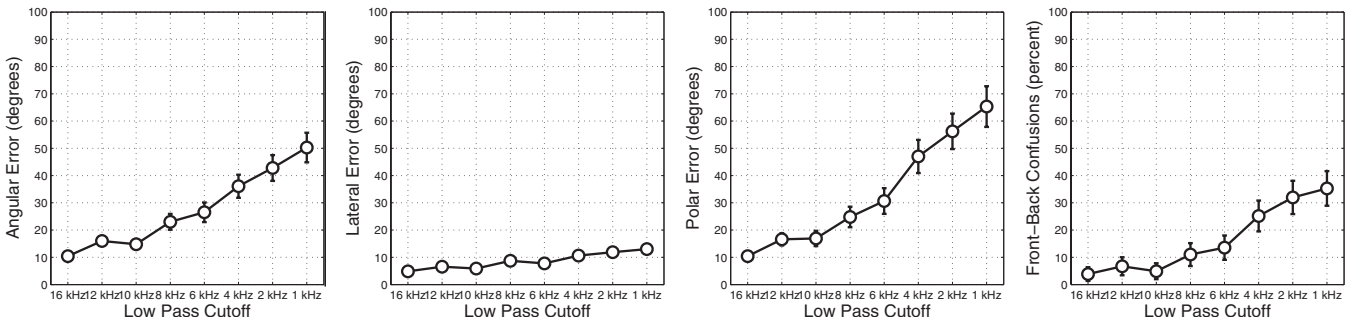


FIG. 5. Localization errors in the no masker control conditions of the experiment. The error bars show ±2.06 standard errors around each data point. Thus, data points with non-overlapping error bars represent pairs of points where the localization error increased significantly with decreasing bandwidth (at the $p<0.05$ level) in a one-tailed t-test after Bonferroni correction for the 28 possible point-to-point comparisons in each data panel (see footnote [1]).

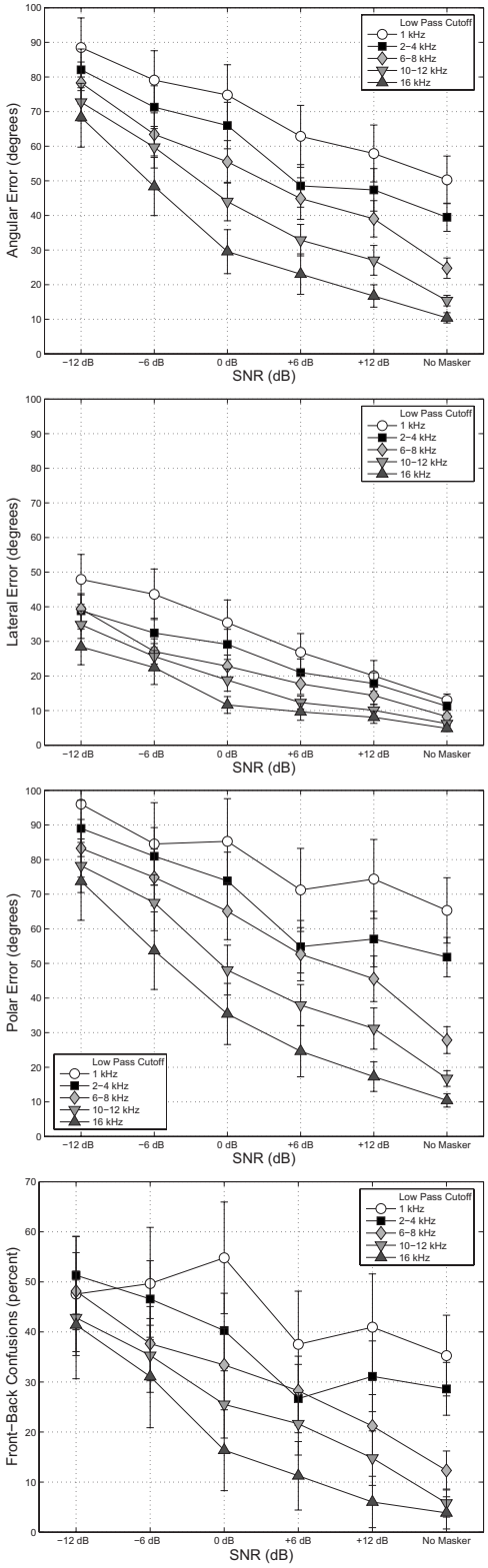


FIG. 6. Localization errors as a function of SNR for ear bandwidth condition of the experiment. The error bars show ±2.61 standard errors around each data point. Thus, data points with non-overlapping error bars represent pairs of points where the localization error increased significantly with decreasing bandwidth or SNR (at the $p < 0.05$ level) in a one-tailed t-test after Bonferroni correction for the 435 possible point-to-point comparisons in each data panel.

the stimuli. At SNR values of 0 dB or higher, the performance curves for the different bandwidth conditions are roughly parallel, suggesting that the decrease in localization performance due to reduced bandwidth is roughly independent of the decrease in localization performance due to reduced SNR in this region. However, at SNR values less than 0 dB, there is a knee in the curve in the 16-kHz bandwidth condition, suggesting that the performance benefits of the 12–16-kHz frequency band decrease more quickly than those in the other frequency bands at SNR values below 0 dB. The significance of this knee in the curve is verified by the significant interaction between bandwidth and SNR that was identified in the ANOVA for angular error shown in the top row of Table I.

At least in part, this result can be explained by the lateral localization errors depicted in the second panel of Fig. 6. This figure shows that left-right localization accuracy in the 16-kHz condition increases steadily as the SNR increases for SNR values below 0 dB, but plateaus at about 10° of error at SNR values above 0 dB. Thus, it seems that decreasing SNR only had a negative impact on polar error at SNR values greater than 0 dB in the 16-kHz bandwidth condition, but that it impacted both lateral and polar errors at SNRs less than 0 dB. This helps explain the knee in the 16-kHz bandwidth curve seen in the top panel of Fig. 6. Note that these lateral angle results are consistent with the findings of Good and Gilkey (1996), which also showed that left-right localization errors are relatively small at positive SNR values but increase with decreasing SNR at negative SNR values.

Although it is clear that bandwidth has some effect on lateral localization errors, this effect is small relative to the effect it has on polar localization errors, shown in the third panel of Fig. 6. These polar errors dramatically increased with decreasing bandwidth, and, in the 1-kHz bandwidth condition, the polar angle errors were as large as approximately 65° even when no masker was present. These polar angle errors also systematically decreased with increasing SNR across the whole range of SNRs tested. Again, this is consistent with the findings of Good and Gilkey (1996), which showed localization in the up/down and front/back dimensions was substantially more sensitive to changes in SNR at positive SNR values than was localization in the left/right dimension.

The final panel of Fig. 6 shows the percentage of front-back confusions for each combination of SNR value and stimulus bandwidth tested in the experiment. As has been the case in previous experiments (King and Oldfield, 1997), these results show that the percentage of front-back confusions increases substantially when the stimulus bandwidth decreases. In general, however, the effects of increased bandwidth were much smaller at negative SNR values than they were at positive SNR values.

Figure 7 shows a more detailed analysis of performance in each condition of the experiment as a function of the bandwidth of the stimulus, with a different SNR condition shown in each column of the figure. From this figure, it is apparent that stimulus bandwidth had a substantial impact on all four metrics of localization performance in almost every SNR condition tested. There was a significant interaction between SNR and the low-pass cutoff frequency in all cases (as indicated by the ANOVAs in Table II), which was primarily due to a floor effect that limited the localization errors at low

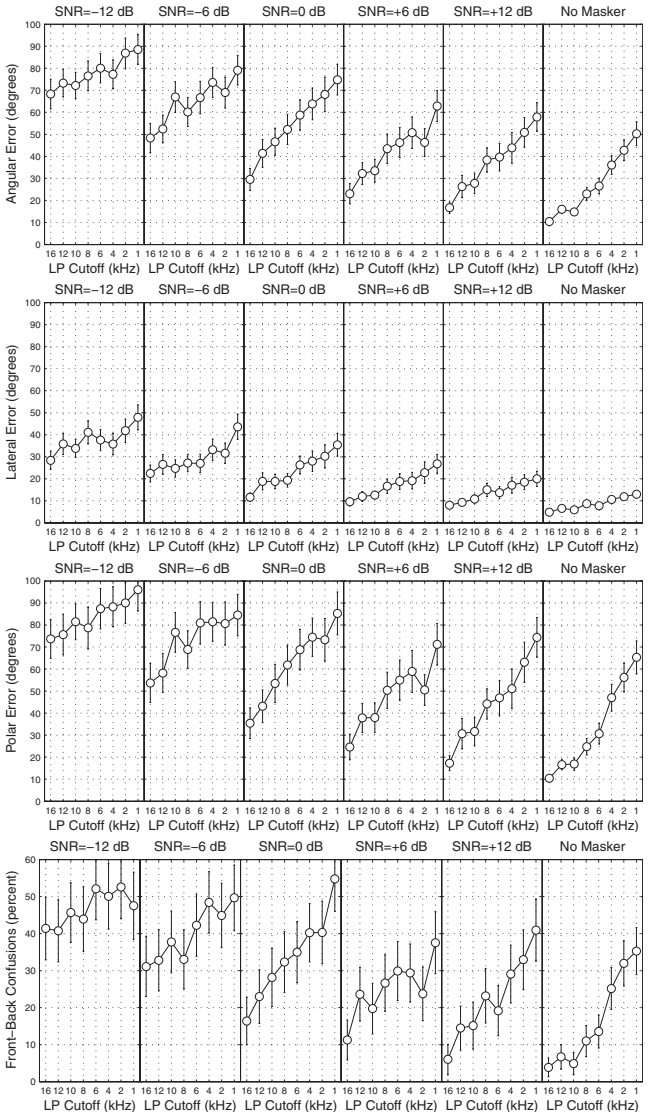


FIG. 7. Localization errors as a function of bandwidth in each SNR condition of the experiment. The error bars show ±2.21 standard errors around each data point. Thus, data points with non-overlapping error bars represent pairs of points where the localization error increased significantly with decreasing SNR (at the $p<0.05$ level) in a one-tailed t-test after Bonferroni correction for the 28 possible point-to-point comparisons in each data panel.

bandwidths to chance performance in angular and polar error (approximately 90°) and chance performance in front-back judgments (50%).

One notable aspect of the data is that the very high-frequency range from 12 to 16 kHz remained important for accurate localization performance even when a low-level masker sound was present in the stimulus. The results of one-tailed, Bonferroni-corrected t-tests reveal that a reduction in bandwidth from 16 to 12 kHz resulted in a significant increase in overall angular error in the 0 and +12 dB SNR conditions, and a significant increase in polar angle error in the +6 and +12 dB SNR conditions. However, the importance of these high-frequency components did diminish as the level of the masker increased. At SNR values below 0 dB, no significant differences were observed between the 12- and 16-kHz conditions.

A final aspect of interest in the data is an analysis of the *types* of errors the listeners made in their responses in each condition of the experiment. King and Oldfield (1997), for example, showed that all three of their listeners tended to compress their mean elevation responses into the region from −40° to −20° when the bandwidth of the stimulus decreased below 9 kHz. Figure 8 provides a similar analysis of the distribution of the polar angles for those trials where the listeners failed to accurately identify the location of the target (defined here by angular errors greater than 15°) in each SNR and bandwidth condition of the experiment. The sizes of the circles in each location represent the proportion of incorrect responses located at that particular location relative to the listener's head (illustrated pictorially by the rightward-facing head at the center of each panel). Each different radius represents a different SNR condition, with the outermost (open) circles representing the NM condition, and the innermost (black) circles representing the −12 dB SNR condition (as indicated in the legend). Because the target and masker locations were randomly located over the surface of the sphere, these distributions give a good indication of overall response bias in the polar dimension for each condition of the experiment.

In the 16-kHz bandwidth condition (leftmost panel of the figure), the majority of the incorrect responses in the no masker condition (open circles) were made at locations directly above and/or slightly behind the head. As the SNR in that condition decreased (moving from the outermost to innermost ring of circles), there was a slight tendency for the responses to shift down and to the rear, toward the horizontal plane. This same shift in responses down and to the rear is also seen as the bandwidth of the stimulus decreases. At bandwidths in the range 2–6 kHz, most of the incorrect responses were at polar locations directly behind the listener, and, in the 1-kHz condition, most of the responses were behind and slightly below the listener, similar to the location reported by King and Oldfield (1997).

It is quite notable that, even at low SNR values, there was a strong correlation between the bandwidth of the signal and the likelihood that listeners would make responses at high elevations: responses at polar angles above the head were common at low-pass cutoff frequencies of 10 kHz and above, but were very rare at lower cutoff frequencies. This trend is illustrated in Fig. 9, which shows how the mean angle between the polar angle of the incorrect responses and the horizontal plane varied as a function of stimulus bandwidth. The solid symbols show performance in the no-masker condition, which varied from +31° in the 16-kHz condition to −22° in the 1-kHz condition [$F_{(7,49)}=20.4$, $p<0.001$]. The open symbols show performance in the masker conditions, where the same general trend is seen but the mean responses varied over a much smaller range than in the no-masker condition. The masker data have been pooled across all SNR values because a within-subjects ANOVA on the individual subject data showed that the main effect of bandwidth was significant [$F_{(7,49)}=10.62$, $p<0.001$], but that SNR was not significant, and there was no significant interaction between SNR and the low-pass cutoff frequency. These results suggest that high-frequency content has a

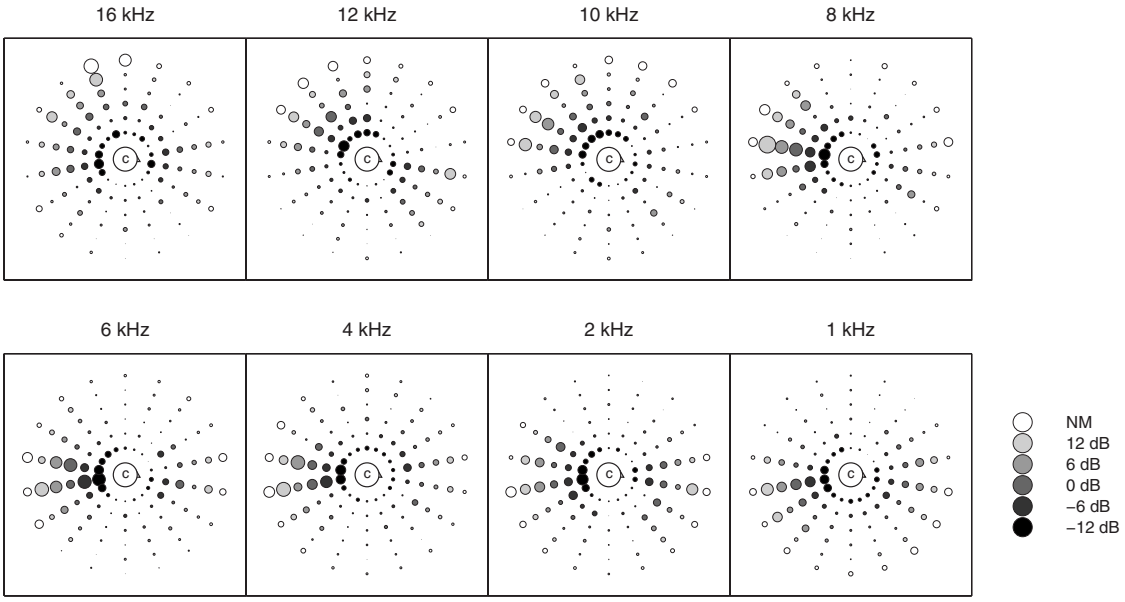


FIG. 8. Distribution of polar error responses in trials where the listener incorrectly identified the location of the target. The diameter of each circle is proportional to the percentage of incorrect responses that occurred in that particular polar error region relative to the listener's head (collapsed across all the lateral angles). The six different radii each represent a different SNR condition, with the outermost (open) circles representing the no masker condition, and the innermost (black) circles representing the −12 dB SNR condition (as indicated in the legend).

strong influence on localization bias in the vertical dimension for unmasked target sounds, and that this effect is weakened somewhat but not eliminated when a masker is added to the stimulus.

IV. DISCUSSION

A. Signal detectability

In choosing the pulsed-noise stimuli for the target signal, we were motivated by a desire to find stimuli that were as robustly localizable as a broadband noise burst but at the same time easy to distinguish from a noise masker. However, because the pulsing attribute that distinguished the target from the masker was, in part at least, reliant on sharp onsets and offsets that may in some way have become blurred when the stimulus was passed through a low-pass filter, there was some concern that the target stimuli may have become harder to detect in some of the low-pass filtered conditions. If this were true, then it is possible that some of the reductions in localization accuracy that occurred with decreasing bandwidth in this experiment might have been the result of increased difficulty in detecting the target signal, rather than increased difficulty in localizing a detectable target signal.

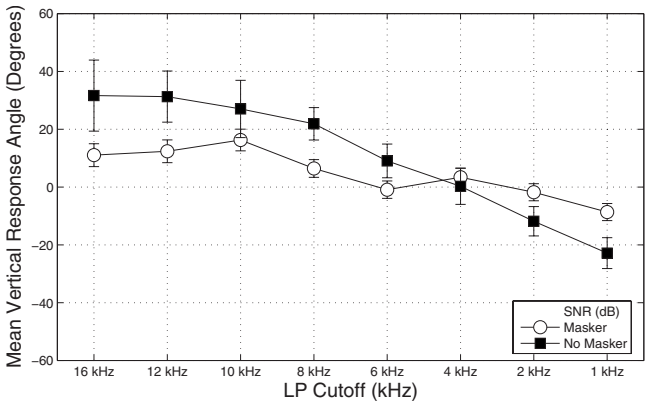


FIG. 9. Mean vertical angle of incorrect responses in the experiment as a function of bandwidth. The error bars show the 95% confidence intervals around each data point.

In order to address this potential concern, a control experiment was conducted to assess the detectability of the target signal as a function of the bandwidth of the combined stimulus. This control experiment was conducted by nine listeners (including the seven from the primary experiment) while wearing headphones (Beyerdynamic DT 990 Pro) in a quiet listening booth. The experiment was a cued, two-interval forced choice experiment, where each listener heard three discrete stimulus intervals on each trial: a reference interval containing only an example of the pulsed-noise target signal, followed by two stimulus intervals; one containing just the 250-ms noise masker and one containing the noise masker plus the pulsed-noise target.

Prior to stimulus presentation, the target and masking signals were convolved with individualized HRTFs corresponding to two different locations that were randomly selected from the 224 speaker locations used to generate the free-field stimuli in the experiment. These HRTFs were measured in the Auditory Localization Facility prior to the experiment, and localization testing conducted prior to the experiment verified that listeners using these individualized HRTFs were able to accurately localize a 250-ms burst of broadband noise with an average angular error of 16° versus 12° for a free-field loudspeaker presentation of the same signal.

These virtual stimuli were presented at a comfortable listening level (roughly 65 dB SPL), with a 250-ms interval of silence between each interval. As in the primary experiment, these stimuli were low-pass filtered at one of eight different cutoff frequencies, ranging from 1 to 16 kHz, and

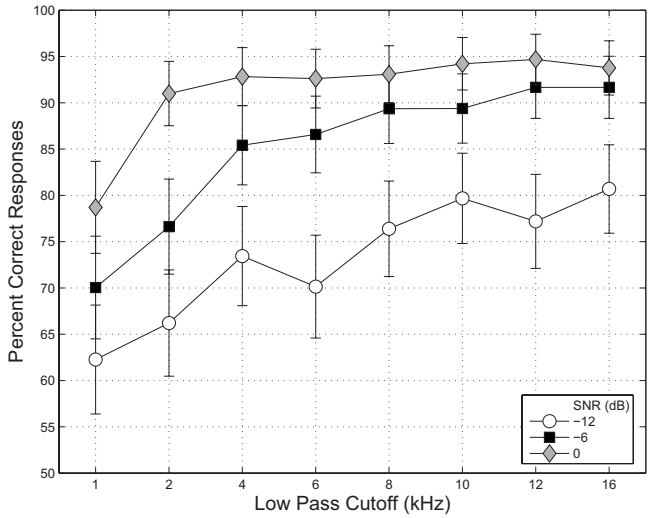


FIG. 10. Results of control experiment examining detectability of each stimulus type as a function of SNR and bandwidth. The error bars show ±2.52 standard errors around each data point. Thus, data points with non-overlapping error bars represent pairs of points where the probability of correct detection decreased significantly with decreasing bandwidth (at the $p<0.05$ level) in a one-tailed t-test after Bonferroni correction for the 235 possible point-to-point comparisons in the figure.

they were presented with the target signal scaled to one of three different SNR values relative to the masker (−12, −6, or 0 dB). After hearing the sequence of three intervals, the listeners were asked to identify which interval (the second or the third) contained the pulsed target. Trial-by-trial feedback was provided.

The stimuli were presented in 96-trial blocks, with a random bandwidth and SNR selected on each new trial. Each of the nine subjects participated in a total of 48 trials at each of the 24 different combinations of target bandwidth and SNR tested in the experiment, for a total of 1152 trials per subject.

The results of the experiment are shown in Fig. 10. A two-factor, within-subjects analysis of variance of the arcsine-transformed results for the individual listeners indicates that the main effects of SNR and bandwidth were both significant at the $p<0.05$ level, but that their interaction was not significant [$F_{(14,112)}=1.706$, $p=0.064$]. However, from the results shown in the figure, it is clear that SNR had a more consistent impact on detection performance than stimulus bandwidth, and that bandwidth really only had a substantial impact on stimulus detectability at the lowest bandwidth conditions tested, namely, those with cutoff frequencies at or below 4 kHz. This is confirmed by the results of one-tailed, Bonferroni-corrected t-tests (as indicated by the error bars in the figure), which reveal that there were no significant improvements in detection performance when the stimulus bandwidth increased above 4 kHz at a fixed SNR value.

These results clearly demonstrate the target stimuli were highly detectable (>90%) in virtually all of the conditions with stimulus SNR values greater than 0 dB. Since this is the region where bandwidth had the greatest impact on localization accuracy, we can safely conclude that the effects of bandwidth on localization performance in this experiment cannot be explained simply by differences in the detectability of the target stimuli. However, there is some evidence that target detectability did have an impact on performance in the lowest bandwidth conditions tested, particularly in the 1- and 2-kHz bandwidth conditions at negative SNR values. In those conditions, it is quite possible that slightly better localization performance could have been obtained with a different target signal that was more easily distinguished from the noise masker. However, the room for improvement is limited, as localization performance in those conditions was only modestly worse than in the 4-kHz bandwidth condition, where the stimulus was nearly as detectable as the full bandwidth stimulus.

B. Response biases introduced by the presence of a masker

Many previous studies that have examined auditory localization accuracy in the presence of a simultaneous masker have reported the presence of a systematic response bias that caused the listener's responses to be shifted either towards the location of the masker (a "pulling" effect) or away from the location of the masker (a "pushing" effect) (see Best *et al.*, 2005 for an overview of the literature on this subject). Figure 11 shows an analysis of this type of response bias across all the masking conditions tested in the experiment. The figure simply plots the proportion of trials in which the angular distance between the response location and the masker location was shorter than the angular distance between the actual target location and the masker location [comparable to the approach taken by Good and Gilkey (1996)]. Note that trials where the target location exactly matched the masker location have been removed from this analysis. Thus a value of 0.5 indicates that there was no response bias, a value greater than 0.5 indicates that the responses were, on average, pulled toward the masker, and a proportion less than 0.5 indicates that the responses were, on average, pushed away from the masker.

The results shown in Fig. 11 suggest that the magnitude of this response bias did indeed change systematically with the bandwidth and SNR. The significance of this effect was confirmed in a within-subjects ANOVA on the individual subject results, which showed that the main effects of bandwidth [$F_{(7,49)}=6.877$] and SNR [$F_{(4,28)}=14.454$] were both significant at the $p<0.001$ level, and that the interaction between the two was significant at the $p<0.05$ level [$F_{(28,196)}=1.537$].

In general, there was a strong response bias toward the masker when the SNR was very low (−12 dB), which may suggest the listeners tended to respond in the masker location when they were unable to clearly hear the location of the target. However, this pulling bias decreased as the SNR of the stimulus increased, particularly in cases where the bandwidth was relatively high (>10 kHz). And, when the SNR was +6 or +12 dB and the bandwidth was 16 kHz, the results indicate that there was actually a significant pushing response bias away from the location of the masker. These results may provide some insights into the discrepancies that have been observed in previous studies that have examined response biases with a noise target and a noise masker, which

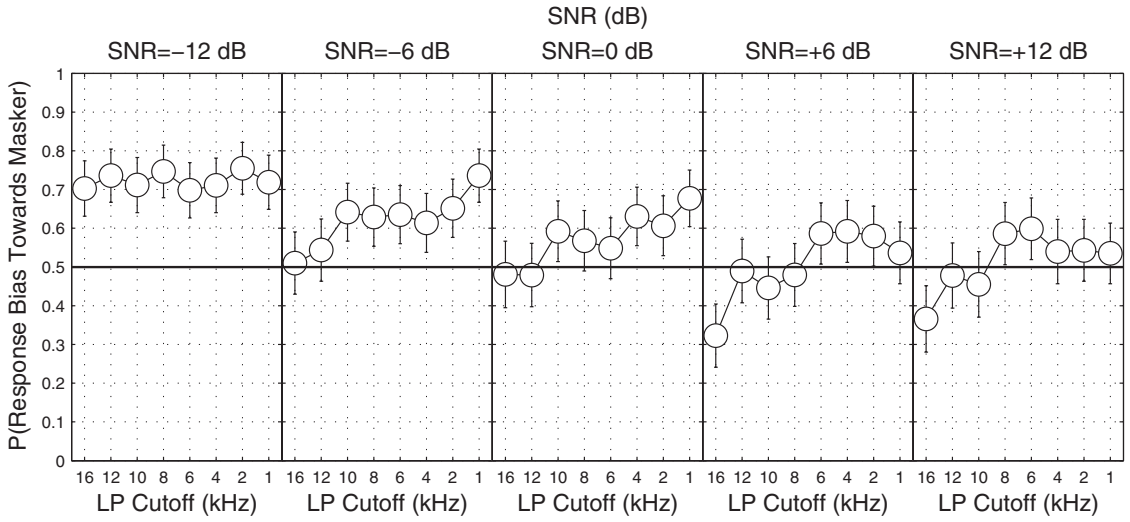


FIG. 11. This figure reports the magnitude of the response bias toward the location of the masker, measured by the proportion of trials in which the angular distance between the response location and the masker was shorter than the angular distance between the target location and the masker. Thus, proportions greater than 0.5 indicate a "pulling" bias toward the masker, proportions less than 0.5 indicated a "pushing" bias away from the masker, and proportions equal to 0.5 indicate no response bias. The error bars show the 95% confidence intervals around each data point.

have variously reported conflicting results of either a pulling effect (Butler and Naunton, 1964), a pushing effect (Best *et al.*, 2005; Braasch and Hartung, 2002; Good and Gilkey, 1996), or no bias effect at all (Langendijk *et al.*, 2001; Lorenzi *et al.*, 1999). The results of this experiment suggest that pushing effects tend to occur in high-bandwidth, high-SNR situations where localization is relatively easy. In contrast, pulling effects are more prevalent in low-SNR, low-bandwidth situations where the localization task is relatively difficult.

C. High-frequency head shadowing and its effect on ipsilateral-ear SNR

When this experiment was conceived, there was an expectation that distortions in the high-frequency peaks and notches of the HRTF would cause the high-frequency spectral components of a target signal to have less impact on localization accuracy in the presence of a noise masker than they do for the localization of a target signal in quiet. However, the results of this experiment suggest that high-frequency components may actually provide *more* benefit in the presence of a noise masker, in part at least due to an apparent ceiling effect in localization accuracy for bandwidths above 10 kHz. One possible explanation for the importance of high-frequency spectral energy in the presence of a masker is related to the much larger head-shadowing effects that occur for high-frequency sounds. These head-shadowing effects can cause the signal at the ear closer to the target sound (known as the ipsilateral ear) to actually have a higher SNR value at high frequencies than at low frequencies (where the effects of head shadowing are less profound.) This notion is supported by the results of Gilkey *et al.* (1995), in which detection thresholds were found to be 6 dB lower with high-frequency stimuli than with lower-frequency stimuli.

The size of this effect is quantified in Fig. 12, which shows the mean SNR at the ipsilateral ear as a function of frequency for a randomly located target sound masked by a randomly located masker sound. These SNR values were calculated from a set of 5000 simulated trials in which target and masking locations were selected randomly in accordance with the distribution of loudspeaker locations used in the experiment. In each case, the level of the signal in the ear closest to the target was calculated from a HRTF measured for that stimulus location (belonging to author D.S.B.), and this level was compared to the level of the masker in that same ear using the HRTF for the masker location. The results in Fig. 12 show the mean SNR resulting from this calculation. These results show that the mean SNR at the ipsilateral ear is indeed 2–3 dB higher in the 12–16-kHz frequency band than in the lower frequency regions.

In light of earlier results that have shown that spectral cues at the ipsilateral ear tend to dominate the perception of stimulus elevation (Macpherson and Sabin, 2007), these re-

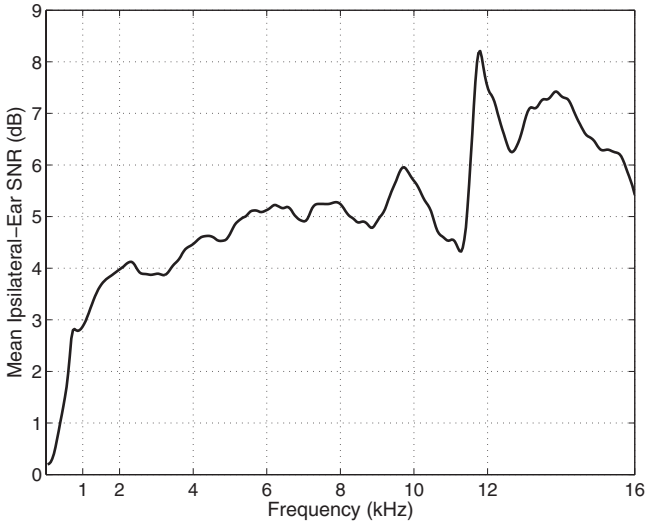


FIG. 12. Mean SNR value in the ipsilateral ear calculated from the HRTFs selected in 5000 simulated trials. See text for details.

sults suggest that the availability of a higher SNR at high frequencies might indeed at least partially account for the relatively large reduction in polar localization error provided by the 12–16-kHz frequency region in the multisource conditions of this experiment.

V. CONCLUSIONS

This experiment examined auditory localization performance as a function of stimulus bandwidth for stimulus conditions that involved a pulsed-noise target and continuous noise masker. In general, the results of the experiment can be summarized as follows.

- Left-right localization accuracy is roughly independent of the bandwidth of the stimulus and masker in all SNR conditions.
- Front-back localization accuracy requires a minimum bandwidth of approximately 10 kHz for target sounds presented in quiet, but is substantially improved by the addition of high-frequency components of up to 16 kHz in the presence of a noise masker.
- Vertical polar angle localization accuracy can benefit from the presence of frequency components up to 16 kHz in quiet, but it is improved much more by the presence of these high-frequency components when a noise masker is present.
- As a general rule, broadband stimuli tend to be much easier to localize in the presence of noise than narrowband stimuli. Lateral angle localization is relatively robust to the presence of noise even when the bandwidth is restricted, but a broadband stimulus seems to be critical in order to obtain accurate front/back and up/down localization in the presence of a noise masker.

All of these findings are important to some extent for advancing our understanding of auditory localization in noise. But as a parting comment, it is worthwhile noting that this last finding has important implications when considering the use of virtual audio displays in applications such as airborne traffic warning indicators for aircraft pilots. In these applications, and particularly in those applications where it may be necessary for the listener to attend to more than one virtual sound simultaneously, it may be critically important for system designers to resist the urge to cut costs by restricting system bandwidth, because even small decreases in bandwidth seem to have the potential to result in large increases in localization error. For example, in the case where two randomly located sounds are presented at the same level (i.e., 0 dB SNR), the results of this experiment suggest that the overall angular error for localizing one of the two signals might increase nearly 50% (from roughly 30° to roughly 45°) when the bandwidth of the system is reduced from 16 to 12 kHz. Further research is now needed to determine whether this additional benefit of high frequencies is restricted to the case where the high-frequency portion of the target signal contains spatial information that exactly matches the HRTFs of the listener's own ears (Wenzel and Foster, 1993). On the one hand, a negative result might imply that high bandwidth signals are not needed in applications that rely on generalized HRTFs. However, on the other hand, such a result may also carry with it the direct implication that individualized HRTFs are even *more* important for obtaining good localization performance in virtual audio display applications with more than one simultaneous sound.

[1]If the mean errors are assumed to be normally distributed and roughly equal across the different bandwidth conditions, then the difference between any two mean values can be roughly estimated by a Gaussian variable with variance σ equal to the square root of two times the standard error of the mean measurement. In a one-tailed t-test, a difference is assumed to be significant at the $p<0.05$ level if the difference between the means exceeds 1.64σ, which is the critical value that a random Gaussian variable will only exceed 5% of the time. However, in each panel of this graph there are 28 possible comparisons between pairs of bandwidth values $(N_*(N-1)/n!)$ where we might want to determine if performance decreased with decreasing bandwidth. Thus we need to find the critical value that a Gaussian variable will only exceed with probability $p=0.05/28$. This value is 2.91σ. Thus, the critical difference in the means between two data points required for significance at the $p<0.05$ level is $2.91*2_*^{0.5}\rho$, where ρ is the standard error of the mean estimate for each data point. Thus, the point where non-overlapping error bars is significant is half this value, or 2.06ρ.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF Database," in Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, (**2001**), pp. 99–102.

Best, V., Carlile, S., Jin, C., and van Shaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am. **118**, 353–363.

Braasch, J., and Hartung, K. (**2002**). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane. I. Psychoacoustical data," Acta. Acust. Acust. **88**, 942–955.

Butler, R., and Naunton, R. (**1964**). "Role of stimulus frequency and duration in the phenomenon of localization shifts," J. Acoust. Soc. Am. **36**, 917–922.

Gilkey, R., Good, M., Ericson, M., Brinkman, J., and Stewart, J. M. (**1995**). "A pointing technique for rapidly collecting localization responses in auditory research," Behav. Res. Methods Instrum. Comput. **27**, 1–11.

Good, M. D., and Gilkey, R. H. (**1996**). "Sound localization in noise: The effect of signal-to-noise ratio," J. Acoust. Soc. Am. **99**, 1108–1117.

Good, M. D., Gilkey, R. H., and Ball, J. M. (**1997**). "The relation between detection in noise and localization in noise in the free field," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. Anderson (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 349–376.

King, R., and Oldfield, S. (**1997**). "The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional auditory displays," Hum. Factors **39**, 287–295.

Langendijk, E. H. A., Wightman, F. L., and Kistler, D. J. (**2001**). "Sound localization in the presence of one or two distracters," J. Acoust. Soc. Am. **109**, 2123–2134.

Lorenzi, C., Gatehouse, S., and Lever, C. (**1999**). "Sound localization in noise in normal-hearing listeners," J. Acoust. Soc. Am. **105**, 1810–1820.

Macpherson, E. A., and Middlebrooks, J. C. (**2003**). "Vertical-plane sound localization probed with ripple-spectrum noise," J. Acoust. Soc. Am. **114**, 430–445.

Macpherson, E. A., and Sabin, A. T. (**2007**). "Binaural weighting of monaural spectral cues for sound localization," J. Acoust. Soc. Am. **121**, 3677–3688.

Middlebrooks, J., and Green, D. (**1990**). "Directional dependence of interaural envelope delays," J. Acoust. Soc. Am. **87**, 2149–2162.

Wenzel, E., and Foster, S. (**1993**). "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," in Applications of Signal Processing to Audio and Acoustics (1993), Final Program and Paper Summaries, 1993 IEEE Workshop pp. 102–105.

Wightman, F., and Kistler, D. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1660.

Amplification of interaural level differences improves sound localization in acoustic simulations of bimodal hearing

Tom Francart[a] and Tim Van den Bogaert
ExpORL, Department of Neurosciences, Katholieke Universiteit Leuven, O & N 2, Herestraat 49 bus 721, B-3000 Leuven, Belgium

Marc Moonen
SCD, Department of Electrical Engineering/ESAT, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10 bus 2446, B-3001 Heverlee, Belgium

Jan Wouters
ExpORL, Department of Neurosciences, Katholieke Universiteit Leuven, O & N 2, Herestraat 49 bus 721, B-3000 Leuven, Belgium

(Received 6 May 2009; revised 7 September 2009; accepted 14 September 2009)

Users of a cochlear implant and contralateral hearing aid are sensitive to interaural level differences (ILDs). However, when using their clinical devices, most of these subjects cannot use ILD cues for localization in the horizontal plane. This is partly due to a lack of high-frequency residual hearing in the acoustically stimulated ear. Using acoustic simulations of a cochlear implant and hearing loss, it is shown that localization performance can be improved by up to 14° rms error relative to 48° rms error for broadband noise by artificially introducing ILD cues in the low frequencies. The algorithm that was used for ILD introduction is described.
 [DOI: 10.1121/1.3243304]

PACS number(s): 43.66.Qp, 43.66.Ts, 43.64.Me [RYL] Pages: 3209–3213

I. INTRODUCTION

For cochlear implant (CI) users with residual hearing in the contralateral ear (bilateral bimodal stimulation), combined electric acoustic stimulation may be a feasible alternative to bilateral cochlear implantation. With clinical devices, it has been demonstrated that combining a CI with a contralateral hearing aid (HA) can improve sound source localization and speech perception performance (Ching *et al.*, 2004, 2007; Dunn *et al.*, 2005; Seeber *et al.*, 2004). While sound localization performance with a single CI improves by adding a contralateral HA, it is still poor compared to normal hearing (NH) listeners. Only a small fraction of the subjects tested in the studies cited above showed performance better than side discrimination.

Francart *et al.* (2008a, 2009) have shown that many users of a bilateral bimodal hearing system are sensitive to the main localization cues with stimuli under direct computer control that are matched in time and frequency. Their just noticeable difference (JND) in interaural level difference (ILD) was in the order of 2 dB and their JNDs in interaural time difference (ITD) were in the order of 100–350 μs. However, current signal processing strategies in clinically available CI speech processors and HAs do not allow the use of these cues by the subject. The ITD cues are weak since (1) the CI speech processor does not encode ITD information in the pulse timing and (2) the CI and HA stimulation does not occur synchronized in time due to different processing delays of the two devices and the additional traveling wave delay at the acoustic side (Francart *et al.*, 2009). The ILD cues are not available because in most subjects residual acoustic hearing is only available at the lower frequencies, and ILD cues are mainly physically present at the high frequencies (Feddersen *et al.*, 1957).

Nevertheless, if sufficiently large ILD cues are available between the low frequencies of the residual acoustic hearing and the broader spectrum of the electric stimulation via the CI, bimodal listeners may improve on localization performance. The current study assesses whether localization performance can improve when ILD cues between the low frequencies of the residual hearing and the broader spectrum of the contralateral electric stimulation are artificially introduced. This is done using acoustic simulations of bimodal hearing, presented to NH listeners. The performance improvement is measured between conditions with and without application of an algorithm that determines the ILD from the full-band signal and introduces it into the low frequencies.

II. METHODS

The steps used to simulate directional hearing using bimodal devices in normal hearing subjects are illustrated in Fig. 1. First, directional hearing was simulated using head related transfer functions (HRTFs) measured with behind the ear (BTE) devices. Second, bimodal hearing was simulated using a noise band vocoder and a low pass filter. The first step is described in Sec. II B and the second step is described in Sec. II C. The signals that were subjected to these transformations contained broadband information and are described in Sec. II A.

Two conditions were evaluated. In the first condition (noamp), stimuli were generated, simulating bimodal hearing (see Sec. II C). In the second condition (amp), a custom ILD

[a]Author to whom correspondence should be addressed. Electronic mail: tom.francart@med.kuleuven.be

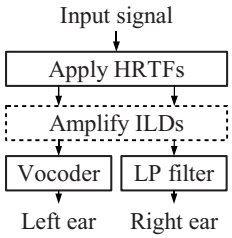


FIG. 1. Signal flow for the simulation of directional hearing for NH subjects.

introduction algorithm was applied to the signal before simulating bimodal hearing. The ILD introduction algorithm is described in Sec. II D.

Six normal hearing subjects were tested (see Sec. II E) using the procedures described in Sec. II F.

A. Stimuli

Two broadband signals were used as input for the processing described in Secs. II B and II C. The first signal (designated as noise14000) was created by filtering a 400 ms long white noise signal with an eighth order low pass Butterworth filter with a cutoff frequency of 14 000 Hz. The second, more realistic, signal (designated as telephone) was a 1000 ms long alerting signal of an old-fashioned telephone. An important feature of this signal is the prominent modulation of about 16 Hz. Its main acoustic energy is situated between 500 and 3000 Hz. Its properties are extensively described by Van den Bogaert *et al.* (2006). A cosine gate of 50 ms was applied to the start and end of both signals.

B. Simulation of directional hearing

To simulate directional hearing, the method of headphone simulation as described by Wightman and Kistler (1989) was applied. Since localization with bimodal hearing is investigated, HRTFs were measured using omnidirectional microphones (type Knowless TM-6436-D46) positioned in BTE HA shells. This BTE configuration is typically used in high power hearing aids and CI speech processors. To avoid the time consuming process of measuring HRTFs for each subject, HRTFs were measured using an artificial head of type Cortex MK2, and the same set of HRTFs was used for each subject. HRTFs were measured for each angle of incidence in an anechoic chamber using the same loudspeaker configuration as in the testing room. The input signal was processed by filtering it with the corresponding HRTFs for each angle of incidence.

C. Simulation of bimodal hearing

To simulate the amount of spectral information that would be perceived by a CI user in optimal circumstances, an eight-channel noise band vocoder (Shannon *et al.*, 1995) was used. It does not provide an exact model for loudness perception using a CI. With proper settings of the CI and HA signal processing, loudness growth between electric and acoustic stimulation is linear (Francart *et al.*, 2008a). Therefore, for the current purpose of demonstrating improvement of localization performance by ILD introduction, it suffices to have equal loudness growth with the simulated electric and acoustic stimuli.

The noise band vocoder mimics the characteristics of a typical CI speech processor by sending the input signal through an eight-channel analysis filter bank, performing envelope detection in each channel and finally adding together bands of noise with different cutoff frequencies, after modulating them with the corresponding envelopes.

Severe hearing loss was simulated in the contralateral ear by the use of a sixth order low pass Butterworth filter with a cutoff frequency of 500 Hz. The signal was calibrated at 65 dBA, such that for our NH subjects the frequencies up to 500 Hz were clearly audible, 1000 Hz was just audible, and higher frequencies were inaudible. This filter simulates the hearing loss of an average bimodal system user from our clinic, as for most of these patients the frequencies below 500 Hz and in some cases up to 1000 Hz can be sufficiently amplified by a HA.

No systematic ITDs could be found in the resulting stimuli using cross correlations on either the full-band stimuli or per critical band.

D. ILD introduction algorithm

An ILD introduction algorithm was developed that measures the ILD on the full-band signals from the microphones of both the HA and CI devices (in the current study microphones mounted in BTE HA shells). It then accordingly increases or decreases the level of the low-frequency signal to be emitted by the HA. If A_{CI} is the root mean square (rms) amplitude of the signal at the microphone of the CI device and A_{HA} the rms amplitude of the signal at the microphone of the HA, then the ILD is defined as $\mathrm{ILD}=A_{HA}/A_{CI}$. The ILD is then introduced into the acoustical signal by multiplying it by ILD/2. The idea behind this algorithm is that with bimodal stimulation the head shadow effect is only reduced in the acoustically stimulated ear, and therefore the ILD needs to be increased by half the ILD present in the broadband signal.

Note that if the subject has more residual hearing than in the current simulations, it can be useful to amplify only the low frequencies (e.g., using a shelving filter) instead of amplifying the entire frequency range of the acoustic signal.

The effect of our simulations and the ILD introduction algorithm is illustrated in Fig. 2. The "before sim" lines show the levels of the unprocessed signals at the left and right ears for different angles. Around 0° the ILD (difference in level between both ears) varies approximately linearly with angle while at larger angles the curve flattens. The "vocoder L" and "LP filter R" lines show the levels of the signals at both ears after a simulation of bimodal hearing. The curve for the left (CI) ear remains approximately the same before and after simulation, but the curve for the acoustical signal is severely flattened because of limited ILD cues at low frequencies.

The "LP filter + amp R" line in Fig. 2 shows the same low pass filtered acoustic signal as before, but now the ILD introduction algorithm was applied before the simulation of bimodal hearing. The overall ILD after processing is now as prominent as in the before sim stimuli.

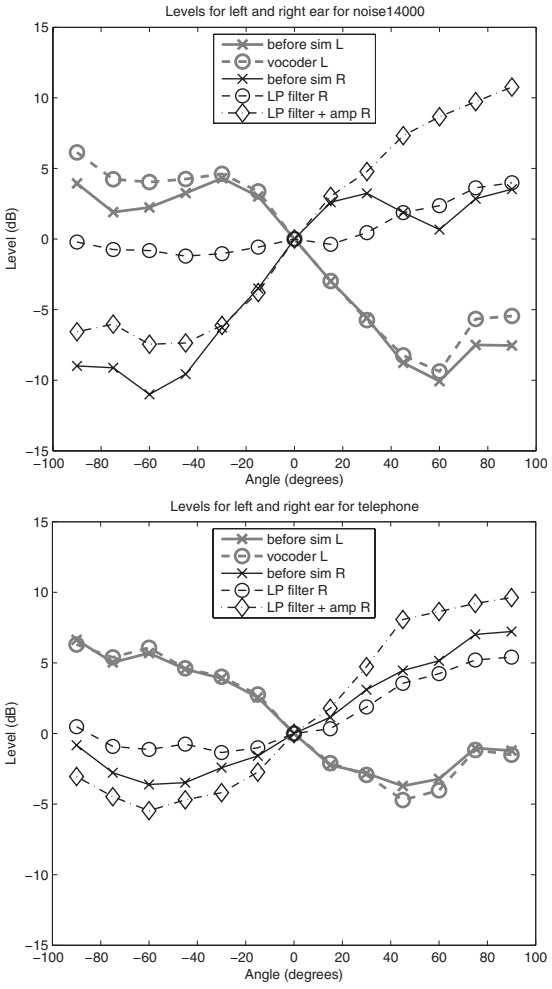


FIG. 2. Levels of the wide band signals (noise14000 and telephone) after filtering with BTE HRTFs, with and without simulation of bimodal hearing, and with and without application of the ILD introduction algorithm. The noise band vocoder simulation (CI) was applied to the left ear signal, and the low pass filtering (HA) was applied to the right ear signal. The ILD at a certain angle can be obtained by subtracting the respective levels in decibels for the left and right ears.

E. Subjects

Six normal hearing subjects aged 21–27 years participated in the current study. Their pure tone thresholds were better than 20 dB hearing level at the octave frequencies between 125 and 8000 Hz. The subjects came to the laboratory for at least two sessions of about 1.5 h, including training and excluding breaks.

F. Procedure

The subject was seated in a chair in the middle of an array of loudspeakers placed at a distance of 1 m from the subject. The chair was adjusted such that the cones of the loudspeakers were at ear height. Identical speakers were positioned at 15° intervals, yielding a total of 13 speakers and spanning 180° in front of the subject. The loudspeakers were labeled with numbers 1–13 and served only as visual cues. Stimulation was delivered using headphones of type Sennheiser HD650 connected to an RME Hammerfall DSP sound card. The experiment was controlled by the APEX 3 program developed at ExpORL (Francart *et al.*, 2008b).

Localization performance was assessed with and without introduction of ILD cues into the low pass acoustic signal using the ILD introduction algorithm described in Sec. II D. A single run consisted of three presentations of a stimulus from each angle, resulting in a total of 39 presentations per run. The chance level for the rms error in this setup is 79°. For each stimulus presentation a random angle was selected. Stimulus levels were roved trial by trial by ±6 dB to reduce monaural level cues.

As the ILD introduction algorithm introduces artificial ILD cues, subjects needed some training before being able to associate the ILD cues with the correct angles. Therefore, for each combination of stimulus and condition, some training runs were performed. A training run was the same as a normal run, but after the subject's response, feedback was shown on the screen. At least three training runs were done for each stimulus/condition before performing a normal run. After this training period, three or four normal runs were done for each subject, stimulus, and condition. Before each normal run, a training run was performed. Only the normal runs were included in the reported results. If T is a training run and N a normal run, an example test sequence for a certain stimulus/condition could be TTTNTNTN.

Calibration was done separately for each stimulus type using the stimulus from angle of 0°. The left and right outputs were calibrated separately using an artificial ear such that the level for the left and right headphone channels was 65 dBA. This resulted in a stimulus that was approximately balanced in loudness at 0°. This roughly reflects a CI and HA fitting strategy in which both devices are balanced for a sound source in front of the subject.

III. RESULTS

The results of the localization experiments are shown in Fig. 3. In the noamp condition the rms errors were 48° and 41° for, respectively, the noise14000 and telephone signals. In the amp condition the rms errors were 34° and 39°, respectively. The rms errors for each subject and each normal test run were used in a repeated measures analysis of variance (ANOVA) with main factors signal (noise14000 and telephone), condition (amp and noamp), and repetition (test run number). No main effect of or interactions with the factor repetition were found, indicating the absence of learning effects during normal test runs. Since the ANOVA showed a significant interaction between condition and signal $[F(1,67)=17.49, p<0.01]$, separate ANOVAs were performed per stimulus type with factors condition and repetition. As expected, neither of these ANOVAs showed significant main effects of or interactions with the factor repetition. For the telephone signal, no significant difference was found between the amp and noamp conditions. For the noise14000 signal, the rms error improved significantly by 14° rms error when amplifying ILDs $[F(1,32)=46.92, p<0.01]$.

The bottom panel of Fig. 3 shows how the total rms error is generated by the combination of bias and variability. In the top row, the standard deviation of the response per stimulus angle of incidence is shown. It is an indication of the variability of the subject's responses, excluding the bias.

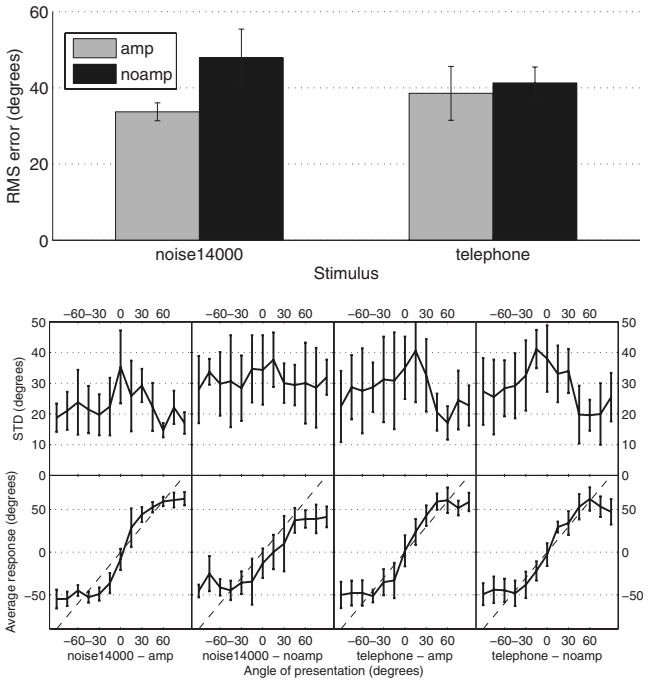


FIG. 3. Results of the localization experiments averaged per signal and condition. amp is the condition with application of the ILD introduction algorithm and noamp is the condition without ILD introduction. The error bars are between subjects' standard deviations. The top panel shows the global rms localization error. The bottom panel shows the standard deviation and the average response per presented angle.

The bottom row shows the average response. It is an indication of the local bias and the extent of laterality. The main observation from this panel is that the decrease in rms error for the noise14000 signal after ILD amplification is caused by both a decrease in bias and a decrease in variability. The bias is largest for stimuli at either side of the head.

Visual inspection of the results of the training runs shows an improvement in performance over the first few training runs of the first condition that was tested. No sufficient data were available to assess differences in learning across conditions.

The global bias was very small for each condition and stimulus (average $-0.7°$). This is likely due to the dBA calibration of the stimulus at 0° and the subjects' training. When considering change in bias separately for stimuli either presented from the left or right hand side, there is a slight, non-significant tendency of decrease with increasing run number.

IV. DISCUSSION

For both the white noise and telephone signals, the conditions without ILD introduction (noamp) yield worse performance than for NH listeners in the free field and virtual free field conditions reported by Van den Bogaert *et al.* (2006) and Francart (2008). This can be explained by the severe signal degradation caused by the simulation of bimodal hearing. The processed left and right signals are largely uncorrelated, yielding a diffuse sound image, and do not contain consistent ITDs, disabling the use of ITDs for localization. Moreover, the reduction in spectral detail by the noise band vocoder and the bandwidth restriction by the low pass filter restrict the use of ILDs, such that only limited spectral cues and very small ILD cues remain, yielding low localization performance. While rms error values cannot be compared directly, qualitatively our results are comparable to those for bimodal listeners found in the literature (Ching *et al.*, 2001; Dunn *et al.*, 2005; Seeber *et al.*, 2004), which confirms the validity of the used simulations.

The use of ILD amplification improved localization performance significantly for the noise14000 stimulus. While level roving was done over a range of 12 dB, it is still possible that monaural level cues were used to some extent. However, considering the white noise signal, in the noamp condition the maximal monaural level difference in the CI simulation (left ear) was 15 dB. In the amp condition, the maximal monaural level difference was 18 dB (see Fig. 2, right ear). Even under the assumption that the listener focuses on the ear with the largest range of level fluctuations, it seems unlikely that this 3 dB difference would have been the cause of the 14° performance improvement of the amp condition over the noamp condition.

One of the reviewers pointed out that the increased monotonicity of either monaural or binaural cues in the amp condition could have been the cause of the performance improvement. However, if the performance improvement for the noise14000 signal would have been based on monaural cue monotonicity, one would expect a larger improvement in performance after ILD amplification for stimuli from the right hand side than for the left hand side. From an analysis of error patterns per angle of presentation (not shown), it seems that this was not the case. We therefore assume that mainly binaural cues were used for localization in the current study.

The lack of a significant effect of ILD introduction for the telephone signal is probably due to the following reasons. (1) The difference in ILD between the amp and noamp conditions is smaller for the telephone signal than for the noise14000 signal (see Fig. 2). (2) Monotonicity of the ILD cues in the noamp condition is better for the telephone signal than for the noise14000 signal and does not improve with ILD introduction, while monotonicity for the noise14000 signal does improve with ILD introduction. (3) Signals with clear modulations merge better between the ears (Francart *et al.*, 2009), increasing the performance in the condition without ILD introduction [repeated measures ANOVA with factors stimulus and repetition, $F(1,36)=9.83, p<0.01$] and therefore reducing the potential improvement. However, better merging for the telephone signal is not evident in our data and therefore warrants further research.

The current simple ILD introduction algorithm suffices for showing feasibility of localizing sounds using artificial ILD cues. For use in practice, however, it should be further optimized. There are two main steps in the optimization. The first is to optimize the amplification function. The optimal ILD versus angle function could be a natural ILD versus angle function or a modified version without ambiguities and with enlarged ILDs. Enlarging ILD cues may further improve localization performance, which is necessary because ITD cues are largely unavailable, both for bilateral CI users

and for users of a bilateral bimodal hearing system. A second step in the optimization of the algorithm is to use a signal processing system to determine the location of the most prominent sound source and then use the optimal ILD versus angle function to introduce an unambiguous and sufficiently large ILD into the signal. This would make the algorithm also useful for stimuli containing only low frequencies, which do not contain sufficient ILDs to be amplified.

V. CONCLUSIONS

We investigated the effect of a practical ILD introduction algorithm in NH subjects. When introducing ILDs determined on the broadband signal into the low-frequency signal, there was no significant effect on localization performance for a telephone alerting signal, but for broadband noise, performance improved by 14° rms error relative to 48° rms error. While the usability of a similar algorithm still needs to be tested with bimodal listeners and while aspects of combined fitting of both devices are still to be considered, the current results demonstrate that it is perceptually feasible to use ILD introduced into the low frequencies.

ACKNOWLEDGMENTS

This research was carried out in the frame of the IWT (Institute for the Promotion of Innovation by Science and Technology in Flanders) Project No. 050445 and was partly sponsored by Cochlear Ltd. We thank our test subjects for their patient and enthusiastic participation in the experiments and Ying Zhao for carrying out the experiments. We acknowledge the advice of Koen Eneman on the signal processing used in the current study. We also thank Dr. Bernhard Laback and two anonymous reviewers for their constructive remarks.

Ching, T., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.

Ching, T., Psarros, C., Hill, M., Dillon, H., and Incerti, P. (**2001**). "Should children who use cochlear implants wear hearing aids in the opposite ear?," Ear Hear. **22**, 365–380.

Ching, T., van Wanrooy, E., and Dillon, H. (**2007**). "Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review," Trends Amplif. **11**, 161–192.

Dunn, C., Tyler, R., and Witt, S. (**2005**). "Benefit of wearing a hearing aid on the unimplanted ear in adult users of a cochlear implant," J. Speech Lang. Hear. Res. **48**, 668–680.

Feddersen, W., Sandel, T., Teas, D., and Jeffress, L. A. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988–991.

Francart, T. (**2008**). "Perception of binaural localization cues with combined electric and acoustic hearing," Ph.D. thesis, KU Leuven, Leuven, Belgium.

Francart, T., Brokx, J., and Wouters, J. (**2008a**). "Sensitivity to interaural level difference and loudness growth with bilateral bimodal stimulation," Audiol. Neuro-Otol. **13**, 309–319.

Francart, T., Brokx, J., and Wouters, J. (**2009**). "Sensitivity to interaural time differences with combined cochlear implant and acoustic stimulation," J. Assoc. Res. Otolaryngol. **10**, 131–141.

Francart, T., van Wieringen, A., and Wouters, J. (**2008b**). "APEX 3: A multi-purpose test platform for auditory psychophysical experiments," J. Neurosci. Methods **172**, 283–293.

Seeber, B., Baumann, U., and Fastl, H. (**2004**). "Localization ability with bimodal hearing aids and bilateral cochlear implants," J. Acoust. Soc. Am. **116**, 1698–1709.

Shannon, R., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Van den Bogaert, T., Klasen, T., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). "Horizontal localization with bilateral hearing aids: Without is better than with," J. Acoust. Soc. Am. **119**, 515–526.

Wightman, F. and Kistler, D. (**1989**). "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Using a signal cancellation technique involving impulse response to assess directivity of hearing aids[a)]

Yu-Hsiang Wu[b)] and Ruth A. Bentler
Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, Iowa 52242

(Received 25 January 2009; revised 23 August 2009; accepted 25 August 2009)

The directional microphone systems of modern digital hearing aids are capable of changing their spatial directivity pattern and/or the microphone mode in response to changes in the properties of environmental sounds. These adaptive/automatic features make measurement of a hearing aid's directivity in a given test environment very difficult. Assessing the directivity of such systems requires a signal that can record the system's response while not changing the system's directivity. This paper proposes a method using a signal cancellation technique involving impulse responses to acoustically assess a hearing aid's directivity (referred to as the IR method). The impulse is presumed to be undetectable to the adaptive/automatic system because it contains little energy and a short response could be recorded before the system actually reacts. In the current study, the IR method was evaluated by testing five adaptive/automatic directional hearing aids in noise of various intensities. The results revealed that the IR method was an accurate and repeatable way to assess slow-acting directional systems in noise of varying intensities and fast-acting systems in noise of high intensities. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3238161]

PACS number(s): 43.66.Ts, 43.66.Yw [MW] Pages: 3214–3226

I. INTRODUCTION

It is well established that people with hearing loss experience greater difficulty understanding speech in noisy environments (Dirks *et al.*, 1982; Dubno *et al.*, 1984). Modern hearing aids use several technologies to improve speech perception within such environments. Among these technologies, the directional microphone is one of the most efficacious in that it is designed to have a relatively high sensitivity to signals arriving from one direction (on-axis), while suppressing signals coming from other directions (off-axis) (Kates, 2008). If directional microphone hearing aid (DMHA) users can place the talker facing them and noise behind them, DMHAs can improve signal-to-noise ratio and speech intelligibility (Ricketts and Dittberner, 2002).

In recent years, a variety of signal processing schemes have been incorporated into the directional systems employed by digital hearing aids (Kates, 2008). These schemes allow the system to change its spatial directivity pattern and/or the microphone mode in response to changes in the properties of the acoustic environment. For example, an adaptive directional system can steer the null, which is the least sensitive angle, of the directivity pattern to the noise source to achieve a higher degree of noise attenuation (e.g., Bentler *et al.*, 2004) or adapt its directivity pattern to minimize the total input level while keeping the response from the front constant (e.g., Blamey *et al.*, 2006). An automatic system can switch between the omnidirectional and the directional modes in environments having different noise levels (e.g., Olson *et al.*, 2004). A fully adaptive system has both adaptive and automatic features (e.g., Kuk *et al.*, 2005). Some directional systems are even capable of estimating a signal's temporal and spectral characteristics and using them to determine directivity; that is, the system will switch to the directional mode if a rear hemisphere signal is more like noise and switch to the omnidirectional mode if the signal is more like speech (Chung, 2004). By using different directivity patterns in different noisy environments, either adaptively or automatically, speech intelligibility and user satisfaction may be increased. Although the effectiveness of these advanced DMHAs has not been observed in the real world (Palmer *et al.*, 2006), their efficacy has been proven in some laboratory environments (e.g., Bentler *et al.*, 2004; Ricketts and Henry, 2002).

Although adaptive/automatic directional systems may improve speech intelligibility amid noise by responding to the acoustic environment, experiments regarding such systems are difficult to validate. That is, researchers cannot always accurately determine the directivity pattern of an adaptive/automatic DMHA in a laboratory test environment. For example, some directional hearing aids may respond differently when random noise is used as a masker than when speech is used as a masker. This could be important in understanding the results of experiments designed to assess directional benefit for speech and speech-like maskers compared to non-speech maskers (e.g., Hornsby and Ricketts, 2007). Another example is given in the study by Kuk *et al.* (2005). These investigators tested a DMHA by presenting words in a quiet environment to listeners from 0° to 180° in 45° increments. They found that the listeners' performances did not differ significantly between the (fully adaptive) directional mode and the omnidirectional mode at all azimuths, including the azimuths in which speech signals were presented from the listener's back. To explain these results, Kuk *et al.* (2005) assumed that the fully adaptive system used in the study remained in the omnidirectional mode during the testing because (1) the ambient noise of the test environment

[a)]Portions of this work were presented at the International Hearing Aid Research Conference, Lake Tahoe, CA, August, 2006.
[b)]Author to whom correspondence should be addressed. Electronic mail: yu-hsiang-wu@uiowa.edu.

was too low to shift the system into the directional mode and (2) the duration of the word signals were shorter than the time needed for the system to switch from the omnidirectional mode to the directional mode. Without knowing the exact directivity status of the DMHA during the test, the validity of the experimental design cannot be ensured.

Although knowing the directivity status of DMHAs is desirable, it is not easy for researchers to obtain this information. Some general rules about how an adaptive/automatic system reacts to the properties of sounds may be available from the literature (e.g., Chung, 2004; Olson *et al.*, 2004) or from the manufacturer. However, these rules may not predict a device's directivity in a given test environment, which is one of a large number of combinations of speech and noise signals having a variety of spectral/temporal characteristics, numbers, levels, and presentation angles. Another way to determine directivity is to read the directivity information directly from the DHMA's integrated circuit chips. Yet, the software needed to do so is typically proprietary to the manufacturer and may not be available to independent researchers. Methods used to assess the DMHA that has a fixed directivity pattern, such as presenting signals subsequently to the device from different angles at a constant distance, are invalid in assessing the adaptive DMHA because the aid's directivity pattern varies during the measurement. Puder (2006) suggested a method to evaluate the performance (i.e., the signal-to-noise ratio enhancement) of an adaptive/automatic DMHA by deriving the frequency response of the DMHA to each of simultaneous input signals located at various angles. However, this method does not provide the device's spatial directivity pattern (i.e., the polar pattern). It is therefore necessary to develop a measurement system that can assess DHMA directivity. This method has to be (1) accurate, (2) repeatable, (3) able to assess most adaptive/automatic DMHAs, (4) able to assess directivity in most test environments which consist of various noise configurations, and (5) easy for most hearing aid researchers to implement. Metrics that evaluate the accuracy of this measurement system should be developed as well.

As the first step to achieve this goal, we have developed a system that uses a signal cancellation technique to acoustically obtain an adaptive DMHA's planar polar pattern (Wu and Bentler, 2007). In that article, the method is referred to as the signal cancellation (SC) method. In the SC method, the directivity pattern is derived by presenting two signals to the hearing aid simultaneously at each azimuth. A probe signal, which is a low-intensity wideband noise, is presented from a fixed-position loudspeaker to the hearing aid that is rotated to different azimuths. The probe signal is used to record the response of the aid. A jammer signal, which is a high-intensity wideband noise, is presented from a fixed azimuth to the hearing aid by a second loudspeaker that rotates with the aid. The jammer signal serves as the noise source. Because the intensity of the probe signal is much lower than that of the jammer signal, the directivity pattern of the hearing aid is determined by the jammer signal and will not change when the probe is presented. Measurement at each azimuth is obtained by two sequential recordings from the hearing aid, one using the input of a probe and a jammer, and the other using the input of the same probe and a phase-inverted jammer. After canceling out the jammer signal, the remaining response to the probe signal is used to derive the directivity pattern. The SC method has been shown to be accurate and repeatable in assessing adaptive DMHAs (for details, see Wu and Bentler, 2007).

The SC method has two limitations, however. First, the polar pattern obtained by the SC method is accurate only if the jammer level is high [e.g., 75-dB sound pressure level (SPL)]. If lower level (e.g., 60-dB SPL) jammers are used, the probe signal will alter the directivity pattern. Therefore, the SC method cannot be used to assess DMHAs that respond to low-intensity noises or to signals that have quieter intervals, such as speech. Second, because the system has to be time-invariant during the presentation of the probe signal, which is 12 s in the SC method, the SC method cannot be used to "pinpoint" the polar pattern at a given moment during the presentation of signals that change over time. To solve the first limitation, an "undetectable" probe signal, which leaves the directivity pattern unchanged while still being able to record the system's response, is needed. To solve the second limitation, a very short probe signal is required. In this article, we propose a new method that uses the signal cancellation scheme of the SC method employing the impulse as a probe signal to assess the planar directivity pattern of DMHAs.

A. Theory of measurement

The signal cancellation technique is summarized below. Two signals are presented to the hearing aid sequentially. They are

$$A_{\text{in}} = u(t) + v(t) \tag{1}$$

and

$$B_{\text{in}} = u(t) - v(t), \tag{2}$$

where $u(t)$ is the probe signal, $v(t)$ is the jammer signal, and t is time. The probe signals in A_{in} and B_{in} are the same, whereas the jammer signals in these two input signals are phase-inverted. The probe signal $u(t)$ is presented from different angles from the probe loudspeaker, and the jammer signal $v(t)$ is presented from the jammer loudspeaker at a fixed angle (with regard to the hearing aid). The outputs of the hearing aid will be

$$A_{\text{out}} = u'(t) + v'(t) + e_1(t) \tag{3}$$

and

$$B_{\text{out}} = u'(t) - v'(t) + e_2(t), \tag{4}$$

where $u'(t)$ and $v'(t)$ are the outputs of the probe and jammer signals, respectively, if the device's status is held constant and if the probe and jammer are presented in isolation. The terms $e_1(t)$ and $e_2(t)$ are error signals. The output probe signal can be extracted as follows:

$$(A_{\text{out}} + B_{\text{out}})/2 = u'(t) + [e_1(t) + e_2(t)]/2. \tag{5}$$

Subtracting the level of the output probe signal (i.e., the probe response, expressed in decibels) at 0° (the on-axis response) from the probe response obtained at each of the off-axis angles derives the directional response, which can then be used to construct the polar pattern. This signal cancellation technique is modified from the technique first described by Hagerman and Olofsson (2004) who successfully used this technique to investigate the effectiveness of hearing aids' noise reduction algorithms.

To solve the limitations of the SC method mentioned above, an impulse (i.e., a click) can be used as the probe signal $u(t)$. Three reasons support the use of impulses. First, the impulse is widely used as an exciting signal to assess a system and the impulse response can be used to derive the system's frequency response. Second, the directional system may not respond to an impulse because it contains little energy. Third, the durations of the impulse and impulse response are short. Because adaptive/automatic DMHAs typically take from several milliseconds to seconds to change their directivity (Chung, 2004), it is likely that even if the impulse changes the directivity pattern, the impulse response can be recorded before the change. Further, the short-duration impulse response can "snapshot" the directivity pattern at a given moment. For these reasons, we hypothesized that the impulse would be undetectable to DMHAs and would be an ideal probe signal to assess the directivity of adaptive/automatic directional systems. Because the impulse response is used, this new method is referred to in the present article as the IR method.

II. EXPERIMENT 1

A. Method

Since one assumption of the IR method is that the impulse response can be recorded before the DMHA changes its directivity, the purpose of this experiment was to measure the directivity adaptation time of DMHAs.

Five commercially available hearing aids equipped with adaptive/fully adaptive directional systems were tested, here denoted as HA1–HA5. All hearing aids were the behind-the-ear style. The aids were programed to fit a flat 50 dB hearing loss using the manufacturers' fitting software and then modified into a linear gain mode (Ricketts, 2000). The maximum power output was set to the maximum. The adaptive or fully adaptive directional mode was enabled. Adaptive features other than the directional microphone, such as noise reduction and feedback cancellation algorithms, were disengaged. Except for HA1, all the directional systems were first-order systems; that is, two microphones were used to achieve directivity (Chung, 2004). HA1 consisted of two different systems: a first-order directional system in low-frequency bands, and a second-order directional system (three microphones used) in high frequency bands. The crossover frequency between the first- and second-order systems was around 1 kHz. Only the first-order system could be programed to the adaptive mode. The directivity of the second-order system was fixed. The microphone ports of HA1 were isolated to the right side of the hearing aid case; that is, it was a right-ear hearing aid.

TABLE I. Directivity adaptation times of the five hearing aids under test.

	HA1 (ms)	HA2 (ms)	HA3	HA4 (s)	HA5 (s)
From silence to 90°	<10	75	18 s	8.5	3
From 90° to 180°	<10	40	<10 ms	4.5	3

The measurements were made in an anechoic chamber that adheres to ANSI S3.35 (ANSI, 2004). Two loudspeakers were placed at 90° and 180° relative to the hearing aid under test. The hearing aid was placed at the center of the anechoic chamber without being mounted on a manikin's head. After 1 min of silence, a 75-dB SPL pure tone at 1 kHz was presented from the loudspeaker at 90° for 1 min, and then was shifted to the loudspeaker at 180° instantaneously. The output of the hearing aid was recorded. For each signal change (from silence to 90° or from 90° to 180°), the amplitude of the hearing aid's output was initially large, representing the directivity pattern at the moment of the signal change, and was then decreased, which represented the directivity pattern adapting to the change of the pure tone. The directivity adaptation time of a DMHA was determined as the time taken for the output level to decrease to within 3 dB of its final value.

The pure tone signals were generated by a RME DIGI96/8 PAD sound card on a personal computer (PC) with a 44.1-kHz sampling rate and 16-bit resolution. From the sound card, the signals were routed to two Tannoy i5 AW point source loudspeakers via an ART 355 equalizer and an ADCOM GFA5002 amplifier. The distance between the loudspeakers and the hearing aid was 1 m. A G.R.A.S. Sound and Vibration IEC 126 2-cc coupler and a Larson-Davis 2560 1/2" random incidence microphone were used to collect the hearing aids' outputs. The recorded signals were routed through a Larson-Davis preamplifier and a LISTEN SoundConnect microphone power supplier to the RME sound card. The sound card digitized the signals with a 44.1-kHz sampling rate for analysis.

B. Results

The results first revealed that for all the hearing aids tested, the directivity adapted to the change in input signal and reduced the output level by more than 3 dB. The directivity adaptation times of the five DMHAs are shown in Table I. The times showed a wide range, from less than 10 ms to several seconds. For the time shown in the first row, which indicates the time for a system to change from its default pattern in a quiet condition (in which the system could be in the directional mode or the omnidirectional mode) to the dipole pattern, all the adaptation times were longer than 10 ms except for HA1. For the time shown in the second row, which indicated the time for a system to change from the dipole pattern to the cardioid pattern, HA1 and HA3 had adaptation times of less than 10 ms. Note that the adaptation times of HA3 are very different for the two conditions of measurement.

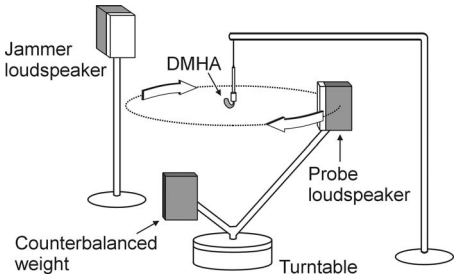


FIG. 1. Equipment set-up (DMHA: directional microphone hearing aid).

III. EXPERIMENT 2: METHODOLOGY

In this section, the IR method and the metrics to evaluate the IR method are described.

A. Equipment set-up of the IR method

All measurements in Experiment 2 were conducted in the same anechoic chamber as that used in Experiment 1. The equipment set-up is shown in Fig. 1. A probe loudspeaker was mounted on an Outline ET1.1 turntable so that it could present the probe signal to the hearing aid at different azimuths. The reason for not testing the hearing aids on a manikin's head is that the shape of the polar pattern and the polar index (PI) (see below) generated in this free-field measurement could be used to judge the accuracy of the IR method (see below). A jammer loudspeaker was fixed in a position so as to present to the hearing aid jammer signals from a constant azimuth angle. The aid under test was hung upside-down at the center of the horizontal circle in which the probe loudspeaker moved. Note that this set-up differs from that used in the SC method in which the probe loudspeaker is in a fixed position and both the hearing aid and the jammer loudspeaker are installed on the turntable (Wu and Bentler, 2007). The set-up shown in Fig. 1 is regarded as superior because the number and the position of the jammer loudspeaker is not limited by the turntable, allowing the easy creation of different noise configurations. The distance between the aid and the probe loudspeaker was set at 1 m. Although it was desirable to place the jammer loudspeaker on the same horizontal plane as the hearing aid and the probe loudspeaker, it was not possible to do so because one loudspeaker would block the other when both loudspeakers were at the same azimuth. Therefore, the jammer loudspeaker was placed approximately 7° above the horizontal plane (7° below the horizontal plane, with regard to the hearing aid). The distance between the aid and the jammer loudspeaker was 1.1 m. Two Tannoy i5 AW loudspeakers served as the probe and jammer loudspeakers, respectively. The turntable was placed at the center of the anechoic chamber. The heights of the hearing aid and the probe loudspeaker (with regard to the catwalk floor) were both 1.3 m. The turntable rotation was controlled via an external controller and a PC. The signal generation, routing, and recording were identical to those in Experiment 1.

B. Signal presentation and analysis

During the measurement, the probe signal was presented to the hearing aid from 0° to 360° in 10° increments. At each azimuth, the first signal set (A_{in}) was presented: probe signal [$u(t)$] from the probe loudspeaker and the jammer signal [$v(t)$] from the jammer loudspeaker. Before the probe signal presentation, the jammer signal was presented for 20 s to ensure the directivity pattern. The data from Table I suggested that a 20-s interval was long enough for the directional pattern of all DMHAs to stabilize. The probe signal was presented at the 20th second without shutting off the jammer signal. After the probe signal had been presented, the jammer continued for 1 s, followed by a 10-s silent interval. After the silent interval, the second signal set with the inverted jammer phase (B_{in}) was presented in the same order. The outputs of the hearing aid (A_{out} and B_{out}) were recorded. Because the durations of the stimuli (A_{in} and B_{in}) and that of the silent interval were known, A_{out} and B_{out} could be time-aligned accurately and most jammer signals could be canceled out by averaging these two output signals [Eq. (5)]. In order to improve the signal-to-noise ratio and to increase the repeatability of the IR method, the stimuli A_{in} and B_{in} were repeated four times at each azimuth and the eight output signals were averaged. Five-tenths seconds of the averaged signal with the output probe $u'(t)$ in the middle was then converted from the time domain to the frequency domain using the fast Fourier transform (FFT) function of the Mathworks MATLAB 7.2.0 (22,050 points; uniform windowing). The frequency spectrum generated by the FFT was apportioned among the one-third-octave bands across 0.4–6.3 kHz with center frequencies specified in ANSI S1.6 (ANSI, 1984) to produce the corresponding one-third-octave probe response. The probe response at each azimuth was then used to construct the planar polar pattern and compute the polar index (see below). To increase data repeatability, the measurement at the 0° azimuth was performed twice: before and after the rotation of the probe loudspeaker. The average of the two probe responses was used as the on-axis response.

White noise was chosen as the jammer signal to drive directivity at all frequencies simultaneously. A monophasic, rectangular click was used as the probe signal. The duration of the click was 2/44,100 s (~0.045 ms). The amplitude of the click was determined individually for each jammer signal condition (see below) and for each hearing aid. Specifically, the level of the click has to be as low as possible in order not to change the hearing aid's directivity pattern, whereas it has to be high enough to provide sufficient signal-to-noise ratios across frequencies. Generally, a 20-dB signal-to-noise ratio is required to obtain the null of a polar pattern. To minimize the possibility of the click changing the directivity pattern, the lowest intensity that can achieve the 20-dB signal-to-noise ratio was selected. To obtain the signal-to-noise ratio, the jammer was presented from the angle the jammer loudspeaker would be in during the measurement when the probe was presented from 0°. As described previously, four pairs of A_{in} and B_{in} were presented and the output signals were averaged. The one-third octave levels of the averaged signal between 1 and 0.5 s before the presentation of the probe was computed as the noise floor. Note that the noise floor was the summation of the residual jammer signal after accounting for cancellation, ambient noise, and the internal electric noise of the hearing aid. As a result, the signal-to-noise ratio varied

among jammer levels and hearing aids, and was frequency-dependent. For each hearing aid in each jammer signal condition in this experiment, the probe was obtained by spectrally shaping an 80-dB peak-to-peak equivalent sound pressure level (peSPL) click, with this spectrally shaped click having a 20-dB signal-to-noise ratio in each of the one-third octave bands across 0.4–6.3 kHz. The shaping was achieved using the graphic equalizer function of the software sound-manipulation tool ADOBE AUDITION 1.0. The duration of the shaped click was longer than the unshaped one (~0.8 ms for the main peak, and ~8 ms if it included the small fluctuations before and after the main peak). Both the probe and jammer signals were created by the AUDITION.

C. Procedures

The procedures were designed to examine the IR method's accuracy, repeatability, and generalizability. The same five DMHAs used in Experiment 1 were tested. The hearing aid settings were identical to those in Experiment 1.

1. Accuracy

Ideally, the easiest way to assess the accuracy of the IR method is to freeze the directivity of the adaptive/automatic system in a given jammer condition using proprietary software from the manufacturer, and then to compare this polar pattern to the results obtained using the IR method. However, since the IR method would be used only when the proprietary software is unavailable, it is necessary to develop methods to assess its accuracy without having to use the software.

Four metrics, which have been used to evaluate the accuracy of the SC method (Wu and Bentler, 2007), were modified and used here. The first metric is the shape of the polar pattern. For a first-order directional system, the directional response pattern can be described by the equation

$$R(\theta) \approx |\alpha + (1-\alpha) \times \cos(\theta)|, \tag{6}$$

where θ is the angle at which the sound arrives, R is the directional response, and $0 \leq \alpha \leq 0.5$. The value of α is determined by the separation between the two microphone ports and the rear microphone delay time (Kates, 2008). For $\alpha=0$, the pattern is the dipole pattern with nulls at 90° and 270°. For $\alpha=0.5$, the pattern is a cardioid pattern with a null at 180°. Since the adaptive/automatic directional systems of most DMHAs on the market are first-order (Chung, 2004), the shape of the theoretic pattern can be used to evaluate the accuracy of the IR method. If the shape of the measured pattern is significantly different from those described by Eq. (6) and is not the omnidirectional pattern (i.e., a circle), it can be inferred that the probe signal may change the pattern and the result may be inaccurate.

The second metric is the PI. The PI is

$$\mathrm{PI} = 10 \log_{10}\left[\frac{1}{\pi/(2n)\sum_{j=1}^{n} 10^{R(\theta_j)/10}|\sin\ \theta_j|}\right], \tag{7}$$

where n is the number of measurements in the 360° azimuth, θ_j is the azimuth angle, and $R(\theta_j)$ is the directional response, which is the difference (in decibels) between the probe response at azimuth θ_j and that at 0°. In fact, Eq. (7) is also the equation for the planar directivity index (DI) (ANSI, 2004). Note that although the PI and DI share the same equation, they are different conceptually. The DI is the ratio (in decibels) of the sound pressure of the microphone's on-axis response to a sound source in a free sound field to the sound pressure of the microphone's off-axis response in a diffuse sound field (Beranek, 1949), which represents a device's directivity in a diffuse field. However, the PI derived from the IR method represents the DMHA's response to one or more discrete jammers and a single low-intensity probe signal.

The PI can be used to assess the accuracy of the IR method. Specifically, during the IR method measurement, if the hearing aid responds to the probe signal and changes directivity pattern, the measured directional response will be decreased. If the PI is calculated based on the falsely decreased response, the PI will show a false increase [Eq. (7)]. Theoretically, the maximal free-field DI for a first-order directional system is around 6 dB in a hypercardioid polar pattern that has nulls at 110° and 250° (Valente *et al.*, 2000). Therefore, if a PI derived from the directional responses measured by the IR method is higher than 6 dB, it can be inferred that the result of the IR method is inaccurate. Because the shape of the polar pattern and the theoretical PI value were used to evaluate accuracy, the DMHAs were tested in a free field without being mounted on the head and a jammer signal was presented at 110° to cause the directional system to shift into the hypercardioid pattern.

The third metric is the error index, which is derived from the error value. The error value is defined as the change in probe response (in decibels) due to the presence of the probe at a given azimuth and frequency. In the SC method, the error value is obtained by comparing the amplitude at the beginning of the 12-s output probe and that at the end of the output probe signal. If the system responds to the probe signal, the initial amplitude will be larger than that at the end of the output probe signal (see Wu and Bentler, 2007 for details). However, in the IR method the error value cannot be derived in this manner because of the short impulse response. Alternatively, the error value obtained by the IR method is estimated using two consecutive clicks as the probe signal [$u(t)$]. After canceling out the jammer, the two extracted impulse responses are compared. It is assumed that the two impulse responses will be the same if the click does not change the directivity pattern, given that the interval between the two clicks is short. On the contrary, if the click causes the directivity to change, the first response will be larger than the second response. However, a difference between the two responses does not always indicate that the probe changes the polar pattern. This is because the system may not respond to a single click but respond to two consecutive clicks, which contain twice the energy of a single click.

To implement the error value estimation, an additional pair of stimuli (A_{in} and B_{in}) that used two clicks with an interval of 10 ms as the probe signals were presented to the hearing aid at each azimuth. Because of the short interval, the two extracted impulse responses overlapped. Instead of the whole impulse response, only the responses' peaks, which were not overlapped, were used to derive the error

value. Specifically, the extracted impulse responses were filtered by an octave filter bank centered at 0.5, 1, 2, and 4 kHz using the MATLAB filter function (third-order Butterworth filter). The waveform around the peak of each impulse response was extracted and the peak-to-peak amplitude of each response was then calculated in each frequency band. The amplitude of the second response was subtracted from that of the first response (in decibels) to obtain the frequency-specific error value at each azimuth. The error index (EI) is then derived by the equation

$$\mathrm{EI} = \sqrt{\frac{\sum_{j=1}^{n} E_j^2}{n}}, \tag{8}$$

where n is the number of measurements in the 360° azimuth, and E_j is the error value at each azimuth angle θ_j. The EI is the root-mean-square level of the error values across all azimuth angles in a certain frequency band.

The fourth metric, which is referred to as the PI difference, is the difference between the measured PI and the corrected PI. Specifically, if the error value is the amount the probe response changes due to the probe signal at a given azimuth, the true response not altered by the probe can be derived by adding the error value back to the measured probe response. These corrected responses then can derive the corrected PI by Eq. (7). If the measured and corrected PIs are very close, it is reasonable to assume that the IR method is accurate. If the probe changes the directivity pattern and results in a positive error value, the corrected PI will be smaller than the measured PI [Eq. (7)]. Because both EI and PI differences are derived from the error value, EI or PI difference with a small value indicates that the IR method is accurate, while a large value cannot prove or disprove the accuracy of the IR method. In the SC method, the EIs obtained by comparing the start and the end of the output probe in a 75-dB SPL jammer are smaller than 1.5–4 dB, and PI difference values under the same condition are within the ranges of ±0.3–±0.6 dB, depending on frequency (Wu and Bentler, 2007). In this experiment, the IR method was considered accurate if the EI and PI difference values were not greater than those obtained in the SC method.

2. Repeatability

To examine the repeatability of the IR method, five consecutive measures of each hearing aid were performed. The hearing aid was removed from the coupler, replaced, and repositioned for each of the repeated measurements. The standard deviation of PI values and the standard deviation of directional responses at each azimuth were calculated for each aid. The repeated measurements were done in the 75-dB jammer condition. If the standard deviations were small (e.g., smaller than 0.1 dB), the IR method was considered repeatable.

3. Generalizability

The five DMHAs were tested using jammers of various levels: 75-, 60-, and 45-dB SPL, and quiet. The generalizability of the IR method was considered good if accurate results could be obtained from all hearing aids in all jammer conditions.

TABLE II. Probe levels [dB peSPL (peak-to-peak equivalent sound pressure level)] of different hearing aids and jammer conditions. Each probe signal was created by spectrally shaping an 80-dB peSPL click, with this click having a 20-dB signal-to-noise ratio across frequencies in each hearing aid under each jammer condition.

	Jammer condition			
	75 dB	60 dB	45 dB	Quiet
HA1	86.2	84.6	82.7	81.9
HA2	85.4	82.8	82.3	81.7
HA3	81.4	80.6	78.5	78.2
HA4	81.9	80.7	79.3	79.1
HA5	82.1	80.4	78.5	76.5

In summary, five DMHAs were tested in a free field following the IR method to assess its accuracy, repeatability, and generalizability under four conditions: 75-, 60-, and 45-dB SPL white noise jammer at 110°, and quiet. The 75-dB jammer condition was repeated five times for each hearing aid. The intensities and spectra of the probes were adjusted to achieve a 20-dB signal-to-noise ratio in each aid and each test condition. The probe levels are listed in Table II. In addition to the IR method, the five hearing aids were tested using the SC method under the 75-dB 110° jammer condition to compare the results obtained from the two methods. The details of the SC method are described elsewhere (Wu and Bentler, 2007). Before the measurement, each hearing aid's input/output function was obtained by a hearing aid analyzer (Audioscan Verifit) to ensure that the hearing aid worked linearly. An algorithm written in the MATLAB platform controlled the signal presentation, turntable rotation, and signal extraction and analyses.

IV. EXPERIMENT 2: RESULTS

A. Accuracy

1. 75-dB jammer condition

Each DMHA was tested five times in this condition. The planar directivity patterns, which were constructed by the directional response obtained in each azimuth, are shown in first column of Fig. 2. The plot displays the spatial attenuation pattern at frequencies of 0.5, 1, 2, and 4 kHz. The gray, thick curve represents the theoretical hypercardioid pattern derived by Eq. (6) with α=0.25 (Kates, 2008). Typical hypercardioid polar patterns were found in all DMHAs except for HA1 at all frequencies and HA2 at 500 Hz. The polar patterns of HA1 differed from those of the other hearing aids in two ways. First, the polar patterns at frequencies below 1 and above 2 kHz were different. This difference in patterns is due to HA1's first-order system in the lower frequency bands and second-order system in the higher frequency bands. Second, although a hypercardioid-like pattern was found in the low-frequency bands, it was not symmetric. The whole pattern rotated approximately 10° counterclockwise. The polar patterns in the high frequency bands were also asymmetric. This was because this hearing aid's microphone ports were

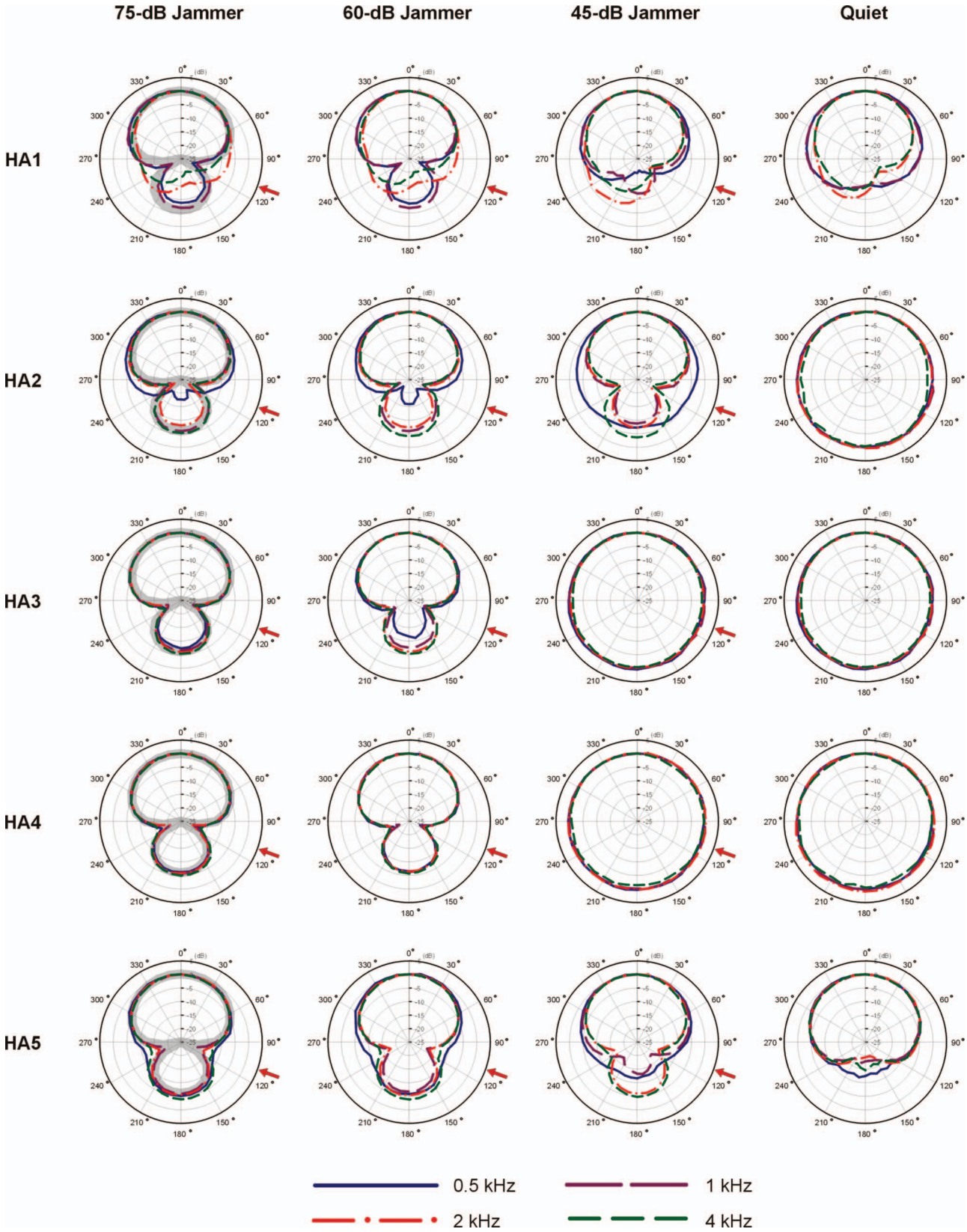


FIG. 2. Planar spatial directivity patterns of the five hearing aids in the four test conditions. The center of the circle corresponds to −25 dB and the radius of the outermost circle corresponds to +5 dB. The gray, thick curve shown in the first column represents the theoretical hypercardioid polar pattern. The arrow indicates the azimuth of the jammer signal.

on the right side instead of on the top. The rear lobe of HA2's pattern at 500 Hz was smaller than that of the theoretical pattern. The measured pattern was the transition between the hypercardioid pattern at frequencies above 500 Hz and the omnidirectional pattern at frequencies below 500 Hz (i.e., 400 Hz, not shown in Fig. 2; also see the omnidirec-

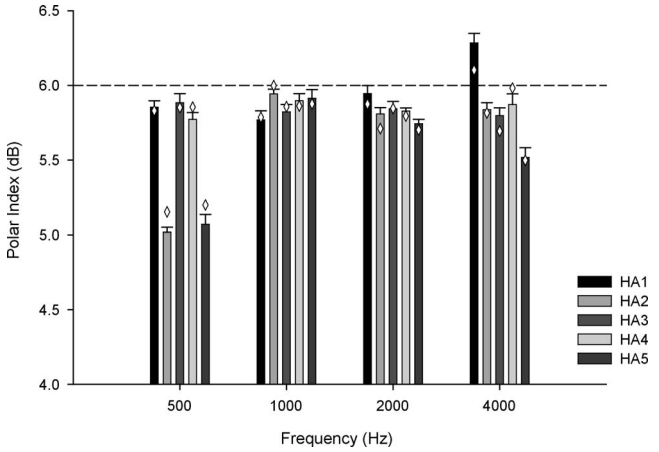


FIG. 3. Mean polar index measured in the 75-dB condition as a function of frequency. The error bars represent 1 standard deviation. The dashed line represents the theoretical upper limit of the first-order directional microphone's free-field polar index. The diamond symbol represents the polar index measured using the SC method under the same jammer condition.

tional pattern of HA2 in the quiet condition in Fig. 2). In follow-up measurement, it was found that a jammer with higher low-frequency intensity, such as an 80-dB SPL pink noise, would be needed to shift the HA2's microphone mode to the directional mode in the low-frequency bands.

The mean PIs at 0.5, 1, 2, and 4 kHz of five consecutive measurements of each aid are shown in Fig. 3. The PI values measured by the SC method are shown in the same figure. Except for the PI at 4 kHz of HA1, which uses the second-order microphone system in the high frequency bands, all the PIs were smaller than 6 dB. The differences between the mean PI values obtained from the IR and the SC methods were very small, all less than 0.2 dB. As well, there was no systematic difference between the results of the two methods.

The mean error values of the five measurements at each azimuth angle are shown in Fig. 4. A positive value indicates that the peak-to-peak amplitude of the first impulse response was larger than that of the second response. For HA2, HA4, and HA5, the error values were generally small (85% of the values were within ± 0.5 dB). The variations in the error value at and around azimuths 110° and 250° were larger because when the probe approached the null, the probe response was much attenuated and was close to the noise floor. As a result, the measurement was easily contaminated by the noise floor. For HA1 and HA3, the error values between azimuths 110° and 250° tended to be positive, indicating the possibility that the two consecutive clicks used to derive the error value changed the directivity at these azimuth angles.

The EIs of each hearing aid at 0.5, 1, 2, and 4 kHz are shown in the upper panel of Fig. 5. The error bars represent the range of the EIs of the five measures. All the EI values were smaller than 1.5 dB. The differences between the measured PI and corrected PI at 0.5, 1, 2, and 4 kHz are shown in the upper panel of Fig. 6. A positive value represents that the measured PI is larger than the corrected PI. Generally, the differences in the PI values are small: between -0.2 and 0.4 dB.

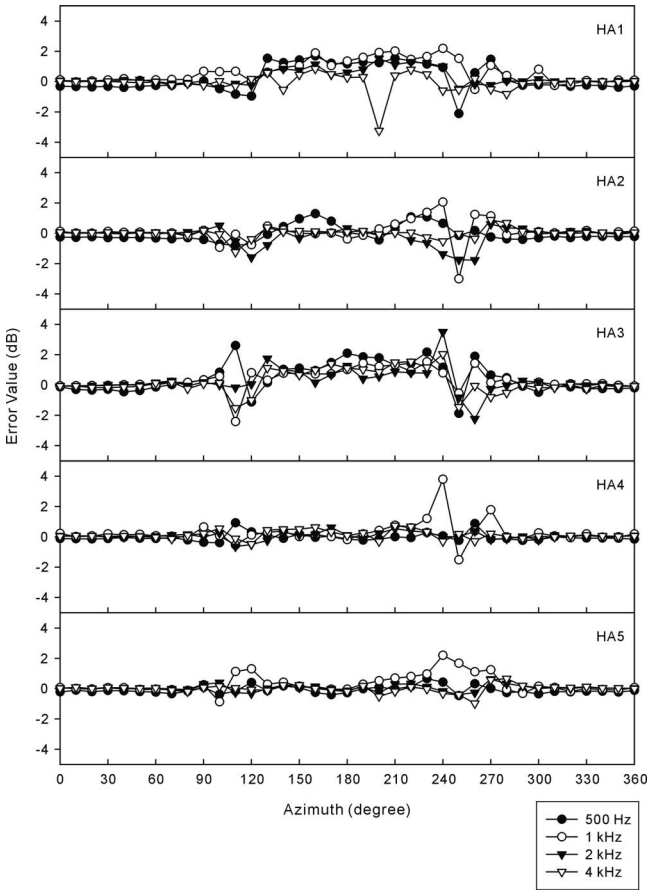


FIG. 4. Mean error value measured in the 75-dB condition as a function of azimuth angle.

2. 60-dB jammer, 45-dB jammer, and quiet conditions

The directivity patterns in the conditions involving the 60-dB and 45-dB SPL jammers and without the jammer are shown in Fig. 2. With the 60-dB jammer, the rear lobe of HA3's directivity pattern at 500 Hz was smaller than that in the 75-dB condition. Except for this, all the polar patterns were similar to those involving the 75-dB jammer. With the 45-dB jammer and with no jammer, the patterns of HA3 and HA4 became circular, indicating the omnidirectional pattern. For HA2 and HA5, when the jammer level decreased from 75-dB SPL to quiet, starting from low-frequency bands, the polar patterns gradually shifted from the hypercardioid pattern to the omnidirectional and cardioid patterns, respectively. For HA1, its polar patterns in the lower frequency bands changed from the hypercardioid-like pattern to the cardioid-like pattern. In the high frequency bands, especially at 4 kHz, the polar pattern of its non-adaptive system did not change under different jammer conditions.

The PI values are shown in Table III. Except for HA1 at 4 kHz, no PI was larger than 6 dB, although the polar pattern was no longer hypercardioid under many conditions involving lower-intensity jammers (Fig. 2). The EIs in lower-intensity jammer conditions are shown in Fig. 5. The indices of HA1 at 0.5 and 1 kHz in the 60-dB, 45-dB, and quiet conditions and the indices of HA 3 in the 60-dB conditions across frequencies were larger than those of other hearing

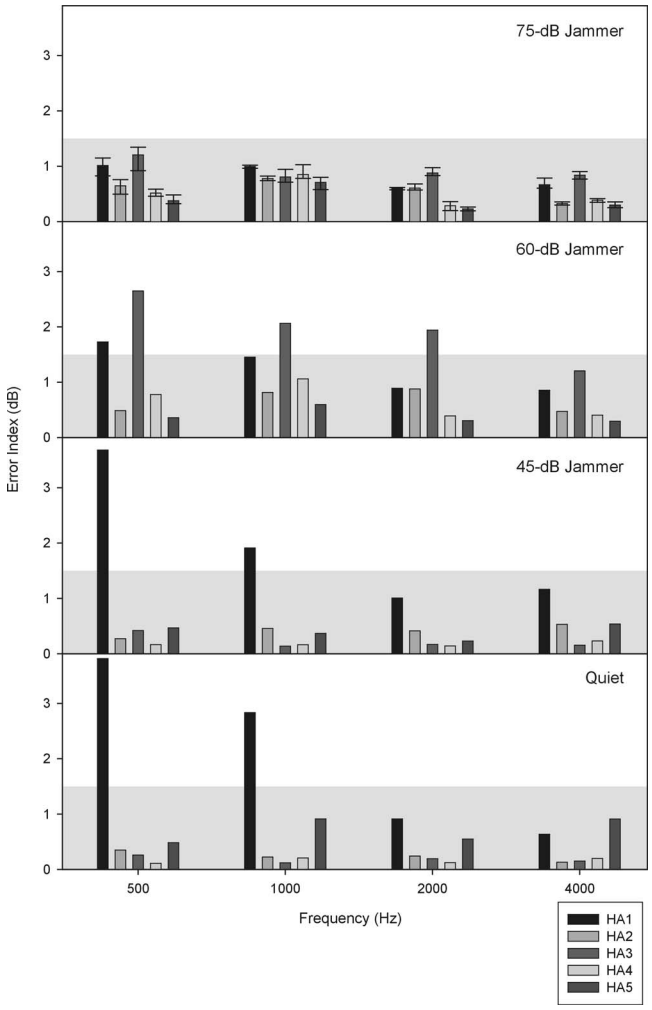


FIG. 5. EI measured in each of the test conditions as a function of frequency. The error bar in the upper panel represents the data range from five measurements. The shaded area represents the norm (see text).

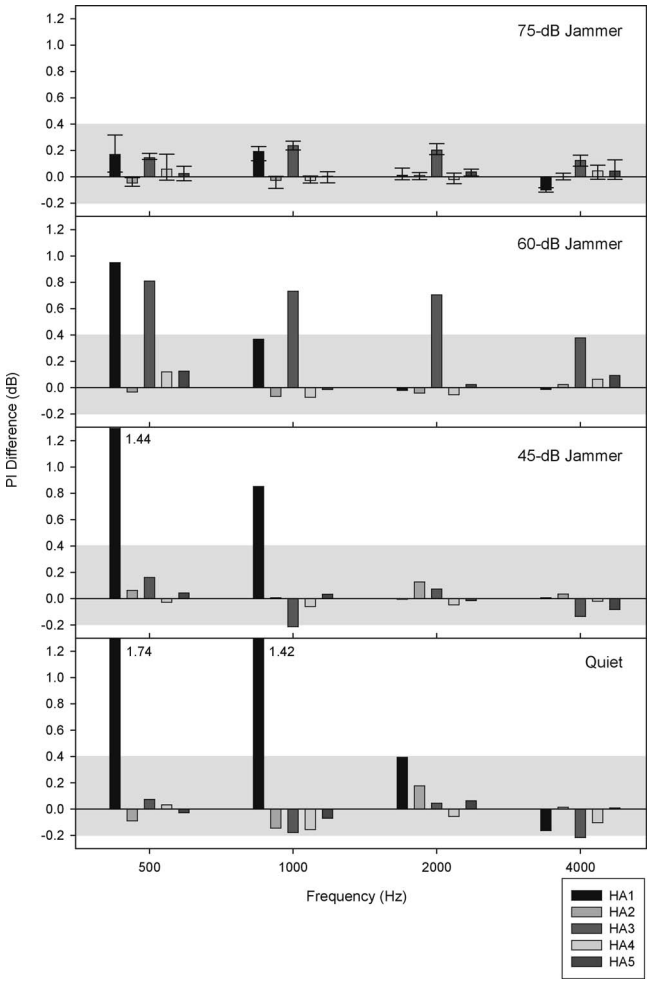


FIG. 6. Difference between measured and corrected polar index (i.e., PI difference) as a function of frequency. The error bar in the upper panel represents the data range from five measurements. Numbers shown in the lower panels indicate the PI difference values, which are beyond the scale of the *y*-axis and are truncated. The shaded area represents the norm (see text).

aids. The rest of the EIs were close to or smaller than those obtained in the 75-dB condition. Similar trends in the PI difference can be found in Fig. 6.

B. Repeatability

To evaluate the repeatability of the IR method, each aid was tested five times under the 75-dB jammer condition. The standard deviations of the PI and directional response at each azimuth were computed. In Fig. 3, the error bar represents 1 standard deviation of the measured PI. All standard deviations were smaller than 0.1 dB. Standard deviations of the directional response as a function of azimuth angle at 0.5, 1, 2, and 4 kHz are shown in Fig. 7. Generally, the variations in the response were small (88% of the standard deviations $<$0.3 dB), except for those of HA1 in the high frequency bands at the rear azimuths.

V. DISCUSSION

A. General discussion

In Experiment 1, the directivity adaptation time was measured for the five DMHAs. A comparison of the data shown in Table I and Figs. 5 and 6 shows that the EI values and the PI difference values were larger for the hearing aids whose adaptation times were shorter than 10 ms. Therefore, the hearing aids tested in this study were classified as fast-acting (shorter than 10 ms) and slow-acting (longer than 10 ms). HA1 was fast-acting, while HA2, HA4, and HA5 were slow-acting systems. Note that HA 3 had two different adaptation times. It was fast-acting in steering the null within the directional mode, while in a quiet condition it was slow-acting in switching from the default mode, which is the omnidirectional mode (Fig. 2), to the directional mode.

The large range of the times shown in Table I suggests that different manufacturers choose the directivity adaptation time based on different philosophies. Systems with a short adaptation time may achieve the maximal noise attenuation by always aiming the null at a moving noise source or at a transient noise. This fast-changing directivity also may compensate for head movement. However, switching directivity too fast may cause artifacts that could annoy the hearing aid user (Olson *et al.*, 2004) and may attenuate desirable signals such as soft speech or warning sounds from behind (Kuk *et al.*, 2005). On the other hand, using a long adaptation time

TABLE III. Polar indices of different hearing aids in lower-intensity jammer conditions.

	Jammer condition											
	60 dB				45 dB				Quiet			
Frequency (kHz)	0.5	1	2	4	0.5	1	2	4	0.5	1	2	4
HA1	5.9	5.9	6.0	6.4	4.9	5.4	5.4	6.2	3.8	4.0	5.6	6.1
HA2	5.2	6.0	5.9	5.9	2.8	5.7	5.8	5.6	−0.1	−0.1	−0.1	0.9
HA3	5.8	6.0	5.9	5.8	−0.1	0.0	0.3	0.7	−0.1	0.1	0.2	0.7
HA4	5.8	5.8	5.8	6.0	−0.1	−0.1	−0.3	0.6	−0.1	−0.1	−0.6	0.3
HA5	4.6	5.8	5.7	5.5	4.5	5.2	5.6	5.6	4.4	4.8	4.5	4.7

may preserve the audibility of the signal so that the listener will have a chance to decide whether or not the signal is desirable (Kuk *et al.*, 2005).

In Experiment 2, the accuracy, repeatability, and generalizability of the IR method were evaluated. The IR method can be considered accurate in the 75-dB condition for all hearing aids tested in this study. This is supported by several observations. First, all the directivity patterns in this condition were consistent with the theoretical pattern, except for the asymmetric patterns of HA1 at all frequencies and the transitional pattern of HA 2 at 500 Hz (Fig. 2, first column). Second, all the PIs were smaller than 6 dB, except for HA1 in the high frequency band (Fig. 3). The larger-than-6-dB PI is due to HA1's non-adaptive second-order microphone system. Third, the PIs obtained by the IR and SC methods were very close and no systematic difference can be seen between them (Fig. 3). Fourth, error values were mostly smaller than 0.5 dB, except for those around the azimuth of the null (Fig. 4). Fifth, all the EIs were smaller than 1.5 dB, which were smaller than the EI range of the SC method under the same jammer condition (from 1.5 to 4 dB, Wu and Bentler, 2007). Sixth, the PI difference values were between −0.2 and 0.4 dB, which were also smaller than those obtained by the SC method under the same jammer condition (±0.3–±0.6 dB, Wu and Bentler, 2007). Although these observations support the accuracy in the 75-dB condition, it should be noted that the error values of HA1 and HA3 tended to be positive when the probe was presented from the rear of these two hearing aids (Fig. 4). Also note that although the PI difference values were small, the values of the difference of fast-acting HA1 and HA3 were still larger than those of other slow-acting aids, especially at 0.5 and 1 kHz (Fig. 6, upper panel). These data suggest that the probe might have changed the directivity pattern of these two aids, even under the presentation of the 75-dB jammer. However, recall that the EI values and the PI difference values were derived by presenting two consecutive clicks. These metrics with small values suggest that the IR method is accurate, while metrics with large values do not necessarily suggest inaccuracy. Even if the single click did change the directivity, the amount of change did not significantly alter the shape of the pattern or the PI value by 0.4 dB (i.e., the PI difference). This accuracy should be sufficient for hearing aid researchers to judge the directivity pattern of DMHAs. Thus, this leads us to conclude that directivity patterns obtained in the 75-dB condition were accurate. If so, the EI of 1.5 dB and the PI difference ranging from −0.2 to +0.4 dB can be used as the norm of the IR method. If a measured EI is smaller than 1.5 dB and a PI difference is within the −0.2 to +0.4 dB range in a given noise configuration, the result would be as accurate as the result obtained in the 75-dB jammer condition. The norms are shown in Figs. 5 and 6 by the shaded areas.

In the conditions involving jammer levels lower than 75-dB SPL, the accuracy metrics of fast- and slow-acting systems showed different results. It appears that the IR method generated accurate directivity patterns when testing slow-acting systems at all jammer levels. Specifically, for HA2, HA4, and HA5, most of their polar patterns were consistent with the theoretical patterns (Fig. 2) and the EI and PI differences were within the norms in all jammer conditions (Figs. 5 and 6). Note that some directivity patterns shown in

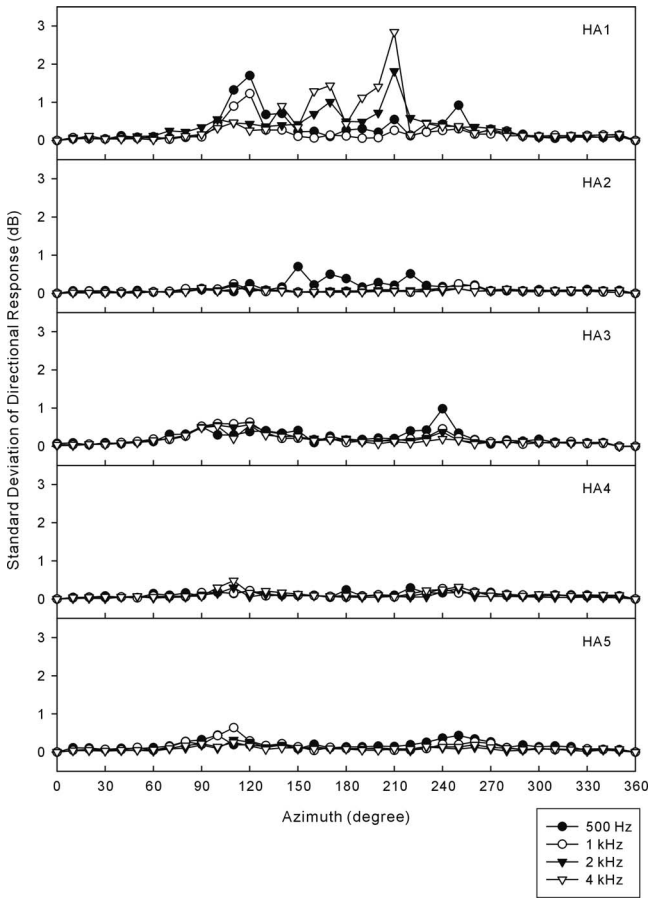


FIG. 7. Standard deviation of the directional responses in the 75-dB condition as a function of azimuth angle.

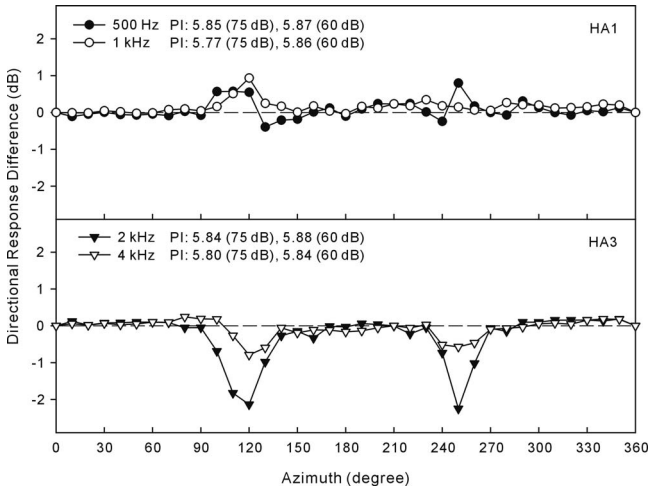


FIG. 8. Differences in directional responses between the 75- and 60-dB jammer conditions for HA1 (0.5 and 1 kHz) and HA3 (2 and 4 kHz), respectively (the mean response in the 75-dB condition minus the response in the 60-dB condition). PIs in these two jammer conditions are also shown.

Fig. 2 do not look like theoretical patterns, such as those of HA2 in the 75-, 60-, and 45-dB conditions at 500 Hz. As mentioned previously, these patterns are believed to be the transitions of different polar patterns in higher and lower frequency bands. The same explanation can be applied to the patterns of HA5 in the 45-dB condition at 0.5 and 1 kHz. For HA3 in the 45-dB and the quiet conditions, the results were also accurate because all the metrics were within the normal limit. Note that the microphone mode of HA3 was omnidirectional in these two conditions, needing longer than 10 ms to transit to the directional mode (Table I).

At first glance it appears that the IR method did not work very well with fast-acting systems in low-level jammer conditions. Specifically, for HA1's adaptive system in the low-frequency band, the EI values and the PI difference values were larger than the norms in jammer conditions with lower levels. The same trends can be observed for HA3 at 60-dB conditions across frequencies (Figs. 5 and 6). These data are consistent with the fact that fast-acting systems can respond to transient signals. However, it should be noted that the shapes of HA1's polar patterns in the 75- and 60-dB conditions were very similar (Fig. 2). The patterns of HA3 at 2 and 4 kHz were also similar in these two conditions. In the lower-intensity jammer conditions, similar to HA2, the patterns of HA3 at 0.5 and 1 kHz have started to shift into the omnidirectional pattern (Fig. 2). In addition to the similar patterns, the PIs of these two aids were very close in the 75- and 60-dB conditions (shown in Fig. 8). The similarities in the shape of the polar pattern and PI value suggest that the metrics EI value and PI difference value might overestimate the error in the 60-dB condition. To examine this possibility, the directional responses of HA1 (0.5 and 1 kHz) and HA3 (2 and 4 kHz) measured in 75- and 60-dB conditions are compared, as shown in Fig. 8. This figure indicates that the directional responses of these two aids are very close in the two jammer conditions, except for those around the null. The measurement around the null was easily contaminated by the noise floor and the positioning of the hearing aid in the measurement. The similarities of directional responses/PI values in the 75- and 60-dB conditions (Fig. 8) and the differences of EI values/PI differences in those two conditions (Figs. 5 and 6) indicate the likelihood that the directivity pattern of the two fast-acting DMHAs measured in the 60-dB condition was accurate and the EI value and PI difference value overestimated the errors. For HA1 in the 45-dB and quiet conditions, the probe probably has changed the directivity pattern because the EI and PI difference values were very large (Figs. 5 and 6).

In summary, the IR method was accurate for slow-acting systems at all the jammer levels tested in this study. The IR method was accurate for fast-acting systems in the 75-dB jammer condition, probably accurate in the 60-dB condition and probably inaccurate in the 45-dB and quiet conditions. Unfortunately, the accuracy in the latter two conditions cannot be proved or disproved by the metrics used in this study. Therefore, it is concluded that the generalizability of the IR method is fair because the accuracy in assessing fast-acting systems under low-intensity jammer conditions is unknown. More research is needed to explore better metrics that can evaluate the accuracy of the IR method in low-intensity conditions.

The IR method was considered repeatable because of the small standard deviations of the PI (Fig. 3) and directional response (Fig. 7), at least in the 75-dB condition. The directional responses shown in Fig. 7 were more variable around the azimuth of the null (e.g., between 120° and 240° for HA1 at 2 and 4 kHz, and 110° and 250° for the rest of hearing aids). This is because the measurement was more likely to be contaminated by the noise floor around these azimuths, as mentioned previously.

B. Implementation and limitations

The IR method was designed to help hearing aid researchers assess the directivity of a DMHA in a given laboratory test condition. This method can be implemented in most hearing aid laboratories/anechoic chambers by converting the equipment set-up required for regular directivity measurement (ANSI, 2004). Although it was not tested in this study, the directivity pattern of a DMHA in a reverberant environment can be obtained using the IR method-derived impulse response and the quasi-anechoic technique (Everest, 2000), given that the space of the test environment is large enough. The IR method also can be used with a manikin. However, the equipment set-up needs to be modified. One of our turntables had been modified to incorporate a Knowles Electronic Manikin for Acoustic Research (KEMAR) in the measurement (Fig. 9). A hole was created at the center of the turntable and the KEMAR was installed at the turntable's base. The KEMAR did not move when the turntable was rotating.

Before implementing the IR method, it would be helpful to measure the hearing aid's directivity adaptation time using the methods suggested in Experiment 1. The IR method may not accurately assess a fast-acting aid, especially in the low-intensity jammer condition. The norms obtained in Experiment 2 can then be used as criteria to judge if the result is not

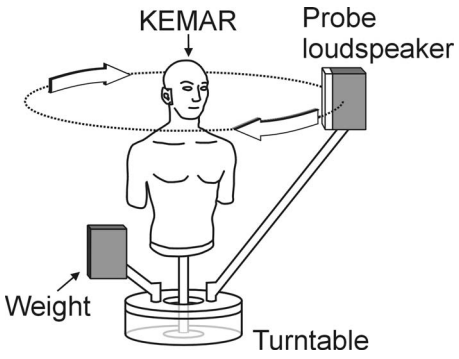


FIG. 9. The equipment set-up for using the IR method with a manikin (KEMAR: Knowles Electronic Manikin for Acoustic Research).

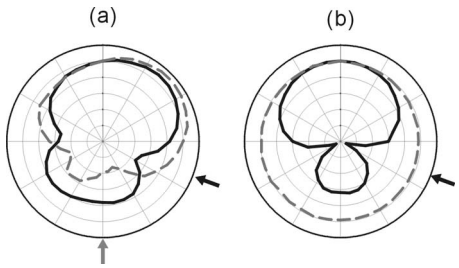


FIG. 10. (a) The 1-kHz spatial directivity pattern of HA 4 on a manikin's right ear responding to a jammer presented from 110° (solid curve) and that responding to a jammer from 180° (dashed curve). The jammer was white noise at 75-dB SPL. (b) The free-field spatial directivity pattern at 1 kHz of HA3 responding to the six-talker babble of the connected speech test (solid curve) and that responding to the homemade six-talker babble (dashed curve). The babbles were presented from a 110° azimuth at a level of 62 dB SPL. The arrow indicates the azimuth of the jammer or babble signal.

accurate and needs to be rejected. Specifically, an EI value larger than 1.5 dB (Fig. 5) or a PI difference value larger than 0.4 dB (Fig. 6) indicates that the result may be inaccurate. An example is given in Fig. 10(a). This plot reports the polar pattern at 1 kHz of HA4 on the KEMAR responding to a single jammer presented from 110° (solid curve) and that responding to a single jammer from 180° (dashed curve). The EI (0.3 and 0.6 dB for 110° and 180°, respectively) and PI difference values (0.1 dB for both azimuths) were within the suggested criteria, indicating accurate results. Note that the shape of the polar pattern and the PI value could not be used to judge the accuracy of the result in this example because the device's directivity pattern on the KEMAR's head could not be described by Eq. (6).

Note that using the IR method to assess hearing aids in a nonlinear mode or with their adaptive features (such as noise reduction algorithms and adaptive feedback cancelers) enabled should be done with caution. Although the signal cancellation technique described by Hagerman and Olofsson (2004) has been successfully used on hearing aids that are nonlinear (Bentler *et al.*, 2006) or with an enabled noise reduction algorithm (Hagerman and Olofsson, 2004), these features may have very short attack or engaging times (e.g., <10 ms) so that they can respond to the probe signal, resulting in inaccurate results. The anti-shock algorithm that is designed to control loud transients (e.g., Hayes, 2006) may also make the IR method invalid.

C. Examples of application

Two examples of the IR method's applications are given below. Multi-talker babble is widely used in tests examining speech perception amid noise. For some advanced DMHAs that can recognize speech signals and change directivity patterns accordingly, it is important to know how they respond to babble noise. Figure 10(b) shows the free-field polar pat-

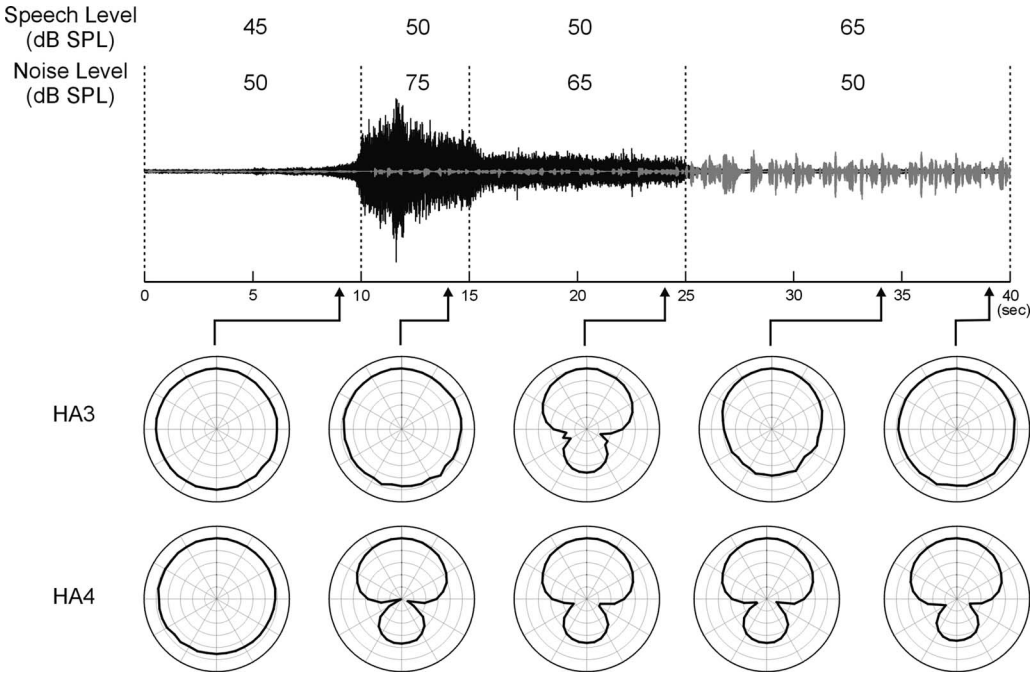


FIG. 11. Spatial directivity patterns at 1 kHz of HA3 and HA4 responding to speech (the gray waveform) amid car noise (the black waveform) with different levels presented from a 110° azimuth. The numbers represent the energy equivalent sound level (Leq) of speech and car noise in each time period indicated by the dashed lines. The directivity patterns were measured at the 9th, 14th, 24th, 34th, and 39th seconds.

tern at 1 kHz of HA3 responding to the six-talker babble of the Connected Speech Test (Cox *et al.*, 1987) (solid curve) and that responding to the "homemade" six-talker babble (dashed curve) presented from a 110° azimuth. The homemade babble was created by superimposing six different continuous discourses of the Speech Intelligibility Rating Test (Cox and McDaniel, 1989). Apparently HA3 treated the former babble as noise and shifted into the directional mode, and treated the latter one as speech and shifted into the omnidirectional mode. This example emphasizes the importance of verifying the directivity of adaptive/automatic directional systems before the research commences. Another example is shown in Fig. 11 in which a series of speech and car noise signals of different levels were presented to HA3 and HA4 from a 110° azimuth. The directivity patterns were measured at five different time points. It can be observed that (1) both hearing aids stayed in the omnidirectional mode at low presentation levels, (2) HA4 responded faster than did HA3 to the car noise at a rapidly increasing level, and (3) the two hearing aids responded differently to speech signals amid noise. This figure indicates that the IR method can obtain a snapshot of the directivity pattern at a given moment. For example, the directivity pattern of HA3 at the 34th second shows the transition from the directional mode to the omnidirectional mode. This information will help investigators design studies and predict the results of experiments.

VI. CONCLUSIONS

The IR method is an accurate and repeatable method to evaluate the directivity of most adaptive/automatic DMHA in most jammer conditions. This method may not be able to assess fast-acting systems in low-level conditions (i.e., lower than 60-dB SPL) because it tends to be inaccurate and the involved metrics, such as the EI and PI difference, tend to overestimate the error. For research purposes, the IR method provides a powerful way to assess and verify the adaptive/automatic directional microphone system of hearing aids.

ACKNOWLEDGMENTS

This research was in part supported by Grant No. H133E080006 from the Department of Education/National Institute on Disability and Rehabilitation Research. The authors would like to acknowledge three anonymous reviewers for providing a variety of insightful comments and suggestions regarding this work.

ANSI, (**1984**). "American National Standard: American National Standard preferred frequencies, frequency levels, and band numbers for acoustical measurements" (American National Standards Institute, New York), ANSI S1.6-1984 (R2006).

ANSI, (**2004**). "American National Standard: Method of measurement of performance characteristics of hearing aids under simulated real-ear working conditions" (American National Standards Institute, New York), ANSI S3.35-2004.

Bentler, R. A., Palmer, C., and Dittberner, A. B. (**2004**). "Hearing-in-noise: Comparison of listeners with normal and (aided) impaired hearing," J. Am. Acad. Audiol **15**, 216–225.

Bentler, R. A., Wu, Y. H., and Jeon, J. (**2006**). "Effectiveness of directional technology in open-canal hearing instruments," Hear. J. **59**, 40–47.

Beranek, L. L. (**1949**). *Acoustic Measurements* (Wiley, London).

Blamey, P. J., Fiket, H. J., and Steele, B. R. (**2006**). "Improving speech intelligibility in background noise with an adaptive directional microphone," J. Am. Acad. Audiol **17**, 519–530.

Chung, K. (**2004**). "Challenges and recent developments in hearing aids. Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms," Trends Amplif. **8**, 83–124.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). "Development of the connected speech test (CST)," Ear Hear. **8**, 119S–126S.

Cox, R. M., and McDaniel, D. M. (**1989**). "Development of the speech intelligibility rating (SIR) test for hearing aid comparisons," J. Speech Hear. Res. **32**, 347–352.

Dirks, D. D., Morgan, D. E., and Dubno, J. R. (**1982**). "A procedure for quantifying the effects of noise on speech recognition," J. Speech Hear. Disord. **47**, 114–123.

Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). "Effects of age and mild hearing loss on speech recognition in noise," J. Acoust. Soc. Am. **76**, 87–96.

Everest, F. A. (**2000**). *Master Handbook of Acoustics* (McGraw Hill, Columbus).

Hagerman, B., and Olofsson, A. (**2004**). "A method to measure the effect of noise reduction algorithms using simultaneous speech and noise," Acta. Acust. Acust. **90**, 356–361.

Hayes, D. (**2006**). "Transient impulse control for hearing aids," Hear. Rev. **13**, 56–59.

Hornsby, B. W., and Ricketts, T. A. (**2007**). "Directional benefit in the presence of speech and speechlike maskers," J. Am. Acad. Audiol **18**, 5–16.

Kates, J. M. (**2008**). *Digital Hearing Aids* (Plural, San Diego).

Kuk, F., Keenan, D., Lau, C. C., and Ludvigsen, C. (**2005**). "Performance of a fully adaptive directional microphone to signals presented from various azimuths," J. Am. Acad. Audiol **16**, 333–347.

Olson, L., Ioannou, M., and Trine, T. D. (**2004**). "Appraising an automatically switching directional system in the real world," Hear. J. **57**, 32–38.

Palmer, C., Bentler, R. A., and Mueller, H. G. (**2006**). "Evaluation of a second-order directional microphone hearing aid: II. Self-report outcomes," J. Am. Acad. Audiol **17**, 190–201.

Puder, H. (**2006**). "Adaptive signal processing for interference cancellation in hearing aids," Signal Process. **86**, 1239–1253.

Ricketts, T. (**2000**). "Directivity quantification in hearing aids: Fitting and measurement effects," Ear Hear. **21**, 45–58.

Ricketts, T., and Dittberner, A. B. (**2002**). "Directional amplification for improved signal-to-noise ratio: Strategies, measurement, and limitations," in *Strategies For Selecting And Verifying Hearing Aid Fittings*, edited by M. Valente (Theme, New York), pp. 274–346.

Ricketts, T., and Henry, P. (**2002**). "Evaluation of an adaptive, directional-microphone hearing aid," Int. J. Audiol. **41**, 100–112.

Valente, M., Hosford-Dunn, H., and Roeser, R. J. (**2000**). *Audiology Treatment* (Thieme, New York).

Wu, Y. H., and Bentler, R. A. (**2007**). "Using a signal cancellation technique to assess adaptive directivity of hearing aids," J. Acoust. Soc. Am. **122**, 496–511.

Analysis of acoustic reduction using spectral similarity measures

Annika Hämäläinen,[a] Michele Gubian, Louis ten Bosch, and Lou Boves
Centre for Language and Speech Technology (CLST), Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

(Received 6 December 2008; revised 10 September 2009; accepted 11 September 2009)

Articulatory and acoustic reduction can manifest itself in the temporal and spectral domains. This study introduces a measure of spectral reduction, which is based on the speech decoding techniques commonly used in automatic speech recognizers. Using data for four frequent Dutch affixes from a large corpus of spontaneous face-to-face conversations, it builds on an earlier study examining the effects of lexical frequency on durational reduction in spoken Dutch [Pluymaekers, M. *et al.* (2005). J. Acoust. Soc. Am. **118**, 2561–2569], and compares the proposed measure of spectral reduction with duration as a measure of reduction. The results suggest that the spectral reduction scores capture other aspects of reduction than duration. While duration can—albeit to a moderate degree—be predicted by a number of linguistically motivated variables (such as word frequency, segmental context, and speech rate), the spectral reduction scores cannot. This suggests that the spectral reduction scores capture information that is not directly accounted for by the linguistically motivated variables. The results also show that the spectral reduction scores are able to predict a substantial amount of the variation in duration that the linguistically motivated variables do not account for. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3243291]

PACS number(s): 43.70.Fq, 43.72.Ar, 43.72.Lc, 43.72.Ne [DOS] Pages: 3227–3235

I. INTRODUCTION

It has long been known that words in normal speech—in particular, in spontaneous speech—are frequently pronounced in a more reduced form than their canonical phonetic transcriptions would suggest (e.g., Ernestus, 2000; Ernestus *et al.*, 2006; Jespersen, 1922; Lindblom, 1963; Pluymaekers *et al.*, 2005; Zipf, 1929). Weak forms of reduction may become manifest in the acoustic signal as shortened segments with flatter spectral envelopes, while strong reduction may result in the deletion of phonemes or whole syllables (Greenberg, 1999; Johnson, 2004). It has been hypothesized that the degree of reduction could be explained by the amount of information carried by the word in question. This has resulted in competing theories, such as the smooth signal redundancy hypothesis (Aylett and Turk, 2004), the probabilistic redundancy hypothesis (Jurafsky *et al.*, 2001), and the speech efficiency hypothesis (van Son and Pols, 2003). Different theories seem to invoke different cognitive and physiological processes, such as the compression of motor routines as a result of practice (Bybee, 2001), as well as adaptation to the needs of the listener (e.g., Jurafsky *et al.*, 2001). All theories aim to explain reduction phenomena that are manifest in both the temporal and spectral domains.

It has, however, proved difficult to design experiments for investigating the causes of reduction in detail. This is because it is difficult to exert enough experimental control for a fair comparison of reduction when words do not only differ in frequency, but also in their intrinsic phonemic and morphological complexity, such as the number and type of phonemes they consist of (Pluymaekers *et al.*, 2005). To avoid these difficulties, Pluymaekers *et al.* (2005) investigated reduction by focusing on affixes, i.e., on morphemes that can occur in a large number of different words with varying frequencies. More specifically, they studied the role of various linguistically motivated predictors (e.g., word frequency, speech rate, and the age and regional origin of the speaker) in explaining reduction observed in syllable-sized affixes.

Pluymaekers *et al.* (2005) chose to use a correlate of reduction that is relatively easy to measure in the acoustic speech signal: duration. They showed that regression models based on linguistically motivated variables could, at best, predict moderate proportions of variance in duration (i.e., the dependent variable). Reduction is, however, known to manifest itself in many different ways, and duration only reflects part of the reduction phenomenon. Therefore, it is worthwhile investigating other indices of reduction in the acoustic speech signal, as well. Because of the relation between the gestures of the articulators and the spectrum of the resulting speech signal, *spectral* reduction measures are particularly interesting. In this paper, we propose an automatically derived measure of spectral reduction and test it using the same data as Pluymaekers *et al.* (2005). The resulting spectral reduction scores reflect the reduction phenomenon in a different way than duration does. In this paper, we therefore investigate the relation between the newly developed spectral reduction measure, the duration-based reduction measure, and the linguistically motivated context variables employed by Pluymaekers *et al.* (2005).

Scholars agree that reduction must be interpreted as the deviation of an observed pronunciation from some reference pronunciation. Since speech production involves multiple ar-

[a]Author to whom correspondence should be addressed. Electronic mail: a.hamalainen@let.ru.nl

ticulators and results in acoustic trajectories in a high-dimensional acoustic space, deviation from a reference pronunciation can take place along several different dimensions. Using duration as the only measure of reduction would leave open the option of reduction being limited to a time compression of otherwise "unreduced" articulatory gestures. However, most studies on reduction imply that, in addition to being shorter, the articulatory gestures are simplified. This "simplification" should manifest itself in the spectral structure of the signals. A spectral measure of reduction captures the deviation between an actual trajectory and a "reference trajectory" in the acoustic space. In our case, this is the deviation between an observed acoustic token (e.g., a particular instance of an affix) and the reference model of the token. Coarticulation is a pervasive phenomenon in speech, and its effects could be interpreted as just another, unavoidable manifestation of reduction. We should point out that our definition of reduction also holds in the presence of coarticulation effects; spectral reduction can always be interpreted as the deviation between the observed and the reference trajectories in the acoustic space, with the reference trajectories including coarticulation effects. The goal of this paper is to investigate whether duration and spectral reduction are overlapping or complementary indices of the underlying articulatory simplification in the case of syllable-sized affixes. To that end, we carry out experiments using the same data as Pluymaekers *et al.* (2005).

This paper is further organized as follows. In Sec. II, we introduce our approach to quantifying spectral reduction. We recapitulate the speech material in Sec. III, and describe the statistical variables used in this study in Sec. IV. We discuss the design and results of our first experiment in Sec. V, and do the same for a follow-up experiment in Sec. VI. Finally, we discuss the results and suggest directions for future research in Sec. VII, and conclude the paper in Sec. VIII.

II. QUANTIFYING SPECTRAL REDUCTION

The question how to quantify spectral reduction can be made more precise by asking how to quantify the amount of (dis)similarity between the reduced and reference realizations of a speech unit—in our case, syllable-sized affixes. To that end, speech decoding and alignment techniques developed for automatic speech recognition provide powerful tools. Speech recognizers based on hidden Markov models (HMMs) are able to provide estimates of the degree of (dis-)similarity between a particular stretch of speech signal and a model of the acoustics of the corresponding speech unit(s) [e.g., phoneme(s), syllable(s), or word(s)] derived from some corpus of training data. One such estimate is the log-likelihoods (usually referred to as *acoustic scores*) that HMM-based speech recognizers compute as a by-product of forced alignment. Forced alignment is a technique in which a speech signal is aligned with a predefined sequence of acoustic models associated with speech units (e.g., phonemes, syllables, or words). The output of the alignment is a score for the goodness of the fit between the speech signal and the models, usually in combination with a corresponding segmentation. Forced alignment can also be used for estimating the best transcription for a word token: If a word is represented by more than one phonemic transcription in the recognizer lexicon, the forced alignment procedure is able to select the most likely one. The result of the forced alignment then depends on the available phonemic transcriptions ("candidate transcriptions") of the word in the lexicon and the quality of the acoustic models corresponding to these phonemic transcriptions. Should the recognizer lexicon only contain *one* possible transcription per word, the acoustic score for each token of that word would express how well the signal matches that single transcription. Should that single transcription be a canonical transcription (which is the closest we can get to a reference pronunciation of the word), the total acoustic score would express how well the signal matches the reference. Below, we argue why the acoustic scores obtained from forced alignment with a sequence of HMMs corresponding to a canonical transcription are viable estimators of spectral deviation and, consequently, viable estimators of spectral reduction. By using just a single scalar to represent the distance between the models and the actual acoustic signals, we clearly lose information about the details (temporal and spectral) of the deviation between the token and the model. However, the spectral reduction measure obtained in this way provides information that reflects the deviation from the reference in the articulatory and acoustic space better than a plain duration measure can do. Both measures reflect differences between acoustic trajectories, but focus on different kinds of differences between these trajectories.

The rationale underlying our approach to computing spectral reduction is as follows. Suppose $X=\{x_1,x_2,\ldots,x_N\}$ is a sequence of observed acoustic feature vectors, and $S=\{s_1,s_2,\ldots,s_K\}$ is the sequence of HMM states used in the forced alignment between the speech signal and the corresponding acoustic models. The alignment procedure returns the log-likelihood $\log P(X|S)$ defined by

$$\log P(X|S) = \log \prod_{n,j} P_e(x_n|s_j) \prod_{j,i} P_t(s_j|s_i), \tag{1}$$

in which P_e and P_t denote the emission and transition probabilities and the (n,j) and (j,i) pairs are uniquely determined by the alignment path (the indices i and j specify the indices of the states, and n specifies the frame index, along the path resulting from the alignment).

To justify that Eq. (1) leads to a viable estimator of acoustic reduction, please notice that, for a single feature vector x and a HMM state s, the distance (dissimilarity) between the feature vector and the HMM state can be written as

$$d^2 = -\log(P_e^M(x|s)) - \log(P_t(s|s_{\text{previous}})), \tag{2}$$

where P_e^M denotes the emission probability modeled by a mixture of M Gaussians. To obtain a measure of dissimilarity between a vector sequence and an acoustic model represented by a sequence of HMM states, the dissimilarity scores d^2 along the best path through the trellis must be accumulated as

TABLE I. The number of tokens, the number of speakers, the maximum number of tokens uttered by each speaker, and the broad phonetic transcriptions of the uttered tokens for each affix.

Affix	No. of tokens	No. of speakers	Maximum (No. of tokens/speaker)	Phonetic transcriptions
ge-	427	132	12	/x@/, /x/, /G@/, /G/
ver-	137	80	8	/v@r/, /v@/, /vEr/, /vr/, /v/, /f@r/, /f@/, /f/
ont-	101	63	4	/Ont/, /Ond/, /Omp/, /Od/, /Om/, /On/, /Ot/, /@nd/, /@nt/, /@n/, /@t/;/l@k/, /l@g/, /lEk/, /lIk/, /lYk/,
-lijk	157	87	6	/l@/, /lk/; /@k/, /@/, /g/, /k/

$$D = \sum_{n=1}^{N} d_n^2 = -\sum_{n,j} \log(P_e^M(x_n|s_j)) - \sum_{i,j} \log(P_t(s_j|s_i)). \quad (3)$$

In this expression, the sum over $\log(P_e)$ represents the spectral distance between the token and the models, while the sum over $\log(P_t)$ represents the total scores associated with the state-to-state transition probabilities. The dissimilarity score D depends on the duration of the speech segment (represented by the number of frames in the sequence of input frames). To be able to compare the results of Eq. (3) *across* tokens of different durations, we obtain an *average frame-to-state dissimilarity* by normalizing the score D for the number of frames

$$D_{\text{norm}} = \frac{-\sum_{n,j} \log(P_e^M(x_n|s_j)) - \sum_{i,j} \log(P_t(s_j|s_i))}{N}. \quad (4)$$

Equation (4) is the expression used in this paper to compute the final spectral reduction scores.

In this study, we use the HMM toolkit (Young *et al.*, 2002), which actually outputs *similarity scores* instead of dissimilarity scores. Therefore, we use $-D_{\text{norm}}$ from Eq. (4) as the spectral reduction score in this paper.

III. SPEECH MATERIAL

We re-used the affix data that were selected and measured by Pluymaekers *et al.* (2005). These data originate from spontaneous face-to-face conversations between speakers of Dutch (as spoken in The Netherlands) in the Spoken Dutch Corpus [Corpus Gesproken Nederlands (CGN)] (Oostdijk *et al.*, 2002).

We investigated the prefixes ge-, ver-, and ont-, and the suffix -lijk. ge- is commonly used to create the perfect participle in Dutch [e.g., *gespeculeerd* (the perfect participle form of the verb "to speculate")], and can also appear as a nominal or a verbal prefix {e.g., *gebak* ["cake(s)"]; *gebeuren* ("to happen")}. However, we only investigated the participial instances of ge-. ver- and ont- are verbalizing prefixes expressing change in state [e.g., *verplaatsen* ("to move")] and reversal or inchoation [e.g., *onteigenen* ("to disown")]. The suffix -lijk appears in adverbs and adjectives {e.g., *natuurlijk* ["natural(ly)"]; *eigenlijk* ["actual(ly)"]}. The canonical phonetic transcriptions (using the Speech Assessment Methods Phonetic Alphabet) of the four affixes are /x@/, /v@r/, /Ont/, and /l@k/, respectively. (Pluymaekers *et al.*, 2005.)

Pluymaekers *et al.* (2005) provide a detailed description of the selection of the affix tokens that were analyzed. To summarize, they selected one token for each word type containing a target affix. As word types, they did not only consider words belonging to different lemmas but also different word forms of the same lemma (e.g., the sample for the affix ont- included both *ontwikkelt* "develops" and *ontwikkelde* "developed"). The recordings contained the complete utterances in which the affixes were embedded. Table I presents an overview of the affix samples used in the study.

IV. STATISTICAL VARIABLES

The statistical variables we used in this study included the spectral reduction scores, which we used both as a dependent variable and as a predictor; duration, which we used as a dependent variable, and the linguistically motivated variables from Pluymaekers *et al.* (2005), which we used as predictors. In this section, we describe these variables in more detail.

A. Spectral reduction scores

We obtained the spectral reduction scores by carrying out forced alignment on the stretches of speech that Pluymaekers *et al.* (2005) had manually labeled as the target affixes. When carrying out the forced alignment, we used a single sequence of HMM states for each affix. This sequence was formed by concatenating the triphone models underlying the canonical transcription of the affix in question.

As the model topology for the triphone models, we used standard three-state left-to-right HMMs with no state skips allowed. We carried out feature extraction of the affix data and of the data used for training the triphone models at a frame rate of 5 ms using a 25-ms Hamming window and applied first order pre-emphasis to the signal using a coefficient of 0.97. Using the "default" frame rate of 10 ms in combination with the chosen model topology would have required the ge- tokens to have a minimum duration of 60 ms (i.e., two phone models × three states per model, at least one frame per state) and the ver-, ont-, and -lijk tokens to have a minimum duration of 90 ms to allow alignment. Reducing the frame rate to 5 ms allowed us to obtain acoustic scores for the vast majority of the very short affix tokens as well. We calculated 12 mel frequency cepstral coefficients and log-energy with first and second order derivatives. We applied channel normalization using cepstral mean normalization over the complete recordings.

We carried out forced alignment using two different sets of triphone models. The first set of triphones (*manual triphones*) comprised 8-Gaussian HMMs trained with the *manu-*

ally verified transcriptions of the read speech in the core set of CGN. The training data contained 45 172 orthographic word tokens (4 h, 51 min, 27 s of speech). The second set of triphones (*canonical triphones*) comprised 64-Gaussian HMMs trained with *canonical* transcriptions of a much larger part of the read speech data in CGN. The training data contained 396 187 orthographic word tokens (37 h, 20 s of speech). The (standard) triphone training procedure is described in Hämäläinen *et al.*, 2007 for the manual triphones, and in Hämäläinen *et al.*, 2009 for the canonical triphones. For this study, we carried out state tying such that both sets of triphones had about 3400 physically distinct triphones. While the amount of training data and the number of Gaussian mixtures were different for the two sets, the number of data points (frames) used to define each diagonal-covariance Gaussian after tying was almost equal.

The reason to use triphones trained on *read speech* was that we wanted to base the spectral reduction scores on the dissimilarity between an individual affix token and a maximally unreduced form of the affix. Such maximally unreduced form can be considered maximally similar to the canonical pronunciation of the affix. Triphones trained on carefully read speech provided us with a reference that was as unreduced as possible. The triphones trained with manually verified transcriptions were arguably the "cleanest" models in this sense. However, as manually verified transcriptions are not always available in speech corpora because of their expensiveness, we also tested triphones trained with canonical transcriptions of read speech.

Unlike Pluymaekers *et al.* (2005), who also fitted models to predict the durations of the individual segments of the affixes, we only carried out statistical analyses on the affix level. This is because the acoustic scores obtained for individual segments using forced alignment are not necessarily meaningful due to differences between manual and automatic segmentations. The acoustic scores that the forced alignment process computes for each affix are sums of the acoustic scores of the constituent triphones. In addition to the acoustic scores, the alignment process provides a segmentation of the triphones. However, this automatic segmentation of the triphones might differ considerably from the manual segmentation of the corresponding phonemes. This is because the speech recognizer is forced to align the speech signal with the full sequence of constituent triphones and because the minimum duration of each triphone is 15 ms (with a frame rate of 5 ms and three emitting states per triphone). In the case of very short or deleted phonemes, the recognizer uses parts of the previous or the following phoneme to satisfy the minimum length criterion. This renders the acoustic scores for the individual segments of the affixes potentially meaningless.

B. Duration

For all target words, Pluymaekers *et al.* (2005) measured the duration of the affix and the durations of the individual segments in the affix in milliseconds. They placed the segment boundaries where they found clear formant transitions in the spectrogram supported by visible changes in the waveform pattern.

C. Linguistically motivated control variables

We took over the linguistically motivated control variables investigated by Pluymaekers *et al.* (2005). These include both probabilistic and non-probabilistic variables. The probabilistic variables comprise word frequency; the number of times the target word, or a word from the same inflectional paradigm had occurred earlier in the conversation; the number of times the target affix had occurred earlier in the conversation; mutual information; and word-stem ratio. The non-probabilistic variables include the rate of speech; the gender, age, and regional origin of the speaker; the location of the target word in the utterance (utterance-initial/utterance-final); the presence of disfluencies directly before and after the target word; the segment following the affix (consonant/vowel); the number of consonants in the onset of the stem of the prefixed word (onset complexity); and the absence of segments in the affix. Pluymaekers *et al.* (2005) describe the motivations for using the above-listed control variables, and detail the ways they obtained their values.

V. EXPERIMENT 1

In experiment 1, we investigated whether our spectral reduction scores capture the same information about acoustic reduction as duration. To achieve our goal, we repeated the experiments described by Pluymaekers *et al.* (2005) with the spectral reduction scores as the dependent variable (instead of duration) and using the same linguistically motivated variables as the predictors. We experimented with the spectral reduction scores based on both the manual triphones and the canonical triphones as the dependent variable (*the manual score models* and *the canonical score models*, respectively). If our spectral reduction scores and duration (the duration models, referred to as *Pluymaekers models* in the remainder of this paper) captured essentially the same information about reduction, the models for the different dependent variables should be very similar.

For the results to be comparable across the three models, we first removed the one to three tokens per affix for which we were not able to generate acoustic scores because of their exceptionally short duration. We then determined the outlier tokens for the different models and removed them from all of the models (i.e., the final data sets used for the analyses were the same). Following Pluymaekers *et al.* (2005), we used leverage and Cook's distance values to determine the outliers. The resulting sets of affixes were slightly different from the selection used by Pluymaekers *et al.* (2005). Therefore, in order to allow a fair comparison, we recomputed the models for duration with the same data as used for the spectral reduction scores.[1]

In other words, we fitted three different linear multiple regression models to the data. The Pluymaekers model had affix duration as the response variable, while the manual score model had the spectral reduction scores based on the manual triphones as the response variable, and the canonical score model had the spectral reduction scores based on the canonical triphones as the response variable. Eight data points were removed from each of the models for ge- because they were outliers for the Pluymaekers model, the

TABLE II. The amount of variance explained (R^2) by the Pluymaekers model, the manual score model, and the canonical score model in experiment 1.

Affix	Pluymaekers model	Manual score model	Canonical score model
ge-	0.09	0.04	0.03
ver-	0.10	0.02	0.01
ont-	0.22	0.04	0.04
-lijk/non-final	0.13	0.01	0.01
-lijk/final	0.45	0.02	0.01

manual score model, and/or the canonical score model. For the same reason, seven data points were removed from the models for ver- and ont-. For -lijk, seven data points were removed from the models for words in non-final position, whereas six data points were removed from the models for words in final position. Table II summarizes the results of experiment 1 by presenting the amount of variance explained (R^2) by the three different models fitted for the different affixes. It becomes immediately clear from Table II that the spectral reduction scores cannot properly be predicted by the linguistically motivated variables. This would seem to suggest that the hypothesis of spectral reduction and duration representing the same information about reduction does not hold true. We return to this finding in Sec. VII.

VI. EXPERIMENT 2

Considering the results of experiment 1, experiment 2 was designed to test the hypothesis that reduction is a complex phenomenon of which temporal and spectral reduction measures each deal with different and incomplete aspects. Given this hypothesis, it would be unlikely that these two measures would capture exactly the same aspects of reduction. The second experiment, therefore, aimed to investigate the extent to which the more complex spectral reduction measure can help to *explain* duration as a measure of reduction over and above the contribution of the linguistically motivated variables (cf. Sec. I). Again, we first fitted the statistical models described by Pluymaekers *et al.* (2005) (*the Pluymaekers models*). We then extended the Pluymaekers models with the spectral reduction scores based on both the manual triphones and the canonical triphones as another predictor (*the manual score models* and *the canonical score models*, respectively). For the results to be comparable across the different models, we again excluded the very short affix tokens and determined and excluded the outlier tokens. Because the data set used for experiment 2 was a bit larger than the data set used for experiment 1 (in experiment 1, we had to remove outliers for when duration and the spectral reduction scores were the dependent variables), the results we report with the Pluymaekers models also differ somewhat from the ones reported for experiment 1.

We used least-squares regression for the statistical analyses in this study. The proportion of variance accounted for by a model is expressed by the coefficient R^2. The signs of the reported $\hat{\beta}$ coefficients indicate whether there is a positive or a negative correlation between a predictor (independent) variable and the response (dependent) variable [for a more elaborate explanation of multiple regression models, see Izenman (2008), Chap. 5]. Before embarking on model building, we checked the distributions of the continuous variables (duration and the spectral reduction scores) for deviations of normality that would necessitate some kind of transformation of the data. No such transformation appeared to be necessary.

In other words, we used the duration of the prefix as the response variable and fitted three different linear multiple regression models to the data for each of the prefixes ge-, ver-, and ont-: the Pluymaekers model, the manual score model, and the canonical score model. In the case of the suffix -lijk, we followed Pluymaekers *et al.* (2005) by carrying out the analysis separately for suffix tokens originating from words in non-final and final positions. The number of data points removed as outliers was six for ge-, four for ver-, three for ont-, four for -lijk in the case of words in non-final position (114 observations), and five for -lijk in the case of words in final position (43 observations). Sections VI A–VI D present and discuss our results. To evaluate the significance of our results, we report the outcome of *t*-tests (*t*-statistics) for each response variable. The *p*-value is the probability of obtaining a statistical result (in this case, the result of a *t*-test) at least as extreme as the one that was actually observed, assuming that the null hypothesis (the response variable is *not* significant) is true.

A. ge-

For the Pluymaekers model, we found the following effects: frequency [$\hat{\beta}=-3.5, t(417)=-2.65, p<0.01$], onset complexity [$\hat{\beta}=-6.7, t(417)=-1.88, p<0.1$], and speech rate [$\hat{\beta}=-8.3, t(417)=-5.56, p<0.0001$]. The amount of variance (R^2) explained by this model was 9%. For the manual score model, we found the following effects: frequency [$\hat{\beta}=-4.1, t(416)=-3.28, p<0.01$], onset complexity [$\hat{\beta}=-3.3, t(416)=-0.98, p\approx0.33$], speech rate [$\hat{\beta}=-7.3, t(416)=-5.16, p<0.0001$], and manual score [$\hat{\beta}=3.1, t(416)=7.49, p<0.0001$]. The R^2 of this model was 20%. For the canonical score model, we found the following effects: frequency [$\hat{\beta}=-4.0, t(416)=-3.21, p<0.01$], onset complexity [$\hat{\beta}=-3.5, t(416)=-1.04, p\approx0.30$], speech rate [$\hat{\beta}=-7.6, t(416)=-5.34, p<0.0001$], and canonical score [$\hat{\beta}=3.1, t(416)=7.13, p<0.0001$]. The R^2 of this model was 19%. Words with a higher frequency had shorter realizations of ge-. The prefix was also shorter if the speech rate was high, or if the prefix was followed by a large number of consonants (onset complexity). The prefix was longer if the manual score or the canonical score was high.

Unlike in the Pluymaekers model, onset complexity was not significant as a predictor in the manual score model or in the canonical score model. In the Pluymaekers model, onset complexity was only significant at the 0.1 level, so the additional predictors may actually have turned it insignificant in the manual score model and in the canonical score model. Because the most complex onsets all start with a fricative, it

may also be that onset complexity lost its significance because the spectral reduction scores account for its effect by capturing onset-specific coarticulation.

The observed effects of manual score and canonical score went in the expected direction. The shorter, i.e., the more reduced, the token, the worse one would expect it to match the sequence of models corresponding to the canonical transcriptions and the lower one would expect the score to be.

An analysis of variance showed that both the manual score model [$F(1,416)=56.13, p<0.0001$] and the canonical score model [$F(1,416)=50.85, p<0.0001$] differed from the Pluymaekers model significantly. (The F-statistic used in an analysis of variance is similar to the t-statistic described earlier in this section, and the p-value is interpreted the same way as in the case of t-tests.) There was virtually no difference in the R^2 of the manual score model and the canonical score model.

B. ver-

For the Pluymaekers model, we found the following effects: onset complexity [$\hat{\beta}=-16.8, t(130)=-3.09, p<0.01$] and the year of birth [$\hat{\beta}=-0.5, t(130)=-2.49, p<0.05$]. The R^2 of this model was 12%. For the manual score model, there were significant main effects of onset complexity [$\hat{\beta}=-17.4, t(129)=-3.38, p<0.001$], the year of birth [$\hat{\beta}=-0.5, t(129)=-2.55, p<0.05$], and manual score [$\hat{\beta}=2.3, t(129)=4.08, p<0.0001$]. The R^2 of this model was 22%. For the canonical score model, there were significant main effects of onset complexity [$\hat{\beta}=-17.4, t(129)=-3.34, p<0.01$], the year of birth [$\hat{\beta}=-0.5, t(129)=-2.51, p<0.05$], and canonical score [$\hat{\beta}=2.2, t(129)=3.54, p<0.001$]. The R^2 of this model was 20%. Younger speakers produced shorter prefixes. The prefix was also shorter if the number of consonants in the onset of the stem was high, or if the manual score or the canonical score was low.

An analysis of variance showed that both the manual score model [$F(1,129)=16.62, p<0.0001$] and the canonical score model [$F(1,129)=12.56, p<0.001$] differed from the Pluymaekers model significantly. The manual score model and the canonical score model did not, however, differ from each other much. Unlike in the case of ge-, onset complexity (which was significant at the 0.01 level in the Pluymaekers model) was not overridden by the spectral reduction scores. Apart from the fact that onset complexity was a more robust variable to begin with, it may well be that cross-syllable coarticulation is weaker and less systematic for the closed syllable /v@r/ than for the open syllable /x@/.

C. ont-

For the Pluymaekers model, there were significant main effects of the interaction between frequency and speech rate [$\hat{\beta}=-3.1, t(94)=-3.66, p<0.001$], the interaction between frequency and the year of birth [$\hat{\beta}=0.3, t(94)=3.24, p<0.01$], and the year of birth [$\hat{\beta}=-1.4, t(94)=-5.06, p<0.0001$]. The R^2 of this model was 25%. For the manual score model, there were significant main effects of the interaction between frequency and speech rate [$\hat{\beta}=-2.9, t(93)=-3.38, p<0.01$], the interaction between frequency and the year of birth [$\hat{\beta}=0.3, t(93)=3.03, p<0.01$], the year of birth [$\hat{\beta}=-1.4, t(93)=-4.96, p<0.0001$], and manual score [$\hat{\beta}=1.1, t(93)=1.24, p\approx 0.22$]. The R^2 of this model was 26%. For the canonical score model, there were significant main effects of the interaction between frequency and speech rate [$\hat{\beta}=-3.0, t(93)=-3.43, p<0.001$], the interaction between frequency and the year of birth [$\hat{\beta}=0.3, t(93)=3.06, p<0.01$], the year of birth [$\hat{\beta}=-1.4, t(93)=-4.99, p<0.0001$], and canonical score [$\hat{\beta}=0.8, t(93)=0.98, p\approx 0.33$]. The R^2 of this model was 26%. Younger speakers produced shorter prefixes. The prefix was also shorter if the manual score or the canonical score was low.

An analysis of variance showed that neither the manual score model [$F(1,93)=1.53, p\approx 0.22$] nor the canonical score model [$F(1,93)=0.95, p\approx 0.33$] differed from the Pluymaekers model significantly. The manual score model and the canonical score model did not differ from each other either. It is unclear why spectral reduction was not a significant predictor for /Ont/. It could be that the degree of nasalization in the vowel varies independently from reduction proper. It could also be that the variance induced by uncontrolled factors, such as between-speaker differences, limits the maximum proportion of variance that can be explained with the variables in the model.

D. -lijk

In the case of words in non-final position, there were significant main effects of frequency [$\hat{\beta}=-7.0, t(107)=-3.48, p<0.001$] and the year of birth [$\hat{\beta}=-0.8, t(107)=-3.45, p<0.001$] for the Pluymaekers model. The R^2 of this model was 19%. For the manual score model, there were significant main effects of frequency [$\hat{\beta}=-6.8, t(106)=-3.45, p<0.001$], the year of birth [$\hat{\beta}=-0.8, t(106)=-3.63, p<0.001$], and manual score [$\hat{\beta}=1.9, t(106)=2.20, p<0.05$]. The R^2 of this model was 22%. For the canonical score model, there were significant main effects of frequency [$\hat{\beta}=-6.9, t(106)=-3.46, p<0.001$], the year of birth [$\hat{\beta}=-0.8, t(106)=-3.60, p<0.001$], and canonical score [$\hat{\beta}=1.6, t(106)=1.89, p<0.1$]. The R^2 of this model was 21%. Words with a higher frequency had shorter realizations of -lijk. The prefix was also shorter if the speakers were young, or if the manual score or the canonical score was low.

In the case of words in final position, there were significant main effects of the presence of the plosive [$\hat{\beta}=144.9, t(35)=-3.32, p<0.01$] and speech rate [$\hat{\beta}=-32.8, t(35)=-3.92, p<0.001$] for the Pluymaekers model. The R^2 of this model was 45%. For the manual score model, there were significant main effects of the presence of the plosive [$\hat{\beta}=154.9, t(34)=-3.65, p<0.001$], speech rate [$\hat{\beta}=-29.0, t(34)=-3.48, p<0.01$], and manual score [$\hat{\beta}$

$=6.4, t(34)=1.88, p<0.01]$. The R^2 of this model was 50%. For the canonical score model, there were significant main effects of the presence of the plosive $[\hat{\beta}=157.1, t(34)=-3.69, p<0.001]$, speech rate $[\hat{\beta}=-29.9, t(34)=-3.65, p<0.001]$, and canonical score $[\hat{\beta}=6.5, t(34)=1.89, p<0.1]$. The R^2 of this model was 50%. The prefix was shorter if the speech rate was high, the plosive was absent, or if the manual score or the canonical score was low.

For the words in non-final position, an analysis of variance showed that both the manual score model $[F(1,106)=4.84, p<0.05]$ and the canonical score model $[F(1,106)=3.55, p<0.1]$ differed from the Pluymaekers model significantly. Also for the words in final position, an analysis of variance showed that both the manual score model $[F(1,34)=3.53, p<0.1]$ and the canonical score model $[F(1,34)=3.57, p<0.1]$ differed from the Pluymaekers model significantly. Again, there was virtually no difference between the manual and canonical score models in either case. It is interesting to note that spectral reduction does not subtract from the predictive power of the categorical variable "plosive present." This should not be taken to mean that the absence or presence of /k/ does not affect the spectral reduction scores. Rather, these results are due to the mechanics of the model fit: If two or more predictors explain the same part of the variance, the most powerful variable will take it all—only leaving the residuals for its competitors. Thus, it seems that the categorical absence or presence of /k/ is a stronger predictor of the duration of the suffix than the spectral reduction scores.

VII. GENERAL DISCUSSION

In this study, we investigated the use of log-likelihoods (normalized for duration) from a HMM-based forced alignment procedure as a correlate of acoustic reduction in the speech signal as an alternative for, or as an addition to duration as a correlate of reduction. We referred to these normalized log-likelihood values as spectral reduction scores. The results of our study suggest that the spectral reduction scores capture different aspects of reduction than duration—at least in the sense that the spectral reduction scores cannot be explained by the same linguistically motivated variables as duration. However, they do explain part of the duration variance unaccounted for by the linguistically motivated variables for three of the four Dutch affixes under investigation: ge-, ver-, and -lijk. This is supported by the finding that, for these affixes, the spectral reduction scores only weakly correlate with the durations of the affixes [the correlation between duration and the canonical scores is 0.33 ($R^2=0.11$) for ge-, 0.29 ($R^2=0.08$) for ver-, 0.12 ($R^2=0.01$) for ont-, 0.10 ($R^2=0.01$) for non-final -lijk, and 0.34 ($R^2=0.12$) for final -lijk without any outliers removed]. Except for final -lijk, the increase in the proportion of variance in the multiple regression models explained by the spectral reduction measures is close to the R^2 for the bivariate correlation between spectral reduction and duration. This corroborates the conclusion that our measure of spectral reduction is largely orthogonal to the linguistic measures. At the same time, it is interesting to note that all correlations between spectral reduction and duration predict that shorter tokens correspond with larger spectral reduction. Since our spectral reduction measure is normalized for duration, this suggests that reduction is not limited to time compression, but that there is an additional effect on articulatory simplification.

In our first experiment, we tried to predict the spectral reduction scores of the affixes using the linguistically motivated variables from Pluymaekers *et al.* (2005). None of the "linguistic" models that we fitted explained more than 4% of the variance in the data. Considering the fact that duration *can* (partially) be predicted using the said linguistically motivated variables, and the fact that there is a weak correlation between duration and the spectral reduction scores, this finding is rather interesting. There are at least two potential explanations for it. First, it may be difficult for linguistically motivated variables to predict the spectral reduction scores because the latter are based on a complex measure that combines spectral and time-warp differences in the acoustic space into a single number [as opposed to duration, which is rather a simple, one-dimensional correlate of reduction (see Sec. I)]. Second, the spectral reduction scores are subject to token-by-token variation due to a large number of uncontrolled factors, such as speaker identity and phonetic context from the preceding and following morphemes. This may have added "noise" to the spectral reduction scores. The same holds for duration but the variance contributed by the uncontrolled variables can again be expected to be smaller because of duration being a simpler correlate of reduction. While random variation should not affect the outcome of linear regression models if the number of observations is very high, the number of observations may have been an issue for all models except for ge-, which had more than 420 observations (see Table I). Then again, in the case of ge-, the impact of the first phoneme of the following morpheme may have been particularly strong because the affix ends with a vowel.

As one can see from Eq. (4), the distance between an observed token of an affix and the maximally unreduced pronunciation not only depends on the properties of the token itself, but also on the representation of the unreduced reference. We defined the reference as the sequence of the triphones underlying the canonical phonetic transcription of the affix. We investigated triphones trained with both manual(ly verified) and canonical transcriptions of read speech. The spectral reduction scores obtained using the two sets of triphones were almost identical (the correlation coefficients between the manual and the canonical scores were 0.98 for ge-, ont-, and -lijk, and 0.93 for ver-). However, it must be pointed out that both sets of acoustic models were based on the same type of training data. In other words, the distance from the canonical transcription is not a purely linguistic measure; it is actually the distance from the training data.

Our spectral reduction measure is susceptible to the well-known trajectory folding problem (Han *et al.*, 2007); different tokens taking different trajectories through the acoustic space may end up with identical log-likelihoods, even if their trajectories make very different auditory impressions. This is yet another reason why it may not be appropriate to map multidimensional acoustic reduction to a real

number. While it is difficult to imagine how reduction could be described in terms other than deviation from some reference, it is not obvious that there is one unique reference or one correct way of defining it. In this paper, we used context- and speaker-independent statistical models as the reference. This implies the assumption that all effects of context, speech style, regional background, gender, age, etc., are accounted for by the models. As we have seen, this assumption may not be warranted. Including "context" and "speaker" as random factors in the regression models might be one way around this problem. However, this would require a data set that is orders of magnitude larger than the data set we had available for our research. Similarly, building a mixed model would not be possible with the amount of data that we had.

If we blame the failure to model spectral reduction on the inherent uncontrolled variation in the scores, the question arises what makes duration a measure of reduction that is so much easier to model. We believe that the answer lies in duration being less sensitive to factors such as phonetic context and speaker identity than the trajectories in the spectral space. In addition, while spectral reduction is a result of a trajectory in a multidimensional space, duration is inherently a scalar variable.

In passing, it may be interesting to note that the relation between the "predictability of a linguistic unit" and its duration in a spoken utterance is not as clear-cut as one might think. In a recent study, Kuperman *et al.* (2007) found that infixes in Dutch (/@/, /@n/, or /s/ connecting two nouns that together form a compound) are *longer* if they are more predictable from the nouns that make up the compound. This finding is explained as a tendency to gloss over sounds of which the speaker is not very confident that they should be there.

Both in this study and in the paper of Pluymaekers *et al.* (2005), the proportion of variance in the affix durations that could be explained by the linguistically motivated variables ranged from the low $R^2=0.09$ for ge- to the high $R^2=0.45$ for -lijk in final position; the R^2 values for ver- and -lijk in non-final position were almost as low as the value for ge-, while the value for ont- ($R^2=0.25$) was in the middle. The original paper does not offer an explanation for the wide range of explained variance, and we are not in the position to offer a convincing explanation either. For ge-, ver-, and non-final -lijk—i.e., for the affixes with a low R^2 in the Pluymaekers model—spectral reduction scores raised the proportion of explained variance to about 20%. For ont-, spectral reduction scores were unable to increase to proportion of explained variance much. We speculate this to be due to the effect of the nasal that is likely to cause substantial variance in the spectral reduction measure (over and above the variance introduced by deletions of /t/ and/or /n/).

Because extending the linguistically motivated variables with the spectral reduction scores as predictors increase R^2 for almost all models, one might ask if a similar effect would hold for models that predict spectral reduction scores with the combination of linguistically motivated variables and duration. This appears not to be the case; the explained variance for such models is much lower than the explained variance for models predicting duration with the combination of linguistically motivated variables and spectral reduction scores. Although this may seem surprising, it is an effect that is frequently encountered in regression studies that involve more than two variables (Langford *et al.*, 2001).

In this study, we opted for a measure of spectral reduction that does not rely on the descriptive concepts of acoustic phonetics (e.g., formant frequencies). By doing so, we may seem to ignore previous research on the acoustic reduction in vowels (van Bergem, 1995) and consonants (van Son and Pols, 1999) in Dutch. However, we argue that an approach along the lines of conventional acoustic phonetics is not feasible for capturing spectral reduction in the four affixes under investigation. Three of the affixes have a schwa in their canonical transcription; this raises the question how one could represent vowel reduction in terms of formant frequencies. Furthermore, the formant values of the /O/ in the prefix ont- may be affected both by the final phonemes in the preceding word and by spectral reduction in the affix proper; the potentially disturbing effects of the nasal have already been alluded to. As for consonant reduction, a representation in terms of formant frequencies is inherently questionable; the formant concept only applies with strong restrictions. Moreover, formants in the consonants occurring in spontaneous conversations defy any attempt at automatic measurement. Finally, known reduction measures from acoustic phonetics would only apply to individual phonemes in an affix, leaving us with the problem of incorporating these phoneme-based measures into a measure of acoustic reduction on the affix level.

VIII. CONCLUSIONS

In this study, we proposed a measure of spectral reduction that might either replace or add to duration as a measure of reduction in speech. It appeared that the proposed spectral reduction scores capture other aspects of reduction than duration: While duration can—to a moderate degree—be predicted by a number of linguistically motivated variables, spectral reduction scores cannot. At the same time, spectral reduction scores are able to predict a substantial amount of the variation in duration that the linguistically motivated variables do not account for. We discussed why spectral reduction measures are difficult to express in the form of a scalar. It appears that powerful models of spectral reduction require modeling techniques that can handle factors such as phonetic context, speaker, and speaking style as random variables. Such models will only be feasible when very large corpora are available.

ACKNOWLEDGMENTS

We would like to thank Mark Pluymaekers, Mirjam Ernestus, and Harald Baayen for sharing their data, and Mirjam Ernestus for the discussions we have had about the topic and for her help with the R software. M.G. is funded through the Marie Curie Research Training Network Sound2Sense. L.t.B. is funded through the FP6 FET project ACORNS.

[1]This explains the slight differences between our "Pluymaekers models" and the numbers in the original paper.

Aylett, M., and Turk, A. (**2004**). "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," Lang Speech **47**, 31–56.

Bybee, J. L. (**2001**). *Phonology and Language Use* (Cambridge University Press, Cambridge, UK).

Ernestus, M. (**2000**). *Voice Assimilation and Segment Reduction in Casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface* (LOT, Utrecht, The Netherlands).

Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. H. (**2006**). "Lexical frequency and voice assimilation," J. Acoust. Soc. Am. **120**, 1040–1051.

Greenberg, S. (**1999**). "Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation," Speech Commun. **29**, 159–176.

Hämäläinen, A., Boves, L., de Veth, J., and ten Bosch, L. (**2007**). "On the utility of syllable-based acoustic models for pronunciation variation modelling," EURASIP J. Audio Speech Music Process **2007**, 46460.

Hämäläinen, A., ten Bosch, L., and Boves, L. (**2009**). "Modelling pronunciation variation with single-path and multi-path syllable models: Issues to consider," Speech Commun. **51**, 130–150.

Han, Y., de Veth, J. M., and Boves, L. (**2007**). "Trajectory clustering for solving the trajectory folding problem in automatic speech recognition," IEEE Trans. Audio, Speech, Lang. Process. **15**, 1425–1434.

Izenman, A. J. (**2008**). *Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning* (Springer, New York).

Jespersen, O. (**1922**). *Language: Its Nature, Development and Origin* (George Allen & Unwin Ltd., London, UK).

Johnson, K. (**2004**). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis*, edited by K. Yoneyama and K. Maekawa (The National Institute for Japanese Language, Tokyo, Japan), pp. 29–54.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (John Benjamins, Amsterdam), pp. 229–254.

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2007**). "Morphological predictability and acoustic salience of interfixes in Dutch compounds," J. Acoust. Soc. Am. **121**, 2261–2271.

Langford, E., Schwertman, N., and Owens, M. (**2001**). "Is the property of being positively correlated transitive?," Am. Stat. **55**, 322–325.

Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.

Oostdijk, N., Goedertier, W., Van Eynde, F., Boves, L., Martens, J. P., Moortgat, M., and Baayen, H. (**2002**). "Experiences from the spoken Dutch corpus project," in Proceedings of the LREC '02, Vol. **1**, pp. 340–347.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005**). "Lexical frequency and acoustic reduction in spoken Dutch," J. Acoust. Soc. Am. **118**, 2561–2569.

van Bergem, D. (**1995**). *Acoustic and Lexical Vowel Reduction* (IFOTT, University of Amsterdam, The Netherlands).

van Son, R. J. J. H., and Pols, L. C. W. (**1999**). "An acoustic description of consonant reduction," Speech Commun. **28**, 125–140.

van Son, R. J. J. H., and Pols, L. C. W. (**2003**). "Information structure and efficiency in speech production," in Proceedings of the Eurospeech '03, pp. 769–772.

Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2002**). *The HTK Book (for HTK Version 3.2.1)* (Cambridge University, Cambridge).

Zipf, G. (**1929**). "Relative frequency as a determinant of phonetic change," Harv. Stud. Classic. Philol. **15**, 1–95.

The dynamic range of speech, compression, and its effect on the speech reception threshold in stationary and interrupted noise

Koenraad S. Rhebergen,[a] Niek J. Versfeld, and Wouter. A. Dreschler
Department of Clinical and Experimental Audiology, Academic Medical Center, Amsterdam, Meibergdreff 9, 1105 AZ, The Netherlands

(Received 5 June 2008; revised 27 September 2009; accepted 30 September 2009)

Changes in the speech reception threshold (SRT) after amplitude compression of speech or speech in noise may be due to changes in the dynamic range of the speech signal. However, current models set up to predict the speech intelligibility consider the dynamic range of speech to be fixed regardless of the type of compression. The present paper describes two experiments with normal-hearing subjects to examine the effect of the dynamic range on the SRT in stationary and interrupted noise after wide dynamic range compression. The dynamic range has been varied by compression or expansion of only the speech signal, leaving the masking noise unaltered, or by compression or expansion of the mixed speech-in-noise signal. The results show that compression affects the SRT, both in a positive or a negative direction, not only due to dynamic range but also due to distortion of the speech signal.
 [DOI: 10.1121/1.3257225]

PACS number(s): 43.71.An, 43.66.Ba, 43.71.Gv, 43.72.Kb [MSS] Pages: 3236–3245

I. INTRODUCTION

The articulation index (AI) (ANSI, 1969) and its successor the speech intelligibility index (SII) (ANSI, 1997) are able to predict well the speech intelligibility in normal-hearing listeners in stationary noise. Rhebergen and Versfeld (2005) and Rhebergen *et al.* (2006) extended the SII model to predict the speech reception threshold (SRT) in non-stationary background noises in normal-hearing listeners. The SII model can also be used to some extent to predict the speech intelligibility for hearing-impaired listeners (Rhebergen, 2006) and hearing-impaired listeners aided with simple linear amplification (e.g., Ching *et al.*, 2001; Amlani *et al.*, 2006). So far, however, neither the SII nor other models have been evaluated with respect to the effects of non-linear signal processing on speech intelligibility. In particular, modeling the effects of amplitude compression on speech intelligibility is highly relevant, since most hearing aids make use of this type of signal processing.

Amplitude compression changes the dynamic range (or level distribution) of the speech signal, which may be beneficial to intelligibility, since the weaker parts of the speech signal can be made audible (French and Steinberg, 1947; Villchur, 1989). Moreover, especially in nonstationary background noises, compression may have a substantial positive effect on the signal-to-noise ratio (SNR) (Souza *et al.*, 2006; Rhebergen, 2006; Naylor *et al.* 2006; Rhebergen *et al.*, 2008b). However, at the same time compression causes speech modulations to be less pronounced (van Buuren *et al.*, 1999) and introduces other types of non-linear distortion of the speech signal, which may be detrimental to intelligibility (e.g., Kates and Kozma-Spytek, 1994; Crain and van Tasell, 1994; Stelmachowicz *et al.*, 1999; Versfeld *et al.*, 1999; Boike and Souza, 2000; Jenstad and Souza, 2005).

The first part of the present paper deals with two experiments that have been conducted to examine the effect of the dynamic range of speech on speech intelligibility for conditions with speech in stationary or interrupted noise. In the first experiment, the dynamic range of clean speech is varied utilizing a wide dynamic range compression (WDRC) scheme. By compressing the dynamic range (i.e., level distribution) of the speech signal alone, weaker parts of the speech signal might be easier to detect in background noise or are elevated above the threshold of audibility. Therefore, it is expected that compression on the speech signal might improve the SRT in noise. Speech intelligibility in noise is measured with the SRT test, as described by Plomp and Mimpen (1979). In the second experiment, the SRT is measured for speech in noise, with the difference that, instead of speech only, the dynamic range of the speech-in-noise signal is varied, using the same WDRC scheme. The second part deals with a perceptually relevant method of analyzing the dynamic range of speech, followed by a part that deals with a method to separate speech from noise after a non-linear transformation (such as compression), and a part that deals with the quantification of distortion due to compression and its effect on speech quality and intelligibility. These methods will be used to analyze and describe the results obtained in the two experiments.

II. EXPERIMENT

In Experiment 1, the dynamic range of speech is varied by means of single-channel fast-acting WDRC, where WDRC is applied to the speech signal even before the noise

[a]Author to whom correspondence should be addressed. Electronic mail: k.s.rhebergen@amc.uva.nl

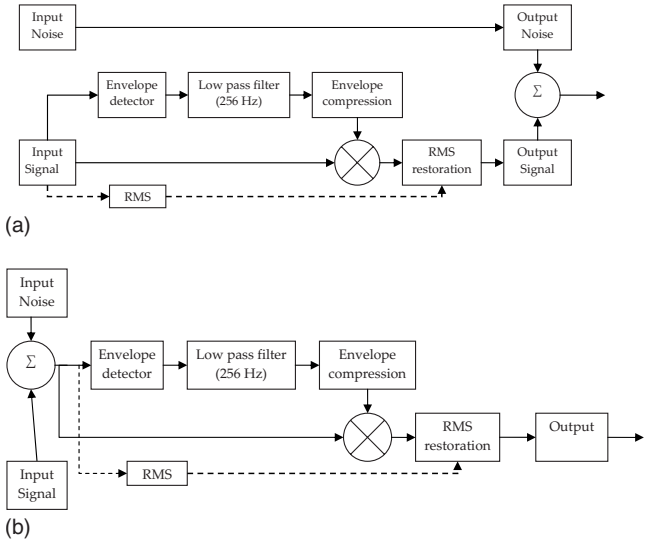


FIG. 1. The upper panel (a) shows a block diagram of WDRC signal processing of Experiment 1. The bottom panel (b) shows a block diagram of WDRC signal processing of Experiment 2. Dashed lines indicate operations on the rms level of the speech signal.

signal is added. In Experiment 2, the same compression scheme is used as in Experiment 1, but now WDRC is applied to the speech-in-noise signal.

A. Subjects

Eight subjects (three males, five females) with normal hearing participated in both experiments. Their age ranged from 24 to 42 years (average 30.9 years). Subjects were native speakers of the Dutch language. They had at least high school education. Individual subjects had pure-tone thresholds of 15 dB HL or better at octave frequencies from 125 to 8000 Hz (ANSI, 2004), measured according to the Hughson–Westlake up-down procedure.

B. Stimuli

The target speech material consisted of short every-day meaningful sentences, uttered by a female speaker (Versfeld *et al.*, 2000). Noise conditions comprised one condition with 8-Hz, 100% modulated interrupted noise with a duty cycle of 50% and one condition with stationary noise. Both masking noises had a spectrum equal to the long-term average spectrum of the female target speech.

1. Signal processing Experiment 1

In Experiment 1, the dynamic range of the target speech was varied by means of fast-acting WDRC with compression ratios (CRs) of 4:1, 2:1, and 1:1, or with an expansion ratio of 1:1.5. With CR=1:1, the speech signal remained unaltered. A block diagram of the WDRC signal processing on the target speech is shown in the top panel of Fig. 1 and is based on the scheme used by van Buuren *et al.* (1999).

The envelope of the speech signal was determined by means of a Hilbert transform. In order to prevent higher frequency envelope modulations from controlling the gain, the envelope was filtered with a 96 dB/oct low pass Bessel filter with a cut-off frequency of 256 Hz. In contrast to van Buuren *et al.* (1999), who used a 32 Hz low pass filter, here a 256 Hz low pass filter was used to obtain a higher effective compression ratio (Apoux *et al.*, 2004). The low pass filter was applied twice: once to the envelope and once again to the filtered time-reversed envelope to remove the phase shifts introduced in the first filtering. The low passed speech envelope was compressed or expanded (compression ratios 4:1, 2:1, and 1:1 or expansion ratio 1:1.5) and next divided, sample by sample, by the original envelope, resulting in a multiplication factor (i.e., gain) for every sample of the speech signal. The multiplication factor was applied to the original speech signal and the long-term rms level of the input signal was restored. Because the aim of Experiment 1 is to examine the effect of the dynamic range of speech on its intelligibility, the WDRC signal processing operates only on the speech signal. The envelope was compressed or expanded down to 60 dB below the rms (knee point). Lower envelope levels were amplified linearly. The noise had a fixed level of 65 dBA *before* the signal processing. Speech and noise were summed *after* the signal processing, but with the same SNR as before signal processing. Furthermore, the overall level of the speech and noise mix after the signal processing was taken equal to the overall level of the speech and noise mix before signal processing.

2. Signal processing Experiment 2

In Experiment 2, the same WDRC signal processing was used as in Experiment 1, except the compression now acted on the speech-in-noise mix (see bottom panel of Fig. 1). The overall level of the speech-in-noise mix after signal processing was taken equal to the overall level of the speech-in-noise mix before signal processing. Note that, due to compression, in both experiments the spectral content of speech and noise may differ. Therefore, the signal-to-noise ratio here is defined as the difference in sound level of speech and noise in rms.

C. Procedure

Subjects were tested individually in a sound-insulated booth. The monaural SRT was measured at the better ear. Signals were played out via an Echo soundcard (Gina 24/96) on a personal computer at a sample frequency of 44.1 kHz. The signals were fed through a TDT Microphone Amplifier (MA2) and a TDT Headphone Buffer (PA4) and were played out via TDH 39P headphones. The SRT was estimated according to the up-down procedure described by Plomp and Mimpen (1979). The SRT, or "critical signal-to-noise ratio," is defined as the SNR at which 50% of presented sentences are repeated entirely correctly. With each sentence presentation, a fixed sample of the interfering masking noise was taken (frozen noise). The masking noise started 1200 ms before the start of the sentence and finished at least 800 ms after the end of the sentence. In total, eight conditions were tested (all conditions with a test and a retest): four conditions with interrupted noise and four with stationary noise, where the four conditions were different with respect to the degree of compression (or expansion) applied to the speech signal. To avoid order effects and to cancel out a possible learning

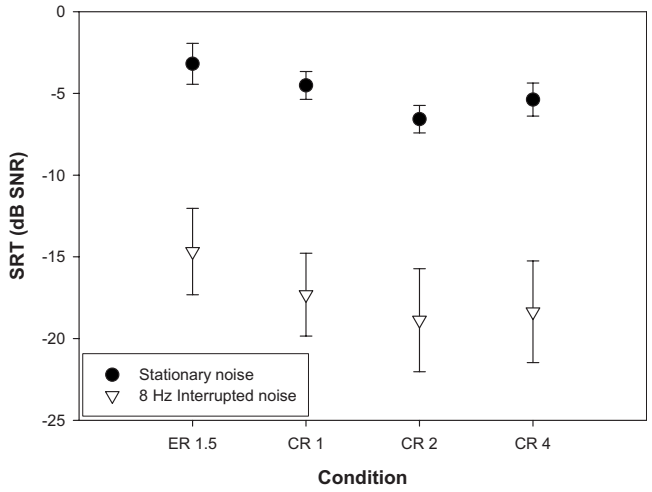


FIG. 2. Speech reception threshold (dB SNR) as function of compression ratio and noise condition. CR 1, CR 2, CR 4, and ER 1.5 denote compression ratios 1, 2, and 4, and expansion ratio 1.5, respectively. Error bars denote the standard deviation between subjects.

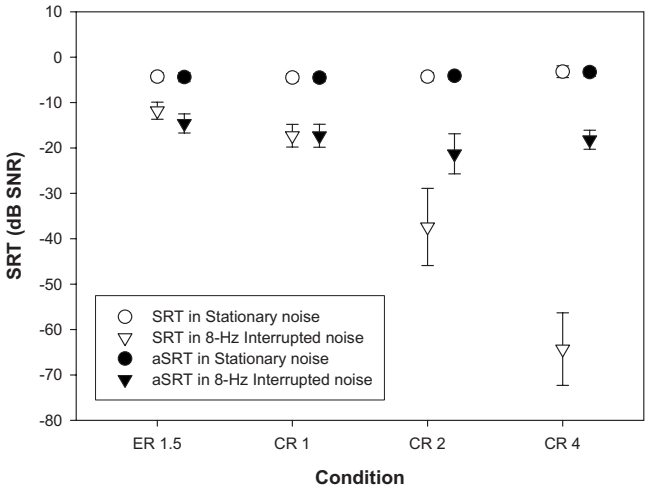


FIG. 3. SRT as a function of compression ratio and noise condition. Open symbols denote the SRT (dB SNR) of the input SNR; filled symbols that of the "apparent" SRT, aSRT (dB SNR). CR 1, CR 2, CR 4, and ER 1.5 denote compression ratios 1, 2, and 4, and expansion ratio 1.5, respectively. Bars denote the standard deviation between subjects.

effect, the order of conditions and sentence lists was counterbalanced across subjects according to a Latin square method. To further account for learning effects (Rhebergen *et al.*, 2008a), the experiment was preceded by six practice lists, to practice the SRT in 8-Hz interrupted noise and stationary noise. In total, each subject received 28 lists; 16 lists of 13 sentences in Experiment 1 (2 noise conditions by 4 compression conditions by 2 SRT measurements) and 12 lists of 13 sentences in Experiment 2 (2 noise conditions by 3 compression conditions by 2 SRT measurements). The reference condition CR=1 in Experiment 2 was omitted because the condition has already been measured in Experiment 1.

D. Results

1. Results of Experiment 1

Figure 2 displays the SRT-values obtained in Experiment 1, averaged across subjects as function of compression ratio in stationary noise or in 8-Hz interrupted noise. A 2[noise type]×4[compression condition]×2[test/retest] analysis of variance (ANOVA) was performed on the data set. Of the main effects, "noise type" ($F[1,112]=1025.8$, $p<0.0001$) and "compression condition" ($F[3,112]=17.7$, $p<0.0001$) were significant. Differences between "test" and "retest" were not significant ($F[1,112]=1.8$, $p>0.05$). There were no significant interactions. On average, SRTs in interrupted noise were much lower (i.e., better) than in stationary noise. *Post hoc* tests (Tukey HSD) showed that the compression condition with ER=1.5 significantly differed from the compression conditions with CR=1, CR=2, and CR=4. Furthermore, compression conditions with CR=1 and CR=2 differ also significantly from each other.

2. Results of Experiment 2

Figure 3 displays the SRT-values obtained in Experiment 2 averaged across subjects as a function of compression ratio in 8-Hz interrupted noise or in stationary noise (open symbols, filled symbols will be discussed below). A 2[noise type]×4[compression condition]×2[test/retest] ANOVA was performed on the data set. Of the main effects, noise type ($F[1,112]=1373.3$, $p<0.001$) and compression condition ($F[3,112]=226$, $p<0.001$) were significant. Differences between test and retest were not significant ($F[1,112]=1.3$, $p>0.05$). SRTs in interrupted noise were much lower (i.e., better) than in stationary noise.

Also, separate ANOVAs were performed on the data obtained with stationary noise and with interrupted noise. For the stationary noise conditions, the main effect of compression condition was significant ($F[3,56]=5.1$, $p<0.005$). *Post hoc* tests (Tukey HSD) showed that the compression condition with CR=4 significantly differed from all other compression conditions. For the interrupted noise conditions, the main effects of compression condition was significant ($F[3,56]=242.8$, $p<0.001$). *Post hoc* tests (Tukey HSD) showed that the compression conditions with CR=1, CR=2, and CR=4 differed significantly, and the compression condition with ER=1.5 differed significantly from the compression conditions with CR=2 and CR=4.

III. SIGNAL ANALYSIS

In the current section, three methods of signal analysis are described. They will be used to analyze the results described in Sec. II. The first part deals with a perceptually relevant method of analyzing the dynamic range of speech. The second part deals with a method to separate speech from noise after a non-linear transformation, and the third part deals with the quantification of distortion due to compression and its effect on speech quality and intelligibility.

A. The dynamic range of speech

In current intelligibility models, such as the SII, the dynamic range of speech is kept fixed to 30 dB, regardless of the type of speech signal or signal processing (e.g., Cox *et al.*, 1988; Rankovic, 1997, 1998; Boothroyd, 1990, 2000; van Tasell, 1993; Drullman, 1995a, 1995b; Studebaker *et al.*, 1999; Studebaker and Sherbecoe, 2002; Zeng *et al.*, 2002;

Molis and Summers, 2003; Rhebergen, 2006; Moore *et al.*, 2008). Also, the peak-to-rms ratio of speech is kept fixed, although it has been changed from 12 dB (ANSI, 1969) to 15 dB (ANSI, 1997). This change was not based on statistical analyses of the speech signal, but rather on a best fit of the SII and speech transmission index (STI) (Steeneken and Houtgast, 1980) model to the Studebaker *et al.* (1993) data (Studebaker and Sherbecoe, 2002). Yet, in several studies it was pointed out that the peak-to-rms ratio of speech was on average about 12 dB (e.g., French and Steinberg, 1947; Kryter, 1962; Studebaker *et al.*, 1993; Pavlovic and Studebaker, 1984; Phatak and Allen, 2007). Similarly, the 30 dB dynamic range of speech has its origins in the work of Beranek (1947). He proposed the 30 dB dynamic range on his estimation of the short-term (125 ms) speech spectrum measurements reported by Dunn and White (1940) (Beranek, 1988). Recently, Lobdell and Allen (2007) showed that 99% of speech levels fall over a range of 40 dB using either a software Volume-Unit (VU) meter or levels in 125 ms intervals, which is broader than the 30 dB used in the present speech intelligibility models. In retrospect, the differences in calculated speech dynamic range of the Lobdell and Allen study compared to Dunn and White (1940) most likely is due to the limitations of the recording equipment of Dunn and White (1940) (Moore *et al.*, 2008): their obtained dynamic range was not the difference between the 1% (L1) and 99% (L99) level distributions of speech (see Byrne *et al.*, 1994), but rather between 1% and approximately 80% (French and Steinberg, 1947; Lobdell and Allen, 2007). Additional support for the larger dynamic range comes from behavioral measurements which indicate that the effective dynamic range of speech is 40 or even 50 dB, depending on the measurement method or speech material used (e.g., Drullman, 1995a, 1995b; Studebaker *et al.*, 1999; Studebaker and Sherbecoe, 2002; Hilkhuysen, 2007; Boothroyd, 2008).

Usually, the analysis of the dynamic range of speech is carried out while using the "fast" integration time (125 ms) on a sound level meter. However, normal-hearing listeners can detect gaps of 22 ms at the low frequencies up to about 2 ms at the high frequencies (e.g., Shailer and Moore, 1983). If one considers that the behavioral dynamic range is about 45 dB (e.g., Studebaker and Sherbecoe, 2002), it is clear that the fast integration time constant is too large. The effect of integration time on the dynamic range of speech is shown in Fig. 4: A 30-s sample of running speech (female speech of Versfeld *et al.*, 2000, as used in the present experiments) without any pauses was analyzed. Figure 4 shows the dynamic range (defined as the level that is exceeded in 1% of the time, L1, minus the level that is exceeded in 99% of the time, L99). Instead of analyzing the broadband signal, speech was first filtered in 29 frequency bands [29 gammatone equivalent rectangular bandwidth (ERB) filters, with center frequencies of 163–7793 Hz; ERBs 5–33; Glasberg and Moore, 1990], mimicking the human frequency resolving power. The output of each individual filter went through a sliding rectangular temporal window with an integration time ranging from 1 ms to 8192 ms. Figure 4 shows that the calculated dynamic range ranges from 59.1 dB for a window of 1 ms to 4.9 dB at 8192 ms. Bars denote the standard deviation between the dynamic range of the 29 ERB filters. The dynamic range thus is highly dependent on the integration time. Because human temporal disciminability increases with increasing frequency, it seems to be perceptually more relevant to use frequency-dependent integration times, in order to calculate the effective dynamic range of speech.

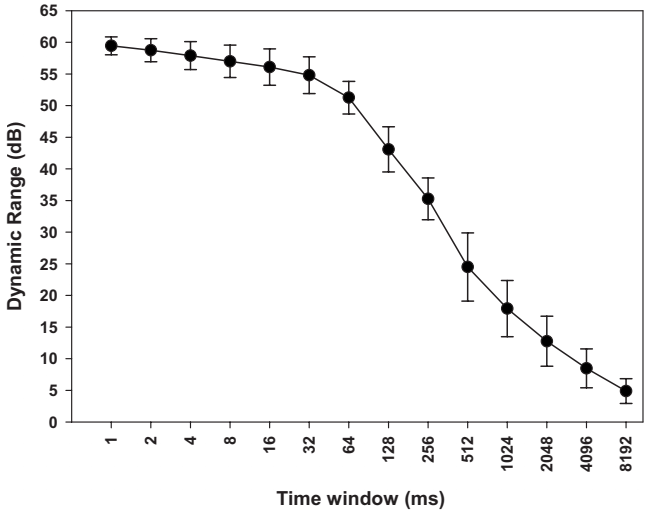


FIG. 4. The dynamic range of speech averaged across 29 ERB filters (163–7793 Hz; ERBs 5–33; Glasberg and Moore, 1990), as obtained with a sliding rectangular temporal window with an integration time ranging from 1 to 8192 ms. The bars indicate the standard deviation between ERBs.

van Schijndel *et al.* (1999) used wavelets (Gabor, 1947) to examine the shape of the auditory frequency-time window. For a detailed explanation to obtain the optimal time-frequency resolution, see van Schijndel *et al.*, 1999. In short, the effective time window can be calculated as follows: The effective duration of the time window (Δt) at a specific center frequency (CF) is $\Delta t = 1/(\alpha \mathrm{CF})$. The shape factor ($\alpha$) is the effective bandwidth divided by CF ($\alpha = \Delta f/\mathrm{CF}$). This equation is used to determine the effective duration associated with each of the 29 ERB filter. It ranges from about 24 ms (lowest band) to about 1 ms (highest band). These values were taken to determine the level distribution and dynamic range of the speech material under investigation. Figure 5(a) shows, for the four conditions used in the listening Experiment 1, the level distribution (averaged across 29 ERB bands) relative to the rms, obtained with a rectangular frequency-dependent temporal window. Figure 5(b) shows the cumulative level distribution under the same conditions.

B. Separating speech from noise after non-linear transformation

Signal processing, either linear or non-linear, may cause the SNR of the output signal to be different from that of the input signal. A change in the SNR may affect intelligibility (although other factors also may play a role). It thus is important to know the SNR of the output signal, the so-called "apparent" SNR (aSNR) (Hagerman and Olofsson, 2004). With linear systems, it is relatively easy to calculate the aSNR, since speech and noise do not interact. Compression,

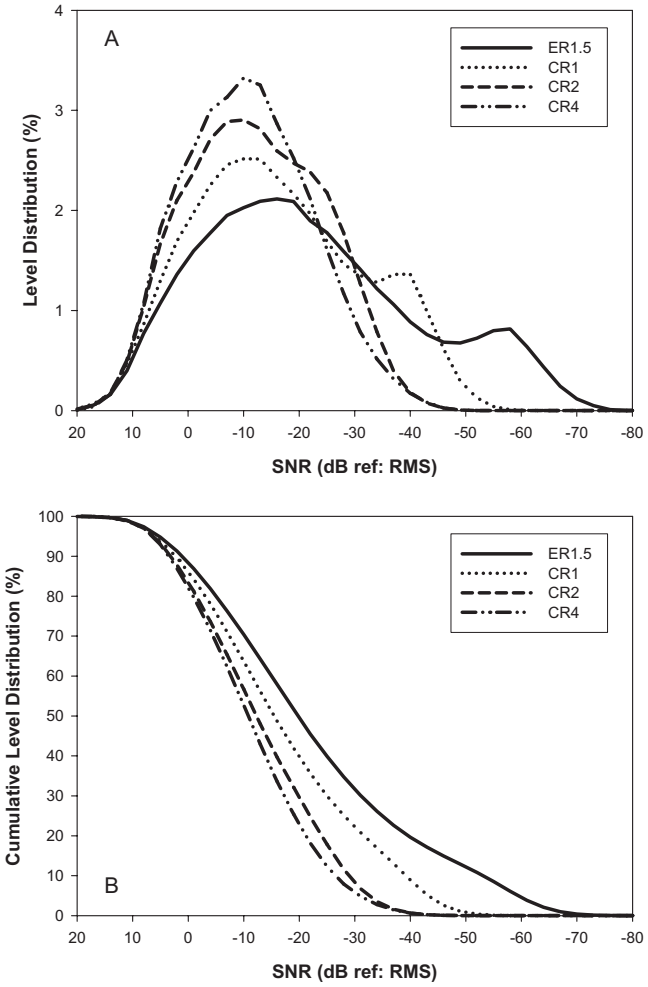


FIG. 5. The average level distribution (a) and cumulative level distribution (b) of female speech for the four compression ratios in Experiment 1. CR 1, CR 2, CR 4, and ER 1.5 denote compression ratios 1, 2, and 4, and expansion ratio 1.5, respectively.

or any other non-linear signal processing for that matter, does cause speech and noise to interact, so that it is difficult to separate speech from noise at the output.

Hagerman and Olofsson (2004) introduced a technique to calculate the aSNR, which they used to study the effect of noise reduction systems. The basic idea is that a system is fed with a speech in noise mixture and with a similar mixture where the speech or the noise signal is phase inverted. Summation or subtraction of the output signals then eliminates the speech or the noise portion, respectively, thus separating the output into a part due to speech and a part due to noise. With this so-called inversion technique, Souza *et al.* (2006) and Rhebergen (2006) showed that the aSNR for speech in stationary noise is changed after WDRC. While this approach is only a linear approximation and compression causes (non-linear) distortion terms, the technique is only allowed as long as these distortion products are relatively small. However, Rhebergen (2006) and Rhebergen *et al.* (2008b) showed that the inversion technique of Hagerman and Olofsson (2004), due to phase shifts, does not always produce valid results. As an alternative to the inversion technique in the case of WDRC, Rhebergen *et al.* (2008b) calculated the aSNR with a method where the speech signal and the noise signal are separately compressed with the compression gain factor for every sample of the speech-in-noise mix. The method is as follows: (1) the speech and noise are mixed at a given SNR, (2) the speech-in-noise mix is fed to a compressor, (3) the compression gain factor for every sample of the speech-in-noise mix is stored, (4) the compression gain factor is applied on the speech signal and noise signal separately with their given input levels, and (5) the difference in level after compression is the aSNR. Rhebergen *et al.* (2008b) showed with aid of the quality model of Arehart *et al.* (2007) that the reconstructed output speech signal was less distorted compared to that obtained with the inversion technique. Below, the alternative method of Rhebergen *et al.* (2008b) is used in the analysis of the present data.

C. Effects of distortion

In addition to background noise, speech intelligibility can be affected by distortion due to non-linear signal processing. Different researchers have studied the relationship between the physical measured non-linear distortion due to peak clipping, and the perceived quality of the speech signal (e.g., Kates, 1992; Crain and van Tasell, 1994; Stelmachowicz *et al.*, 1999; Versfeld *et al.*, 1999; Kates and Arehart; 2005; Jenstad and Souza, 2005; Olofsson and Hansen, 2006; Stone and Moore, 2007). In most studies, the coherence and signal-to-distortion-ratio (SDR) described by Kates (1992) or a modification was used to quantify the effects of distortion on the speech signal. Recently, Kates and Arehart (2005, 2007), Anderson *et al.* (2006), and Arehart *et al.* (2007) proposed a modification to this measure, extending its range of application. Now, it is able to predict the quality rating for peak clipping, center clipping, WDRC, and spectral subtraction with sufficient accuracy (r=0.98, Kates and Arehart, 2007). To analyze the results of the present paper, the Q3 calculation scheme of Arehart *et al.* (2007) was used. The Q3 is a value between 0 and 1 and represents a subjective quality rating of 0 (poor) to 10 (good), respectively. Simply put, Q3 is related to the coherence between the input and the output signal, where the coherence is a measure that indicates the proportion of the output signal that is linearly related to the input signal. The lower the coherence, the lower the SDR, and finally the lower the quality, hence Q3. For a detailed description of this measure, see Arehart *et al.*, 2007. Because there is a relation between the quality of a speech signal and the observed speech intelligibility (e.g., Crain and van Tasell, 1994; Preminger and van Tasell, 1995; Eisenberg *et al.*, 1998), one might expect worse SRTs with lower Q3 values.

IV. DISCUSSION

A. Experiment 1

The results of Experiment 1 showed that best SRTs were obtained for CR=2, resulting in a SRT being about 2.0 dB better than the uncompressed signal (CR=1). The outcomes of the present study are comparable with those from other studies with WDRC speech (Dirks *et al.*, 1986; Boothroyd *et al.*, 1988; Dubno and Dirks, 1989; Kamm *et al.*, 1985). Due to the compressed dynamic range of the speech signal alone, weaker parts (notably consonants) of the speech signal are

easier detectable in background noise (stationary noise) or are elevated above the threshold of audibility (fluctuating noise). The increase in SRT when the compression ratio is further increased to CR=4 probably is due increased distortion of the speech signal. The calculated Q3 values for the WDRC speech of Experiment 1 are 0.91 for ER 1.5, and 1.0, 0.89, and 0.68, for CR equal to 1, 2, and 4, respectively. The highest Q3 value is obtained for the unprocessed speech signal (CR=1) and the lowest for CR=4. Apparently, there is a trade-off between effective audibility and distortion for compressed speech. With mild compression, the increase in audibility has a stronger effect on intelligibility than distortion. With increased compression ratios, distortion dominates.

Speech with an expanded dynamic range (expansion ratio 1.5:1) yields poorer SRTs compared with uncompressed speech, both in stationary and interrupted noise. Contrary to compression, expanding the dynamic range of speech, together with increased distortion of the speech signal, has a clearly negative effect on the speech intelligibility. These outcomes are in line with Clarkson and Bahgat, 1991; Freyman and Nerbonne, 1996; van Buuren *et al.*, 1999; Apoux *et al.*, 2004. Expansion causes the weaker parts of the speech signal to become inaudible. The ratio between the vowels and consonants increases and thus important (consonant) information disappears. Indeed, subjects reported that although clearly audible the speech signal was often unintelligible.

B. Experiment 2

The present experiment shows that, given the same masking condition, the four different compression ratios result in considerably different SRTs. On average, SRTs in interrupted noise are much better than in stationary noise. The exact size of the effects is complicated by the fact that the SRTs displayed in Fig. 3 (open symbols) are expressed in SNRs before compression was applied to the speech-in-noise. It is clear that, in the case of speech in interrupted noise, the SNR at the output is different that at the input, since the level of the speech signal is greatly elevated in the silent periods. For all conditions of Experiment 2, the aSNR was determined, using the method of Rhebergen *et al.* (2008b), as described in Sec. IV A. Figure 3 displays, as filled symbols, the aSRT (i.e., the aSNR at threshold) averaged across subjects as a function of compression ratio in 8-Hz interrupted noise or in stationary noise. The lowest aSRT is obtained for CR=2. As expected, at CR=1 the aSRT is identical to the SRT, but for ER=1.5, the aSRT is lower than the SRT. Compression apparently increases the SNR; expansion, on the other hand, decreases the SNR.

The aSRT shows less variability than the SRT before compression (i.e., the standard deviation between subjects is smaller, see Fig. 3). A 2[noise type]×4[compression condition]×2[test/retest] ANOVA was performed on the aSRT. Of the main effects, noise type ($F[1,112]=1217.3$, $p<0.001$) and compression condition ($F[3,112]=22.3$, $p<0.001$) were significant. Differences between test and retest were not significant ($F[1,112]=1.1$, $p>0.05$). An interaction effect was found between noise type and compression condition ($F[3,1]=14.4$, $p<0.001$). Also, separate ANOVAs were performed on the aSRT data obtained with stationary noise and with interrupted noise. For the stationary noise conditions, only the main effect of compression condition was significant ($F[3,56]=3.3$, $p<0.05$). *Post hoc* tests (Tukey HSD) showed that the condition with CR=2 significantly differed from the condition with ER=1.5. For the interrupted noise conditions, the main effect of compression condition was significant ($F[3,56]=21.1$, $p<0.001$). *Post hoc* tests (Tukey HSD) showed that only the conditions with CR=1 and CR=4 did not differ significantly from each other.

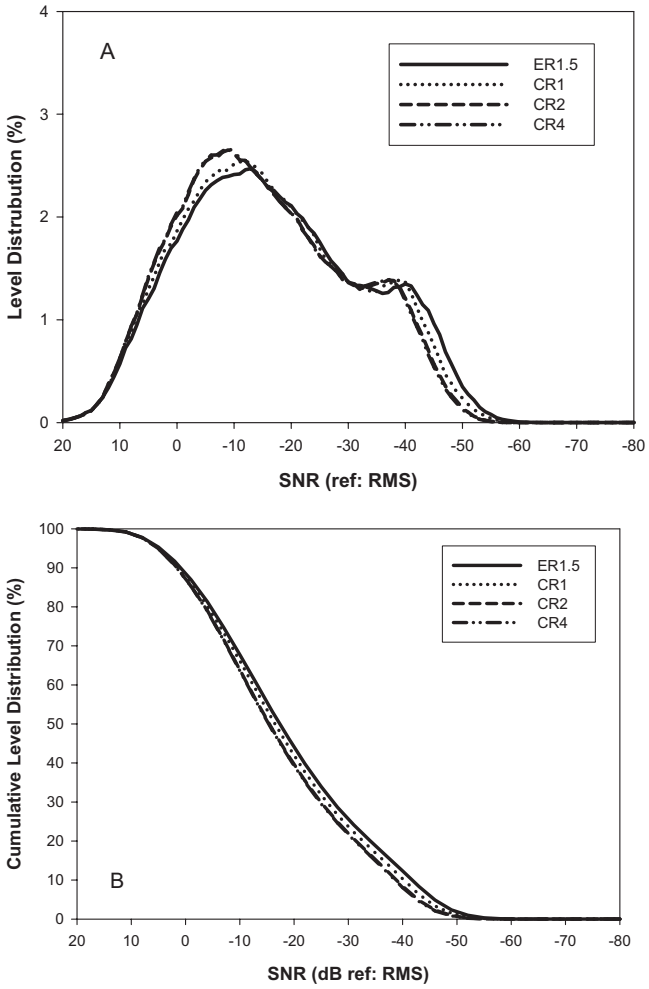


FIG. 6. The average level distribution (a) and cumulative level distribution (b) of female speech for the four compression ratios in Experiment 2. CR 1, CR 2, CR 4, and ER 1.5 denote compression ratios 1, 2, and 4, and expansion ratio 1.5, respectively.

The marginal effect of compression for compressed speech in stationary noise is due to the fact that at threshold, i.e., the SRT, the compressor acts almost exclusively on the stationary noise. Indeed, the level distribution [Fig. 6(a)] and the cumulative level distribution [Fig. 6(b)], calculated as described in Sec. IV A, resemble one another very much.

With interrupted noise, SRTs are such that speech is completely inaudible in presence of the noise bursts. At each silent period, speech is compressed in a manner similar to that in Experiment 1. The Q3 values for these conditions are 0.85, 1.0, 0.83, and 0.63 for ER=1.5, CR=1, 2, and 4, respectively. The highest Q3 value is obtained for the unproc-

essed speech signal (CR=1) and the lowest for CR=4. Again, the decrease in intelligibility at CR=4 is due to the decrease in Q3.

V. GENERAL DISCUSSION

A. Effect of the dynamic range of speech on the Speech Reception Threshold

Within limits, the SRT in stationary noise decreases when the speech dynamic range is compressed, and increases whenever the dynamic range is expanded (*re* unprocessed speech). This also holds for the SRT in interrupted noise. Apparently, audibility of the weaker parts of speech is important to intelligibility (the energy ratio between the vowels and in the consonants decreases with compression). The degree to which the weaker parts contribute to intelligibility is given by the transfer function, the relation between the proportion of audible speech and percent correct score. This transfer function (e.g., Sherbecoe and Studebaker, 1990) is likely to be dependent on the type of speech material (e.g., words and sentences) or test method (e.g., up-down, word score, and sentence score), but should according to articulation theory (e.g., French and Steinberg, 1947; Fletcher and Galt, 1950; Kryter, 1962; Pavlovic and Studebaker, 1984; Pavlovic, 1987; Sherbecoe and Studebaker, 2003) be independent of the presence or absence of any type of background noise. Hence, if the effective dynamic range is changed, the intelligibility score will change as well at a given SNR. This holds as long as the speech signal is not distorted too much.

B. Effect of compression on the dynamic range of speech in noise

The dynamic range of speech in a compressed speech-in-noise signal can be determined by means of the method introduced by Rhebergen *et al.* (2008b). At large SNRs, the compressor acts merely on the speech signal, hence reducing the dynamic range of speech. At large negative SNRs, the compressor acts on the noise, leaving speech relatively unchanged. This is illustrated in Fig. 7, which shows the level distribution (LD) of a speech signal in stationary noise after single-channel WDRC with compression ratio 2. The figure shows that the shape of the LDs is dependent on the SNR. At lower SNRs (−15 and −5 dB), the LDs resemble the shape of the unprocessed speech (CR=1). At higher SNRs, the shape of the LDs is different compared to CR=1. Furthermore, WDRC has an effect on the input SNR. Figure 8(a) shows the aSNR as a function on input SNR after single-channel WDRC with compression ratios 1, 2, and 4 and expansion ratio 1.5. The change in aSNR depends on the input SNR, the noise condition, and compression ratio. This is in line with the results of Souza *et al.* (2006), Rhebergen (2006), Naylor *et al.* (2006, 2008), and Rhebergen *et al.* (2008b) who also used the Hagerman and Olofsson (2004) method. Figure 8(b) shows the aSNR as a function on input SNR in 8-Hz interrupted noise. The change in SNR after WDRC in interrupted noise is even larger compared to stationary noise conditions. Figure 8(c) shows the difference in aSNR (ΔaSNR) between interrupted and stationary noise as

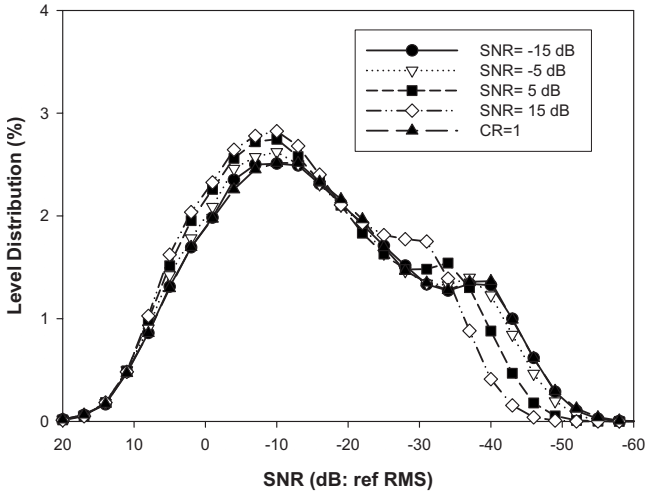


FIG. 7. The average level distribution of female speech for CR=2 in stationary noise at a SNR of −15, −5, +5, and +15 dB. CR=1 denotes the average level distribution for the unprocessed speech.

a function of input SNR and compression ratio. The figure shows that the highest ΔaSNR is measured with compression ratio 4 and the lowest with ER=1.5. The benefit (*re* CR=1) of WDRC in interrupted noise compared to stationary noise is shown in Fig. 8(d). The SNR in interrupted noise with CR=2 and CR=4 improves (higher SNRs *re* CR=1) at low input SNRs and is more or less the similar at higher SNRs. The above shows that it is important to calculate aSNR after a non-linear transformation, such as compression, in order to interpret the results of the speech intelligibility measurements. The Hagerman and Olofsson (2004) method should be used with care. A small spectral mismatch between the speech and noise is enough to bias the results (Naylor *et al.* 2006). Furthermore, phase shifts in the processed speech-in-noise signal due to the non-linear signal processing are detrimental to the inversion technique (Rhebergen *et al.*, 2008b). So far, the method proposed by Rhebergen *et al.* (2008b) can only be used off-line with a known digital WDRC scheme. For analysis of the effects of an (unknown) WDRC system, the Hagerman technique is (to our knowledge) the only appropriate alternative.

C. Implications for modeling the intelligibility of non-linear processed speech

As mentioned in the Introduction, non-linear signal processing changes the input signal in both the time and frequency domain. Furthermore, depending on the type of background noise and SNR, the dynamic range of the speech signal is changed after WDRC. Audibility models such as the AI, SII, and STI are unable to account for all these effects. The aided articulation index (AAI) proposed by Stelmachowicz *et al.* (1994) is a first order approach that accounts for the change in dynamic range of clean speech after WDRC. The intensity importance function (IIF) of the AAI is obtained by dividing the IIF of the SII model by the effective compression ratio (ECR) of the compressed speech signal (i.e., 15 dB speech peaks/ECR; 30 dB dynamic range/ECR). As a result, the AAI IIF ranges from −15 dB/ECR to +15 dB/ECR *re* the rms of speech. However, calculations

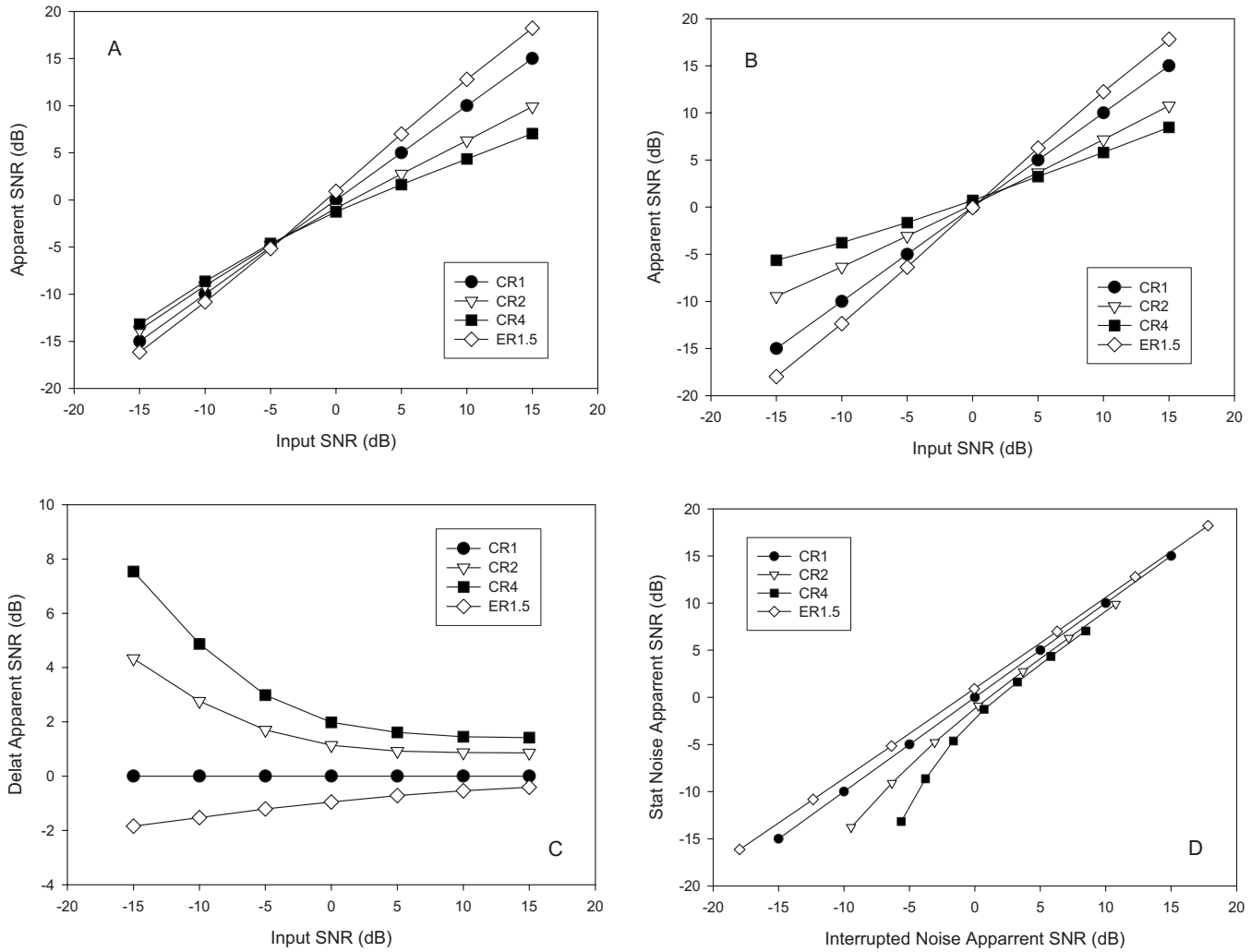


FIG. 8. (a) aSNR (dB) as a function of input SNR (dB) for compressed speech in stationary noise; (b) aSNR (dB) as a function of input SNR (dB) for compressed speech in 8-Hz interrupted noise; (c) ΔaSNR (dB) as a function of input SNR (dB) for compressed speech in stationary noise; (d) aSNR (dB) of compressed speech in 8-Hz interrupted noise as a function of aSNR (dB) of compressed speech in stationary noise. CR 1, CR 2, CR 4, and ER 1.5 denote compression ratios 1, 2, and 4, and expansion ratio 1.5.

on the level distribution after WDRC in Experiments 1 and 2 suggest that a cumulative normal distribution related speech dynamic range might be more realistic. This idea was already suggested by French and Steinberg (1947) (p. 114), but they did not use this approach due to its laborious procedure. Rhebergen (2006) and Rhebergen *et al.* (2008b) attempted to include such a function into the SII model, but due to complicated interactions between dynamic range, change in SNR, and compression induced distortion, they had difficulties in testing the surplus value of this modification of the SII model.

VI. SUMMARY AND CONCLUSIONS

The present paper describes two SRT experiments with normal-hearing subjects to examine the following.

(a) The effect of the dynamic range of the speech signal on SRT in stationary and interrupted noise (Experiment 1). Expansion of the speech signal has a negative effect on the SRT in stationary and interrupted noise. Mild compression of the speech signal has a positive effect on the SRT in stationary and interrupted noise.

(b) The effect of a compressed mixed signal of speech and noise on SRT in stationary and interrupted noise (Experiment 2). In stationary noise, expansion has a negative effect on the SRT. The SRT for the compressed speech-in-stationary noise mix does not change much. However, the SRT does improve significantly for the compressed speech-in-interrupted noise mix.

ACKNOWLEDGMENTS

We are grateful to Professor Dr. Ir. T. Houtgast for many useful discussions and for his comments on earlier versions of this paper. The editor Professor Dr. Mitch Sommers is acknowledged for this detailed and constructive comments. We thank Professor Dr. G. A. Studebaker and Professor Dr. A. Boothroyd for many useful information related to the dynamic range of speech. László Körössy is acknowledged for his help with the SRT computer program and Rolph Houben for his technical support.

Amlani, A. M., Punch, J. L., and Ching, T. Y. C. (**2002**). "Methods and applications of the audibility index in hearing aid selection and fitting," Trends Amplif. **6**, 81–129.

Anderson, M. C., Arehart, A. H., Kate, J. M., and Harvey, L. O. (**2006**). "Predictions of speech quality under conditions of noise and distortion," poster presentation at the International Hearing Aid Research Conference, Lake Tahoe, CA.

ANSI (**1969**). "Methods for the calculation of the articulation index" (American National Standards Institute, New York), ANSI S3.5-1969.

ANSI (**1997**). "American National Standard methods for calculation of the speech intelligibility index" (American National Standards Institute, New York), ANSI S3.5-1997.

ANSI (**2004**). "American National Standard methods for specification for audiometers (American National Standards Institute, New York), ANSI S3.6-2004.

Apoux, F., Tribut, N., Debruille, X., and Lorenzi, C. (**2004**). "Identification of envelope-expanded sentences in normal-hearing and hearing-impaired listeners," Hear. Res. **189**, 13–24.

Arehart, K. H., Kates, J. M., Anderson, M. C., and Harvey, L. O., Jr. (**2007**). "Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 1150–1164.

Beranek, L. L. (**1947**). "The design of speech communication systems," Proc. IRE **35**, 880–890.

Beranek, L. L. (**1988**). *Acoustic Measurements*, revised ed. (American Institute of Physics, New York).

Boike, K. T., and Souza, P. E. (**2000**). "Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification," J. Speech Lang. Hear. Res. **43**, 456–468.

Boothroyd, A. (**1990**). "Articulation index: Importance function in the intensity domain," J. Acoust. Soc. Am. **88**, S31.

Boothroyd, A. (**2000**). "Thar's gold in them thar hills: Mining the P/I function Carhart Memorial," lecture delivered to the Annual Convention of the American Auditory Society, Scottsdale, AZ.

Boothroyd, A. (**2008**). "The performance/intensity function: An underused resource," Ear Hear. **29**, 479–491.

Boothroyd, A., Springer, N., Smith, L., and Schulman, J. (**1988**). "Amplitude compression and profound hearing loss," J. Speech Lang. Hear. Res. **31**, 362–376.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Heto, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westermann, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Ching, T. Y. C., Dillon, H., Katsch, R., and Byrne, D. (**2001**). "Maximising effective audibility in hearing aid fitting," Ear Hear. **22**, 212–224.

Clarkson, P. M., and Bahgat, S. F. (**1991**). "Envelope expansion methods for speech enhancement," J. Acoust. Soc. Am. **89**, 1378–1382.

Cox, R. M., Matesich, J. S., and Moore, J. N. (**1988**). "Distributions of short-term rms levels in conversational speech," J. Acoust. Soc. Am. **84**, 1100–1104.

Crain, T. R., and van Tasell, D. J. (**1994**). "Effect of peak clipping on speech recognition threshold," Ear Hear. **15**, 443–453.

Dirks, D. D., Bell, T. S., Rossman, R. N., and Kincaid, G. E. (**1986**). "Articulation index predictions of contextually dependent words," J. Acoust. Soc. Am. **80**, 82–92.

Drullman, R. (**1995a**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

Drullman, R. (**1995b**). "Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level," J. Acoust. Soc. Am. **98**, 1796–1798.

Dubno, J. R., and Dirks, D. D. (**1989**). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1666–1675.

Dunn, H. K., and White, S. D. (**1940**). "Statistical measurements on conversational speech," J. Acoust. Soc. Am. **11**, 278–288.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., and Martinez, A. S. (**1998**). "Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," J. Speech Lang. Hear. Res. **41**, 327–339.

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.

Freyman, R. L., and Nerbonne, G. P. (**1996**). "Consonant confusions in amplitude-expanded speech," J. Speech Hear. Res. **39**, 1124–1137.

Gabor, D. (**1947**). "Acoustical quanta and the theory of hearing," Nature (London) **159**, 591–594.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notch-noise data," Hear. Res. **47**, 103–138.

Hagerman, B., and Olofsson, A. (**2004**). "A method to measure the effects of noise reduction algorithms using simultaneous speech and noise," Acta. Acust. Acust. **90**, 356–361.

Hilkhuysen, G. L. M. (**2007**). "Intelligibility trade-off between multi-band peak-limiting and masking by noise," in Workshop Temporal Dynamics in Speech and Hearing, Park Plaza Astrid Hotel, Antwerp, Belgium.

Jenstad, L. M., and Souza, P. E. (**2005**). "Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility," J. Speech Hear. Res. **48**, 651–667.

Kamm, C. A., Dirks, D. D., and Bell, T. S. (**1985**). "Speech recognition and the articulation index for normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 281–288.

Kates, J. M. (**1992**). "On the use of coherence to measure distortion in hearing aids," J. Acoust. Soc. Am. **91**, 2236–2244.

Kates, J. M., and Arehart, K. H. (**2005**). "Coherence and the speech intelligibility index," J. Acoust. Soc. Am. **117**, 2224–2237.

Kates, J. M., and Arehart, K. H. (**2007**)."Characteristics of nonlinear distortion in hearing aids: Relating signal characteristics to auditory perception," poster presentation at the 13th Association for Research in Otolaryngology, Denver, CO.

Kates, J. M., and Kozma-Spytek, L. (**1994**). "Quality ratings for frequency-shaped peak-clipped speech," J. Acoust. Soc. Am. **95**, 3586–3594.

Kryter, K. D. (**1962**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.

Lobdell, B. E., and Allen, B. A. (**2007**). "A model of the VU (volume-unit) meter, with speech applications," J. Acoust. Soc. Am. **121**, 279–285.

Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low-and high-frequency speech," ARLO **4**, 124–128.

Moore, B. C. J., Stone, M. A., Füllgrabe, C., Glasberg, B. R., and Puria, S. (**2008**). "Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss," Ear Hear. **29**, 907–922.

Naylor, G., Johannesson, R. B., and Lunner, T. (**2006**). "Fast-acting compressors change the effective signal-to-noise ratio—both upwards and downwards!," presentation at the International Hearing Aid Research Conference, Lake Tahoe, CA.

Naylor, G., Johannesson, R. B., and Rønne, F. (**2008**). "An investigation of effective SNR-change through amplitude-compression hearing aids," in *Auditory Signal Processing in Hearing Impaired Listeners*, edited by T. Dau, J. M. Bucholz, J. M. Harte, and T. U. Christiansen (Center Tryk A/S, Helsingør), pp. 595–600.

Olofsson, A., and Hansen, M. (**2006**). "Objectively measured and subjectively perceived distortion in nonlinear systems," J. Acoust. Soc. Am. **120**, 3759–3769.

Pavlovic, C. V. (**1987**). "Deviation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.

Pavlovic, C. V., and Studebaker, G. A. (**1984**). "An evaluation of some assumptions underlying the articulation index," J. Acoust. Soc. Am. **75**, 1606–1612.

Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Preminger, J. E., and van Tasell, D. J. (**1995**). "Quantifying the relation between speech quality and speech intelligibility," J. Speech Hear. Res. **38**, 714–725.

Rankovic, C. M. (**1997**). "Prediction of speech reception by listeners with sensorineural hearing loss," in *Modeling sensorineural hearing loss*, edited by W. Jesteadt (Lawrence Erlaum Associates, Mawah, NJ).

Rankovic, C. M. (**1998**). "Factors governing speech reception benefits of adaptive linear filtering for listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **103**, 1043–1057.

Rhebergen, K. S. (**2006**). "Modeling the speech intelligibility in fluctuating noise," Ph.D. thesis, University of Amsterdam, Amsterdam.

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. (**2008a**). "Learning effect observed for the speech reception threshold in interrupted noise with normal-hearing listeners," Int. J. Audiol. **47**, 185–188.
Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (**2008b**). "Quantifying and modeling the acoustic effects of compression on speech in noise," J. Acoust. Soc. Am. **123**, 3167.
Shailer, M. J., and Moore, B. C. (**1983**). "Gap detection as a function of frequency, bandwidth, and level," J. Acoust. Soc. Am. **74**, 467–473.
Sherbecoe, R. L., and Studebaker, G. A. (**1990**). "Regression equations for the transfer function of ANSI S3.5–1969," J. Acoust. Soc. Am. **88**, 2482–2483.
Sherbecoe, R. L., and Studebaker, G. A. (**2003**). "Audibility-index predictions of normal-hearing and hearing-impaired listeners' performance on the connected speech test," Ear Hear. **24**, 71–88.
Souza, P. E., Jenstad, L. M., and Boike, K. (**2006**). "Measuring the acoustic effects of compression amplification on speech in noise," J. Acoust. Soc. Am. **119**, 41–44.
Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech transmission quality," J. Acoust. Soc. Am. **67**, 318–326.
Stelmachowicz, P., Lewis, D., Hoover, B., and Keefe, D. H. (**1999**). "Subjective effects of peak clipping and compression limiting in normal and hearing-impaired children and adults," J. Acoust. Soc. Am. **105**, 412–422.
Stelmachowicz, P., Lewis, D., Kalberer, A., and Creutz, T. (**1994**). *Situational Hearing Aid Response Profile User's Manual,* SHARP, *v 2.0* (Boys Town National Research Hospital, Omaha, NE).
Stone, M. A., and Moore, B. C. J. (**2007**). "Quantifying the effects of fast-acting compression on the envelope of speech," J. Acoust. Soc. Am. **121**, 1654–1664.
Studebaker, G. A., and Sherbecoe, R. L. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422–36.
Studebaker, G. A., Sherbecoe, R. L., and Gilmore, C. (**1993**). "Frequency importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test," J. Speech Hear. Res. **36**, 799–807.
Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normalspeech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.
van Buuren, R. A., Festen, J. M., 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 Schijndel, N. H., Houtgast, T., and Festen, J. M. (**1999**). "Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **105**, 3425–3435.
van Tasell, D. J. (**1993**). "Hearing loss, speech, and hearing aids," J. Speech Lang. Hear. Res. **36**, 228–244.
Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.
Versfeld, N. J., Festen, J. M., and Houtgast, T. (**1999**). "Preference judgments of artificial processed and hearing-aid transduced speech," J. Acoust. Soc. Am. **106**, 1566–1578.
Villchur, E. (**1989**). "Comments on "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function" [J. Acoust. Soc. Am. 83, 2322–2327 (1988)]," J. Acoust. Soc. Am. **86**, 425–427.
Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., and Segel, P. (**2002**). "Speech dynamic range and its effect on cochlear implant performance," J. Acoust. Soc. Am. **111**, 377–386.

Speech identification in noise: Contribution of temporal, spectral, and visual speech cues

Jeesun Kim and Chris Davis[a)]
MARCS Auditory Laboratories, University of Western Sydney, NSW 1797, Australia

Christopher Groot
Department of Psychology, The University of Melbourne, VIC 3010, Australia

(Received 20 March 2008; revised 24 September 2009; accepted 25 September 2009)

This study investigated the degree to which two types of reduced auditory signals (cochlear implant simulations) and visual speech cues combined for speech identification. The auditory speech stimuli were filtered to have only amplitude envelope cues or both amplitude envelope and spectral cues and were presented with/without visual speech. In Experiment 1, IEEE sentences were presented in quiet and noise. For in-quiet presentation, speech identification was enhanced by the addition of both spectral and visual speech cues. Due to a ceiling effect, the degree to which these effects combined could not be determined. In noise, these facilitation effects were more marked and were additive. Experiment 2 examined consonant and vowel identification in the context of CVC or VCV syllables presented in noise. For consonants, both spectral and visual speech cues facilitated identification and these effects were additive. For vowels, the effect of combined cues was underadditive, with the effect of spectral cues reduced when presented with visual speech cues. Analysis indicated that without visual speech, spectral cues facilitated the transmission of place information and vowel height, whereas with visual speech, they facilitated lip rounding, with little impact on the transmission of place information.
 [DOI: 10.1121/1.3250425]

PACS number(s): 43.71.Es, 43.66.Ts [ADP] Pages: 3246–3257

I. INTRODUCTION

Teasing apart the pattern recognition process that underpins speech perception is not straightforward because speech is redundantly represented by multiple cues. One approach has been to use variously filtered speech signals in an attempt to establish the minimal time-frequency properties that can support good speech recognition performance. This research strategy has the potential to contribute to theories of speech pattern recognition as well as to the design of the signal processing component of hearing prostheses (by helping to pin-point which cues are important for speech recognition).

A number of earlier studies examining the perception of reduced auditory speech did so in the context of determining whether a much reduced auditory signal could facilitate visual speechreading. For example, Erber (1972) tested hearing impaired and normal hearing participant's ability to identify speech envelope-shaped white-noise derived from a set of 240 frequent nouns. Two conditions were tested, visual speech alone or visual speech plus noise. The performance of both participant groups (correct word recognition) improved in the visual speech plus noise condition by around 7%. Erber's study is important for several reasons: it established the potential for speech-envelope shaped noise as a useful probe of speech recognition ability and examined this within an auditory-visual speech processing environment (a situation that is particularly relevant for hearing-impaired people, Walden *et al.*, 1990 and cochlear-implant listeners, Rouger *et al.*, 2007; Desai *et al.*, 2008).

Several subsequent studies investigated the contribution of auditory time-intensity variation to speech perception without accompanying visual speech. For example, Van Tasell *et al.* (1987) examined the effect of envelope bandwidth on consonant recognition by having three different low pass frequency cutoffs (20, 200, 2000 Hz) for the speech used to shape the noise (in this case pink noise was used). This manipulation of increasing bandwidth had the effect of increasing the temporal detail of the noise waveform envelope. Significant improvements were found in consonant recognition when the bandwidth was increased from 20 to 200 Hz (but not for that of 200–2000 Hz) suggesting that amplitude modulation (AM) rates of about 200 Hz may be more useful for speech recognition. Multidimensional scaling identified response classes based on the features of voicing, amplitude, and burst envelop. Based on their results, Van Tassel *et al.* (1987) proposed that auditory envelope features provided complementary information to that specified by visual speech (specifically visemes) and that the two combined would theoretically permit almost all consonant information to be transmitted (although the combination of reduced auditory cues and visual speech was not actually tested).

The way in which amplitude envelope cues combine with visual speech has been investigated in a number of other studies (e.g., Risberg, 1974) with the key issue being the degree to which speech recognition performance with

[a)]Author to whom correspondence should be addressed. Electronic mail: chris.davis@uws.edu.au

different auditory cues will be facilitated by visual speech. In a comprehensive study, Grant *et al.* (1991) examined factors that had varied across previous studies, i.e., the spectral region of the speech signal from which the envelope was derived, the effect of different width smoothing filters, the carrier frequency employed (using pure tones that had different spectral centers so the signal was transposed to different frequency regions), and carrier type (pure tones or noise carriers, although this was not systematically tested). In general, it was found that the wideband speech derived envelope did not provide the greatest speechreading benefit compared to envelopes derived from selected octave bands of speech. These selected amplitude envelope cues provided substantial benefit to speechreading but not if these signals were transposed to regions remote from their original spectral locations. A follow-up study by Grant *et al.* (1994) (see also Breeuwer and Plomp, 1984, 1986), that again examined how visual and auditory speech cues assist in speech perception, showed that two speech envelopes from different spectral regions resulted in better speechreading than did the envelope from a single channel.

In sum, the results from the above-mentioned studies have indicated that amplitude envelope cues from certain spectral bands are more critical to speech recognition than others and that information derived from two (or more) of these bands supports better recognition than does that from a single band. The broad progression of these studies has been to move beyond the demonstration that greatly reduced auditory signals can facilitate speech processing to probing which aspects of the speech signal best support speech recognition. In this regard, one additional study is worth mentioning because it showed that besides amplitude envelope cues, fundamental frequency information can also assist speech recognition in accompany with visual speech. Grant *et al.* (1985) examined the impact of low frequency speech cues (e.g., a fixed frequency tone modulated by the speech intensity envelope or a fixed amplitude tone frequency-modulated in a one-to-one correspondence with the talker's fundamental frequency, or both F0 and cues to amplitude modulation) on speech reading performance. It was shown that the greatest improvement in speech reading was with the addition of both the AM and the frequency modulation (FM) auditory information (i.e., those cues related to rhythm, stress, and intonation pattern).

Given that it was later shown that high levels of intelligibility could be achieved by combinations of time-intensity envelopes from different speech bands (e.g., Shannon *et al.*, 1995), many subsequent studies focused on what could be achieved with these amplitude envelope signals alone rather than in combination with visual speech (or F0 information). For example, it was found that very high levels of word recognition were possible with speech amplitude shaped noise envelopes from as few as four frequency bands (e.g., Dorman *et al.*, 1997; van der Horst *et al.*, 1999; Shannon *et al.*, 1995). Demonstrations that amplitude modulations from a limited number of speech bands were sufficient to support speech recognition have played a role in the widespread adoption of cochlear implant speech processing strategies that extract and represent such temporal envelope cues.

Recently, it has become clear that although temporal envelope cues provide sufficient information for speech recognition in quiet conditions, they do not do so when there is background noise (e.g., Dorman *et al.*, 1998; Fu and Shannon, 1999; Nelson *et al.*, 2003; Zeng *et al.*, 2005). This susceptibility to competing noise has been demonstrated either by showing that recognition performance with a fixed number of channels is degraded by noise (Stickney *et al.*, 2005) or by showing that more channels are required to maintain the same level of performance in the presence of noise (Dorman *et al.*, 1998; Fu *et al.*, 1998).

In response to the apparent shortcomings of amplitude modulated processing schemes, Nie *et al.* (2005, 2006) showed that the addition of slowly varying FM cues to those associated with the speech amplitude envelope leads to better recognition performance for both words and phonemes in the presence of a competing talker (Nie *et al.*, 2005; Zeng *et al.*, 2005; Stickney *et al.*, 2007). The FM extraction algorithm that was developed [the frequency amplitude modulation encoding (FAME) algorithm] was designed to suit cochlear implant listening by decomposing a signal into slowly varying amplitude and frequency modulations (for the FM component, band-limited instantaneous frequency was extracted from each bandpass filtered signal and used as the carrier frequency for each signal source). To verify the properties of the extracted AM and FM components after FAME processing, Nie *et al.* (2005) performed a comparison that demonstrated that, compared to the AM only stimuli, in the AM +FM stimuli both formant transition and fundamental frequency cues were more clearly extracted and encoded. Indeed, it was recently concluded that the FAME algorithm results in an F0 contour that is very similar to the original (with slight differences due to small errors in the instantaneous frequency calculation, Morton *et al.*, 2008) and that compared to AM coding schemes, FAME enhances temporal envelope information and adds temporal fine structure (Nie *et al.*, 2006).

In explaining why identification in noise is better with FAME processed signals compared to the standard implant simulations, Stickney *et al.* (2005, 2007), Nie *et al.* (2005), and Zeng *et al.* (2005) adopted a two stage account. First, it was proposed that because FAME processing delivers more acoustic information (e.g., both the temporal envelope and temporal fine structure cues) than AM only schemes, FAME processed auditory signals would transmit more information relevant to speech processing. Second, because FAME preserves cues to F0 and access to the target speech's intonation patterns, it may facilitate the segregation of perceptual streams in a multi-source environment this leading to better speech recognition (Bregman, 1990), allowing the target and masker signals to be segregated into separate perceptual streams, which in turn leads to better speech identification in noise.

The issue, raised by Stickney *et al.* (2005, 2007) and Nie *et al.* (2005, 2006), of how cochlear implant speech signal processing strategies perform under more realistic listening (noisy) conditions leads to the question of how such processing strategies interact with visual speech (since under normal listening conditions the perceiver also has information avail-

able from viewing the talker). The reintroduction of visual speech as relevant to the problem of how best to achieve noise tolerance for reduced speech signals not only makes contact with earlier research (e.g., Grant *et al.*, 1985) but also provides a bridge to recent research concerned with developing sophisticated speech processing algorithm for cochlear implants. As such, the current research, using the processing algorithm developed by Nie *et al.* (2005, 2006) and Zeng *et al.* (2005), examined degraded speech recognition in noise in conjunction with visual speech. Here, the issues raised by Grant *et al.* concerning the degree to recognition performance with different auditory cues will be facilitated by visual speech are of interest. In this regard, the basis of the current inquiry is straightforward: auditory signals that provide cues that are complementary with speechreading should result in greater auditory-visual (AV) benefit than those that are less complementary with speechreading (see Grant and Walden, 1996; Walden *et al.*, 1974).

In considering possible interactions between FM and visual cues, the explanation (outlined above) for why FAME processing facilitates speech recognition in noise suggests some interesting possibilities. First, if the FM information provided by FAME provides additional phonetic information (e.g., spectral cues) about the identity of the target speech segments, then the effectiveness of these might depend on how such cues interact with those provided by visual speech. For example, it might be that in some cases auditory and visual cues are complementary (e.g., cues for manner transmitted by the auditory channel; cues for place by the visual one, Walden and Montgomery, 1975). Or such cues might provide very similar information, e.g., place of articulation may be redundant with speechreading (indeed, it has been shown that there is generally a high correlation between amplitude in the F2 region and changes in mouth opening, Grant and Seitz, 2000; Kim and Davis, 2003).

Second, if the provision of FM cues assists in segregating the target talker from the competing one (Stickney *et al.*, 2007), then this function might be more useful for auditory only (AO) presentations compared to AV ones. That is, it is clear that visual speech can also have a powerful effect on auditory speech segregation, with the usefulness of visual speech information for source separation demonstrated in behavioral experiments by Driver (1996) and more recently by Halfer and Freyman (2005). For instance, Helfer and Freyman (2005) showed that visual speech cues provided benefit for both recognition and detection of speech when the masker consisted of other voices compared to steady-state noise and proposed that speechreading cues help to segregate a target voice from competing voices. Moreover, Sodoyer *et al.* (2004) and Rivet *et al.* (2007) demonstrated that visual speech can provide a useful constraint for blind source separation. Here, proposals concerning speaker segregation based on visual speech have also adopted a scene analysis framework in which the coherence of audio and visual speech features is exploited to either extract a speech source or to use the visual speech information of a talker as a voice activity detector.

The above considerations of the different ways in which FAME processing might interact with visual speech provided the motivation for the two experiments of the current study. The basic aim was to examine the size of the FAME processing advantage in auditory only and auditory-visual presentation conditions. Although FM and visual speech signals provide cues both to phonetic features and for segregating target speech from the background, it seems reasonable to expect that sentences will provide more opportunity for F0 cues to assist in speech segregation compared to meaningless syllables. Thus, Experiment 1 used sentence stimuli to investigate AV combination in a presentation situation more suited to of the operation of segregation cues, whereas Experiment 2 used consonant and vowel stimuli to investigate the combination of auditory and visual cues for specific phonetic features.

II. EXPERIMENT 1

This experiment tested word identification for sentences presented in AO and in AV conditions. Adopting the method of Nie *et al.* (2005) and Zeng *et al.* (2005) auditory speech was represented by traditional amplitude modulated envelope cues only or by envelope cues in addition to those associated with the slowly varying frequency modulated signal (using the FAME processing algorithm developed by Nie *et al.*, 2005; Zeng *et al.*, 2005). These AM and AM+FM stimuli were tested in quiet and noise in AO and AV conditions. The target voice was male and the masker used was multi-talker babble (consisting of three female talkers and one male). It should be pointed out that the use of multiple talkers as competing noise is different from Nie *et al.* (2005), who used a single different male talker uttering a competing IEEE sentence, or Zeng *et al.* (2005) who used either a competing single male talker (same sex) or a single competing female talker (different sex), with the difference between four talker babble and a single competitor speaker possibly reducing the benefit of F0 cues for stream segregation.

A. Methods

1. Listeners

Fifty-seven first year undergraduate university students participated in the experiment. Participants were native speakers of Australian English, 18 years of age or over and had self-reported normal or corrected-to-normal vision and no history of reported hearing loss.

2. Test materials

Eighty sentences were selected from the 1969-revised IEEE/Harvard list of phonetically balanced sentences. IEEE sentences were employed in order to avoid ceiling effects due to their complicated sentence structure and information content that have been shown to elicit lower recognition performance than either City University of New York (CUNY) sentences or Hearing in Noise Test sentences (Zeng *et al.*, 2005).

The sentences were spoken by a male native Australian English speaker and recorded in a double walled, sound attenuated room. Both audio and video were recorded using a Sony DCR-DVD602E Digital DVD handycam. The speaker was recorded against a uniform gray background, facing the

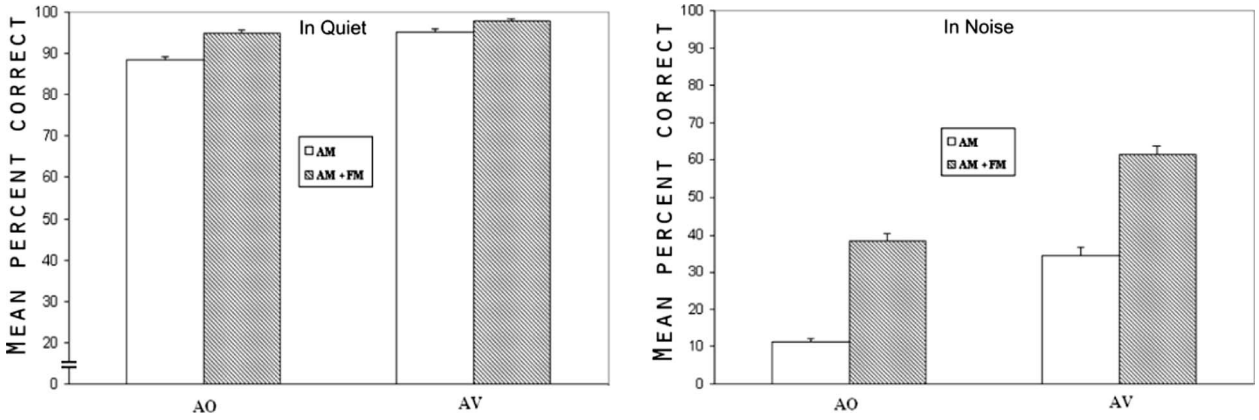


FIG. 1. Left panel: Mean percent correct word recognition scores for each condition presented in quiet. Right panel: Mean percent correct word recognition scores for each condition presented in noise (AM=AM cue; AM+FM=AM+FM cue; AO=auditory only; AV=auditory-visual speech). Error bars are ±1 standard error.

camera and the recording showed the head and shoulders. The video was digitized at 29.9 frames/s with a resolution of 352×240 pixels. The video was sampled at 44 000 Hz. A commercially available multi-talker babble track (Auditec, St. Louis, MO) was used as competing noise stimuli.

3. Signal processing

Audio signal processing was performed using the FAME processing algorithm (Zeng *et al.*, 2005). The following presents a brief description of this algorithm (for more detailed information, see Nie *et al.*, 2005). The wide-band signal is separated into n sub-bands by band-pass filters (with cochlear filters mimicked by having the center frequencies distributed equally on a log scale). Two separate pathways are then used to extract AM and FM in each band. The AM path extracts the slowly varying envelope and is obtained by full-wave rectification of the band-pass filter output and then low-pass filtered. The slow varying FM signal is obtained by the removal of each band's center frequency (by phase-orthogonal demodulators) and a subsequent low-pass filter (with a cutoff frequency of 500 Hz) and rate (with a cutoff value of 400 Hz). The delay between the AM and FM signals is adjusted and the AM and FM signals are each combined into their respective sub-bands. These signals are further bandpassed filtered to remove frequency components outside the original analysis filter's bandwidth. The band-passed signals are then summed to form the synthesized AM+FM signal.

In the current study, eight frequency bands were used as this condition showed the largest FM advantage for sentence recognition both in quiet and noise (Zeng *et al.*, 2005). In the noise condition, the audio portion of both the target IEEE sentences and competing "babble" noise were mixed at −5 dB signal to noise ratio (SNR) and subjected to FAME processing (i.e., FAME processing was applied after the target and masker were mixed). For all stimuli containing competing noise, the onset of noise occurred prior to that of speech stimuli and had a longer duration. For visual speech conditions, the processed speech signal was dubbed back onto the video of the talker (using Virtualdub).

4. Procedure

All participants were tested individually in a separate booth. Auditory stimuli were presented through Sennheiser HD580 headphones. The video clips (compressed to MPEG 2) were played back using the DMDX software (Forster and Forster, 2003) on a ViewSonic G810 21 in. monitor at a resolution of 1024×768. In conditions for which there was no visual speech, the participants viewed a blank screen. All participants heard a total of 80 sentences, i.e., 10 sentences presented in 8 conditions, i.e., 2 (AO/AV)×2 (quiet/noise)×2 (AM/AM+FM). Sentences within each condition were presented in a block but the presentation of sentences within each block was randomized, as was the presentation order of the condition blocks themselves. Eight versions of the experimental list were prepared such that no item was repeated in any version but each version contained all conditions (i.e., sentences for all conditions were fully rotated, resulting in ten sentences in each condition). At the commencement of each testing condition, participants were provided with a set of four practice items to familiarize them with the various conditions of degraded speech. After each stimulus presentation, participants typed their responses. In scoring these data, all words were scored with credit only given if the typed word exactly matched the spoken word (except where the response was an obvious typo). The percentage correct word identification was calculated as the measure of speech recognition for each condition.

B. Results and discussion

Mean percent correct word recognition scores for each condition in quiet (left panel) and in noise (right panel) are presented in Fig. 1.

It is clear from comparing the left and right panels of Fig. 1 that presenting the sentences mixed with multi-talker babble had a deleterious effect on speech recognition. There was a significant drop in correct identification scores for AO AM speech in-quiet versus in-noise (from 89% correct to 11%), $F(1,50)=1977.05$, $p<0.05$.

1. Word recognition in quiet

As can be seen in Fig. 1 (left panel), performance was near ceiling for all conditions. This was particularly the case for the AV conditions. The participant data were analyzed with a 2 (AO versus AV) ×2 (AM versus AM+FM) ANOVA. There was a FM effect, as more words were correct in the AM+FM condition compared to the AM condition, $F(1,50)=43.92$, $p<0.05$. This result shows that even in quiet, the addition of FM resulted in superior performance. Likewise, there was an AV effect with more words correctly identified in the AV condition than in the AO condition, $F(1,50)=43.73$, $p<0.05$. The AV effect was significant both in the AM condition (6.6%), $F(1,50)=37.7$, $p<0.05$, and in the AM+FM condition (2.6%), $F(1,50)=15.95$, $p<0.05$.

2. Word recognition in noise

As with the in-quiet data, the noise data were analyzed with a 2 (AO versus AV) ×2 (AM versus AM+FM) ANOVA. Similar to the in-quiet data, there was a FM effect, more words were correctly recognized in the AM+FM condition compared to the AM condition, $F(1,50)=273.13$, $p<0.05$. There was also an AV effect; more words correctly identified in the AV condition than in the AO condition, $F(1,50)=243.60$, $p<0.05$. There was no interaction between these effects, $F<1$.

As can be seen in Fig. 1 (right panel), there was a clear FM effect in both the AO condition (23.9%), $F(1,50)=158.7$, $p<0.05$ and the AV condition (23.0%), $F(1,50)=127.9$, $p<0.05$; also a clear AV effect in the AM condition (27.4%), $F(1,50)=186.0$, $p<0.05$ and in the AM+FM condition (26.5%), $F(1,50)=127.3$, $p<0.05$. A step-wise increase in correct word recognition occurred as the different conditions were combined, i.e., adding the FM cues to the AM AO condition (a 23.9% increase), adding visual cues to the AM AO condition (a 27.4% increase), or adding both (a 50.5% increase).

The results showed that FAME processing enabled better speech recognition in noise in both the AO and the AV conditions. This finding clearly supports the position that FM cues can play an important role in speech perception in noise. Contrary to the expectation that the FM effect might be reduced for AV presentation (because visual speech would provide a strong cue for speech segregation), the size of the effect (as measured by subtracting the AM result from the AM+FM one) remained constant across the AO and AV conditions. This suggests that, at least in this experiment, the cues provided by FAME processing and those by visual speech were not redundant. However, given that there are numerous speech cues provided by AM, FM, and visual speech sources of information, determining the degree of redundancy between these sources is not a simple matter and a more detailed examination of the contribution of specific phonetic features is required.

To examine the way in which the FM and AV effects combined, the pattern of errors made across the experimental conditions was inspected. Unfortunately, participant responses did not reveal clear error patterns since typically in the AO and AM noise conditions participants simply reported fewer words than in the AV and AM+FM conditions. One way to explore in more detail how the FM and AV effects combine is to investigate how consonants and vowels (presented in the context of nonsense syllables such as a Ca and hVd) are recognized in the presence/absence of the cues. Separating consonant and vowel recognition provides a way of examining how visual speech interacts with the spectral and temporal aspects of speech, since it appears that, to some degree, the recognition of consonants and vowels depends to different degrees on spectral and temporal cues (Nie *et al.*, 2006; Xu *et al.*, 2005).

TABLE I. Feature classification for vowels and consonants.

Vowels					
Label	Features				
Vowel height	Closed		Mid		Open
Place	Front		Center		Back
Lip rounding	Rounded		Unrounded		
Consonants					
Label	Features				
Manner	Plosive	Fricative	Affricate	Nasal	Glide
Place	Labial	Dental	Alveolar	Palatal	Back
Voicing	Voiceless	Voiced			

III. EXPERIMENT 2

Experiment 2 examined how AM, AM+FM, and visual cues would combine to affect phoneme identification in noise. As with Experiment 1, the aim was to characterize which FM and visual speech cues appear to be redundant and which appear more complementary. Consonant and vowel recognition were examined separately with vowel identification tested in hVd syllables (Hillenbrand *et al.*, 1995) and consonant identification in aCa syllables (Shannon *et al.*, 1999).

In what follows, we briefly consider how vowels and consonant recognition might be affected by FAME processing, visual cues, and then as a function of the combination of both. Following Grant and Walden (1996), specific features are considered, for vowels: lip-rounding, vowel height, and place (backness); for consonants: voicing, manner, and place (see Table I for details).

As mentioned above, the choice to contrast vowels and consonants was due to likely differences in the acoustic and visual properties that underpin their recognition in noise. For instance, based on the experiments that used babble and speech-shaped noise maskers, Parikh and Loizou (2005) proposed that vowel identification in noise is supported by information about F1 frequency along with some information about F2. More recently, Xu and Zheng (2007) examined vowel and consonant identification in noise by manipulating the presented spectral information (i.e., varying the number of channels in a noise-excited vocoder). It was found that an increase in the number of channels in noise conditions beyond the point of in-quiet plateau performance improved vowel (but not consonant) recognition. Xu and Zheng (2007) suggested that this result might have been due to vowel rec-

ognition in noise relying on spectral cues (whereas consonant recognition does so to a far lesser extent). If this is the case, then FAME processing should assist recognition of all of the vowel features (since expression of these features are typically correlated with formant frequencies). That is, lip rounding typically lowers formant frequencies, with the second formant particularly affected (Harrington and Cassidy, 1999). Lip height tends to have an inverse relationship with F1 and place (backness) tends to be inversely related with F2.

In regard to consonant features, Rosen (1992) suggested that manner is mainly transmitted by envelope cues, voicing by cues for periodicity and place by spectral cues (with Parikh and Loizou, 2005, proposing that in noise, formant transitions may be the most prominent cue). Given the above analysis, it would be expected that the FM cues provided by FAME processing will assist vowel more than consonant recognition (perhaps with the exception of the consonant place cue). This expectation is consistent with the results of Nie *et al.* (2005), who tested vowels and consonant recognition at several noise levels and numbers of processing bands and showed the FM information tended to provide a larger benefit for vowel compared to consonant recognition.

What might be expected for vowels and consonants features for AV compared to AO presentation? In regard to vowel features, Robert-Ribes *et al.* (1998) presented individual vowels in white noise at several SNRs in visual-Only, AO, and AV conditions. They found evidence for a complementarity of the auditory and visual signals at the level of features, such that for visual transmission the most information was transmitted by the cue for lip rounding, followed by the cue for vowel height, and then the place cue (backness). This contrasted with the AO transmission results where vowel height was best, then place, and then rounding. The AV results showed a synergy between the complementary features such that the rounding feature was identified best, at all noise levels, then the height feature, and then place.

For consonants, Grant and Walden (1996) examined the recognition of aCa consonant syllables in speech-shaped noise (mixed at various SNRs) for AO, visual-only, and AV presentations as a function of voicing, manner, and place. They showed that for AV recognition place of articulation appeared to be determined mainly by the visual modality (the AV scores for place-of-articulation were relatively constant despite of changes in AO performance). The recognition of voicing was determined by the auditory modality (there was no clear difference between AO and AV conditions). Whereas both modalities appeared to contribute to the recognition of manner of articulation, however, when chance performance was taken into account, it appeared that cues for manner were derived predominantly from the AO condition. Grant and Walden (1996) also examined the effects of filtering the auditory signal and found that low-frequency bands of speech provided more AV benefit than did high-frequency ones. It was suggested that this was because low-frequency speech bands present cues that are somewhat complementary with speechreading (such as voicing and manner of articulation) whereas high-frequency bands present cues that are relatively redundant (e.g., place of articulation). This finding gave rise to the proposal that limited AV benefit will occur when listening conditions result in auditory responses that are redundant with speechreading ones.

In summary then, FAME processing should facilitate vowels more than consonants with all vowel features showing a benefit. Furthermore, compared to AM processed stimuli, auditory recognition responses to FAME processed stimuli might be more similar to speechread ones (particularly for the vowel height feature). This would result in smaller FAME effects with AV presentation. Such a reduction in the effect of FAME processing might not be as evident with consonants.

A. Methods

1. Listeners

Thirty-one undergraduate university students participated in the experiment. Participant characteristics were similar to those of Experiment 1; none had participated in Experiment 1.

2. Test materials

The speech stimuli consisted of 37 syllables (21 consonants in aCa context; C=/p,t,k,f,θ,s,ʃ,tʃ,h,b,d,g,v,z,ʤ,j,m,n,r,l,w/ and 16 vowels in hVd context; V=/iː,I,æ,e,ɐ,ɔ,a,aː,əʉ,ʊ,ʉ,ɑe,oI,ɐɔ,ɜː/) These syllables were selected such that the different consonants and vowels covered the range of place and manner of articulations in English. The selected syllables were spoken by a male native Australian English speaker in a double walled, sound attenuated room. Other aspects of stimuli preparation including recording the stimuli, audio signal processing and dubbing the processed signal to the video were similar to those of Experiment 1.

3. Procedure

All the items were presented under four conditions (each in noise): AO AM, AO AM+FM, AV AM, and AV AM+FM. Four versions of the item list were prepared such that no item was repeated in any version but each version contained all conditions (i.e., syllables were fully rotated over conditions).

All the participants were allocated to two versions such that they were presented with the same item in the AM and AM+FM conditions. Note that items in the AO condition were not presented in the AV condition. This procedure was used in an attempt to reduce learning effects as these may influence the scores from one presentation to another when visual speech is involved and such learning effects may vary considerably across individual participants (Breeuwer and Plomp, 1986). In regard to the AM and AM+FM conditions, in which items were repeated, the order of presentation was randomized (in an attempt to off-set any systematic effects of learning).

The AO and AV conditions were presented in blocks but presentation of items within each block was randomized, as well as the presentation of the condition blocks themselves. The rest of the testing procedure was the same as Experiment

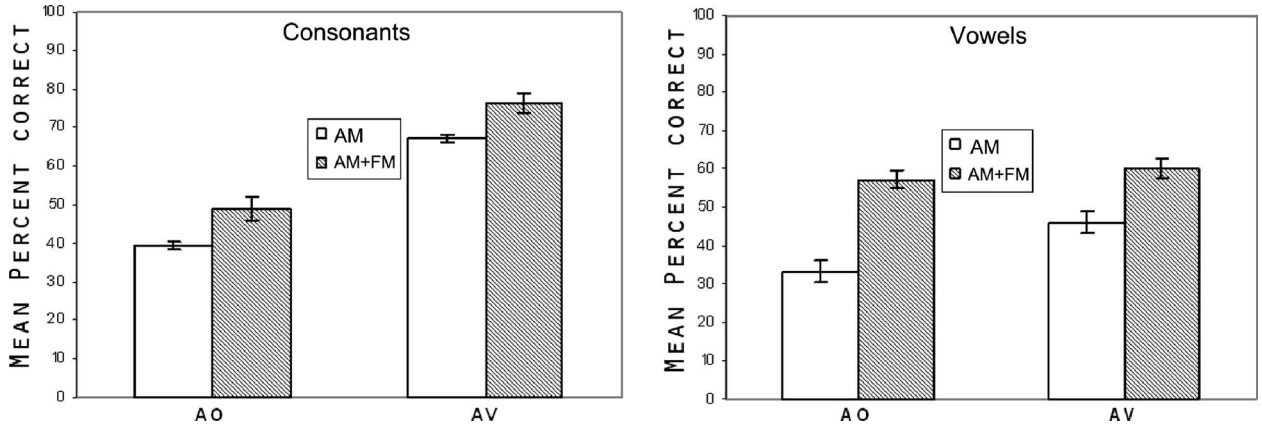


FIG. 2. Mean percent correct consonant recognition scores (left panel) and vowel recognition scores (right panel) for each condition presented in noise (AM=AM cue; AM+FM=AM+FM cue; AO=auditory only; AV=auditory-visual speech). Error bars are ±1 standard error.

1. As in Experiment 1, we used an open-response task as Sommers *et al.* (1997) suggested that closed set tests might encourage atypical perceptual strategies.

B. Results and discussion

The results will be presented in the following order. First, overall levels of identification performance (mean percent correct) will be reported, followed by a data on which features were correctly identified and finally the results of a sequential information analysis (SINFA) (Wang and Bilger, 1973) will be presented.[1]

1. Recognition of consonants and vowels

Mean percent correct scores for vowel and consonant identification in each condition are presented in Fig. 2. The participant data were analyzed with a 2 (AO versus AV) ×2 (AM versus AM+FM) ×2 (vowel versus consonant) ANOVA. More phonemes were correctly identified in the AV condition than in the AO condition, $F(1,29)=77.08$, $p<0.05$ (a visual enhancement effect). Likewise there were more phonemes correct in the AM+FM condition compared to the AM condition,$F(1,29)=165.08$, $p<0.05$ (a FM effect). Consonants were identified better than vowels, $F(1,29)=14.30$, $p<0.05$. There was a significant interaction between these effects, such that the FM effect was greater in the AO condition compared to the AV condition, $F(1,29)=4.48$, $p<0.05$. Furthermore, this FM effect was larger for vowels than consonants, $F(1,29)=14.31$, $p<0.05$, whereas the visual enhancement effect was larger for the consonants than vowels, $F(1,29)=22.15$, $p<0.05$. The three-way interaction was not significant, $F(1,29)=2.49$, $p>0.05$.

For vowels the size of the FM effect was different for AO and AV conditions, i.e., there was a significant interaction between the visual and the FM effects, $F(1,29)=5.26$, $p<0.05$ with the FM effect larger in the AO than the AV condition. Indeed, for vowels, the mean percent correct scores in the AM+FM conditions were the same for the AO and AV presentation conditions, $F<1$. For consonants, the size of the AV effect was similar for AM and AM+FM conditions (i.e., there was no interaction, $F<1$).

In sum, FM cues enhanced vowel identification more than consonant identification and the AV facilitation effect was greater for consonants than vowels. The effects of the visual and FM cues were additive for consonants but subadditive for vowels. Although the overall correct identification rates provide some information about the effect of the combination of speech cues, the main interest of the current experiment was to explore the possible basis of these effects. To this end, the data were reanalyzed in terms of feature identification and an examination of feature information transmission was conducted. An examination of the transmission of phonetic features may clarify the basis for some of the results, for example, the similar performance levels of vowel perception for the AO and AV AM+FM conditions.

2. Perceptual confusions of consonants and vowels

To ascertain the relative contribution of phonetic features to the various conditions, the identification data were re-scored in terms of features. Table I shows how the 16 vowels and 21 consonants used in this study were classified. Feature scores were obtained from AO and AV confusion matrices by scoring each response according to the stimulus feature classifications listed in Table I. For example, if the response to the stimulus /aba/ was /aga/, then this was scored as correct for voicing and manner but incorrect for place. Figure 3 shows the mean percent correct response scores for each of these features measured for AO and AV conditions for vowels and consonants, respectively.

The item scores for the vowels were analyzed with a 3 (features: rounding, height, place) ×2 (AO versus AV) ×2 (AM versus AM+FM) ANOVA. There was a significant AV effect, with more features correctly identified in the AV (75.5%) compared to the AO (66.4%) condition, $F(1,20)=12.10$, $p<0.05$. There was also a significant FM effect, with more features correctly identified in the AM+FM condition (77.4%) compared to the AM (64.5%), $F(1,20)=14.61$, $p<0.05$. There was also a significant interaction between these effects, with a larger FM effect (18.4%) in the AO condition compared to the AV condition (7.3%), $F(1,20)=5.46$, $p<0.05$. The different vowel feature scores

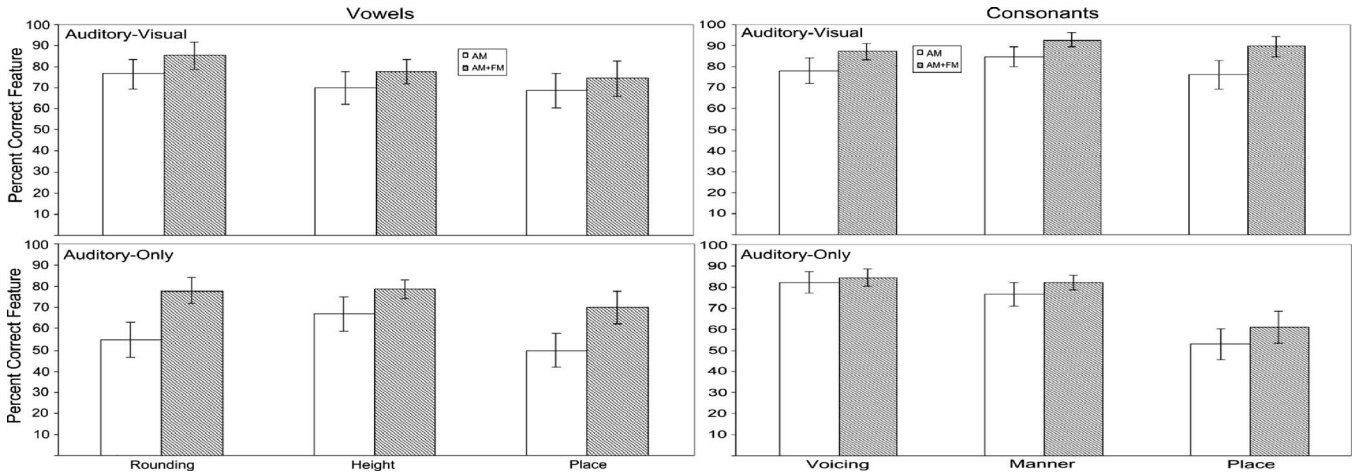


FIG. 3. Percent correct feature judgments as a function of vowels (left panel) and consonants (right panel) in the auditory-visual (top panels) and auditory-only (bottom panels) presentation conditions for the AM and AM+FM conditions. Error bars are ±1 standard error.

did not differ from each other, $F<1$ nor did the feature variable interact with the AV effect, $F(2,20)=1.63$, $p>0.05$ or the FM effect.

The percent correct scores for each feature were analyzed in a series of ANOVAs. For the lip-rounding feature, there was a significant AV effect (14.5%), $F(1,10)=6.62$, $p<0.05$. The FM effect (15.8%) was also significant, $F(1,10)=10.77$, $p<0.05$. Likewise, the interaction between these effects was significant, with a FM effect being larger for the AO (23.1%) than for the AV (8.6%) conditions, $F(1,10)=4.92$, $p=0.057$. The analysis of the vowel height feature showed that the AV effect (1.1%) was not significant, $F<1$. There was a significant FM effect (9.8%), $F(1,10)=5.52$, $p<0.05$. There was no interaction between these effects, $F<1$. For the place feature, overall there was an AV effect, i.e., identification accuracy was higher by 13.3% in the AV compared to the AO condition, $F(1.9)=23.38$, $p<0.05$. There was also an FM effect, with better performance in the AM+FM condition (by 14.75%) than in the AM alone, $F(1,9)=14.37$, $p<0.05$. There was a significant interaction between these effects, with the FM effect being larger in the AO (21%) than the AV condition (8.5%), $F(1,9)=12.31$, $p<0.05$. Further analyses showed that there was a larger AV effect if the feature was a front vowel compared to a centre or back vowel, $F(1,9)=4.7$, $p=0.054$. This was not the case for the FM effect, where the size of the effect was the same for the front and centre/back vowels, $F(1,9)=2.5$, $p>0.05$.

The feature scores for consonants were also analyzed with a 3 (features: voicing, manner, place) × 2 (AO versus AV) × 2 (AM versus AM+FM) ANOVA. There was a significant AV effect, with more features correctly identified in the AV (84.7%) compared to the AO (73.1%) condition, $F(1,40)=13.08$, $p<0.05$. There was also a significant FM effect, with more features correctly identified in the AM+FM condition (82.7%) compared to the AM (75.1%), $F(1,40)=7.75$, $p<0.05$. There was no significant interaction between these effects, $F(1,40)=1.15$, $p>0.05$. The different consonant feature scores differed significantly from each other, $F(2,40)=3.83$, $p<0.05$. There was also a significant interaction between the feature variable and the AV effect, with a negative effect for voicing (–0.7%) a reasonable effect for manner (9.4%) and a large effect for place (26%), $F(2,40)=9.38$, $p>0.05$. The feature variable did not interact with the FM effect, $F(2,40)=1.19$, $p<0.05$.

As with the vowels, the percent correct scores for each consonant feature were analyzed in a series of ANOVAs. The analysis of voicing feature showed that there was no AV effect, $F<1$. There was a FM effect, with identification of the correct voicing feature 5.7% better in the AM+FM compared to the AM condition, $F(1,20)=4.25$, $p=0.05$. There was no interaction between the AV and FM effects, $F(1,20)=2.17$, $p>0.05$. For the place feature, there was an AV effect, with performance in the AV condition better than in the AO condition by 31.6%, $F(1,19)=29.8$, $p<0.05$. There was also a significant FM effect, with performance in the AM+FM condition 9.5% better than in the AM condition, $F(1,19)=4.86$, $p<0.05$. These effects did not interact, $F(1,19)=1.91$, $p>0.05$. Further analyses revealed that there was a larger AV effect for the front consonants than for the back ones, $F(1,19)=5.2$, $p<0.05$. The FM effect was not different for the front and back consonants, $F<1$. There was no three-way interaction between the front-back consonants and the AV and FM effects, $F(1,19)=1.42$, $p>0.05$. For the manner feature, the AV effect was almost significant, $F(1,20)=3.55$, $p=0.07$. This was also the case with the FM effect, $F(1,20)=3.59$, $p=0.07$. There was no interaction between the AV and FM effects, $F<1$.

In sum, for vowels there was a smaller FM effect for AV presentation compared to AO. This interaction occurred for individual features (lip-rounding and place). There was no such interaction for consonants. For vowels, there were no interactions between features and the AV or FM effects. For consonants, there was an interaction between the type of feature and the AV effect with the voicing feature showing no AV effect.

So far, the analysis has examined the amount of facilitation for each condition by simply comparing the correct feature scores; however, it is possible to take into account the redundancy of features by conducting a SINFA that in a se-

ries of iterations identifies then partials out the feature with the highest percentage of information transmitted. In this way, SINFA helps to eliminate the redundancy of specific features in their contribution to phoneme recognition. To reveal the type of information that was transmitted with different cues, perceptual confusion matrices for vowels and consonants were constructed and a set of SINFA was conducted. The SINFA procedure consists of a sequence of iterations. In the first iteration, the unconditional information transmission (IT) analysis is performed that specifies the relative information transfer results according to standard IT analysis (Miller and Nicely, 1955). In successive iterations, conditional IT analyses are performed across the remaining features with the most salient feature (i.e., with the highest percentage of information transmitted in the previous iteration) held constant in order to partial out its contribution. As with the above feature analysis, three features were used for the vowels [lip-rounding, vowel height, and place of articulation (backness)] and three features for the consonants (voicing, manner, and place of articulation), see Table I.

The results of the SINFA for the vowels are presented in Table II. Note that only monophthongs (for which feature categorization was more straightforward) were included so that the overall correct responses for vowels in the results of the SINFA are not the same, as shown in the Fig. 2. Note also that since participant's responses were not limited to the stimulus set, the confusion matrices (for both vowels and consonants) are not square. Any other phonemes that received fewer than 4% of responses to any single stimulus phoneme were collected into a "miss" column (such misses were very rare, overall less than 0.05%).

The matrix for the most reduced speech condition (AO AM) shows that confusions occurred across the range of vowels presented and that these were mitigated in the other presentation conditions. The SINFA results provide a formal quantification of these patterns (Table II). One of the main interests of the study is in the effect of FAME processing (adding FM cues) and how this might change across AO and AV presentations. For the AO condition, the saliency order was vowel height, place and rounding, and information transmission for the first two of these features increased with FM cues. One of the main findings for vowels has been the reduction in the FM effect for AV compared to AO presentation. This reduction was also apparent in the SINFA, especially for the place feature. That is, for AO presentation, adding FM increased the transmission of place information by 8.5% but for AV presentation there was no increase at all. The addition of FM cues resulted in the re-ordering of the feature transmission hierarchy (in the AV AM condition, vowel height was best transmitted whereas in the AV FM condition, lip-rounding was best). It is interesting to note that this latter ordering was the same as that found by Robert-Ribes *et al.* (1998) for AV presentation of unfiltered vowels (suggesting that in regard to information transmission, the adding of FM cues has the effect of making the signal closer to its unfiltered counterpart).

The analysis of transmitted vowel information suggests that the AO AM+FM and AV AM+FM conditions likely involved different processing despite their similar overall correct response rates. Comparison of the two conditions showed that they differed in the weight given to the features transmitted. The transmission of information about the place of vowel articulation was drastically reduced for AV presentation (the opposite was true for consonants, see below). This suggests that for vowels, the place cue was not visually distinctive, whereas lip rounding obviously was. Given that lip-rounding and place of vowel articulation tend to be correlated (front vowels more often unrounded; back vowels rounded) it could be that for AV presentation, lip-rounding served also as a cue for place (this would have effectively reduced the impact of the place feature in the SINFA).

The results of the SINFA for the consonants are presented in Table III. The SINFA results for consonant identi-

TABLE II. Results of SINFA for vowel identification. The "Feature" column shows the feature that was the most salient at each iteration. The "Transmission" column indicates the relative amount of information transmission contributing to the pattern of observed phoneme identification performance. The "Percentage" column shows the percentage of feature transmission as a total amount of transmitted information. Note that in this and subsequent tables, the first numerical column (Transmission) is the number of bits of information transmitted for the feature.

AV AM vowels

Total transmitted information=1.915

Correct responses=47.9%

Iteration	Feature	Transmission	Percentage
1	Height	0.629	32.9
2	Rounding	0.571	29.9
3	Place	0.000	0
Sum		1.200	62.7

AO AM vowels

Total transmitted information=1.619

Correct responses=40.2%

Iteration	Feature[a]	Transmission	Percentage
1	Height	0.590	36.5
2	Place	0.351	21.7
3	Rounding	0.033	2.0
Sum		0.974	60.2

AV AM+FM vowels

Total transmitted information=2.165

Correct responses=58.5%

Iteration	Feature	Transmission	Percentage
1	Rounding	0.816	37.7
2	Height	0.726	33.6
3	Place	0.013	0.6
Sum		1.556	71.9

AO AM+FM vowels

Total transmitted information=1.956

Correct responses=58.7%

Iteration	Feature	Transmission	Percentage
1	Height	0.773	39.5
2	Place	0.590	30.2
3	Rounding	0.015	0.8
Sum		1.377	70.4

[a]The features for Rounding (lip rounding) were Rounded, Unrounded; Height (Vowel height) features were Close, Mid & Open; Place (place of articulation or backness) features were Front, Central and Back.

TABLE III. Results of SINFA for consonant identification. The "Feature" column shows the feature that was the most salient at each iteration; the "Transmission" column indicates the relative amount of information transmission that contributed to the pattern of observed phoneme identification performance and the "Percentage" column shows the percentage of the feature transmission as a total amount of transmitted information.

AV AM consonants

Total transmitted information=3.113

Correct responses=65.9%

Iteration	Feature	Transmission	Percentage
1	Manner	1.566	50.3
2	Place	1.086	34.9
3	Voicing	0.127	4.1
Sum		2.779	89.3

AO AM consonants

Total transmitted information=2.457

Correct responses=41.5%

Iteration	Feature	Transmission	Percentage
1	Manner	1.203	49.0
2	Place	0.582	23.7
3	Voicing	0.149	6.0
Sum		1.934	78.7

AV AM+FM consonants

Total transmitted information=3.552

Correct responses=76.7%

Iteration	Feature	Transmission	Percentage
1	Manner	1.818	51.2
2	Place	1.236	34.8
3	Voicing	0.215	6.0
Sum		3.268	92.0

AO AM+FM consonants

Total transmitted information=2.680

Correct responses=50.5%

Iteration	Feature	Transmission	Percentage
1	Manner	1.347	50.3
2	Place	0.577	21.5
3	Voicing	0.190	7.1
Sum		2.113	78.9

fication showed that the order of the salient features of information transmitted was constant across the testing conditions: manner of articulation, then place of articulation, and voicing. Beginning with the AO condition, the addition of FM cues affected the total transmission of information but hardly affected the relative makeup of feature information transmitted (manner and voice increased slightly).

With the addition of visual cues to the AO signal, there was a dramatic increase in the transmission of the place information (approximately 65% more). The further addition of FM cues once again increased the total transmission of information, but little to alter the proportions of feature transmission. There was a hint of an interaction between the effects of the auditory and visual cues on the transmission of voicing information, the effect of the visual cues was slightly detrimental but this tendency disappeared when visual and FM cues were presented together.

Overall, the SINFA results indicated that the FM and visual cues appeared to perform different functions in the process of consonant and vowel perception. For consonant perception, the FM cues acted to slightly boost the transmission of manner information (and voicing) whereas the provision of visual cues dramatically improved the transmission of place of articulation information. Since these cues were complementary, their effects were additive. The vowel perception results suggested that the FM cues enhanced place of articulation and vowel height information, whereas visual cues acted to enhance the transmission of lip-rounding information. Even though FAME processing provided more information about place of articulation, there was no increase in the percentage of correct responses when this cue was presented in the AV condition (when compared to the AO AM +FM one). We suggest that this was because visual speech provided lip-rounding information and through this also place of articulation information. Given this overlap in cue information and the prominence of visual cues, any additional place information gained via FM cues may have been made redundant by visual information and so no additional benefit was achieved for the cue combination.

Having found that the FM and AV effects do not combine in an additive way, at least for vowels, begs the question of why the FM and AV effects were additive for the sentence stimuli of Experiment 1. This problem is similar to the one faced by Grant and Walden (1996) in reconciling their result of disproportionate AV benefit scores for aCa syllables observed for low frequency bands with that of Grant *et al.* (1991) who found approximately equal AV intelligibility for sentences across filter bands. Grant and Walden (1996) proposed that the most likely explanation for the discrepancy lies in the many differences between nonsense syllable and meaningful sentence identification. For instance, meaningful words can be recognized with less than complete phonemic information due to lexical redundancy and sentence context. Such supporting context would most likely have obscured the non-additive effect found for vowels, particularly since such was not evident for consonants.

IV. CONCLUSION

Early work investigating the type and amount of information that can support speech recognition concentrated on the issue of how reduced auditory information (AM envelope) supplemented speechreading. It was found that temporal (AM) envelope cues from several speech bands combine better with visual speech than cues from a single wide-band AM envelope. Subsequent studies showed that high levels of speech recognition were supported by amplitude envelope information from multiple frequency bands without the need for visual speech. More recent work has shown that amplitude enveloped derived coding is susceptible to noise. The FAME processing strategy uses slow varying FM information and offers more noise resistance than AM coding only. Just as earlier work examined how AM coding combined with visual speech, the current study examined how FAME encoding combines with visual speech. Experiment 1 demonstrated that for sentence recognition, the size of the FAME

processing advantage for AO presentation was the same as that found with AV presentation. This was a somewhat surprising result given that FAME processing and visual speech both can act as a means of segregating target speech from the background. It was concluded that because AM, FM, and visual speech provide numerous speech cues, more controlled stimuli and a detailed examination specific phonetic features were needed to determine the degree of redundancy between these sources.

Experiment 2 examined the interaction of FM and visual speech cues in finer detail by examining consonant and vowel identification and how these cues affected phonetic features. These cues had a differential impact on consonants and vowels. The facilitation from visual cues was greater for consonants than for vowels, whereas the facilitation from FM cues was greater for vowels than that for consonants. For consonants, analysis of information transmission showed that visual cues had a robust facilitatory effect on the transmission of place of articulation information. For vowels, visual speech had the largest influence on lip-rounding. For consonants, FAME processing and visual speech cues produced additive effects, but this was not that case for vowels where the FM effect was smaller for AV presentation. This latter result indicates that it cannot necessarily be assumed that the addition of visual speech will provide a fixed facilitatory effect for an auditory processing scheme. It should be noted, however, that the simple notion of defining redundancy and complementarity in terms of degree of AV benefit, although straightforward, provides only a first cut at specifying the affect of different sources of information on performance. Clearly, more complete models of AV integration, (e.g., Massaro, 1987; Braida, 1991) should provide a more formal quantification of the interaction of information sources.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Fang-Gang Zeng for use of the FAME signal processing algorithm. They also thank Darren Cullinane and Nicole Lees for help in preparing the test stimuli and Erin Cvejec for help in data collection. The first and second authors acknowledge the support of the Australian Research Council Grant Nos. DP0666857 and TS0669874.

[1]We caution that the sample size (31 trials each for consonants and vowels) might be considered to be relatively small for the purpose of information analysis. The SINFA was conducted using the Analysis Suite "FIX" (Feature Information Xfer, Mike Johnson, Department of Phonetics and Linguistics, University College London).

Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments," Q. J. Exp. Psychol. **43**, 647–677.
Breeuwer, M., and Plomp, R. (**1984**). "Speechreading supplemented with frequency-selective sound-pressure information," J. Acoust. Soc. Am. **76**, 686–691.
Breeuwer, M., and Plomp, R. (**1986**). "Speechreading supplemented with auditorily presented speech parameters," J. Acoust. Soc. Am. **79**, 481–499.
Bregman, A. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).
Desai, S., Stickney, G., and Zeng, F. G. (**2008**). "Auditory-visual speech perception in normal-hearing and cochlear-implant listeners," J. Acoust. Soc. Am. **123**, 428–440.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.
Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.
Driver, J. (**1996**). "Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading," Nature (London) **381**, 66–68.
Erber, N. P. (**1972**). "Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing," J. Speech Hear. Res. **15**, 413–22.
Forster, K. I., and Forster, J. C. (**2003**). "DMDX: A windows display program with millisecond accuracy," Behav. Res. Methods Instrum. Comput. **35**, 116–124.
Fu, Q. J., and Shannon, R. V. (**1999**). "Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant," Ear Hear. **20**, 332–344.
Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.
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**). "The spectral distribution of prosodic information," J. Speech Hear. Res. **39**, 228–238.
Grant, K. W., Braida, L. D., and Renn, R. J. (**1991**). "Single band amplitude envelope cues as an aid to speechreading," Q. J. Exp. Psychol. A **43**, 621–645.
Grant, K. W., Braida, L. D., and Renn, R. J. (**1994**). "Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech," J. Acoust. Soc. Am. **95**, 1065–1073.
Grant, K. W., Ardell, L. H., Kuhl, P. K., and Sparks, D. W. (**1985**). "The contribution of fundamental frequency, amplitude envelope, and voicing duration cues to speechreading in normal-hearing subjects," J. Acoust. Soc. Am. **77**, 671–677.
Harrington, J., and Cassidy, S. (**1999**). *Techniques in Speech Acoustics* (Kluwer Academic, The Netherlands).
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.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Kim, J., and Davis, C. (**2003**). "Hearing foreign voices: Does knowing what is said affect masked visual speech detection?" Perception **32**, 111–120.
Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Lawrence Erlbaum, Hillsdale, NJ).
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Morton, K. D., Torrione, P. A., Throckmorton, C. S., and Collius, L. M. (**2008**). "Mandarin Chinese tone identification in cochlear implants: Predictions from acoustic models," Hear. Res. **244**, 66–76.
Nelson, P., Jin, S.-H., Carney, A., and Nelson, D. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normalhearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Nie, K., Stickney, G., and Zeng, F. G. (**2005**). "Encoding frequency modulation to improve cochlea implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.
Nie, K., Barco, A., and Zeng, F. G. (**2006**). "Spectral and temporal cues in cochlear implant speech perception," Ear Hear. **27**, 208–217.
Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.
Risberg, A. (**1974**). "The importance of prosodic speech elements for the lipreader," Scand. Audiol. Suppl. **4**, 253–264.
Rivet, B., Girin, L., and Jutten, C. (**2007**). "Visual voice activity detection as a help for speech source separation from convolutive mixtures," Speech Commun. **49**, 667–677.
Robert-Ribes, J., Schwartz, J. L., Lallouache, T., and Escudier, P. (**1998**). "Complementarity and synergy in bimodal speech: Auditory, visual and audio-visual identification of French oral vowels in noise," J. Acoust. Soc. Am. **103**, 3677–3689.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., and Barone,

P. (**2007**). "From the cover: Evidence that cochlear-implanted deaf patients are better multisensory integrators," Proc. Natl. Acad. Sci. U.S.A. **104**, 7295–7300.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.
Sodoyer, D., Girin, L., Jutten, C., and Schwartz, J. L. (**2004**). "Developing an audio-visual speech source separation algorithm," Speech Commun. **44**, 113–125.
Sommers, M. S., Kirk, K. I., and Pisoni, D. B. (**1997**). "Some considerations in evaluating spoken word recognition by normal hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format," Ear Hear. **18**, 89–99.
Stickney, G. S., Nie, K., and Zeng, F. G. (**2005**). "Contribution of frequency modulation to speech recognition in noise," J. Acoust. Soc. Am. **118**, 2412–2420.
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.
van der Horst, R., Leeuw, A. R., and Dreschler, W. A. (**1999**). "Importance of temporal-envelope cues in consonant recognition," J. Acoust. Soc. Am. **105**, 1801–1809.
Van Tasell, D., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152–1161.
Walden, B. E., and Montgomery, A. A. (**1975**). "Dimensions of consonant perception in normal and hearing-impaired listeners," J. Speech Hear. Res. **18**, 444–455.
Walden, B. E., Prosek, R. A., and Worthington, D. W. (**1974**). "Predicting audiovisual consonant recognition performance of hearing-impaired adults," J. Speech Hear. Res. **17**, 270–278.
Walden, B. E., Montgomery, A. A., Prosek, R. A., and Hawkins, D. B. (**1990**). "Visual biasing of normal and impaired auditory speech perception," J. Speech Hear. Res. **33**, 163–173.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.
Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
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.

Effects of electrode separation between speech and noise signals on consonant identification in cochlear implants[a)]

Bom Jun Kwon[b)]
Department of Communication Sciences and Disorders, University of Utah, 390 S 1530 E, Salt Lake City, Utah 84112

(Received 18 October 2007; revised 31 August 2009; accepted 25 September 2009)

The aim of the present study was to examine cochlear implant (CI) users' perceptual segregation of speech from background noise with differing degrees of electrode separation between speech and noise. Eleven users of the nucleus CI system were tested on consonant identification using an experimental processing scheme called "multi-stream processing" in which speech and noise stimuli were processed separately and interleaved. Speech was presented to either ten (every other electrode) or six electrodes (every fourth electrode). Noise was routed to either the same (the "overlapped" condition) or a different set of electrodes (the "interlaced" condition), where speech and noise electrodes were separated by one- and two-electrode spacings for ten- and six-electrode presentations, respectively. Results indicated a small but significant improvement in consonant recognition (5%–10%) in the interlaced condition with a two-electrode spacing (approximately 1.1 mm) in two subjects. It appears that the results were influenced by peripheral channel interactions, partially accounting for individual variability. Although the overall effect was small and observed from a small number of subjects, the present study demonstrated that CI users' performance on segregating the target from the background might be improved if these sounds were presented with sufficient peripheral separation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3257200]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.71.Es, 43.71.Gv [KWG] Pages: 3258–3267

I. INTRODUCTION

Everyday sounds typically consist of several spectral components in a wide frequency range. Grouping these components is one of the first steps necessary to capture the meaning of an auditory stimulus. Two primary factors in auditory grouping are the harmonic relations and the resolvability of components by the auditory filters (e.g., Moore *et al.*, 1986) and the temporal arrangement of components (such as synchronous onset; Bregman, 1990). Grouping is particularly important when multiple sound stimuli exist; in the context of speech recognition, decoding the fundamental frequency (F0) of each sound and/or tracking the contour of F0 facilitates segregation of one from the other and improves speech recognition in background sounds (Assmann and Summerfield, 1990; Brokx and Nooteboom, 1982). In addition, amplitude fluctuations of the background also aid speech recognition through masking release (i.e., "listening-in-the-valley;" e.g., Bacon *et al.*, 1998). These mechanisms of grouping/segregation in the context of *auditory scene analysis* (Bregman, 1990) partially account for the robust speech understanding in noisy environments enjoyed by normal hearing (NH) listeners compared to the less impressive results obtained from automatic speech recognition by machine based on pattern-matching algorithms (Allen, 1994).

Speech understanding by individuals implanted with cochlear implants (CIs) is often comparable to that of NH listeners in quiet listening environments. However, CI users' performance sharply drops compared to that of NH listeners if there is a noise or a competing background masker (Skinner *et al.*, 1994; Battmer *et al.*, 1997; Dorman *et al.*, 1998a; Fu *et al.*, 1998; Fetterman and Domico, 2002). For over a decade there have been a myriad of studies of speech recognition in noise by CI users or by NH listeners through CI simulations under a variety of conditions (Nelson *et al.*, 2003; Stickney *et al.*, 2004, 2007; Fu and Nogaki, 2005; Cullington and Zeng, 2008). The findings in the literature can be summarized as follows. First, compared to NH listeners, CI users show a reduced ability to segregate the target voice from a background voice based on F0 differences such as voice gender difference (Stickney *et al.*, 2004, 2007). Second, CI users may not experience or utilize masking release; in contrast, the opposite effect, modulation interference, can sometimes occur (i.e., poorer speech recognition in a modulated background; Nelson *et al.*, 2003; Stickney *et al.*, 2004). In short, the ability of CI users to segregate the target voice from the background based on the cues known in NH listening has not been demonstrated. However, several observations suggest the possibility that CI users may still use auditory grouping or segregation to process sounds. Fu and Galvin (2001) showed similar recognition of spectrally desynchronized sentences by CI users and NH listeners, implying that CI and NH listeners might utilize common temporal grouping mechanisms. In addition, recent studies have suggested that CI users may use stream segregation to process

[a)] A portion of the pilot data for this study was presented in poster sessions at the 149th Meeting of the Acoustical Society of America, Vancouver, Canada, May, 2005 and at the Conference on Implantable Auditory Prostheses, Pacific Grove, CA, July, 2005.

[b)] Author to whom correspondence should be addressed. Present address: Department of Otolaryngology-Head and Neck Surgery, Eye and Ear Institute, Ohio State University, 915 Olentangy River Rd., Columbus, OH 43221. Electronic mail: bjkwon@gmail.com

non-speech stimuli (Chatterjee *et al.*, 2006; Hong and Turner, 2006; Carlyon *et al.*, 2007). Interestingly, in these studies the separation between stimulating electrodes influenced the results in a way that suggests that stream segregation in CI users, like in NH listeners, appears to be affected by peripheral channeling (Hartmann and Johnson, 1991; Beauvois and Meddis, 1996), although the mechanisms underlying these observations may still be a matter of debate (see Cooper and Roberts, 2007).

If CI users can indeed perform auditory grouping or segregation, the underlying cause behind CI users' poor understanding of speech in noise deserves further attention. To this end, it is worth noting that, in commonplace CI settings multiple signals where are acoustically mixed and processed by a set of conventional clinical parameters, the issues of auditory grouping/segregation are not directly addressed; in other words, potentially important attributes of CI users' abilities are not revealed due to the limitations in conventional signal processing. For example, while it is reasonably well accepted that CI users are unable to segregate a male target voice from a female voice background (Stickney *et al.*, 2004, 2007), this effect might be due to the lack of resolvability in the filterbank of the CI device rather than an inability to segregate per se. The filter slopes in the filterbank used for CI speech processing are generally shallow so that when the input is a pure tone, actual stimulation takes place on multiple channels. For example, as can be easily seen in clinical programming equipment and/or software in the nucleus system, three or four electrodes are activated even for a pure tone input. Thus, CI users are not provided with the signal with a sufficient resolvability. In addition to the issue of filter resolution, the degradation of signal quality due to only limited spectral channels imposes a further challenge to CI users. Speech is represented in CIs by modulated pulse trains delivered to a small number of electrodes; fine structure cues are not available in each channel. In acoustic hearing, when noise is added to speech, the auditory periphery is presented with the mixture of the two signals; it is the task of the auditory system to parse the mixture and segregate one from the other with the available cues mentioned above. In the electric representation of speech in CIs, however, the consequence of noise is not just an addition of a distracting or masking signal, but rather a severe distortion or alteration in the pattern of speech pulses. This distortion involves a high compression in loudness mapping between the acoustic input and its electric output, which is necessary to compensate for a substantially reduced electric dynamic range (Nelson *et al.*, 1996). Such distortion is usually nonlinear and irrecoverable, yet another reason noise negatively impacts speech intelligibility more through a CI system than in acoustic hearing. In summary, due to the technical limitations in clinical CI systems, examination of speech understanding by CI users in competing backgrounds through conventional processing mechanisms does not provide sufficient insight with regard to auditory grouping or segregation in electric hearing.

Given this background, the present study aimed to examine CI users' ability to understand speech in the noise background with a setting where both signals are presented in their entirety without the limitations imposed in conventional clinical processing such as resolvability in the filterbank or non-linear compression. In order to achieve this, as shown in Fig. 1, the speech and noise signals were processed separately. Then two separate pulse trains, the output of processing, could be simply interleaved in time, i.e., the speech and noise signals would be presented as two interleaved signals at the periphery (the electrode-nerve interface) rather than one distorted signal as in conventional processing. In the present study, this experimental method is referred to as "multi-stream processing" (Kwon and van den Honert, 2006). Admittedly this method might have only limited practical implications with the current device architecture, as the presumption that each of the multiple input signals is available separately for parallel processing is often violated in real applications. However, at least for a laboratory testing, the signal presentation through multi-stream processing provides an advantage when studying CI users' ability to organize multiple inputs (e.g., Deeks and Carlyon, 2004). The effects of peripheral separation between speech and noise electrodes were of particular interest in the present study and examined by comparing the following two-electrode assignment layouts: "overlapped" and "interlaced," akin to the "same" and "different" conditions in Deeks and Carlyon (2004). In the overlapped layout the speech and noise pulses were presented on the same set of electrodes, whereas in the interlaced layout they were presented on different sets of electrodes. In both layouts, speech and noise pulses were interleaved in time, and the only difference was how the electrodes were assigned. Therefore, if speech understanding improves in the interlaced layout compared to the overlapped layout, it would indicate that CI users would be potentially capable of performing segregation of multiple signals (such as segregating voices with F0 cues) if those signals were separately delivered to the auditory periphery. If not, it would mean that the peripheral manipulation of the separate presentation of speech and noise is not capitalized upon by CI users, implying that any applications built upon an expectation of improved segregation via separately controlled use of electrodes would be only minimal. The present study is

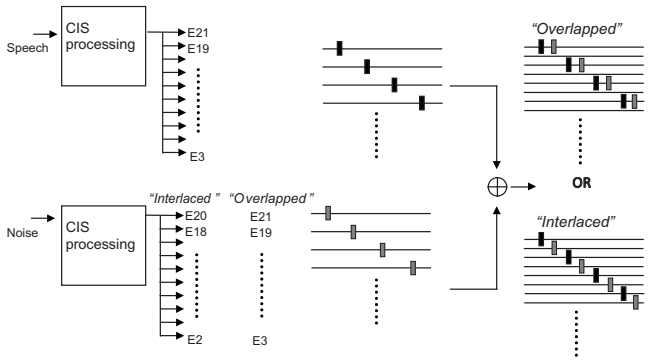


FIG. 1. A block diagram of multi-stream processing with an example of ten-channel CIS processing. Two CIS blocks, one for speech and the other for noise, are processed in parallel and interleaved. For the "interlaced" condition, the speech and noise sequences are delivered on different channel sets (odd-even). For the "overlapped" condition, speech and noise are delivered on the same channel sets (odd-odd). Multi-stream processing allows each of the speech and noise signals to be delivered entirely at the auditory periphery without the limitations imposed by signal processing in conventional CI systems.

TABLE I. Subjects' demographic information, history, devices used, CNC word recognition scores in quiet, and clinical MAP parameters. Subjects S1–S6 were tested with the ten-channel CIS, whereas S7–S11 were tested with the six-channel CIS.

ID	Gender	Age at testing (years)	CI experience at testing (years)	Device implanted (ST: straight electrode array; C: contour array)	Etiology	CNC word recognition score (%)	Channel stimulation rate in clinical MAP (Hz)
S1	M	72	6	CI24M-ST	Noise exposure	46	900
S2	M	55	3	CI24R-ST	Otosclerosis	63	900
S3	M	74	4	CI24M-ST	Familial	54	250
S4	M	60	1	CI24R-C	Hereditary	59	1200
S5	M	69	1	CI24RE-C	Noise exposure	34	900
S6	F	60	1	CI24RE-C	Unknown	42	720
S7	M	76	4	CI24R-C	Unknown/progressive	35	1200
S8	F	49	3	CI24R-C	Unknown	38	250
S9	F	81	3	CI24R-C	Ototoxic drug	75	900
S10	F	52	1	CI24RE-C	Ototoxic drug	52	720
S11	F	56	1	CI24RE-C	Hereditary	45	1200

not a feasibility test of a strategy based on multi-stream processing, as it does not address any issue of source separation algorithm (e.g., Kokkinakis and Loizou, 2008) that should be in place in real implementation. Yet, the findings of the present study might provide us the insight of whether such algorithm would be worthwhile to pursue.

II. METHODS AND PROCEDURE

A. Background in CI signal processing and terminology

The continuous interleaved sampling (CIS) strategy (Wilson *et al.*, 1991) consists of the following steps: (i) a spectral analysis of the input signal is done by a series of band-pass filters and the signal envelope (or the momentary fluctuation of spectral energy) in each band is obtained, (ii) a pulse train at a constant repetition rate is modulated by the envelope in each band (with a nonlinear loudness-mapping function), (iii) the pulse trains of all bands are temporally interleaved to avoid simultaneous current field summation, and (iv) the resulting pulse train is delivered to pre-allocated electrodes with the same number of processing bands. While the CIS strategy is an option that is available from all three manufacturers of widely used CI systems, the strategy is implemented in a slightly different way in each brand, emphasizing different aspects of processing due to unique hardware/electrode features.

Before proceeding to the description of the details of signal processing for the present study, some definition of terminologies would be beneficial to avoid confusion. First, the term "channel" in the present paper is mostly used to refer to the processing band of acoustic input, i.e., the ordinal division of discrete frequency regions, or *acoustic* channel. This is in contrast to *electric* channel, another usage of the same word channel in the literature, which refers to the pathway of the current flow (i.e., electric channel) created by the chosen stimulation configuration in the CI system (e.g., monopolar and bipolar), as established by the ordinal division of electrodes. Electric channel is referred to as "electrode" in the present paper, which is proper because in monopolar stimulation each electric channel is implemented by one electrode. This clarification is particularly important in the present paper because the information processed in each (acoustic) channel is to be delivered to a specifically defined set of electrodes. Whereas the concept of electric channel is directly bound with the use of electrodes, the binding is not as strong for an acoustic channel: for example, the lowest frequency band is not necessarily mapped with the apicalmost electrode for various reasons in clinical applications, and relatively unique manipulation for the frequency-to-electrode binding is attempted in the present study. Second, a MAP is a clinical term (note that it is not an acronym) that describes a complete set of parameters for signal processing including the information on not only the basic pulse parameters such as threshold, stimulation rate, and pulsewidth but also the usage of channels such as the total number of acoustic processing bands, the (de)selection of electrodes, and the loudness-mapping function. Third, the total stimulation rate of a MAP is defined by the stimulation rate per channel multiplied by the number of channels delivered to represent the signal in each analysis time window.

B. Subjects

A total of 11 adult CI users participated in the present study. They had been implanted with the Nucleus® CI24M, CI24R, or CI24RE (known as "freedom" implant) device at least 1 year prior to entry into this study. The study was conducted at the laboratories of Cochlear Americas (Englewood, CO) and at the University of Utah (Salt Lake City, UT). The use of human subjects in this study was reviewed and approved by the Western Institutional Review Board and the University of Utah IRB, respectively. Table I provides subject-specific information, including patient history, the implanted device, the channel stimulation rate used in each subject's clinical MAP, and the consonant-nucleus-consonant (CNC) monosyllabic word recognition scores in quiet (Peterson and Lehiste, 1962) measured with the clinical MAP. Six subjects were tested in the ten-channel CIS condition and subsequently five subjects were tested in the six-channel CIS condition (see Sec. II C).

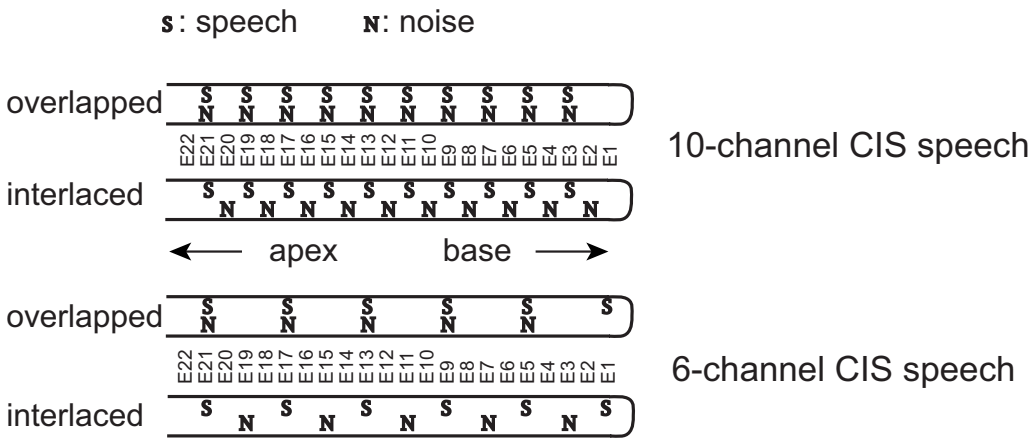


FIG. 2. Electrode layout for the "overlapped" and "interlaced" conditions for each CIS processing condition.

C. Signal processing and experimental conditions

Throughout the present study, many processing parameters were chosen from the default settings in clinical applications. For example, the monopolar stimulation configuration was used, the phasewidth was 25 μs, and the interphase gap was 8 μs. Experimental conditions were designed to create varying degrees of speech-noise separation on electrodes. Given 22 intra-cochlear electrodes in the nucleus system, the following two conditions of speech presentation were chosen: ten-channel CIS (using every other electrode: E21, E19, E17, E15, E13, E11, E9, E7, E5, and E3) and six-channel CIS (using every fourth electrode: E21, E17, E13, E9, E5, and E1). Noise was presented either on the same electrodes as speech in the overlapped condition or different electrodes in the interlaced condition. Therefore, in the interlaced condition, speech and noise electrodes were separated by one-electrode spacing in the ten-channel CIS or by two-electrode spacing in the six-channel CIS, as shown in Fig. 2. It was not possible to use six electrodes for noise in the interlaced condition with a two-electrode spacing margin (note that the number of electrodes is 22, not 24). Therefore, for a fair comparison between the overlapped and interlaced conditions, only five electrodes were used for noise; these electrodes were either E21, E17, E13, E9, and E5 (for the overlapped condition) or E19, E15, E11, E7, and E3 (for the interlaced condition). Note that this configuration might slightly favor the overlapped condition because if the channel spread were wider than two-electrode spacing (but not as wide as four-electrode spacing), the speech information from E1 would be masked in the interlaced, but not in the overlapped condition. Therefore, if speech understanding were substantially better in the interlaced condition, this comparison could serve as evidence, suggesting CI users' capability of perceptual segregation.

In the present study the total stimulation rate was limited to 14 400 Hz due to the hardware/software restrictions in the nucleus system. For example, if the channel stimulation rate was 900 Hz, then up to 16 channels could be used for a MAP ($900 \times n \leq 14\,400$ for $n \leq 16$). Ideally, to maintain the usual speech intelligibility and quality, the same stimulation rate as in the clinical MAP should have been used in all conditions. Unfortunately, this was not possible in some subjects for whom the total stimulation rate in their clinical MAP exceeded 7200 Hz (S4, S7, and S10), one-half of the limit, because the total bandwidth, 14 400 Hz, was shared by the interleaved speech and noise streams. For these subjects, all of whom were tested with ten-channel CIS, a channel stimulation rate of 720 Hz was chosen to make the total stimulation rate of 7200 Hz. Although those subjects noticed some speech quality difference with this rate, subjective speech intelligibility was not particularly degraded. Table II displays the frequency boundaries for processing bands. Frequency boundaries of noise processing in the six-channel CIS condition were different (not shown) from those of speech processing because noise was actually processed with the five-channel CIS, as described above. It is unlikely that the use of different acoustic frequency boundaries for noise in this condition undermined the rationale of the present study because the interaction between the stimuli in the acoustic domain was intentionally avoided and the aim was to compare two conditions manipulating the stimulating electrodes of speech and noise pulses (overlapped or interlaced).

D. Apparatus

Custom-made software written in C++ and MATLAB was used to process the speech material, generate the stimuli, and administer the testing. The software was developed with the NICstream library (Cochlear Ltd., Sydney, Australia), which

TABLE II. Nominal frequency boundaries for the filterbank used for the ten-channel and six-channel CISs.

	Frequency boundaries (Hz)										
Ten-channel	188	438	688	1063	1438	1938	2563	3438	4563	6063	7938
Six-channel	188	563	1063	1813	2938	4813	7938				

included signal processing routines used for the nucleus implant system in the form of MATLAB function files ("m" files) and a series of dynamic link library (DLL) files for communication between a personal computer (PC) and the implant device. The default m files were modified and restructured to implement the multi-stream processing scheme for the present study and were converted to a DLL through MATLAB COMPILER (version 4.7); this DLL was used as the main software that administered the experiment. The pulse trains, generated from speech and noise wave files by the PC software, were transmitted to the subject's internal device transcutaneously through a radio-frequency (rf) signal via one of the following two interfaces: (1) the nucleus clinical programming system with the SPrint speech processor used for a desktop PC or (2) a USB-based POD interface with the L34 speech processor used for a laptop. In both cases SPrint and L34 were used as relays that generated a rf signal from the PC rather than actual processors; neither the microphone connected to the implant nor the processing capability of the implanted processor was used. The speech input level was adjusted at the front-end stage so that the digitized waveform samples would be emulating an acoustic input with 70 dB sound pressure level rms with the sensitivity setting at 8 to ensure sufficient loudness in all conditions. The subject was given a separate PC for making responses and receiving feedback. This PC was connected via the TCP/IP protocol to the operator's PC, which performed signal processing and test administration.

E. Stimuli and procedure

Recognition of the following consonants was measured: b, d, g, p, t, k, z, s, m, n, f, v, θ, $\eth$, $\mathcal{Z}$, and $\int$. The material was recorded in the /aCa/ context by four talkers—two males and two females—and a total of 64 (i.e., 16×4) sound samples (Turner *et al.*, 1995) was used. Each testing session consisted of 128 presentations (two presentations of each sound sample) in a pseudo-random order. The subject was given 16 choices of consonants on a computer screen after each presentation and was asked to respond by a mouse or a finger touch on a touch screen. The average percent-correct score was obtained for each subject after repeating the session at least four times. Feedback was given only prior to data collection during the practice session. Speech-shaped noise was generated by the following method. The long-term spectrum of the aggregate sound file (concatenation of the 64 sound files) was analyzed in third-octave bands in the range 100–8064 Hz. A narrowband noise was generated by filtering a white noise with a bandpass filter for each of the third-octave bands in this range, and the level of each narrowband noise was adjusted to match the relative magnitude of the long-term spectrum of speech materials. The narrowband noise in each band was then summed to generate the speech-shaped noise. Finally, the spectrum of this noise was compared again with the long-term spectrum of speech sounds to ensure that a peculiar summation of noise at transition bands did not occur.

For the ten-channel CIS, noise was combined with speech so that the onset of the noise signal preceded that of the speech signal by a lead time randomly determined in each presentation by a uniform distribution between 0 and 700 ms. For the six-channel CIS, the lead time of the noise signal was fixed at 1 s. The noise level, or the signal-to-noise ratio (SNR), was set such that the identification score reasonably dropped relative to the quiet condition.[1] As the effect of noise on speech understanding is usually variable across subjects, the SNR was selected between 5 and 20 dB for each subject, instead of testing at a fixed SNR. Six subjects were tested in the ten-channel CIS condition and five different subjects were tested in the six-channel CIS condition. In both conditions, the order of overlapped versus interlaced testing sessions was randomized. There was an additional test with two subjects using three different levels of noise after the initial data collection (the rationale and procedure are described in Sec. III B).

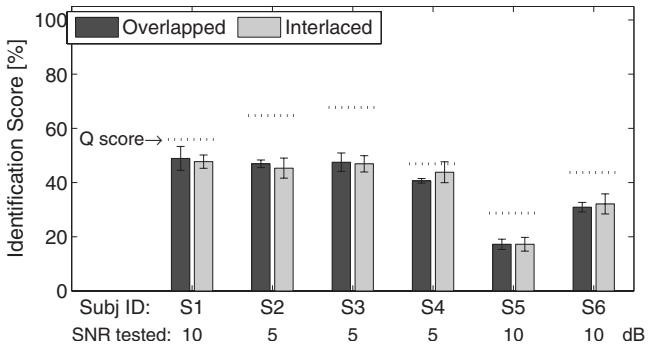


FIG. 3. The percent consonant identification scores of each subject for the "overlapped" (dark bars) and "interlaced" (light bars) conditions in ten-channel CIS. The interlaced condition represented the separation between the speech and noise electrodes by one-electrode spacing. The tested SNR is indicated for each subject. The scores marked with "Q" indicate the scores in quiet. Error bars indicate one standard deviation.

III. RESULTS

A. Effect of electrode separation

The results of consonant identification in the ten- and six-channel CIS conditions are plotted in Figs. 3 and 4, respectively. The percent-correct identification scores of the overlapped (dark bars) and interlaced (light bars) conditions

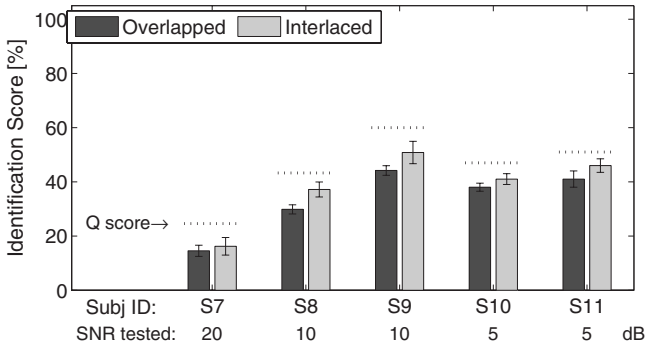


FIG. 4. The percent consonant identification scores of each subject for the "overlapped" (dark bars) and "interlaced" (light bars) conditions in six-channel CIS. The interlaced condition represented the separation between the speech and noise electrodes by two-electrode spacing. The tested SNR is indicated for each subject. The scores marked with "Q" indicate the scores in quiet. Error bars indicate one standard deviation.

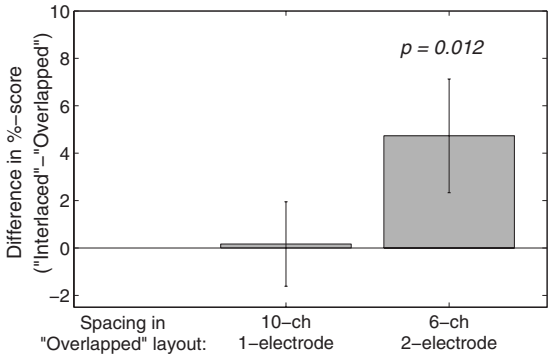


FIG. 5. The mean difference in scores between the "overlapped" and "interlaced" conditions for each electrode separation. Error bars indicate one standard deviation.

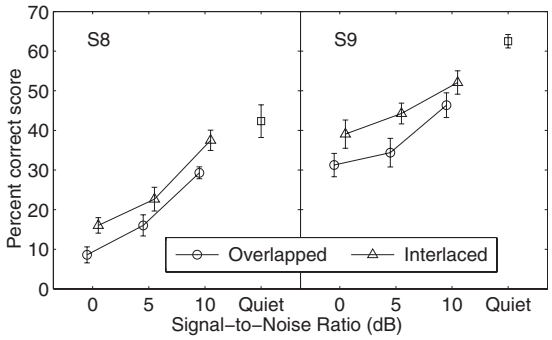


FIG. 6. Additional consonant recognition test for two subjects (S8 and S9) at three SNR levels. Error bars indicate one standard deviation.

are plotted for each subject. Percent-correct scores in a quiet background with the ten-channel CIS ranged from 29% to 68% (mean: 49.2%), and with the six-channel CIS were from 25% to 60% (mean: 45.2%). The observed decrease in the quiet score is generally consistent with previous studies (e.g., Friesen *et al.*, 2001). The effect of noise was variable across subjects: while some subjects showed a relatively higher resilience to noise at 5 dB SNR (e.g., S4), other subjects suffered greatly from an even smaller amount of noise (e.g., S7 at 20 dB SNR). For the ten-channel CIS, the difference in consonant identification scores was negligible between the overlapped and interlaced conditions (0.1% on average; cf. Fig. 5). In this case, the speech and noise signals were presented with a spacing of one electrode in the interlaced condition. In contrast, the difference was greater between the two conditions in the six-channel CIS, in which speech and noise were separated by a two-electrode spacing in the interlaced condition. Using a mixed-ANOVA model with one within-subject factor, the electrode layout (overlapped or interlaced), and two between-subject factors (the two CIS conditions and SNR), a significant interaction was found between the CIS conditions and electrode layout ($p=0.003$), suggesting that the two CIS conditions be examined separately. The effect of SNR was not significant ($p=0.055$). As displayed in Fig. 5, the average difference between the overlapped and interlaced conditions for six-channel CIS, after collapsing across SNR conditions, was 4.7%: this was statistically significant in a paired-t test ($p=0.012$). Thus, the subjects tested with the six-channel CIS speech (S7–S11) on average performed better in the interlaced condition than in the overlapped condition, indicating improved perceptual segregation with the separated presentation. However, due to the large inter-subject variability, the small number of subjects for this analysis ($N=5$), and the small overall effect (4.7%), caution should be exercised when interpreting this statistical analysis. While the statistical trend of an effect of segregation is revealed in the observed group, it is difficult to judge the validity of the effect on an individual basis (except for S7 showing almost no effect). The statistical analysis still suggests that the effects of electrode separation and perceptual segregation be considered in further analyses.

B. Additional experiment

Given that a large inter-subject variability with regard to the observed effect exists, as seen in Fig. 4, which is not unusual in CI studies, perhaps it would be sensible to examine the effect in each individual rather than trying to establish a universal trend. The effect of electrode separation appeared to be almost non-existent in S7, for example, who showed the lowest score without noise and suffered from noise in general most severely (note that the SNR tested was 20 dB). While it is unclear whether the lack of effect was due to the compromise in the peripheral nerve system (such as uneven or simply poor neural survival) or in a more central mechanism in this subject, it is apparent that any potential effect was likely overshadowed by the factor(s) causing poor general performance. Therefore, it would be appropriate to reason that the effect of peripheral separation in the context of speech understanding in noise might not be observed in all CI users. For this reason, instead of trying to demonstrate the effect in a larger number of subjects, it was worthwhile to strengthen the statistical power of the within-subject effect by further testing those who had shown a strong effect of separation. Particularly it was necessary to test whether the effect shown in the main experiment was due to the particular choice of SNR for each subject. Therefore, two subjects (S8 and S9) were called back for post-testing sessions 7 or 8 months after the main experiment. The same speech material in the main experiment was used but only two relatively clear voices were chosen (one male and one female). There was a total of six listening conditions (three SNRs and two layouts: overlapped or interlaced). In order to prevent an order effect from influencing the results, all six conditions were mixed into one session consisting of 192 stimuli (16 stimuli for each of the two voices times 6 conditions). This session was repeated three times to obtain the final results.

Results are displayed in Fig. 6 (triangles and circles indicate the interlaced and overlapped conditions, respectively). In both subjects, the effect of electrode separation, as indicated by better performance in the interlaced condition than in the overlapped condition, was observed at all three SNRs; the difference in score between the overlapped and interlaced conditions was 5.2% or more. A two-way ANOVA tested the effect of separation and SNR as fixed factors with subject as a random factor. Both effects were significant ($p<0.01$), and the effect of subject was not ($p=0.177$). None

TABLE III. Relative information transmitted for each consonant feature (Miller and Nicely, 1955) calculated from aggregate presentation-response matrices across several SNR conditions for each subject.

	S8			S9		
Condition	Voicing	Manner	Place	Voicing	Manner	Place
Quiet	0.59	0.56	0.07	0.72	0.88	0.18
Overlapped	0.1	0.15	0.04	0.31	0.35	0.04
Interlaced	0.45	0.35	0.05	0.68	0.7	0.1

of the interactions were significant. Therefore, at least in these two subjects, the effect of electrode separation was significant regardless of the SNR. The amount of relative information transmitted for voicing, manner, and place features was calculated using feature transmission analysis (Miller and Nicely, 1955). Table III shows these calculations made from aggregate presentation-response matrices in all three SNR conditions for each subject. This analysis reveals that subject performance shows substantial differences between the overlapped and interlaced conditions in voicing and manner features, but not in the place feature. Even in the quiet condition, the transmission of place information was poor, as expected from previous CIS processing studies (Friesen *et al.*, 2001; Donaldson and Kreft, 2006). Therefore, it was not surprising to observe a weak effect of speech-noise electrode separation on the transmission of place information. On the other hand, voicing and manner features are relatively better utilized by CI users in the quiet condition (Donaldson and Kreft, 2006). It was notable that while a substantial reduction in information about these features was seen in the overlapped condition, similar to what was observed in previous studies based on conventional processing (Friesen *et al.*, 2001), the effect of interlaced noise on these features was overall not as severe as had been previously shown. This contrast provides additional evidence of perceptual segregation from electrode separation in these subjects. While the differences in overall percent-correct consonant identification scores are modest (from 5.7% to 9.9%), the effect of electrode separation is further confirmed by the fact that the identification errors made during incorrect trials were still within the same voicing and manner categories more often in the interlaced condition.

IV. DISCUSSION

A. Electrode spacing

The electrode spacing between two center points of adjacent electrodes in the banded (straight) nucleus electrode array is 0.75 mm and is constant across electrodes. On the other hand, the electrode spacing in the nucleus contour array, which is pre-curled toward the modiolus, or the peripheral process of the auditory nerve is not uniform across the array. It is 0.81 mm at the basal end and becomes smaller toward the apical end, where it is 0.4 mm for the three apical-most electrode pairs; the average of 21 electrode-to-electrode distances is 0.56 mm (van den Honert, personal communication). The five subjects tested with six-electrode CIS who demonstrated the effect of perceptual segregation with two-electrode spacing as a group were implanted with the contour array (cf. Table I). Thus, an average electrode distance of 1.12 mm might be considered to be the separation required for perceptual segregation between speech and noise. Although not tested in the present study, a condition testing two-electrode spacing in the straight electrode array (1.5 mm) might have made it easier to demonstrate the beneficial effect of electrode separation. Furthermore, experiments with other CI brands such as the HiFocus electrode array from Clarion II, for which the electrode spacing is 1.1 mm, might provide additional insights regarding electrode spacing and perceptual segregation. There is a caveat, however, in comparing or projecting the results from different electrode arrays because each electrode array has a unique curvature and the electrodes are placed at a different distance from the modiolus. For example, in the straight array, which does not curl inward toward the modiolus, apical electrodes are farther away from the modiolus than basal electrodes. In contrast, apical electrodes in the contour array are closer to the modiolus than basal electrodes. Therefore, the constant electrode spacing in the straight array does not imply that there is a spatially constant tonotopic mapping to the neuronal population excited by each electrode; similarly, the non-uniform spacing in the contour array does not necessarily lead to narrower spacing of nerve populations in the apical region excited by adjacent electrodes. Likewise, although the spacing of the HiFocus array is known to be 1.1 mm and it is clear that this spacing is in general greater than that in nucleus arrays, the implications of this difference are not definitive unless an accurate geometric configuration of the electrode array is available and compared. Thus, in addition to the inter-subject variability, which is common in CI experiments, the actual values of electrode spacing indicated in this paper should be interpreted only within the context of the specific electrode array and should be taken only as a guideline.

B. Number of effective channels and channel interactions in CIs

Although speech processing with only six to eight bands of spectral information provides reasonably good speech intelligibility in quiet for both CI users and NH listeners using CI simulation (Dorman *et al.*, 1998b; Friesen *et al.*, 2001; Fu *et al.*, 1998), it has also been suggested that more bands might be beneficial under adverse listening conditions, such as listening to difficult materials or speech in a competing background (Shannon *et al.*, 2004). While the benefit of an increased number of channels has been observed up to 24 channels in CI simulation with NH listeners (Qin and Oxen-

ham, 2003), there is little evidence demonstrating a similar trend in real CI users. Yet there seems to be a consensus to equip CI systems with more than eight bands, as all CI systems currently manufactured and used for regular clinical applications have at least 12 electrodes. The discrepancy between CI simulations and real implant performance can be often well accounted for by channel interactions limiting CI users' functional spectral resolution. (Note that the word "channel" in the term "channel interactions" refers to electric channel.) Although the term "channel interaction" has originally been used to refer to vector summation among current fields occurring simultaneously in multiple channels (Shannon, 1983; White *et al.*, 1984), interactions may still occur in interleaved stimulation when the interleaved delays are within the range of the neural refractory period of nerve fibers (Eddington *et al.*, 1978; Tong and Clark, 1986; Zeng *et al.*, 2005). Therefore, channel interactions have been identified as one of the primary limiting factors affecting spectral resolution in CIs (Henry *et al.*, 2005; Hughes and Abbas, 2006; Won *et al.*, 2007). The spatial boundaries of the neural populations excited by each electrode are not clearly defined, and there would likely be a substantial overlap in the populations excited by adjacent electrodes, compromising spectral resolution. This might be the basis of an upper limit to the "effective" number of channels available for each CI user. Also, the limited availability of healthy neural populations proximal to CI electrodes could be a further complicating factor. Beyond this number, little additional benefit would be obtained because any detailed spectral information available at this stage of signal processing would not pass through the electrode-neural interface. The number of effective channels in most CI users, presuming that no serious irregularities in neural survival patterns exist, is likely between 8 and 12 (with an average of 10) inferred from the fact that at least eight channels of speech information are required for successful performance in a number of clinical studies, and in certain cases the increase in channels beyond 8 (such as up to 12) leads to additional benefits (Nie *et al.*, 2006). Present-day technology can produce an electrode array with electrodes at a higher spatial density; yet currently CI systems do not utilize more than 22 electrodes (in fact, other devices that were developed later than the nucleus system adopt fewer electrodes, e.g., 16 electrodes for Clarion and 12 for Medel), and there is no evidence of disadvantage associated with fewer electrodes in these systems, as it is difficult to demonstrate an improvement in speech understanding beyond more than 12 channels.

The present results, suggesting that electrodes should be separated by a two-electrode spacing to enhance the effect of perceptual segregation in some subjects, are an interesting coincidence given that the limiting number of effective channels, as discussed above, is about 10. Subjects in the present study demonstrated as a group that they could segregate six speech channels from five different noise channels, while they were unable to segregate ten speech channels from ten different noise channels. Therefore, in the present experiment, subjects were able to effectively use up to 11 channels. This result could be an alternative demonstration of the limiting number of effective channels. Hypothetically, had there been an effect of segregation in the ten-channel CIS, it would have implied that the upper limit would be about 20 channels. This result also suggests that the number of effective channels in S7 (who showed the least effect of separation among those five subjects) might be substantially lower than 11. Furthermore, the present and previous results in speech recognition performance, demonstrating little benefit beyond 12 channels, further suggest that an attempt to present information to substantially more than 12 channels might not necessarily provide a benefit, for example, 43-channel ACE processing using 22 monopolar (MP) modes and 21 dual-MP modes in the nucleus freedom system (Busby and Plant, 2005), unless it used a novel approach that reduced or minimized channel interactions between electrodes. Along the same line of argument, an adjustment of the frequency boundaries between two channels by altering the frequency-to-electrode table or filterbank settings, as described in some studies (e.g., Geurts and Wouters, 2004), might be effective only if used between electrodes with sufficient spacing.

C. Summary, caveats, and further implications

The present study demonstrated a modest improvement in consonant recognition in noise by CI users when speech and noise electrodes were sufficiently separated, such as a two-electrode spacing in the nucleus contour array (equivalent to 1.1 mm on average). This serves as evidence that some CI users could segregate speech from background noise if these signals are presented with sufficient peripheral separation. While it might be problematic to prescribe two-electrode spacing as a necessary and sufficient condition for perceptual segregation, as the clear effect was demonstrated only in two subjects; this measure could serve as a guideline for future investigations. It appears that the performance of perceptual segregation is influenced by channel interactions occurring at the periphery, which might be one source of the individual variability observed in the present study.

1. Two CIS conditions (ten-channel versus six-channel)

The two separation conditions in the present study (one-electrode and two-electrode) were tested with different CIS processing conditions (one in ten-channel and the other in six-channel) due to the layout of nucleus electrode arrays: ten-channel CIS processing could not be tested with two-electrode spacing (a total of 40 electrodes would be required in order to test it); also testing six-channel CIS with one-electrode spacing would not be realistic because then the tonotopic representation of speech information would be highly compressed into a narrow region of the cochlea and the intelligibility of such processing would be severely limited (Baskent and Shannon, 2004). Therefore, the interpretation of the present results is based on a presumption that the effect of electrode separation observed was due to the amount of electrode separation rather than the number of CIS channels per se. On a related note, as the two conditions were tested with different subject groups, the data might have been affected by the grouping of subjects; i.e., had S8 and S9 been tested in the ten-channel CIS condition, a stron-

ger effect might have been observed with one-electrode spacing. Overall, this is in line with the suggestion earlier that the findings in terms of electrode spacing be taken as a guideline and the results be subject to the individual variability.

2. Temporal versus spatial separation

In both the overlapped and interlaced conditions, although pulse trains of speech and noise signals were temporally interleaved, the temporal separation alone does not imply perceptual segregation of those pulse trains because the temporal gap between successive pulses in the present experiment layouts was relatively small (on the order of 0.1 ms). (The temporal gap between successive pulses is the reciprocal of the total stimulation rate: e.g., it is 69 μs with a total stimulation rate of 14 400 Hz). It is unlikely that CI users can segregate streams of input on the same or close electrodes with such a short temporal gap. It was of primary interest in the present study to assess the effect of spatial separation (i.e., separated electrodes) between those two conditions.

3. Implications for application

As mentioned earlier, the conception of multi-stream processing was rationalized for a proper laboratory testing of CI users' ability to manage multiple signals and was not intended to emulate signal enhancement algorithms. The crux of this method is to provide multiple sources of input with as much integrity as possible, thereby allowing CI users to utilize these input streams to their advantage. The present data imply that a multi-stream processing approach might be beneficial to some CI users under some limited circumstances. For example, a dual-microphone system can be utilized, as Deeks and Carlyon (2004) proposed, in which the signals from two microphones are routed to different electrode sets. This application might be useful in applications where there is an uncertainty regarding source locations and selective listening in space is required, as in a cocktail party listening environment. However, a future study of selective listening tasks would be necessary to properly address the issue.

ACKNOWLEDGMENTS

The author is grateful to Wendy Parkinson, Lisa Dahlstrom, and Frank Warren for access to the clinical population. The author is heavily indebted to Chris van den Honert for his guidance and feedback. The following individuals were deeply appreciated for invaluable comments, suggestions, and constructive feedback at early stages of this study and preparation of the manuscript: Chris Turner, Robert Shannon, Monita Chatterjee, Robert Carlyon, John Galvin III, two anonymous reviewers, and Ken Grant. This work was supported by Cochlear Americas, the University of Utah, a seed money grant from the National Organization for Hearing Research Foundation, and the National Institutes of Health (NIDCD Grant No. R03DC009061).

[1]The level of noise was set during the pilot study to decrease the score about 10%–15% from the quiet condition through conventional processing of speech and noise mixture. This reduction might be different from what is shown in Figs. 2 and 3. where the scores were obtained with multi-stream processing.

Allen, J. B. (**1994**). "How do humans process and recognize speech?," IEEE Trans. Speech Audio Process. **2**, 567–577.

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.

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.

Baskent, D., and Shannon, R. V. (**2004**). "Frequency-place compression and expansion in cochlear implant listeners," J. Acoust. Soc. Am. **116**, 3130–3140.

Battmer, R. D., Reid, J. M., and Lenarz, T. (**1997**). "Performance in quiet and in noise with the Nucleus Spectra 22 and the Clarion CIS/CA cochlear implant devices," Scand. Audiol. **26**(4), 240–246.

Beauvois, M. W., and Meddis, R. (**1996**). "Computer simulation of auditory stream segregation in alternating-tone sequences," J. Acoust. Soc. Am. **99**, 2270–2280.

Bregman, A. S., *Auditory Scene Analysis* (The MIT Press, Cambridge, MA, **1990**).

Brokx, J. P. L., and Nooteboom, S. G. (**1990**). "Intonation and the perception of simultaneous voices," J. Phonetics **19**, 23–26.

Busby, P. A., and Plant, K. L. (**2005**). "Dual electrode stimulation using the Nucleus CI24RE cochlear implants: Electrode impedance and pitch ranking studies," Ear Hear. **26**, 504–511.

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.

Chatterjee, M., Sarampalis, A., and Oba, S. I. (**2006**). "Auditory stream segregation with cochlear implants: A preliminary report," Hear. Res. **222**, 100–107.

Cooper, H. R., and Roberts, B. (**2007**). "Auditory stream segregation of tone sequences in cochlear implant listeners," Hear. Res. **225**, 11–24.

Cullington, H. E., and Zeng, F. G. (**2008**). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," J. Acoust. Soc. Am. **123**, 450–461.

Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.

Donaldson, G. S., and Kreft, H. A. (**2006**). "Effects of vowel context on the recognition of initial and medial consonants by cochlear implant users," Ear Hear. **27**, 658–677.

Dorman, M. F., Loizou, P. C., and Fitzke, J. (**1998a**). "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," Ear Hear. **19**, 481–484.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998b**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Eddington, D. K., Dobene, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. L. (**1978**). "Auditory prostheses research with multiple channel intracochlear stimulation in man," Ann. Otol. Rhinol. Laryngol. **87**, 5–38.

Fetterman, B. L., and Domico, E. H. (**2002**). "Speech recognition in background noise of cochlear implant patients," Otolaryngol.-Head Neck Surg. **126**, 257–263.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. J., and Galvin, J. J., III (**2001**). "Recognition of spectrally asynchronous speech by normal-hearing listeners and Nucleus-22 cochlear implant users," J. Acoust. Soc. Am. **109**, 1166–1172.

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Geurts, L., and Wouters, J. (**2004**). "Better place-coding of the fundamental frequency in cochlear implants," J. Acoust. Soc. Am. **115**, 844–852.
Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9**, 155–184.
Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.
Hong, R. S., and Turner, C. W. (**2006**). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," J. Acoust. Soc. Am. **120**, 360–374.
Hughes, M. L., and Abbas, P. J. (**2006**). "Electrophysiologic channel interaction, electrode pitch ranking, and behavioral threshold in straight versus perimodiolar cochlear implant electrode arrays," J. Acoust. Soc. Am. **119**, 1538–1547.
Kokkinakis, K., and Loizou, P. C. (**2008**). "Using blind source separation techniques to improve speech recognition in bilateral cochlear implant patients," J. Acoust. Soc. Am. **123**, 2379–2390.
Kwon, B. J., and van den Honert, C. (**2006**). "Independent and concurrent processing of multiple audio input signals in a prosthetic hearing implant," U.S. Patent Application 20060265061, filed Jan. 20, 2006; published Nov. 23, 2006.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C., 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.
Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). "Intensity discrimination as a function of stimulus level with electric stimulation," J. Acoust. Soc. Am. **100**, 2393–2414.
Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Nie, K., Barco, A., and Zeng, F. G. (**2006**). "Spectral and temporal cues in cochlear implant speech perception," Ear Hear. **27**, 208–217.
Peterson, G. E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear. Disord. **27**, 62–70.
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.
Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.
Shannon, R. V., Fu, Q. J., and Galvin, J. (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Oto-Laryngol., Suppl. **552**, 50–54.
Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antongenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H. J., and Beirer, A. L. (**1994**). "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system," Am. J. Otol. **15**, 15–27.
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.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Tong, Y. C., and Clark, G. M. (**1986**). "Loudness summation, masking, and temporal interaction for sensations produced by electric stimulation of two sites in the human cochlea," J. Acoust. Soc. Am. **79**, 1958–1966.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.
van den Honert, C., and Kelsall, D. C. (**2007**). "Focused intracochlear electric stimulation with phased array channels," J. Acoust. Soc. Am. **121**, 3703–3716.
White, M. W., Merzenich, M. M., and Gardi, J. N. (**1984**). "Multichannel cochlear implants: Channel interactions and processor design," Arch. Otolaryngol. **110**, 493–501.
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.
Won, J. H., Drennan, W. R., and Rubinstein, J. T. (**2007**). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," J. Assoc. Res. Otolaryngol. **8**, 384–392.
Zeng, F. G., Chen, H., and Han, S. (**2005**). "Temporal masking in electric hearing," J. Assoc. Res. Otolaryngol. **6**, 390–400.

Automatic detection of the second subglottal resonance and its application to speaker normalization[a)]

Shizhen Wang[b)]
Department of Electrical Engineering, University of California, Los Angeles, California 90095

Steven M. Lulich
Speech Communication Group, MIT, Cambridge, Massachusetts 02139

Abeer Alwan
Department of Electrical Engineering, University of California, Los Angeles, California 90095

(Received 24 December 2008; revised 24 September 2009; accepted 25 September 2009)

Speaker normalization typically focuses on inter-speaker variabilities of the supraglottal (vocal tract) resonances, which constitute a major cause of spectral mismatch. Recent studies have shown that the subglottal airways also affect spectral properties of speech sounds, and promising results were reported using the subglottal resonances for speaker normalization. This paper proposes a reliable algorithm to automatically estimate the second subglottal resonance (Sg2) from speech signals. The algorithm is calibrated on children's speech data with simultaneous accelerometer recordings from which Sg2 frequencies can be directly measured. A cross-language study with bilingual Spanish-English children is performed to investigate whether Sg2 frequencies are independent of speech content and language. The study verifies that Sg2 is approximately constant for a given speaker and thus can be a good candidate for limited data speaker normalization and cross-language adaptation. A speaker normalization method using Sg2 is then presented. This method is computationally more efficient than maximum-likelihood based vocal tract length normalization (VTLN), with performance better than VTLN for limited adaptation data and cross-language adaptation. Experimental results confirm that this method performs well in a variety of testing conditions and tasks. [DOI: 10.1121/1.3257185]

PACS number(s): 43.72.Ne, 43.72.Fx [DOS] Pages: 3268–3277

I. INTRODUCTION

A major cause of performance degradation in automatic speech recognition (ASR) systems is inter-speaker variations in acoustic characteristics (fundamental and formant frequencies, etc.), which are mostly caused by differences in the supraglottal speech production system (vocal tract and vocal fold apparatus). Accordingly, speaker normalization, which aims to reduce these acoustic variabilities, typically focuses on supraglottal variations. Vocal tract length normalization (VTLN) is one of the most popular methods for reducing the effects of speaker-dependent vocal tract variability through a speaker-specific frequency warping function (linear, bilinear, or piece-wise linear) (Wegman *et al.*, 1996; Lee and Rose, 1998; Pitz and Ney, 2003; Umesh *et al.*, 2005; McDonough *et al.*, 2004; McDonough, 2000). Warping factors are typically estimated based on the maximum likelihood (ML) criterion over the adaptation data through an exhaustive grid search or warping-factor specific models (Wegman *et al.*, 1996; Lee and Rose, 1998). Linear frequency warping can be implemented directly in the power spectrum domain or in the cepstral domain through the linearization of VTLN (Pitz and Ney, 2003; Umesh *et al.*, 2005; McDonough *et al.*, 2004). Along with the linearization of VTLN, the warping factor can be estimated using the expectation maximization algorithm with an auxiliary function (McDonough *et al.*, 2004). Other frequency warping functions have also been studied. A class of transforms, known as all-pass transforms (APTs), was proposed to perform VTLN and studied in detail in McDonough, 2000 for two classes of conformal maps, namely, rational all-pass transforms and sine-log all-pass transforms. It was demonstrated that using multiple-parameter warping functions is more effective than single-parameter ones.

Another way to reduce spectral variability is to explicitly align spectral formant positions or formant-like spectral peaks, especially the third formant (F_3), and to define the warping factors as formant frequency ratios (Eide and Gish, 1996; Gouvea and Stern, 1997; Claes *et al.*, 1998; Cui and Alwan, 2006; Zhan and Westphal, 1997; Wang *et al.*, 2007). In formant-based frequency warping methods, formant positions of different speakers are transformed into a normalized frequency space. Eide and Gish (1996) proposed a nonlinear warping function based on a parameter estimated using F_3 frequency. Zhan and Westphal (1997) extended this formant-based algorithm and compared the performance with ML-based methods. Gouvea and Stern (1997) explored the performance of frequency warping using the first three formant frequencies. Claes *et al.* (1998) proposed a linear approximation of VTLN for reasonably small warping factors estimated based on average F_3 values. Cui and Alwan (2006) proposed

[a)]Portions of this article were presented at ICASSP 2008 and Interspeech 2008.
[b)]Author to whom correspondence should be addressed. Electronic mail: szwang@ee.ucla.edu

a novel spectral formant-like peak alignment method, with a focus on F_3, to reduce spectral mismatch between adults and children's speech (Cui and Alwan, 2006; Wang *et al.*, 2007). Based on the idea of frequency transformation for digital filters, Wang *et al.* (2004) treated formant structures as filters and developed a bilinear transform with parameters estimated using average F_3 frequency and bandwidth values. Due to coarticulation, clarity, speed, and other factors, formant frequencies vary considerably within an utterance and thus make the performance of formant normalization content-dependent.

Besides the effects of the supraglottal system, recent studies show that the subglottal airways also affect spectral properties of speech sounds (Hanson and Stevens, 1995; Stevens, 1998; Sonderegger, 2004; Chi and Sonderegger, 2004; Lulich, 2006; Chi and Sonderegger, 2007; Lulich *et al.*, 2007). The coupling between the supraglottal and subglottal systems has been shown to be non-negligible when a vocal tract formant approaches a subglottal resonance in frequency. At such a point, the formant prominence amplitude will be attenuated and the prominent spectral peak will jump in frequency to skip the subglottal resonance. Many studies have been done to model and analyze the subglottal resonances for adults' speech, particularly focusing on American English, although a few studies in Korean and German have shown similar results (Jung, 2008; Madsack *et al.*, 2008).

Children's speech analysis and recognition have drawn increasing attention for educational purposes, and more efforts have been devoted to ASR's applications on children's speech. Due to developmental changes in vocal tract and vocal ford apparati, children's speech demonstrates high acoustic variabilities, which makes children's ASR more challenging compared to adults' ASR. The performance of an ASR system developed for adult speech decreases drastically when employed to recognize children's speech. Furthermore, recognition performance for children is usually lower than that achieved for adults even when using a recognition system trained on children's speech. Such challenges require further studies on children's speech.

In this paper, we focus on children's speech and explore the hypothesis that subglottal resonances can be used for speaker normalization, much like formant alignment techniques. We first describe the subglottal system and outline the theory of its effects on speech. We then use this theory to implement an automatic detector of the second subglottal resonance. Third, we address the question whether subglottal resonances are constant for a given speaker regardless of the language being spoken. Finally, we implement a speaker normalization scheme based on the second subglottal resonance (Sg2) and evaluate its performance on several tasks. We compare the performance with VTLN and demonstrate the effectiveness of this method.

II. EFFECTS OF COUPLING TO THE SUBGLOTTAL ACOUSTIC SYSTEM

The configuration of the acoustic system below the glottis consists of the trachea, bronchi, and lungs. Similar to the vocal tract, the acoustic input impedance of the subglottal system is characterized by a series of poles (or resonances)

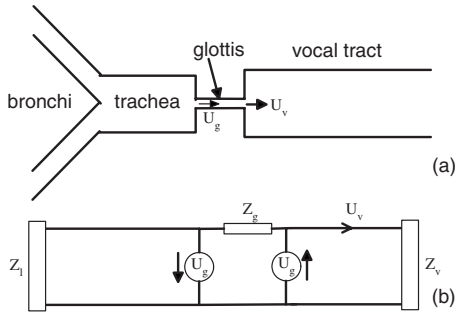


FIG. 1. Schematic model of vocal tract with acoustic coupling to the trachea through the glottis (a) and the equivalent circuit model (b) (Adapted from Stevens, 1998).

and zeros. Unlike the supraglottal system, however, the configuration of the subglottal system is essentially fixed and thus the poles and zeros are expected to remain constant for any given speaker. Like formant frequencies, subglottal resonances are generally higher for female speakers than for male speakers, and there are substantial individual differences from speaker to speaker. It has been shown that the lowest three subglottal resonances are around 600, 1450, and 2200 Hz for adult males, and 700, 1600, and 2350 Hz for adult females (Stevens, 1998).

Figure 1 shows a schematic model of vocal tract coupling to the trachea through the glottis and its equivalent circuit model, where Z_l is the impedance of the subglottal system, Z_g is the glottal impedance, Z_v is the impedance looking into the vocal tract from the glottis, U_g is the volume velocity through the glottis, and U_v is the airflow into the vocal tract. Coupling between the subglottal and supraglottal airways is thought to occur primarily when the glottis is open, such as during a voiceless consonant or the open phase of glottal vibration in a voiced sound, although Lulich (2009) and Lulich *et al.* (2009) suggested that coupling may also occur when the vocal folds are closed, either by means of a posterior glottal opening or the vocal fold tissue itself. During coupling, each subglottal resonance contributes a pole-zero pair to the speech spectrum. The frequency of the zero is the same as that of the subglottal resonance, while the pole is shifted upward in frequency away from the resonance and depends somewhat on the vocal tract configuration. This is because the zero is a function only of the part of the entire system behind the source (that is, the subglottal airways), while the pole is a function of the entire system, including the subglottal and supraglottal airways (Lulich, 2009; Chi and Sonderegger, 2007).

The pole-zero pair introduced in the speech spectrum around Sg2 generally falls within the range of 1300–1500 Hz for adult males, and between 1400 and 1700 Hz for adult females (Stevens, 1998). It is somewhat higher in frequency for children (Jung *et al.*, 2008). When F2 crosses the Sg2 pole-zero pair, F2 can jump in frequency or diminish in amplitude, or both, resulting in a discontinuity in the F2 trajectory (Chi and Sonderegger, 2007). This is illustrated in Fig. 2 for an 8-year-old girl speaking the word *boy*, and it is schematically represented in Fig. 3. In both figures, F2 rises from a low frequency to a high frequency, crossing the Sg2 pole-zero pair along the way. The F2 discontinuity in

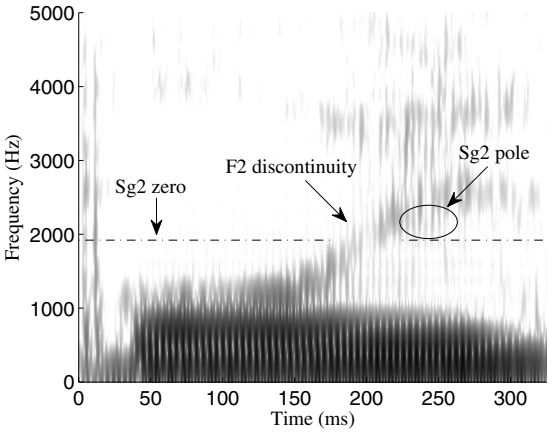


FIG. 2. Spectrogram for the word *boy* from an 8-year-old girl. The second subglottal resonance Sg2 for this speaker is 1920 Hz.

Fig. 2 is marked by a diminished amplitude in the vicinity of the zero. The Sg2 pole has a very low amplitude except during the time when F2 is nearby. In Fig. 2 the diffuse energy between F2 and the zero at 250 ms is likely due to the Sg2 pole, its amplitude decreasing as F2 continues to rise.

Recent studies (Lulich, 2006; Lulich *et al.*, 2007; Sonderegger, 2004) have shown that the acoustic contrasts for some phonological distinctive features are dependent on the subglottal resonances. As illustrated in Fig. 4, for example, the vowel feature [back][1] is dependent on the frequency of Sg2, such that a vowel with F2 > Sg2 is [−back] and a vowel with F2 < Sg2 is [+back]. The ability of Sg2 to underlie the distinctive feature [back] is likely derived from the fact that Sg2 is roughly constant for a given speaker. Subglottal resonances could potentially be affected by lung volume, larynx height, and glottal configuration. Lung volume has been shown not to significantly affect the subglottal resonances in one study (Cheyne, 2001), and reported accelerometer measurements of subglottal resonances across utterances (in which phonetic content was varied and voice quality was uncontrolled—both of which may affect laryngeal height and glottal configuration) have had standard deviations on the order of 30 Hz or less (Chi and Sonderegger, 2007). Thus, although the influence of lung volume, larynx height, and glottal configuration on subglottal resonances invites further research, the available evidence appears to indicate that subglottal resonances are roughly constant under normal speaking conditions.

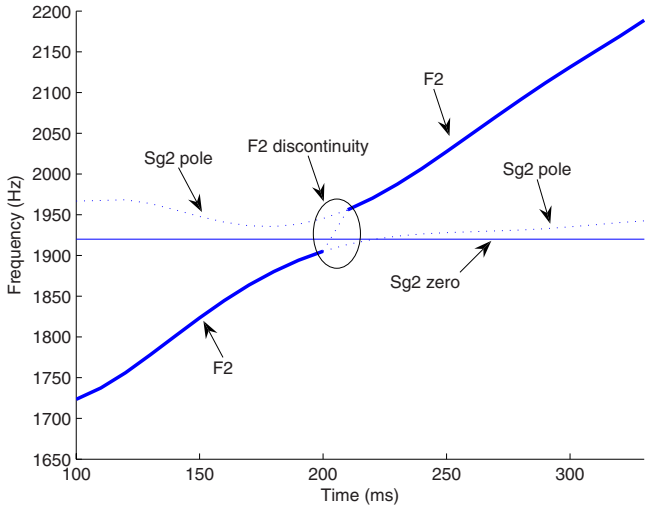


FIG. 3. (Color online) Illustration of the F2 discontinuity caused by Sg2. The bold solid line corresponds to the most prominent spectral peak (F2), which has a jump in frequency and a decrease in amplitude when F2 is crossing the subglottal resonance Sg2. The dotted line represents the Sg2 pole, which varies somewhat in frequency and amplitude when F2 is nearby. The horizontal thin solid line represents the Sg2 zero, which is roughly constant (adapted from Stevens, 1998).

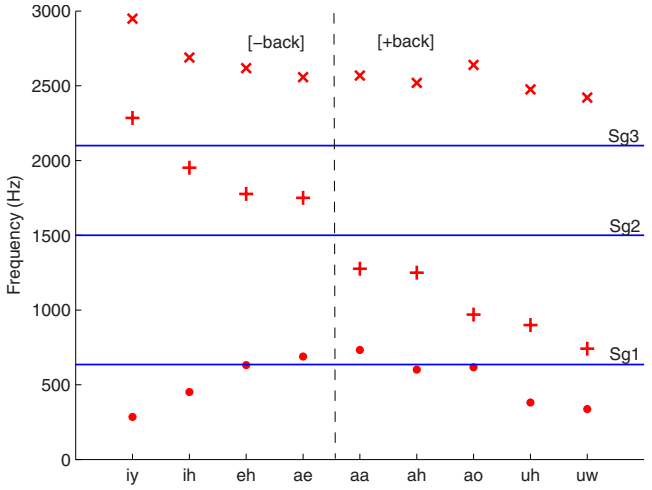


FIG. 4. (Color online) Illustration of the relative positions of vowel formants F1 (·), F2 (+), and F3 (×) and the subglottal resonances (Sg1, Sg2, and Sg3) for an adult male speaker. For the vowels /i, ɪ, ɛ, æ/ F2 > Sg2, and they are therefore [−back]. For the vowels /a, ʌ, ɔ, ʊ, u/ F2 < Sg2, and they are therefore [+back] (adapted from Lulich, 2006).

For this reason, Sg2 might be useful in speaker normalization, since it is context independent but speaker dependent. Sg1 and Sg3 have also been claimed to play a role in distinguishing different classes of speech sounds, but Sg2 has been more thoroughly studied. In this paper, therefore, we focus on Sg2 estimation and its application to speaker normalization.

III. ESTIMATING THE SECOND SUBGLOTTAL RESONANCE

A. Automatic estimation of Sg2 frequency

As noted above, when F2 crosses Sg2, there is a discontinuity in the F2 trajectory. Based on this discontinuity, an automatic Sg2 detector (Sg2D1) was developed in Wang *et al.* (2008a). The Snack sound toolkit (Snack, 2005) was used to generate the F2 trajectory. All experiments were done in clean conditions. The tracking parameters were specifically tuned to provide reliable F2 tracking results on children's speech. Manual verification and/or correction were applied through visually checking the tracking contours against the spectrogram. (Note that this method is limited by the accuracy of the formant tracker, which is known to encounter difficulties in high-pitched speech such as that produced by children.) The F2 discontinuity was detected based on the smoothed first order difference of the F2 trajectory, as shown

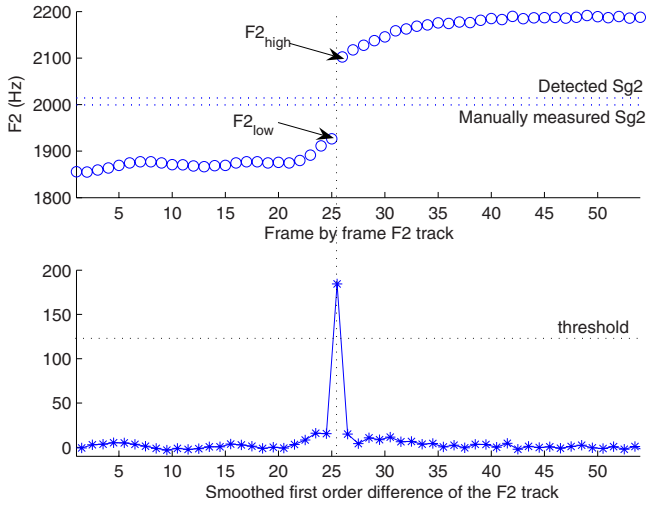


FIG. 5. (Color online) An example of the detection algorithm applied in Wang *et al.*, 2008.

in Fig. 5. If the F2 values on the high and low frequency sides of the discontinuity are $F2_{high}$ and $F2_{low}$, respectively, then Sg2D1 is estimated as

$$\widehat{Sg2} = \frac{F2_{high} + F2_{low}}{2}. \quad (1)$$

If no such discontinuity was detected, Sg2D1 used the mean F2 over the utterance. In many such cases, such as during a monophthong, F2 is consistently above or below Sg2, and the mean F2 value is either too high or too low. Thus, the estimated Sg2 values are dependent on the speech sound analyzed. Furthermore, discontinuities in F2 may arise from other factors beside the subglottal resonances, including pole-zero pairs from the interdental spaces (Honda *et al.*, 2009). These discontinuities occur a few hundred hertz higher than Sg2 discontinuities, but are sometimes more prominent than Sg2 discontinuities and can therefore be mistakenly detected as Sg2.

To address both issues, we developed an improved Sg2 estimation algorithm (Sg2D2) (Wang *et al.*, 2008b). We first detected F3 and obtained an estimate of Sg2 using a formula derived in Lulich, 2009:

$$\widetilde{Sg2} = 0.636 \times F3 - 103. \quad (2)$$

Note that the derivation of this formula was based on a linear regression on children's speech data which have available simultaneous accelerometer recordings, and its extension to adults' speech may still need further refinements.

We then searched for a discontinuity within ±100 Hz of this estimate using the original algorithm. The range ±100 Hz was chosen based on calculated standard deviations of Sg2 on the calibration data. If no discontinuity in this range was found, $\widetilde{Sg2}$ was used. If a discontinuity was found, we estimated Sg2 using the following equation:

$$\widehat{Sg2} = \beta \times F2_{high} + (1 - \beta) \times F2_{low}, \quad (3)$$

where β is a weight in the range (0, 1) that controls the closeness of the detected Sg2 value to $F2_{high}$. The optimal value of β was calibrated over the data described below using the minimum mean square error criterion:

$$\beta = \arg\min E\{(\widehat{Sg2} - Sg2)^2\} \quad (4)$$

and was found to be 0.65 in our experiments.

TABLE I. Comparison of Sg2 estimates for two algorithms over various vowel contents, where Sg2D1 refers to the algorithm used in Wang *et al.* (2008a), Sg2D2 is the new algorithm, Sg2M is the manual measurement from speech spectrum, and Sg2Acc is the "ground truth" measurement from the accelerometer signal. For each algorithm the average Sg2 estimates (Hz) are shown (with standard deviations in parentheses). The two speakers with an " *" symbol are those used for calibration.

Speaker	Sg2D1	Sg2D2	Sg2M	Sg2Acc
1	2135 (531)	2194 (95)	2193 (97)	2176
2	2115 (334)	1766 (137)	1719 (112)	1646
3	2586 (467)	2718 (143)	2634 (135)	2679
4	2098 (358)	1823 (151)	1781 (129)	1614
5*	2065 (267)	2021 (79)	2013 (76)	1970
6*	1612 (251)	1689 (72)	1681 (65)	1648

B. Calibration of the Sg2 estimation algorithm

To verify and calibrate our Sg2 estimation algorithm, acoustic data were collected from six female children aged 2–17 years old (speakers G1–G6 in Lulich, 2009). The children were native speakers of American English and all of them except the youngest were recorded repeating the phrase "hVd, say hVd again" three times for each of the vowels [i], [ɪ], [ɛ], [æ], [a], [ʌ], [o], [ʊ], and [u]. The subjects also recited the alphabet, counted to 10, and recited a few short sentences. The recording list was presented in random order and verbally prompted by the experimenter. The youngest speaker (G1) was recorded counting to 10, reciting the alphabet, and answering questions of the sort "What is this?," in which the experimenter pointed to his hand or head, for instance. All utterances were recorded in a sound-isolated chamber using a SHURE BG4.1 uni-directional condenser microphone and an accelerometer. Both the speech and accelerometer signals were digitized at 16 kHz. Microphone signals of each speaker were used to measure average F3 and the discontinuity in the F2 track. An independent direct measure of the average Sg2 for each speaker was obtained from an accelerometer signal. The accelerometer was attached to the skin of the neck below the larynx so that the measured vibration of the neck skin is directly related to the acoustic pressure variations in the air column at the top of the trachea (Cheyne, 2001; Chi and Sonderegger, 2007). The accelerometer signal can therefore act as a stand-in for the subglottal input impedance, in which the subglottal resonances appear as formants in the accelerometer spectrum.

The detection algorithms Sg2D1 and Sg2D2 were calibrated (to estimate discontinuity thresholds for both Sg2D1 and Sg2D2, and β for Sg2D2) on data from two of the recorded children and tested on the remaining four. The values measured from the accelerometer data were used as the "ground truth" Sg2 frequencies (henceforth denoted by "Sg2Acc"). The average Sg2 estimates (with standard deviations) over various vowel contents are shown in Table I. Compared to Sg2D1, the updated algorithm Sg2D2 estimates

TABLE II. Detailed comparison of Sg2 estimates for the two algorithms on two speakers. For vowels above the double line, there are no discontinuities in the F2 trajectory, and Sg2D1 uses the mean F2 as Sg2, while Sg2D2 uses Eq. (2) ($\widetilde{Sg2}$) to make an estimate; for vowels below the double line, the F2 discontinuity is detectable, and Sg2D1 uses Eq. (1), while Sg2D2 uses Eq. (3). The row "Avg.(std)" shows the mean (and standard deviation) for each algorithm.

	Speaker 1 (age 6) Sg2Acc: 2176 Hz		Speaker 2 (age 13) Sg2Acc: 1646 Hz	
Vowel	Sg2D1	Sg2D2	Sg2D1	Sg2D2
[i]	2987	2312	2563	1971
[ɪ]	2515	2306	2439	1909
[e]	2894	2115	2629	1998
[ɛ]	2799	2291	2378	1867
[æ]	2382	2289	2350	1863
[a]	1599	2020	1796	1700
[ʌ]	1687	2243	1948	1704
[o]	1512	2185	1497	1613
[ʊ]	1578	2228	1964	1717
[u]	1739	2071	1825	1631
[au]	1841	2114	1974	1617
[aɪ]	2103	2170	2072	1709
[ɔɪ]	2115	2183	2063	1659
Avg.(std)	2135 (531)	2194 (95)	2115 (334)	1766 (137)

Sg2 much better with less variance across vowels. The observed standard deviation values of Sg2D2 are similar to those from manually measured Sg2's $(Sg2M)^2$ in this study and those found in other studies (Jung *et al.*, 2008).

The performance of these two algorithms was investigated in more detail for each vowel for two speakers and the results are shown in Table II and Fig. 6. As stated earlier, if no discontinuity in the F2 track is detected (as for the vowels above the double line, Table II), Sg2D1 uses the mean F2 as Sg2 and thus is highly dependent on vowel content. Sg2D2, on the other hand, uses a formula to estimate Sg2 from F3 which is less content-dependent than F2. In such cases, it can be seen that the formula in Sg2D2 gives much closer estimates to the ground truth, especially for mid and back vowels. For the case when there is a discontinuity in the F2 trajectory (as for the diphthongs below the double line), both algorithms work well when the F2 discontinuity is from Sg2, as for speaker 1. In this case, Sg2D1 gave an estimate within about 70 Hz of the true Sg2 value, while the Sg2D2 estimate was within less than 10 Hz. For speaker 2, where the most prominent F2 discontinuity was probably from the interdental space, Sg2D1 gave an estimate hundreds of hertz above the Sg2 value, while Sg2D2 roughly located the correct Sg2 value using Eq. (2). Thus, Sg2D2 is less prone to mistakenly detecting discontinuities not caused by Sg2. In addition to diphthongs, discontinuities in F2 should also be detectable in certain consonant-vowel transitions (Lulich, 2009). Since Sg2D2 performs consistently better than Sg2D1, we will focus only on Sg2D2 in the following experiments. As shown in Tables I and II, and Fig. 6, the proposed detector produces Sg2 estimates close to the ground truth. Also, as will be shown in (Sec. V), the estimated Sg2 helps to improve ASR's performance on children's speech, which is of primary interest to us.

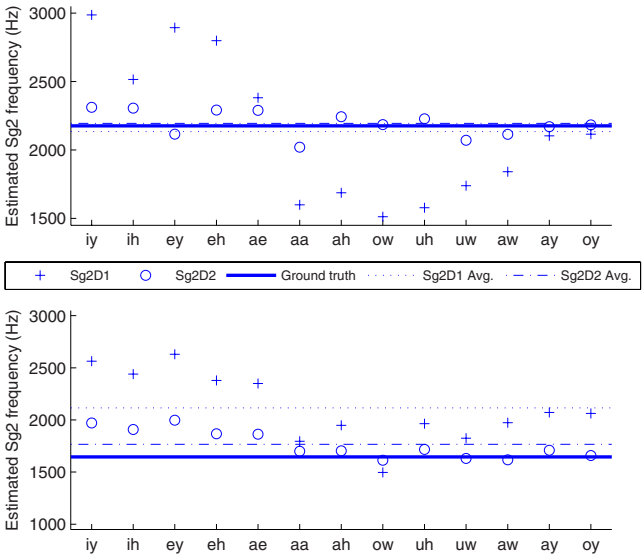


FIG. 6. (Color online) Comparison of Sg2 estimates for the two speakers in Table II: left panel for speaker 1 and right panel for speaker 2.

IV. VARIABILITY OF SUBGLOTTAL RESONANCE Sg2

The acoustic characteristics of children's speech have been shown to be highly different from those of adults' speech, in terms of pitch and formant frequencies, segmental durations, and temporal and spectral intra- and inter-speaker variabilities (Lee *et al.*, 1999; Huber *et al.*, 1999). Studies of subglottal resonances (Hanson and Stevens, 1995; Sonderegger, 2004; Chi and Sonderegger, 2004; Lulich, 2006; Chi and Sonderegger, 2007; Lulich *et al.*, 2007), however, have mainly focused on adults' speech in English with little effort devoted to children's speech or to other languages (but see Jung, 2008; Jung *et al.*, 2008; Madsack *et al.*, 2008). This section analyzes children's speech in English and Spanish, investigating the variabilities of Sg2 under different contents and across languages.

To examine the cross-language variability of Sg2 frequencies, we recorded a database (ChildSE) of 20 bilingual Spanish-English children (ten boys and ten girls) in the first or second grade (around 6 and 7 years old, respectively) from a bilingual elementary school in Los Angeles. The recorded speech consisted of words containing front, mid-, back, and diphthong vowels. There were four English words (*beat, bet, boot*, and *bite*) and five Spanish words (*calle* "street," *casa* "house," *quitar* "to take out," *taquito* "taco," and *cuchillo* "knife"), all of which were familiar to the children. Prior to the recording, children were instructed to practice as many times as they wanted. Both text and audio samples for each target word were available for prompt, and children decided what prompt they needed during recording and what language they wanted to record first. There were three repetitions for each word, and children spoke all the words in one language in a row with 3 s pause between words, and then repeated them. After they finished the recordings in one language, there was about a 1 min pause before they began the recordings in the other language. Recordings were made with 16 kHz sampling rate and 16-bit

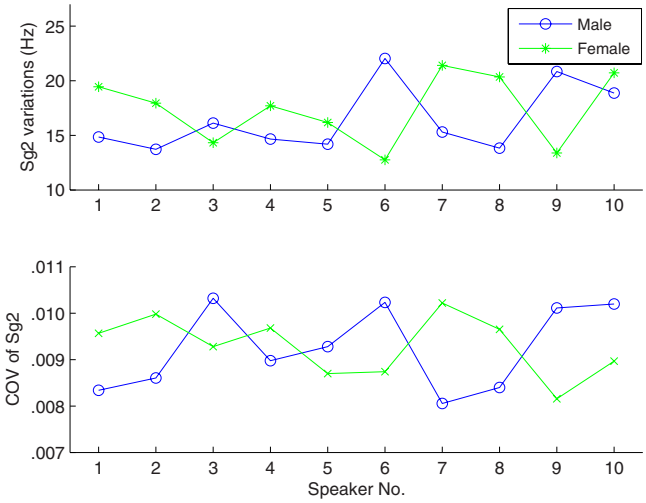


FIG. 7. (Color online) Average within-speaker Sg2 standard deviations and the COVs against contents and repetitions.

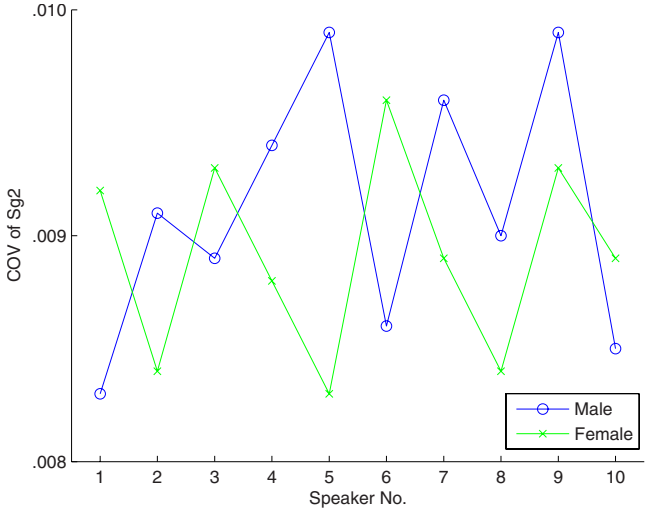


FIG. 8. (Color online) Cross-language within-speaker COV of Sg2 for ten boys and ten girls.

resolution. Like the English word *bite* [baɪt], the Spanish words *calle* [kaje] and *cuchillo* [kutʃijo] had obvious F2 discontinuities. We used these words with diphthongs to estimate Sg2 frequencies. Therefore, for each speaker, there were three English tokens and six Spanish tokens for the Sg2 estimation.

The within-speaker standard deviations were calculated on the Sg2 values estimated from the six Spanish tokens for each speaker. We also calculated the within-speaker coefficients of variation (COVs), a measure of dispersion of a probability distribution, which was computed as the ratio of the standard deviation to the mean Sg2 value for each speaker. As shown in Fig. 7, the within-speaker Sg2 standard deviations are around 20 Hz and the COV is less than 0.01. No significant difference in the COVs is observed between genders. A similar trend is observed for the within-speaker Sg2 standard deviations calculated from the English tokens. Compared to the COV of formant frequencies (Lee *et al.*, 1999), which are usually around 0.10, the COV of Sg2 is about one order of magnitude smaller. Therefore, the within-speaker Sg2 variability is negligible since they are sufficiently small compared to formant variabilities. This means that for a given speaker Sg2 is relatively constant against contents and repetitions.

Since Sg2 frequency for a given speaker does not depend on the contexts, we calculated the Sg2 COV for each speaker over the three English tokens and six Spanish tokens and viewed this as the Sg2 cross-language variabilities, as plotted in Fig. 8. The cross-language Sg2 COVs are less than 0.01, and there is no significant difference between genders. The cross-language COVs are similar to the within-speaker COVs, indicating that the cross-language effects are not significant for Sg2 frequencies and the Sg2 frequency for a given speaker is independent of languages.

Because of its invariability across speech content and language, Sg2 was judged to be applicable to speaker normalization. Since Sg2 is content-independent, we hypothesized that the performance of speaker normalization using Sg2 should be robust and independent of the amount of adaptation data available. This would make the Sg2 normalization method greatly suitable for limited data adaptation, which is often the case in ASR applications.

On the other hand, the language-independent property of Sg2 makes cross-language adaptation possible based on Sg2 normalization. Theoretically, with Sg2 normalization acoustic models trained in one language could be adapted with data in any other language,

V. SPEAKER NORMALIZATION RESULTS

Similar to formant normalization, the warping ratio for Sg2 normalization is defined as

$$\alpha = \mathrm{Sg2}_r/\mathrm{Sg2}_t, \tag{5}$$

where $\mathrm{Sg2}_r$ is the reference Sg2 and $\mathrm{Sg2}_t$ is the Sg2 of the test speaker. The reference Sg2 is defined as the mean value of all the training speakers' Sg2's. The Sg2 values are detected using the Sg2D2 algorithm. In this section, we evaluate the content dependency of Sg2 normalization and also its use for cross-language normalization.

A. Comparison of VTLN and Sg2 frequency warping

Figure 9 shows F1, F2, and F3 values from a 9-year-old girl before and after warping using VTLN (Lee and Rose, 1998) and the Sg2 ratio. The line Sg2 is the reference second subglottal resonance for an adult male speaker (as in Fig. 4). Compared to Fig. 4, unwarped data demonstrate an obviously different pattern as to the relative positions of the formants with respect to the reference Sg2. For instance, the back vowels [ʊ] and [u] have higher F2 values than the reference Sg2, while in Fig. 4 F2's of all the back vowels lie below the Sg2 line. It is necessary to apply frequency warping to achieve the reference formant position pattern. Both VTLN (in circles) and Sg2 (in squares) warp the formants close to the reference pattern, although Sg2 warping yields a formant pattern more similar to the reference speaker's.

To examine the effects of warping in more detail, we plotted in Fig. 10 the reference F1, F2, and F3 values versus

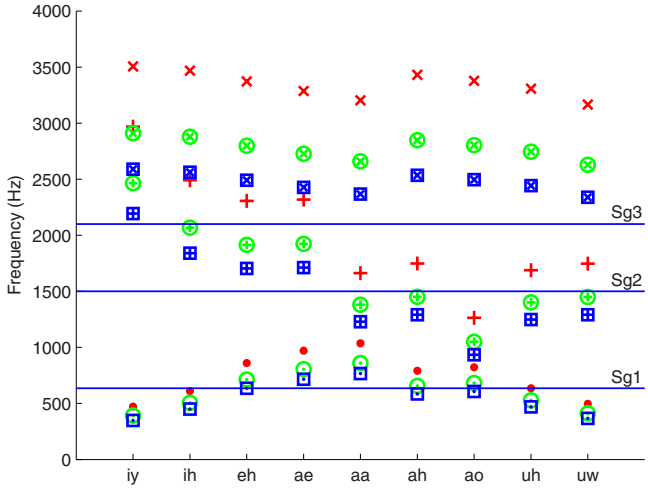


FIG. 9. (Color online) Vowel formants F1 (·), F2 (+), and F3 (×) before and after VTLN (in circles) and Sg2-based (in squares) warping for a 9-year-old girl's vowels. The lines "Sg1," "Sg2," and "Sg3" are the reference subglottal resonances from the same speaker as in Fig. 4.

the normalized values. Sg2 warping aligns the test speaker's formants more closely to the reference speaker's formants (Fig. 4), as indicated by the proximity of the data points to the diagonal line (with slope 1). In ASR such warping results in greatly reduced spectral mismatch between test and reference speakers, and thus can lead to better ASR performance.

B. Effectiveness of Sg2 normalization

Since VTLN has been shown to provide significant performance improvement on children's speech recognition, we first evaluate the subglottal normalization method on a connected digits recognition task of children's speech using the TIDIGITS database. Speech signals were segmented into 25 ms frames, with a 10 ms shift. Each frame was parametrized by a 39-dimensional feature vector consisting of 12 static Mel Frequency Cepstral Coefficients (MFCCs) plus log energy, and their first- and second-order derivatives. Acoustic Hidden Markov Models (HMMs) were monophone-based with three states and six Gaussian mixtures in each state. VTLN was implemented based on a grid search over [0.7, 1.2] with a stepsize of 0.01. The scaling factor producing maximal average likelihood was used to warp the frequency axis (Lee and Rose, 1998).

In this setup, acoustic models were trained on 55 adult male speakers and tested on 50 children. The baseline word accuracy is 55.76%. For each child, the adaptation data, which consisted of 1, 4, 7, 10, 13, or 16 digits, were randomly chosen from the test subset to estimate the Sg2 and VTLN warping factors. For comparison, we also evaluated the performance of manually measured Sg2, which in some sense can be viewed as the upper bound of this Sg2 normalization method. For each speaker, the manual Sg2 was measured from only diphthong words containing obvious F2 discontinuities in the spectrum and, independent of adaptation data, the same Sg2 value was applied for normalization. Figure 11 shows the recognition accuracy for VTLN, F3, and Sg2 warpings with various amounts of adaptation data, where Sg2M represents results using the manually measured subglottal resonance.

When the amount of adaptation data is small, Sg2 normalization offers better performance than VTLN. For instance, with only one digit for normalization, Sg2 normalization outperforms VTLN by more than 2%. VTLN outperforms Sg2D2 when more data are available, while the Sg2M provides slightly better performance to VTLN even with 16 adaptation digits. The improvements of Sg2 normalization over VTLN for up to ten adaptation digits are statistically significant for $p<0.05$. Although automatic detection of Sg2 was fairly accurate, it was not exact and there is thus a gap between the performances of the automatic detection method and that of Sg2M. With more accurate Sg2 detection algorithms, we may expect closer performance to that of the manual Sg2.

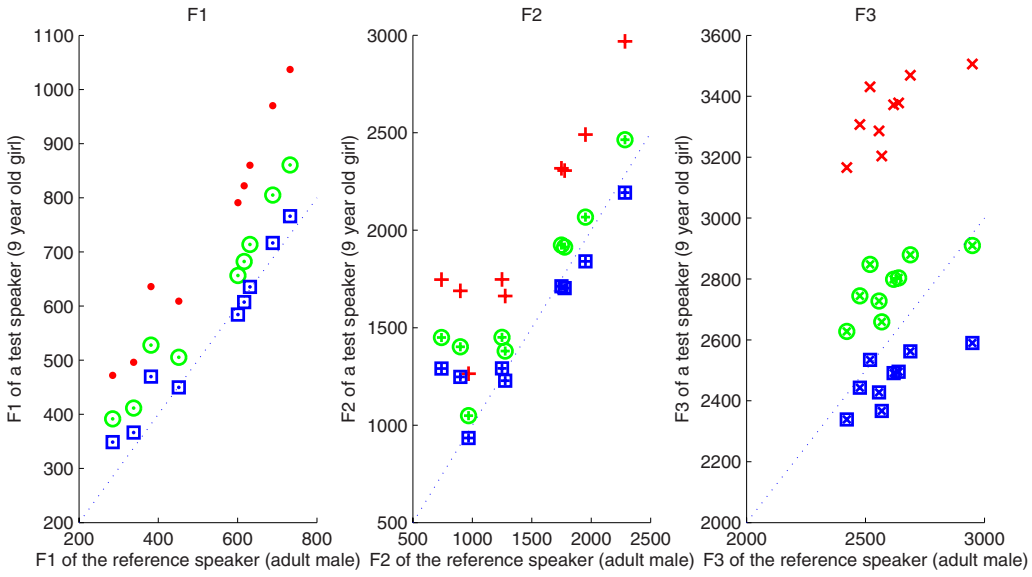


FIG. 10. (Color online) Vowel formants F1 (·), F2 (+), and F3 (×) from the reference speaker (Fig. 4) versus those from the test speaker (Fig. 9) before and after warping (VTLN in circles, Sg2 in squares). The dotted line is $y=x$ which means perfect match between reference and test speakers.

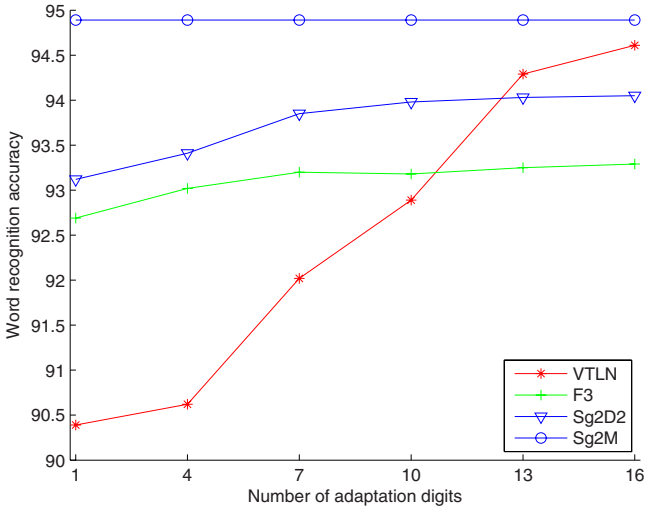


FIG. 11. (Color online) Speaker normalization performance on TIDIGITS with various amounts of adaptation data.

C. Comparison of vowel content dependency

As discussed in Sec. III B, Sg2 is not always detectable from acoustic signals, and thus the Sg2 detectability in adaptation data are important to the normalization performance. It is shown in Wang (2008a) that the normalization performance using Sg2D1 algorithm is highly content-dependent. To investigate the content dependency of the proposed algorithm Sg2D2, we evaluated its normalization performance on TIDIGITS database with one adaptation digit. For each child, the adaptation data were limited to only one digit but with varying vowels from front vowel (e.g., [ɪ] in six), central vowel (e.g., [ʌ] in one), back vowel (e.g., [u] in two) to diphthong (e.g., [aɪ] in five). The adaptation digits were chosen such that F2 discontinuities, if any, come only from vowel contents without any possible interferences from consonant-vowel transitions (Lulich, 2009).

The performance comparison for VTLN, F3, and Sg2 normalizations is shown in Fig. 12. It can be seen that the choice of adaptation data can potentially have an effect on the normalization performance for all three methods. Among them, VTLN is least affected by the choice of adaptation data (the performance standard deviation is 0.55), while F3 normalization is highly data dependent. The performance of Sg2 normalization is less content sensitive compared to F3 normalization, but more content-dependent than VTLN. We expect that the content dependency of Sg2 normalization will decrease with improved Sg2 detection algorithms. In spite of its greater content dependency, on average Sg2 normalization provides better performance than VTLN.

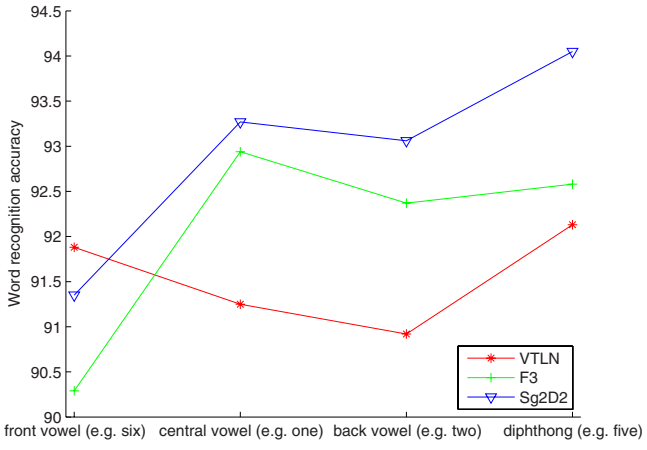


FIG. 12. (Color online) Performance comparison of VTLN, F3 and Sg2D2 using one adaptation digit with various vowel contents.

TABLE III. Performance comparison (word recognition accuracy) on RM1 with one adaptation utterance.

Accuracy	Mismatched	Matched
Baseline	59.10	92.47
F3	79.01	92.58
VTLN	86.65	93.91
Sg2	88.37	94.05

D. Performance on RM1 database

Since the TIDIGITS setup is a highly mismatched case, the experiments demonstrate the effectiveness of subglottal resonance-based speaker normalization. To further verify the effectiveness of this method, we also test the performance on a medium vocabulary recognition task using the DARPA Resource Management RM1 continuous speech database. As a next step, we tested the method on the RM1 database for both a medium-mismatched case and a matched case. Triphone acoustic models were applied with three states and four Gaussian mixtures per state using the same features as in the TIDIGITS experiments. For the mismatched case, HMM models were trained on 49 adult male speakers from the speaker independent (SI) portion of the database, and tested on 23 adult female speakers in the SI portion. The baseline word recognition accuracy was 59.10%. For the regular test on RM1, the HMM models were trained on the SI training portion of the database with 72 adult speakers, and tested on the SI testing set. The baseline performance was 92.47% word recognition accuracy. In both cases, the same utterance was used to estimate the Sg2 and VTLN warping factor for all speakers. Table III shows the results.

For the mismatched case, Sg2 normalization provides better performance than VTLN with about 1.5% absolute improvement. This improvement is statistically significant for $p<0.01$. For the matched case, Sg2 normalization provides comparable performance to that of VTLN. From the computation point of view, Sg2 normalization is more efficient than VTLN, since VTLN relies on an exhaustive grid search over the warping factors to maximize the likelihood of the adaptation data, while for Sg2 normalization the main computational cost comes from formant tracking which can be estimated efficiently.

E. Cross-language speaker normalization

The language-independent property of Sg2 makes cross-language adaptation possible based on Sg2 normalization. In our experiments, training and test data were in English, while the adaptation data were in either English or Spanish. The warping factors were estimated from the adaptation data us-

TABLE IV. Performance comparison (word recognition accuracy) of VTLN and Sg2 normalization using English (four words) and Spanish (five words) adaptation data. The acoustic models were trained and tested using English data.

	Language of adaptation data	
Method	English	Spanish
VTLN	86.61	82.35
Sg2	86.59	85.97

ing Sg2D2 and applied to the test data to warp the spectrum. English adaptation data were collected for comparison.

The performance was evaluated on the Technology Based Assessment of Language and Literacy (TBall) project database (Kazemzadeh *et al.*, 2005), and the English high frequency words for first and second grade students were used in the test. Monophone acoustic models were trained on speech data from native English speakers. The test data were from the same 20 speakers as in the ChildSE. The ChildSE utterances (only one repetition) were used as adaptation data, and for each speaker there were four English words and five Spanish words for adaptation.

The typical text-dependent VTLN method using HMM recognizers for warping factor searching is not quite suitable in this scenario, because decoding Spanish speech with English phoneme models could itself introduce a systematic error due to different phonetic characteristics between these two languages. Instead, for a fair and reasonable comparison, text-independent VTLN is applied, which uses Gaussian mixture models (GMMs) for warping factor searching. A GMM with 512 mixtures was trained on English training set, and then applied to calculate the likelihood for each warping factor in the range [0.8, 1.2] with a step size of 0.01. The warping factor with the highest likelihood was chosen as the VTLN warping factor. Compared to the text-dependent VTLN used in Wang (2008), this text-independent method provides similar performance with English adaptation data, but much better for Spanish adaptation data. The subglottal resonance was estimated using Sg2D2 for each word, and the average was used as the speaker's Sg2 frequency. The Sg2 warping factor was calculated using Eq. (5).

The normalization performance is shown in Table IV for VTLN and Sg2 using English and Spanish adaptation data. When adaptation data are in English, which is the same language as for the acoustic models, Sg2 normalization and VTLN give comparably good results. For Spanish adaptation data, however, the performance of VTLN degrades, while the performance of Sg2 normalization remains similar as for English adaptation data. Sg2 normalization, therefore, produces more robust results than VTLN when performing cross-language adaptation. The performance difference between using Sg2D2 and using VTLN is statistically significant with Spanish adaptation data for $p < 0.01$.

VI. SUMMARY AND DISCUSSION

This paper presents a reliable algorithm for estimating the second subglottal resonance (Sg2) from acoustic signals. The algorithm provides Sg2 estimates very close to actual Sg2 values as determined from direct measurements using accelerometer data. With the proposed algorithm, Sg2 standard deviation over contents and languages was investigated with children's data for English and Spanish. Analysis shows that for a given speaker, the second subglottal resonance does not appear to vary with speech sounds, repetitions, and even across languages. Based on such observations, a speaker normalization method is proposed using the second subglottal resonance. This normalization method defines the warping factor as the ratio of the reference subglottal resonance over that of the test speaker.

A variety of evaluations show that the second subglottal resonance normalization performs better than or comparable to VTLN, especially for limited adaptation data. An obvious advantage of this method is that the subglottal resonances remain roughly constant for a specific speaker. This method is potentially independent of the amount of available adaptation data, which makes it suitable for limited data adaptation.

Cross-language experimental results shows that Sg2 normalization is more robust across languages than VTLN, and no significant performance variations are observed for Sg2 when the adaptation data are changed from English to Spanish. The fact that Sg2 is independent of language should make it possible to adapt acoustic models with available data from any language. The method is also computationally more efficient than VTLN.

The Sg2 variations found in this paper are similar to what has been reported elsewhere. However, given the small number of subglottal resonance studies, more data may need to be collected and analyzed in order to refine the characterization of subglottal resonance variability. For future work, we will further improve the accuracy of the Sg2 detector, evaluate the effectiveness of this method on a large vocabulary database, and test the performance in noisy conditions.

ACKNOWLEDGMENT

This work was supported in part by NSF Grant No. 0326214

[1]The place of articulation feature [+/−back] specifies the tongue positions during speech production: [+back] segments are produced with the tongue dorsum bunched and retracted slightly to the back of the mouth, while [−back] segments are bunched and extended slightly forward.

[2]The manual Sg2's were estimated through visually examining the speech spectrogram, and then applying Eq. (2) or Eq. (3) depending on the existence of F2 discontinuities.

Cheyne, H. A. (**2001**). "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Ph.D. thesis, MIT, Cambridge, MA.

Chi, X., and Sonderegger, M. (**2004**). "Subglottal coupling and vowel space," J. Acoust. Soc. Am. **115**, 2540.

Chi, X., and Sonderegger, M. (**2007**). "Subglottal coupling and its influence on vowel formants," J. Acoust. Soc. Am. **122**, 1735–1745.

Claes, T., Dologlou, I., Bosch, L., and Compernolle, D. V. (**1998**). "A novel feature transformation for vocal tract length normalization in automatic speech recognition," IEEE Trans. Speech Audio Process. **11**, 549–557.

Cui, X., and Alwan, A. (**2006**). "Adaptation of children's speech with limited data based on formant-like peak alignment," Comput. Speech Lang. **20**, 400–419.

Eide, E., and Gish, H. (**1996**). "A parametric approach to vocal tract length normalization," in *Proceedings of ICASSP*, pp. 346–349.

Gouvea, E., and Stern, R. (**1997**). "Speaker normalization through formant-

based warping of the frequency scale," in *Proceedings of Eurospeech*, pp. 1139–1142.
Hanson, H., and Stevens, K. N. (**1995**). "Subglottal resonances in female speakers and their effect on vowel spectra," in Proceedings of 13th International Congress of Phonetic Sciences, Stockholm, Vol. **3**, pp. 182–185.
Honda, K., Takano, S., and Takemoto, H. (**2009**). "Effects of side cavities and tongue stabilization: Possible extensions of quantal theory," J. Phonetics In Press. Doi: 10.1016/j.wocn.2008.11.002.
Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., and Johnson, K. (**1999**). "Formants of children women and men: The effect of vocal intensity variation," J. Acoust. Soc. Am. **106**, 1532–1542.
Jung, Y. (**2008**). "Acoustic articulatory evidence for quantal vowel categories across languages," poster presented at the Harvard-MIT HST Forum.
Jung, Y., Lulich, S. M., and Stevens, K. (**2008**). "Development of subglottal quantal effects in young children," J. Acoust. Soc. Am. **124**(4), 2519.
Kazemzadeh, A., You, H., Iseli, M., Jones, B., Cui, X., Heritage, M., Price, P., Anderson, E., Narayanan, S., and Alwan, A. (**2005**). "TBall data collection: The making of a young children's speech corpus," in *Proceedings of Eurospeech*, pp. 1581–1584.
Lee, L., and Rose, R. (**1998**). "A frequency warping approach to speaker normalization," IEEE Trans. Speech Audio Process. **6**, 49–60.
Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.
Lulich, S. M. (**2006**). "The role of lower airway resonances in defining vowel feature contrasts," Ph.D. thesis, MIT, Cambridge, MA.
Lulich, S. M. (**2009**). "Subglottal resonances and distinctive features," J. Phonetics In press. Doi: 10.1016/j.wocn.2008.10.006.
Lulich, S. M., Bachrach, A., and Malyska, N. (**2007**). "A role for the second subglottal resonance in lexical access," J. Acoust. Soc. Am. **122**, 2320–2327.
Lulich, S. M., Zañartu, M., Mehta, D. D., and Hillman, R. E. (**2009**). "Source-filter interaction in the opposite direction: Subglottal coupling and the influence of vocal fold mechanics on vowel spectra during the closed phase," J. Acoust. Soc. Am. **125**(4), 2638.
Madsack, A., Lulich, S. M., Wokurek, W., and Dogil, G. (**2008**). "Subglottal resonances and vowel formant variability: A case study of high German monophthongs and Swabian diphthongs," Lab. Phon. **11**, 91–92.
McDonough, J. (**2000**). "Speaker compensation with all-pass transforms," Ph.D. thesis, Johns Hopkins University, Baltimore, MD.
McDonough, J., Shaaf, T., and Waibel, A. (**2004**). "Speaker adaptation with all-pass transforms," Sov. Phys. Crystallogr. **42**, pp. 75–91.
Pitz, M., and Ney, H. (**2003**). "Vocal Tract Normalization as Linear Transformation of MFCC," in Proceedings of Eurospeech, pp. 1445–1448.
Snack Sound Toolkit (**2005**). http://www.speech.kth.se/snack/ (Last viewed August, 2008).
Sonderegger, M. (**2004**). "Subglottal coupling and vowel space: An investigation in quantal theory," thesis, MIT, Cambridge, MA.
Stevens, K. N. (**1998**). *Acoustic Phonetics*, MIT, Cambridge, MA.
Umesh, S., Zolnay, A., and Ney, H. (**2005**). "Implementing frequency-warping and VTLN through linear transformation of conventional MFCC," in Proceedings of Interspeech, pp. 269–272.
Wang, X., Wang, B., and Qi, D. (**2004**). "A bilinear transform approach for vocal tract length normalization," in Proceedings of ICARCV, pp. 547–551.
Wang, S., Cui, X., and Alwan, A. (**2007**). "Speaker adaptation with limited data using regression-tree based spectral peak alignment," IEEE Trans. Audio, Speech, Lang. Process. **15**, pp. 2454–2464.
Wang, S., Alwan, A., and Lulich, S. M. (**2008a**). "Speaker normalization based on subglottal resonances," in Proceedings of ICASSP, pp. 4277–4280.
Wang, S., Lulich, S. M., and Alwan, A. (**2008b**). "A reliable technique for detecting the second subglottal resonance and its use in cross-language speaker adaptation," in Proceedings of Interspeech, pp. 1717–1720.
Wegmann, S., McAllaster, D., Orloff, J., and Peskin, B. (**1996**). "Speaker normalization on conversational telephone speech," in Proceedings of ICASSP, Vol. **1**, pp. 339–341.
Zhan, P., and Westphal, M. (**1997**). "Speaker normalization based on frequency warping," in Proceedings of ICASSP, pp. 1039–1041.

Surface response of a fractional order viscoelastic halfspace to surface and subsurface sources

F. Can Meral and Thomas J. Royston[a)]
Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, Illinois 60607

Richard L. Magin
Bioengineering, University of Illinois at Chicago, Chicago, Illinois 60607

(Received 12 June 2009; revised 3 September 2009; accepted 6 September 2009)

Previous studies by the second author published in this journal focused on low audible frequency (40–400 Hz) shear and surface wave motion in and on a viscoelastic material representative of biological tissue. Specific cases considered were that of surface wave motion on a halfspace caused by a finite rigid circular disk located on the surface and oscillating normal to it [Royston *et al.*, J. Acoust. Soc. Am. **106**, 3678–3686 (1999)] and compression, shear, and surface wave motion in a halfspace generated by a subsurface finite dipole [Royston *et al.*, J. Acoust. Soc. Am. **113**, 1109–1121 (2003)]. In both studies, a Voigt model of viscoelasticity was assumed in the theoretical treatment, which resulted in agreement between theoretical predictions and experimental measurements over a limited frequency range. In the present article, the linear viscoelastic assumption in these two prior works is revisited to consider a (still linear) fractional order Voigt model, where the rate-dependent damping component that is dependent on the first derivative of time is replaced with a component that is dependent on a fractional derivative of time. It is shown that in both excitation source configurations, the fractional order Voigt model assumption improves the match of theory to experiment over a wider frequency range (in some cases up to the measured range of 700 Hz). © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3242351]

PACS number(s): 43.80.Cs, 43.80.Ev, 43.20.Jr [FD] Pages: 3278–3285

I. INTRODUCTION

An improved understanding of mechanical wave propagation in viscoelastic materials and structures could fundamentally catalyze technical developments in many areas including medicine, geophysics, infrastructure, and manufacturing. For example, it could lead to improved medical imaging[1] (e.g., dynamic elastography), improved land mine,[2] and other buried object detection,[3] earthquake and tsunami monitoring and analysis,[4–6] intelligent bridge and civil structure assessment,[7] and improved nondestructive testing of man-made materials and components,[8,9] at microscopic and macroscopic scales. In medical imaging, in particular, noninvasive measurement of shear wave motion in soft biological tissues can provide unique spatially localized information about the tissue's material properties. Such information can reflect the development of pathology and in some cases biomechanical integrity. Additionally, it may provide a unique way to nondestructively track the morphogenesis of engineered tissues. Over the past few decades, mechanical wave propagation has been combined with the medical imaging modalities of ultrasound or magnetic resonance (MR) imaging to establish noninvasive means of visualizing shear wave motion for diagnosis or nondestructive tissue assessment.[10–14] These dynamic elastography imaging techniques, as they are commonly referred, have received much attention because shear moduli have a large dynamic range within soft biological tissues, particularly when compared with the variation in material parameters that provide contrast in conventional imaging techniques. For soft biological tissues (e.g., comparing muscle with fat), the x-ray attenuation coefficient varies only by a factor of 2,[15] while MR relaxation times vary by a factor of 3.[16] The shear moduli, on the other hand, can vary by more than a factor of 10.[17–19]

Estimating material properties based on dynamic elastography images is called inversion or reconstruction. Different techniques have been proposed including, for example, "local frequency estimation,"[20] "algebraic inversion of the differential equation,"[21] other "variational" methods,[22,23] and "finite element model reference" based methods[24,25] to name a few. Additionally, focused modulated radiation force of ultrasound at multiple frequencies has been used to remotely drive an embedded hard target[26] and to generate shear waves[27,28] in order to estimate the medium's viscoelastic properties based on measurements of the target's motion or the speed of shear wave propagation as a function of frequency. These techniques often assume a constitutive relation for the tissue viscoelasticity, such as a Voigt model, to aid in interpreting measurements. Studies have shown that such viscoelastic models have limitations in their ability to accurately model dynamic phenomena over multiple time scales and/or with broad spectral content, particularly for biological tissues and tissue mimicking phantoms, and that one way of overcoming such limitations is through the use of fractional order models.[29–33] Fractional order viscoelastic modeling starts with the idea from fractional calculus that the order of the derivative of the strain can be intermediate between 0 and 1 since it is the derivative of the strain that characterizes the material's behavior (assuming a one dimensional stress-strain relation). Such an approach offers a new

[a)]Author to whom correspondence should be addressed. Electronic mail: troyston@uic.edu

viscoelastic stress-strain model. When the order of the derivative is zero, it represents a Hookean solid and when it is of order 1, it describes a Newtonian fluid. Viscoelastic materials occupy the intermediate range with a fractional order "α" between 0 and 1. Using this approach, it is possible to build a multi-component fractional equivalent of the "standard linear solid" (SLS) by replacing one or more springs and dashpots with "springpots." Such models are linear and have shown the potential to yield new disease and treatment specific parameters that more effectively predict underlying changes in tissue associated with developing pathology, such as liver cirrhosis and breast cancer. As an example, in Ref. 33 a relatively simple power law relationship was fitted to the complex shear modulus of human breast tissue and tumors measured by magnetic resonance elastography. The results, when plotted as the fractional power exponent versus the fractional order attenuation, separated benign from malignant tumors with an increase in specificity and sensitivity.

In previous studies by the second author of the present article[34,35] there has been an emphasis on understanding the shear wave field created in a material like biological tissue by canonical vibratory sources. In Ref. 34, a new analytical solution was derived for the problem of surface wave generation on a linear viscoelastic halfspace caused by a finite rigid circular disk located on the surface and oscillating normal to it. While the motivation of the work was to better understand surface wave propagation in biological tissue, the solution approach taken was an incremental advancement of theoretical work reported in seminal articles in the geophysics literature.[5,6] The improved solution was tested experimentally using a viscoelastic phantom with material properties comparable to biological soft tissue. Some agreement could be achieved over a limited frequency range (20–100 Hz) using a Voigt model. Another study[35] analyzed compression, shear, and surface waves in a viscoelastic halfspace generated by subsurface fundamental acoustic sources. Finite and infinitesimal monopole and dipole sources within the low audible frequency range (40–400 Hz) were considered theoretically, computationally and, in some cases, experimentally for a soft tissue phantom material. Again, a Voigt model was assumed, which produced results that matched experiment over a limited frequency band.

In the present article, in order to improve material characterization over a broader frequency range, we revisit the viscoelastic modeling assumption in these two prior works[34,35] to consider fractional order components. Specifically, our objectives are to reconsider wave motion on the surface of a viscoelastic medium (halfspace) created by finite dimension surface and subsurface vibratory sources oscillating normal to the planar surface of the medium. Theoretical predictions for the case of a material with fractional viscoelastic properties are derived and compared with experimental studies. Their ability to match experimental measurements is compared with the conventional (integer order) Voigt and standard linear solid (also known as Kelvin or Zener) models.

II. FRACTIONAL VISCOELASTIC CONTINUUM: GOVERNING EQUATIONS

For an isotropic, homogenous, viscoelastic compressible medium, one can use the following formulation of the equation of motion for small perturbations about an operating point

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

Here, $\mathbf{u}$ is the displacement vector, ρ is the density of the medium, $\partial/\partial t$ denotes a derivative with respect to time, ∇ is the spatial Laplacian operator dependent on the chosen coordinate system, and λ and μ are the Lamé constants of the medium. For a linear viscoelastic Voigt material model, the rate-dependent Lamé "constants" are expressible as $\lambda(t) = \lambda_0 + \lambda_1\partial/\partial t$ and $\mu(t) = \mu_0 + \mu_1\partial/\partial t$, where λ_0, λ_1, μ_0, and μ_1 are coefficients of volume compressibility, volume viscosity, shear elasticity, and shear viscosity, respectively.[36] Other shear viscoelastic models will lead to different rate-dependence relations.

With regard to μ, it has been observed in many materials that the simple two-element Voigt model for shear viscoelasticity (μ_0, μ_1) does not accurately capture material shear dynamic behavior, in terms of its experimentally measured response to various elementary excitation waveforms, such as step inputs or periodic or random inputs with broad spectral content. More complex arrangements of multiple elastic (springs) and viscous (dashpot) components may then be employed empirically in order to more closely match what is observed. For example, the SLS model, also known as the Kelvin or Zener model, consists of a parallel combination of a Maxwell element (spring and dashpot in series) with a spring. The three-element SLS model has more flexibility in representing dynamic viscoelasticity as compared to the Voigt model.[18] Instead of increasing the constitutive model complexity by increasing the number of components that comprise it, an alternative is to consider that the material may exhibit rate-dependent shear deformation that is best described by a single element, comprised of two constants, μ_α and α, whose behavior lies somewhere between Hookean solid and Newtonian fluid. Specifically, fractional order viscoelasticity (a springpot) can be specified as shown in the second term of the following:

$$\mu = \mu_0 + \mu_\alpha\frac{\partial^\alpha}{\partial t^\alpha}, \quad 0 < \alpha \leq 1. \tag{2}$$

Henceforth, Eq. (2) will be referred to as a fractional order Voigt model for $\alpha < 1$. While such a mathematical construction may seem to lack physical meaning, it can be shown that this type of relation results asymptotically when using a ladder-like fractal arrangement of integer-order elastic and viscous components, as depicted in Fig. 1.[37] Indeed, such an arrangement might be rationalized on the grounds that it represents multiscale rate-dependent stress-strain interactions that one would inherently expect in some materials with complex multiscale cellular and extracellular structure, such as biological tissues. Furthermore, suitably defined fractional derivatives do not pose significant difficulty mathematically

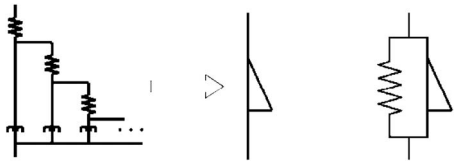


FIG. 1. A tree arrangement of springs and dashpots (left) resulting in a fractional order element, called a springpot (middle). Schematic representation of a fractional order Voigt model (right); the dashpot is replaced with a springpot.

for well-conditioned functions. [In this paper, we have chosen to use the Weyl definition of the fractional order derivative, which for harmonic functions such as $f(t)=e^{j\omega t}$, has the property that $\partial^\alpha/\partial t^\alpha[e^{j\omega t}]=(j\omega)^\alpha e^{j\omega t}$.] The expression in Eq. (2) is still linear in nature and thus all rules and techniques afforded such relations, such as the validity of superposition, reciprocity, the Laplace and Fourier transforms, with associated transfer and frequency response functions, are all still valid.[37] In the Laplace (s) and frequency ($j\omega$) domains where $j=\sqrt{-1}$ and ω is the circular frequency, Eq. (2), respectively, becomes

$$\mu = \mu_0 + \mu_\alpha (s)^\alpha, \tag{3a}$$

$$\mu = \mu_0 + \mu_\alpha (j\omega)^\alpha. \tag{3b}$$

Note that a significant attribute of such fractional representations is that the temporal response takes on characteristics of power-law behavior as opposed to the exponential response that one obtains with the conventional Voigt representation. A power-law response, in fact, has been observed in a number of biological and nonbiological materials, further motivating this type of model.[30,31]

Regardless of whether an "integer order" or fractional order Voigt model or a standard linear solid model is used, wave motion in the *infinite* three-dimensional viscoelastic medium consists of a superposition of dilatational and shear wave displacements, $\mathbf{u}=\mathbf{u_P}+\mathbf{u_S}$, respectively. For the *semi-infinite* halfspace problem, an additional surface (Rayleigh) wave $\mathbf{u_{Su}}$ will exist for the configurations considered in Sec. III.

III. SURFACE WAVE PROPAGATION ON A HALFSPACE DUE TO A SURFACE SOURCE

A. Theory

In Ref. 34, a simplified analytical solution is derived for Rayleigh wave propagation on the surface of an isotropic homogeneous viscoelastic halfspace caused by normal force excitation over a circular region of radius a on the surface of amplitude per unit area P_{in} with harmonic time dependence $e^{j\omega t}$ as depicted in Fig. 2. The analytical solution is

$$\frac{u_z}{P_{\text{in}}} = -\frac{2a}{\mu}\frac{J_1(pak_p)\sqrt{p^2-1}}{F_0'(-p)}\eta^2 K_0(jprk_p)e^{j\omega t}, \tag{4a}$$

where

$$F_0'(-p) = \left.\frac{\partial F_0}{\partial \zeta}\right|_{\zeta=-p}, \tag{4b}$$

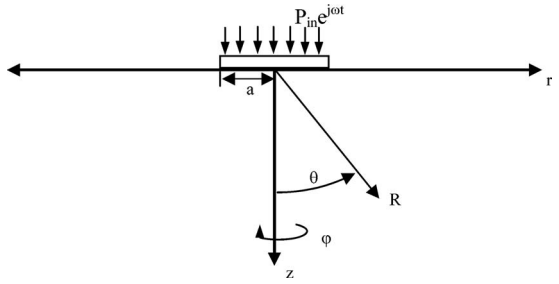


FIG. 2. Ideal viscoelastic halfspace with finite surface source.

$$F_0(\zeta) = (2\zeta^2-\eta^2)^2 - 4\zeta^2\sqrt{\zeta^2-\eta^2}\sqrt{\zeta^2-1}, \tag{4c}$$

$$\eta = \sqrt{(\lambda+2\mu)/\mu}, \tag{4d}$$

$$k_p = \omega\sqrt{\rho/(\lambda+2\mu)}. \tag{4e}$$

Here, u_z is out-of-plane surface displacement, p is the ratio of compression wave speed to surface wave speed and is a root of the function F_0 that is associated with Rayleigh wave motion, k_p is the compression wave number, r is the radial distance from center of the driving disk, J_1 is the Bessel function of the first kind (order 1), and K_0 is the modified Bessel function of the second kind (order 0); K_0 can also be written in terms of Bessel functions of the first and second kinds (order 0) such as $K_0(x)=(\pi/2)i\{J_0(ix)+iY_0(ix)\}$. Equation (4c) links compression, shear, and surface wave behaviors to material viscoelastic properties; the roots of this equation yield compression, shear, and surface wave numbers.

Shear wave speed velocity at frequency ω is related to the real (storage) and imaginary (loss) parts of the shear modulus, μ_R and μ_I respectively, and the material density ρ as[14]

$$c_s = \sqrt{\frac{2}{\rho}\frac{\mu_R^2+\mu_I^2}{\mu_R+\sqrt{\mu_R^2+\mu_I^2}}}. \tag{5}$$

If surface wave speed and attenuation are experimentally measured, material properties that affect shear and surface wave speed can be estimated using Eqs. (4a)–(4e) and (5). Note that both μ_R and μ_I are independent of whether the time derivative part of the Voigt model is of integer or fractional order. They are equal to μ_0 and $\omega\mu_1$, shear elasticity and shear viscosity multiplied with circular frequency, if a conventional integer order Voigt model is used. In the case of a fractional order Voigt model since $(j\omega)^\alpha=\omega^\alpha(\cos[\alpha\pi/2]+j\sin[\alpha\pi/2])$, the storage modulus and loss modulus are defined, respectively, as

$$\mu_R = \mu_0 + \mu_\alpha\omega^\alpha\cos\left(\frac{\pi}{2}\alpha\right), \tag{6a}$$

$$\mu_I = \mu_\alpha\omega^\alpha\sin\left(\frac{\pi}{2}\alpha\right). \tag{6b}$$

Alternatively, the SLS model yields

$$\mu_R = \frac{\mu_0\mu_\omega^2+\omega^2\mu_1^2(\mu_0+\mu_\omega)}{\mu_\omega^2+\omega^2\mu_1^2}, \tag{7a}$$

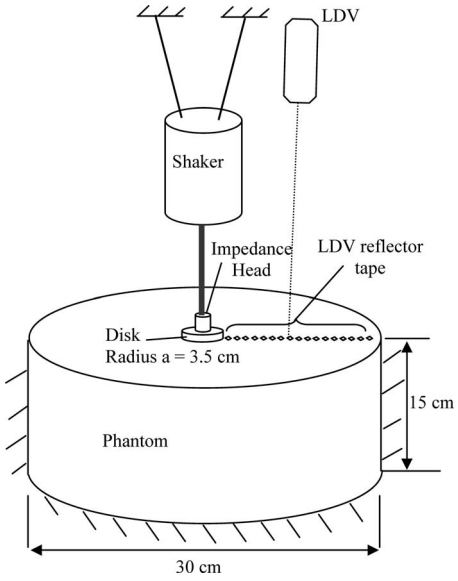


FIG. 3. Experimental schematic for measurement of surface wave motion caused by a surface source using a LDV.

$$\mu_I = \frac{\omega \mu_\omega^2 \mu_1}{\mu_\omega^2 + \omega^2 \mu_1^2}. \tag{7b}$$

Here μ_0 denotes the static stiffness, μ_1 denotes the viscous damping coefficient multiplied with the first order time derivative of the displacement (thus α is equal to 1), and μ_ω denotes the dynamic stiffness, which is only effective when the loading has a non-zero time derivative.

B. Experiment

Surface wave experiments were conducted as depicted in Fig. 3 using a silicone polymer, CF-11 (NuSil Technology, Carpinteria CA), which had a density of 1100 kg/m^3 calculated through basic mass volume measurements of small test specimens. While in liquid form, the material is poured into the container and then cures at room temperature. The container is mounted on a vibration isolated optics bench. A plexiglass disk, driven by a mechanical shaker (ET-132, LabWorks Inc., Mesa Costa, CA) that is supported by a separate structure, is positioned on the surface of the phantom with a sufficient preload to ensure contact during excitation. The shaker is driven via an amplifier (Type 2076, Bruel & Kjaer, Denmark) with a signal input from a dynamic signal analyzer (35670A, Agilent Technologies, Santa Clara, CA). The force and acceleration of the disk are measured with an impedance head (288D01, PCB Piezotronics, Depew, NY), and the out-of-plane velocity at discrete points on the surface is measured using a laser Doppler vibrometer (LDV) (CLV-800, Polytec, Tustin, CA). Small ($\sim$2 mm^2) pieces of 3M retroreflective tape are mounted on the semi-translucent phantom material to aid in LDV measurement. Measurement signals are recorded and the frequency response function (FRF) between the output (vertical velocity of the surface points) and input (motion input of the disk) is calculated by the dynamic signal analyzer. Further analysis of the data is conducted in MATLAB® (Mathworks, Natick, MA).

Additionally, a static measurement of the phantom material stiffness was made by indenting a steel sphere of diameter 9.525 mm into the media. Indentation forces were measured for different indentation depths using a force gauge (Model DPS, Imada, Northbrook, IL). Indentation of the steel ball into the semi-infinite medium was assumed to be a Hertzian contact problem and the solution given by Timoshenko[38] was used:

$$a = \sqrt[3]{\frac{9\pi^2}{16} \frac{P^2 (k_1 + k_2)^2 (R_1 + R_2)}{R_1 R_2}}, \tag{8}$$

where

$$k_i = \frac{1 - \nu_i^2}{\pi E_i}, \quad i = 1,2. \tag{9}$$

Here, P is the indentation force, E_i and ν_i are Young's modulus and Poisson's ratio, respectively, of the two materials in contact, a is the indentation depth, and R_1 and R_2 are the radii of two spherical bodies; in our case one of the bodies was an infinite halfspace, yielding $R_1 = \infty$. Also the indenter ball was very stiff compared to the medium being indented; $E_2 = \infty$. These assumptions simplify the equation so that Young's modulus of the halfspace could be calculated using

$$E_1 = \sqrt{\frac{9}{16} \frac{P^2 (1 - \nu_1^2)^2}{a^3 R_2}}. \tag{10}$$

Given that the halfspace material is nearly incompressible ($\nu \approx 0.5$), we then have $\mu_0 = E_1/3$.

C. Results and discussion

Due to the finite size of the viscoelastic phantom (it is not an infinite halfspace), resonant behavior is observed at lower frequencies. Also, surface velocity measurement data from points further than 10 cm from the source cannot be used due to the high attenuation and/or reflections from the boundaries. The amplitude and phase information of the complex frequency response data are used to calculate the surface wave speed and rate of attenuation at different frequencies from the remaining points. The real part of the surface wavenumber, pk_p, is obtained from the surface wave phase speed. Based on the term $K_0(jprk_p)$ in Eqs. (4a)–(4e), an estimate of complex value p is then obtained as a function of ω. This is then used with Eqs. (4a)–(4e) to calculate the complex shear wave number, ηk_p. From this, one then can calculate the real and imaginary parts of the complex shear modulus, μ_R and μ_I, respectively, as a function of frequency. Results are shown in Fig. 4.

The conventional Voigt model, SLS model, and the fractional order Voigt model for shear viscoelasticity were used to fit the experimental data. Volume elasticity, λ_0, primarily affects compression wave speed and should not affect shear or surface wave speed. Nonetheless, an approximate value was needed for Eqs. (4a)–(4e). Given that the speed of sound (compression waves) in CF-11 was comparable to water, a value of $\lambda_0 = 2.9$ GPa was used. Volume viscosity, λ_1, is assumed negligible for the low frequencies considered in this study.[36] A value for shear elasticity μ_0, which is common to all three shear viscoelasticity models (Voigt, SLS, fractional

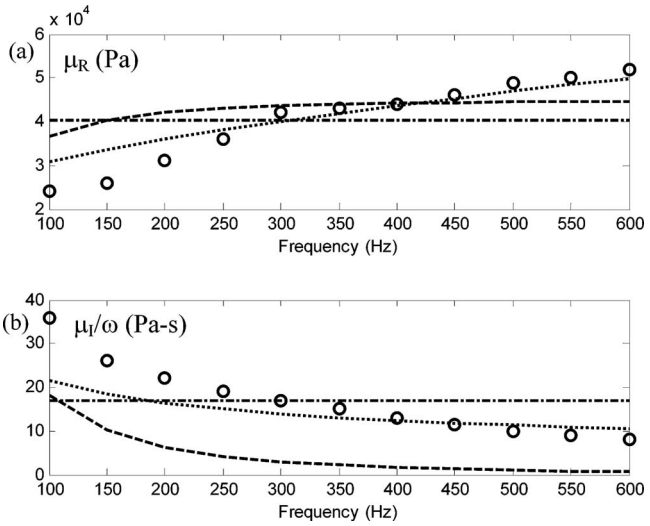


FIG. 4. CF-11 complex shear modulus $\mu=\mu_R+j\mu_I$ as a function of frequency: (a) real part; (b) imaginary part/ω. (○ ○ ○) experimentally derived values; (---) best fit Voigt model; (···) best fit $\alpha=0.6$ fractional order Voigt model; (---) best fit SLS model.

Voigt), was available from the static indentation test described above and is provided in Tables I and II. That value is used in the SLS and fractional order Voigt models, then leaving two parameters in each of those models to be optimized based on the dynamic experimental surface wave data. For the conventional Voigt model, a different value for μ_0 is calculated by determining the average value of the real part of the shear modulus, μ_R, based on the experimental measurements over the frequency range shown in Fig. 4. Similarly, for the conventional Voigt model μ_1 is determined by matching it to the average value of the imaginary part of the shear modulus, μ_I, divided by frequency ω based on the experimental measurements over the frequency range shown in Fig. 4. For the SLS model, note in Eq. (7a) that the value of μ_R approaches $\mu_0+\mu_\omega$ as ω increases; thus, μ_ω was chosen by subtracting μ_0 from the average of μ_R at higher frequencies (>500 Hz). Then, μ_1 was chosen to minimize the least squares error for both μ_R and μ_I based on the experimental measurements over the frequency range shown in Fig. 4. The values of μ_ω and μ_1 yielding the best overall fit are shown in Fig. 4 and in Table I. For the fractional order Voigt model, α was varied in increments of 0.05 between 0 and 1 and μ_α was then chosen to minimize the least squares error for both μ_R and μ_I based on the experimental measurements over the frequency range shown in Fig. 4. The values of α and μ_α yielding the best overall fit are shown in Fig. 4 and in Table I.

TABLE I. CF-11 best fit viscoelastic constants for surface wave experiments for Voigt, fractional order Voigt, and SLS models. (For the fractional order Voigt and SLS models, μ_0 is the value obtained from the static indentation test. For the Voigt model, μ_0 is the average value of real part of complex μ in the frequency range of interest.)

	$\alpha=1$	$\alpha=0.6$	SLS, $\alpha=1$
μ_0 (Pa)	40 300	21 000	21 000
μ_α (Pa s$^\alpha$)	17	351	51
μ_ω (Pa)	···	···	24 000

TABLE II. CF-11 best fit viscoelastic constants for dipole experiments for Voigt, fractional order Voigt, and SLS models. (For the fractional order Voigt and SLS models, μ_0 is the value obtained from the static indentation test. For the Voigt model, μ_0 is the average value of real part of complex μ in the frequency range of interest.)

	$\alpha=1$	$\alpha=0.6$	SLS, $\alpha=1$
μ_0 (Pa)	47 100	21 000	21 000
μ_α (Pa s$^\alpha$)	33.7	553	70
μ_ω (Pa)	···	···	40 000

In Fig. 4, the increasing trend of the storage modulus (μ_R) with frequency is evident; this behavior is not represented by the conventional Voigt model for which $\mu_R=\mu_0$, a constant. The ratio μ_I/ω is also plotted and has a decreasing frequency trend, though for the conventional Voigt model $\mu_I/\omega=\mu_1$ is a constant. While the optimized SLS model does have these frequency-dependent trends in μ_R and μ_I/ω, as shown in Fig. 4, the optimized fractional order Voigt model was better able to track the frequency dependence of both the storage and loss moduli. As noted above, for both SLS and fractional order Voigt models, the value of μ_0 was independently fixed based on the static indentation test, leaving two parameters in each of these models to be optimized based on the dynamic surface wave experimental data. Of course, increasing the number of parameters that can be adjusted in either integer or fractional order models improves the fit to experiment. The point made here is that the fractional order Voigt model, with two parameters, α and μ_α, optimized based on the dynamic data, provides a better fit than the SLS model, with two parameters, μ_1 and μ_ω, optimized based on the same dynamic experimental data.

IV. SURFACE MOTION ON A HALFSPACE DUE TO A SUBSURFACE (DIPOLE) SOURCE

A. Theory

In 1955, Pekeris[39] derived integral expressions for wave motion due to the buried vertical *infinitesimal* dipole (a point force) in a halfspace. With respect to the problem of a *finite* dipole in a semi-infinite viscoelastic or elastic halfspace, as depicted in Fig. 5, we have not found a closed-form analytical solution in the literature. However, an approximate approach is to take the solution for a dipole in an infinite me-

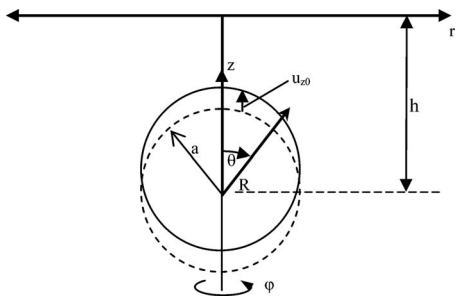


FIG. 5. Ideal viscoelastic halfspace with finite subsurface (dipole) source.

dium and then double its predicted value at the actual location of the free surface. The dipole source produces both dilatational (compression) and shear wave motion; however, in addition to conversions between dilatational and shear wave types, which will occur at the surface, Rayleigh (surface) waves are not accounted for in such an approximation. In prior work of the second author, it was shown that for a material comparable to soft biological tissue, this was an accurate approximation for both infinitesimal and finite dipole sources relatively near the free surface.[35] This approach is adapted here for the finite dipole case, but now with the consideration of the SLS and fractional order Voigt viscoelastic models.

Consider a rigid sphere of radius a embedded in an isotropic viscoelastic medium that is executing rectilinear motion along the z axis given by $u_z = u_{z0}e^{j\omega t}$, as depicted in Fig. 5. This gives rise to dilatational and shear waves in the surrounding medium. Because of axisymmetry, $u_\phi = 0$ and the displacement vector on the sphere surface is $\mathbf{u} = [u_R, 0, u_\theta]^T$, in spherical coordinates where $u_R = u_R(R, \theta, t)$ and $u_\theta = u_\theta(R, \theta, t)$. The resulting displacement fields u_R and u_θ may be expressed as

$$\mu_R = N_1 \cos(\theta)\{[2 + 2jk_pR - (k_pR)^2]e^{-jk_pR} + 2N_2(-jk_sR - 1)e^{-jk_sR}\}e^{j\omega t}/R^3, \quad (11a)$$

$$u_\theta = -N_1 \sin(\theta)\{(-jk_pR - 1)e^{-jk_pR} + N_2[1 + jk_sR - (k_sR)^2]e^{-jk_sR}\}e^{j\omega t}/R^3, \quad (11b)$$

where values for N_1 and N_2 are given in Ref. 35 and are dependent on the assumed boundary condition, welded or slip, at the sphere surface. Here again, while previous studies that have used this equation have assumed a conventional "integer order" Voigt model, a fractional order Voigt model is easily employed.

B. Experiment

An experimental study was conducted for the case of the finite dipole buried in a viscoelastic medium with finite boundaries. The medium is the silicone-based polymeric gel CF-11 described in Sec. III B. A diagram of the experimental apparatus is shown in Fig. 6. A steel sphere is mounted to a mechanical shaker (4808, Bruel & Kjaer, Denmark) via a steel stinger that has a smooth, lubricated surface and via an impedance head (288B02, PCB Piezotronics, Depew, NY). The stinger and sphere are mounted in a container over the shaker via a rubber diaphragm at the base of the container. The silicone material in liquid form is poured into the container until it covers the sphere to the desired depth. The silicone container is mounted on a vibration isolated optics bench. The shaker is mounted on a separate support structure below the bench, with the stinger coming up through a clearance hole drilled in the optics bench below the annular rubber diaphragm at the base of the silicone-filled container. The surface response of the phantom material to shaker driven excitation is measured using a LDV (CLV-800, Polytec, Tustin, CA). Small (~ 2 mm^2) pieces of 3M retro-reflective tape are mounted on the semi-translucent phantom material to aid

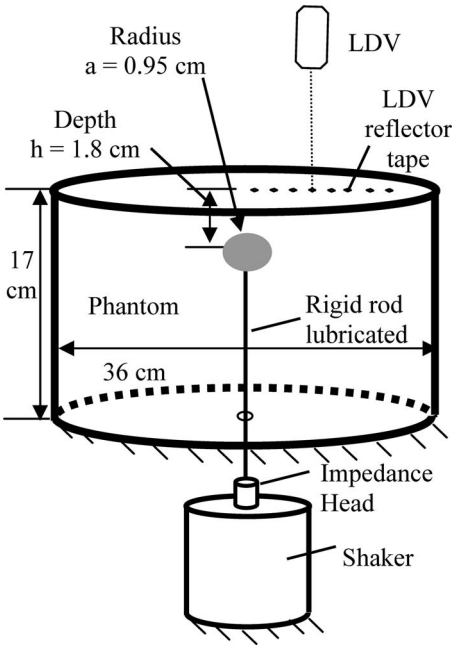


FIG. 6. Experimental schematic for measurement of surface wave motion caused by a subsurface (finite dipole) source.

in LDV measurement. Additionally, the vertical motion of the rigid sphere embedded in the phantom material just below its surface can be measured using the LDV because of the translucent nature of the phantom material. In MATLAB®, the frequency response of the system is calculated, using the rigid sphere velocity as the input and the phantom vertical surface velocity as the output. In theory, this FRF should not be affected by any resonant properties of the rod and sphere connected to the shaker, as these dynamics, present in both the input and output of the FRF, should effectively cancel one another. However, as results show in Sec. III C, a strong resonance of the rod-sphere-shaker system near 200 Hz does seem to alter calculated system values based on this FRF measurement.

C. Results and discussion

Measured and calculated vertical surface motion at four frequencies as a function of radial location is shown in Fig. 7. (These are the same four frequencies used in Ref. 35.) The analytically calculated surface response based on Eq. (11) and the three different viscoelasticity models (integer and fractional order Voigt, and SLS) consists of compression and shear waves, which are also plotted individually to show their relative contribution.

For all three viscoelastic models, at 40 and 80 Hz the shear wave component is dominant and thus its line is nearly indistinguishable from the line for the sum of the compression and shear waves. This is also true for the fractional order Voigt and SLS models at 200 Hz, and only for the SLS model at 400 Hz. The explanation for this is as follows. Both integer and fractional order Voigt models yield a loss modulus μ_I that increases with frequency; for the integer order model, it increases linearly with frequency and for the fractional order model it increases with the α power of frequency where $\alpha < 1$; see Eq. (6b). This increase results in greater

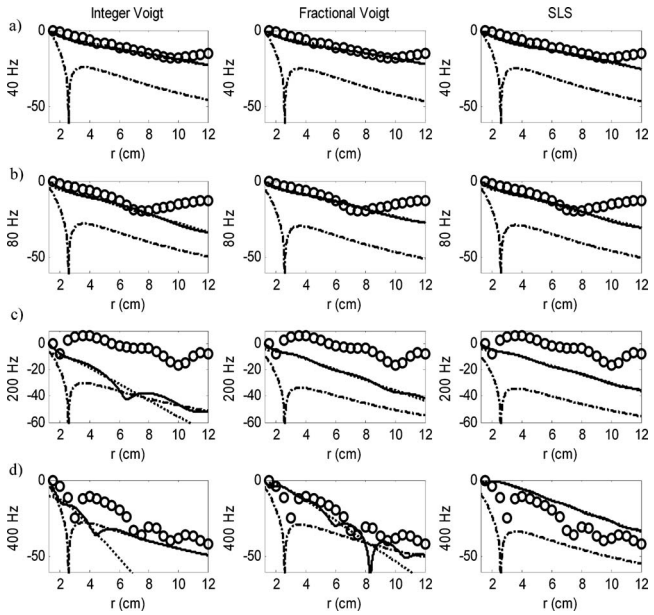


FIG. 7. The surface response of the viscoelastic halfspace to a subsurface (dipole) excitation as a function of radial position r for four different frequencies: (a) 40 Hz, (b) 80 Hz, (c) 200 Hz, and (d) 400 Hz. Columns: (1) Voigt model; (2) Fractional order Voigt model; (3) SLS model. (○○○) experimental measurement; (---) compression wave component; (···) shear wave component; (——) sum of compression and shear waves.

attenuation of shear wave motion as frequency increases, relative to compression waves. Contrary to the Voigt models, for the SLS model the loss modulus μ_I decreases with frequency; see Eq. (7b). Consequently, this model predicts that the shear wave contribution to surface motion will remain larger than the compression wave contribution as frequency increases, as shown in Fig. 7. This is not consistent with experimental studies.

In Fig. 8, the real and imaginary parts of the shear modulus are shown based on the experimental measurements and theoretical models. The measurement-based values were determined at each frequency point by a least squares error

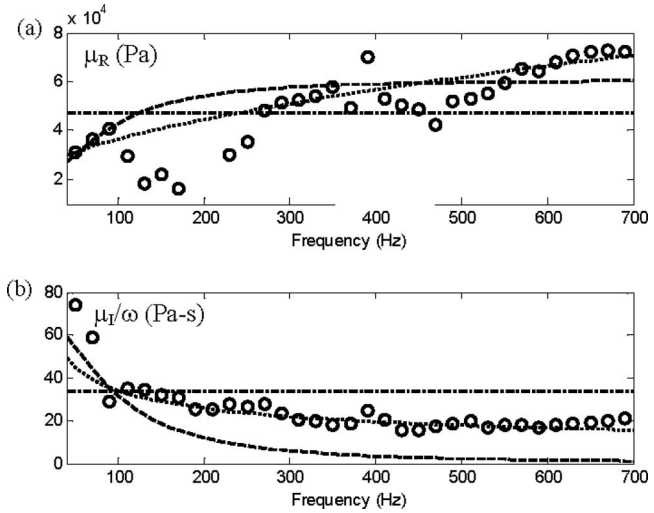


FIG. 8. CF-11 complex shear modulus $\mu=\mu_R+j\mu_I$ as a function of frequency for subsurface (dipole) source study. (a) Real part. (b) Imaginary part/ω. (○○○) experimentally derived values; (---) best fit Voigt model; (···) best fit $\alpha=0.6$ fractional order Voigt model; (---) best fit SLS model.

fit of the measured FRF value at the radial measurement points to the predicted values using Eq. (11), where all other parameters except complex μ are known and given in Sec. III. Once the measurement-based values for μ were determined over the frequency range shown in Fig. 8, least squares error fits of the three viscoelastic models were calculated in the same way as described in Sec. III C. Like before, the statically determined μ_0 was used for the SLS and fractional order Voigt model, leaving two parameters each for these models to be optimized based on the dynamic data. Optimal parameter values for these models are provided in Table II and the real and imaginary parts of the shear modulus based on them are shown in Fig. 8. The analytically calculated surface responses shown in Fig. 7 for the different optimal viscoelastic models could then be determined using Eq. (11).

As noted in Sec. III, the conventional Voigt model assumes constant values μ_R and μ_I/ω. As seen in Fig. 8, experiment-based values for both these parameters change with frequency. The SLS and fractional order Voigt models are able to represent this frequency dependence (though not the variation in the vicinity of 200 Hz, a strong resonance of the dipole source as mentioned in Sec. IV B). At the lower frequencies shown in Fig. 7, 40 and 80 Hz, all three models yield comparable results. However, at the higher frequencies the optimal fractional order Voigt model clearly performs better than the other two, being able to capture some of the unique radially dependent features of the response. Also, note that the optimized values for CF-11 fractional viscoelasticity, in terms of the order α and value of μ_α, are roughly in agreement based on two different types of measurements, surface motion due to a surface source (Sec. III) and due to a subsurface source (this section), given that there is some expected variation between different molds of the material.

V. CONCLUSION

The viscoelastic assumption in two prior studies of surface and shear wave motion on a halfspace has been reconsidered. Studies reported here show that a simple *fractional order* Voigt model, one with a fractional time derivative in its rate-dependent component, significantly improves the match of theory to experiment over a wider frequency range, as compared to the conventional integer order Voigt model and to the standard linear solid model, which has the same number of independent parameters as the fractional order Voigt model. The underlying physical motivation for the use of a fractional viscoelastic assumption and the potential practical benefits, such as improved medical diagnostics, were discussed. Improved model accuracy may be possible by incorporating additional fractional elements; a more thorough investigation of the fundamentals governing viscoelastic behavior at multiple scales in such materials may serve to motivate and direct such an investigation in the future.

ACKNOWLEDGMENTS

The financial support of the National Institutes of Health (Grant Nos. EB004885, EB008373, and EB007537) is acknowledged.

[1]J. F. Greenleaf, M. Fatemi, and M. Insana, "Selected methods for imaging elastic properties of biological tissues," Annu. Rev. Biomed. Eng. **5**, 57–78 (2003).
[2]N. Xiang and J. M. Sabatier, "An experimental study on antipersonnel landmine detection using acoustic-to-seismic coupling," J. Acoust. Soc. Am. **113**, 1333–1341 (2003).
[3]C. H. Frazier, N. Cadalli, D. C. Munson, Jr., and W. D. O'Brien, Jr., "Acoustic imaging of objects buried in soil," J. Acoust. Soc. Am. **108**, 147–156 (2000).
[4]R. A. Sohn and J. A. Hildebrand, "Hydroacoustic earthquake detection in the Artic basin with the Spinnaker array," Bull. Seismol. Soc. Am. **91**, 572–579 (2001).
[5]G. F. Miller and H. Pursey, "The field and radiation impedance of mechanical radiators on the free surface of a semi-infinite isotropic solid," Proc. R. Soc. London, Ser. A **223**, 521–541 (1954).
[6]G. F. Miller and H. Pursey, "On the partition of energy between elastic waves in a semi-infinite solid," Proc. R. Soc. London, Ser. A **223**, 55–69 (1955).
[7]C. R. Farrar, S. W. Doebling, and D. A. Nix, "Vibration-based structural damage identification," Philos. Trans. R. Soc. London **359**, 131–149 (2001).
[8]H. Tohmyoh, M. A. S. Akanda, "Sensitivity of acoustic microscopy for detecting three-dimensional nanometer gaps embedded in a silicon structure," J. Acoust. Soc. Am. **126**, 98–102 (2009).
[9]W. P. Rogers, "Elastic property measurement using Rayleigh–Lamb waves," Res. Nondestruct. Eval. **6**, 185–208 (1995).
[10]L. Gao, K. J. Parker, S. K. Alam, and R. M. Lerner, "Sonoelasticity imaging: Theory and experimental verification," J. Acoust. Soc. Am. **97**, 3875–3886 (1995).
[11]S. Catheline, F. Wu, and M. Fink, "A solution to diffraction biases in sonoelasticity: The acoustic impulse technique," J. Acoust. Soc. Am. **105**, 2941–2950 (1999).
[12]Y. Yamakoshi, J. Sato, and T. Sato, "Ultrasonic imaging of internal vibration of soft tissue under forced vibration," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 45–53 (1990).
[13]R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualization of propagating acoustic strain waves," Science **269**, 1854–1857 (1995).
[14]S. F. Othman, H. Xu, T. J. Royston, and R. L. Magin, "Microscopic magnetic resonance elastography (μMRE)," Magn. Reson. Med. **54**, 605–614 (2005).
[15]F. A. Duck, *Physical Properties of Tissue: A Comprehensive Reference Book* (Academic, New York, 1990).
[16]V. Kuperman, *Magnetic Resonance Imaging: Physical Principles and Applications* (Academic, New York, 2000).
[17]K. Hoyt, T. Kneezel, B. Castaneda, and K. J. Parker, "Quantitative sonoelastography for the in vivo assessment of skeletal muscle viscoelasticity," Phys. Med. Biol. **53**, 4063–4080 (2008).
[18]Y. C. Fung, *Biomechanics: Mechanical Properties of Living Tissues*, 2nd ed. (Springer-Verlag, New York, 1993).
[19]A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435 (1998).
[20]R. Muthupillai, P. J. Rossman, D. J. Lomas, J. F. Greenleaf, S. J. Riederer, and R. L. Ehman, "Magnetic resonance imaging of transverse acoustic strain waves," Magn. Reson. Med. **36**, 266–274 (1996).
[21]T. E. Oliphant, A. Manduca, R. L. Ehman, and J. F. Greenleaf, "Complex-valued stiffness reconstruction for magnetic resonance elastography by algebraic inversion of the differential equation," Magn. Reson. Med. **45**, 299–310 (2001).
[22]A. J. Romano, J. J. Shirron, and J. A. Bucaro, "On the noninvasive determination of material parameters from a knowledge of elastic displacements: Theory and numerical simulation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 751–759 (1998).
[23]A. J. Romano, J. A. Bucaro, R. L. Ehman, and J. J. Shirron, "Evaluation of a material parameter extraction algorithm using MRI-based displacement measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1575–1581 (2000).
[24]E. E. Van Houten, M. M. Doyley, F. E. Kennedy, J. B. Weaver, and K. D. Paulsen, "Initial in vivo experience with steady-state subzone-based MR Elastography of the human breast," J. Magn. Reson Imaging **17**, 72–85 (2003).
[25]E. E. Van Houten, M. M. Doyley, F. E. Kennedy, K. D. Paulsen, and J. B. Weaver, "A three-parameter mechanical property reconstruction method for MR-based elastic property imaging," IEEE Trans. Med. Imaging **24**, 311–324 (2005).
[26]S. G. 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).
[27]S. G. Chen, M. Fatemi, and J. F. Greenleaf, "Quantifying elasticity and viscosity from measurement of shear wave speed dispersion," J. Acoust. Soc. Am. **115**, 2781–2785 (2004).
[28]S. G. Chen, M. Fatemi, and J. F. Greenleaf, "Shearwave dispersion ultrasound vibrometry (SDUV) for measuring tissue elasticity and viscosity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **56**, 55–62 (2009).
[29]R. W. Chan and I. R. Titze, "Effect of postmortem changes and freezing on the viscoelastic properties of vocal fold tissues," Ann. Biomed. Eng. **31**, 482–491 (2003).
[30]D. Craiem and R. Armentano, "A fractional derivative model to describe arterial viscoelasticity," Biorheology **44**, 251–263 (2007).
[31]M. Z. Kiss, T. Varghese, and T. J. Hall, "Viscoelastic characterization of *in vitro* canine tissue," Phys. Med. Biol. **49**, 4207–4218 (2004).
[32]D. Klatt, U. Hamhaber, P. Asbach, J. Braun, and I. Sack, "Noninvasive assessment of the rheological behavior of human organs using multifrequency MR elastography: A study of brain and liver viscoelasticity," Phys. Med. Biol. **52**, 7281–7294 (2007).
[33]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).
[34]T. J. Royston, H. A. Mansy, and R. H. Sandler, "Excitation and propagation of surface waves on a viscoelastic half-space with application to medical diagnosis," J. Acoust. Soc. Am. **106**, 3678–3686 (1999).
[35]T. J. Royston, Y. Yazicioglu, and F. Loth, "Surface response of a viscoelastic medium to subsurface acoustic sources with application to medical diagnosis," J. Acoust. Soc. Am. **113**, 1109–1121 (2003).
[36]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).
[37]R. L. Magin, *Fractional Calculus in Bioengineering* (Begell House, Redding, CT, 2006).
[38]S. P. Timoshenko and J. N. Goodier, *Theory of Elasticity*, 3rd ed. (McGraw-Hill, New York, 1970).
[39]C. L. Pekeris, "The seismic surface pulse," Proc. Natl. Acad. Sci. U.S.A. **41**, 469–480 (1955).

Predictions of angle dependent tortuosity and elasticity effects on sound propagation in cancellous bone

Haydar Aygün[a)]
Medical Physics, PGMI, The University of Hull, Cottingham Road, HU6 7RX Hull, United Kingdom

Keith Attenborough
Department of Design, Development, Environment and Materials, The Open University, Milton Keynes MK7 6AA, United Kingdom

Michiel Postema
Emmy-Noether Research Group, Institute of Medical Engineering, Ruhr-Universität Bochum, 44780 Bochum, Germany

Walter Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Heverlee, Belgium

Christian M. Langton
Medical Physics, Queensland University of Technology, 2 George Street, Brisbane QLD 4001, Australia

(Received 15 April 2009; revised 27 August 2009; accepted 8 September 2009)

The anisotropic pore structure and elasticity of cancellous bone cause wave speeds and attenuation in cancellous bone to vary with angle. Previously published predictions of the variation in wave speed with angle are reviewed. Predictions that allow tortuosity to be angle dependent but assume isotropic elasticity compare well with available data on wave speeds at large angles but less well for small angles near the normal to the trabeculae. Claims for predictions that only include angle-dependence in elasticity are found to be misleading. Audio-frequency data obtained at audio-frequencies in air-filled bone replicas are used to derive an empirical expression for the angle-and porosity-dependence of tortuosity. Predictions that allow for either angle dependent tortuosity or angle dependent elasticity or both are compared with existing data for all angles and porosities. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3242358]

PACS number(s): 43.80.Cs, 43.80.Qf, 43.20.Jr [CCC] Pages: 3286–3290

I. INTRODUCTION

Clinical detection of osteoporosis involves measurement of broadband ultrasonic transmission at peripheral sites containing cancellous bone, which has a highly porous anisotropic cellular network structure filled with fatty bone marrow and including calcified plate-like elements known as trabeculae. The inclinations of the trabeculae vary with the site in the body, possibly as a consequence of mechanical requirements, for example, being somewhat random in the femoral head but more aligned in the calcaneous. Although typical clinical measurements are made normal to the trabeculae, the anisotropic structure of trabecular bone causes wave properties to vary with direction (Hosakawa and Otani, 1998; Hughes *et al.*, 1999; Lee *et al.*, 2007). Some success in modeling sound transmission in cancellous bone has been achieved by means of various forms of Biot theory (Biot 1956a, 1956b) which predicts two types of compressional wave (known as "fast" and "slow") and a shear wave. A basic premise of Biot theory is that the incident sound wavelengths are significantly larger than typical microstructural dimensions. Since the initial application of Biot theory to sound propagation in bone (McKelvie and Palmer, 1991), there has been considerable debate concerning the validity of this application. According to Williams (1992), the pore sizes in cancellous bone vary between 0.5 and 1 mm: a similar range of pore diameters is quoted in Hughes *et al.*, 2003. Also according to Williams (1992), the wavelength of the fast wave in water-saturated cancellous bone at 0.5 MHz is stated to lie between 5 and 7 mm for porosities between 0.1 and 0.4. This corresponds to fast wave speeds of between 2500 and 3500 m/s. In the frequency range from 1 kHz and 1 MHz, Hughes *et al.* (2003) predicted fast wave speeds of between 3700 and 5000 m/s for both water-filled and marrow-filled bones. The higher wave speeds will correspond to wavelengths on the order of 10 mm. In a similar frequency range, Hughes *et al.* (2003) predicted slow wave speeds of approximately 1500 m/s corresponding to wavelengths of between 1.5 m at 1 kHz and 1.5 mm at 1 MHz. Consequently, except at frequencies greater than 1 MHz, the predicted wavelengths in cancellous bone are an order of magnitude greater than the pore size and Biot theory should be applicable. At frequencies higher than 1 MHz, the slow wave should be subjected to a significant degree of scattering and, thereby, there should be higher transmission loss than predicted by Biot theory. However, even if Biot theory underestimates the attenuation of the frequency components of

[a)]Author to whom correspondence should be addressed. Electronic mail: h.aygun@hull.ac.uk

a slow wave pulse above 1 MHz, the influence on predicted waveforms will be small since the bone will act as a low pass filter and the lower frequency content will be more important.

Using isotropic Biot–Allard theory (Allard 1993), Fellah *et al.* (2004) find that tortuosity, defined as the ratio of the average length of the flow path through a porous medium sample to the thickness of the sample, plays an important role in propagation through cancellous bone since it affects the inertial coupling between fluid and solid. The theory employed by Fellah *et al.* (2004) introduces a viscous characteristic length Λ, due originally to Johnson *et al.* (1987), instead of the pore shape parameter originally used by Biot (1956a, 1956b) and, subsequently, by Hughes *et al.* (2007) and Lee *et al.* (2007). The viscous characteristic length depends on the narrowest pore sections where the effects of viscous drag are greatest. Fellah *et al.* (2004) predict that the viscous characteristic length may also have an important influence on wave transmission through bone but less than that of tortuosity.

To model the effects of the anisotropy of cancellous bone, Hughes *et al.* (2007) developed a stratified-Biot (SB) theory. They assumed an idealized microstructure of periodic parallel plates representing the trabeculae. The direction perpendicular to the plate axes, i.e., the dominant structural orientation, was taken to correspond to the zero value for the incidence angle. The resulting theory while giving reasonable agreement with data for large angles ($>30°$) from the normal to the predominant trabeculae direction is found to over-predict the fast wave speed at low angles ($<30°$) and to underestimate the slow wave speed at all angles. Hughes *et al.* (2007) also considered the influence of anisotropic Young's modulus. However, their development results only in a slight improvement in predictions compared with SB theory.

Lee *et al.* (2007) modeled the influence of angle-dependency in the elastic properties on sound propagation in cancellous bone. They considered two formulations of Biot theory and claimed that both give good agreement with data for the variation in fast wave speed with angle and porosity. However, agreement with comparable data for slow wave speeds was less good. Neither of the approaches used by Lee *et al.* (2007) includes an angle dependent tortuosity. Specifically, their tortuosity includes porosity-dependence but exclude angle-dependence, i.e., Lee *et al.* (2007) introduced anisotropy entirely through the elastic properties and ignore the effects of anisotropy in the pore structure. As shown in Fig. 5 of Hughes *et al.* (2007), an angle dependent tortuosity alone can explain some of the variation in fast wave speed with porosity and angle that has been observed. Moreover, unfortunately, in their paper Lee *et al.* (2007) compared predictions for the porosity of 0.65 with data for a porosity of 0.77.

Here, the heuristic form of angle dependent elasticity suggested by Lee *et al.* (2007) is combined in Biot–Allard theory with a heuristic angle and porosity dependent tortuosity function based on data obtained at audio-frequencies with air-filled (human) bone replicas by Attenborough *et al.* (2005). The replicas were 13 times real scale. However, the incident pulses were centered on 1 kHz, so the long wavelength condition for application of Biot theory is easily satisfied. The assumed form of angle-dependence is consistent with the observation that the fast wave speed increases with angle from the normal to the trabeculae. Predictions are explored that (a) only allow for angle dependent tortuosity, (b) only allow for angle dependent elasticity, and (c) allow for both. The predicted angle dependent phase velocities of fast and slow waves are compared with data for bovine bone (Hughes *et al.*, 1999).

II. THEORY

The Biot–Allard model for waves in fluid-saturated poro-elastic media (Allard, 1993) allows for thermal exchange and viscous drag between pore-fluid and the solid framework by introducing two characteristics lengths: the viscous (Λ) and thermal (Λ') characteristic lengths related to pore form factors c and c' by the following relationships:

$$\Lambda = \frac{1}{c}\left(\frac{8\alpha_\infty \eta}{\phi\sigma}\right)^{1/2}, \quad \Lambda' = \frac{1}{c'}\left(\frac{8\alpha_\infty \eta}{\phi\sigma}\right)^{1/2}, \tag{1}$$

where ϕ is porosity, and σ is the flow resistivity (which is equal to η, the dynamic viscosity coefficient, divided by permeability).

Thermal exchange effects between solid and fluid are included through a frequency-dependent bulk modulus of the fluid. This is calculated using (Allard, 1993)

$$K_f(\omega) = \frac{\gamma K_f}{\gamma - (\gamma - 1)\left[1 + \frac{8\eta}{j\Lambda'^2 B^2 \omega \rho_0}\left(1 + j\rho_0 \frac{\omega B^2 \Lambda'^2}{16\eta}\right)^{1/2}\right]^{-1}}, \tag{2}$$

where γ is the fluid specific heat ratio, B^2 is the Prandtl number, and K_f is the isothermal bulk modulus of the fluid. Thermal effects, while fairly important in air-filled porous materials, are expected to be of minor importance in marrow-filled bone. As yet, values for the characteristic lengths in bone have not been evaluated directly. However, Sebaa *et al.* (2006) found that values of Λ between 8 and 10.5 μm are consistent with data. For certain idealized pore structures, it is known that $c' \sim c/2$ (Allard, 1993).

The dependence of tortuosity on angle and porosity assumed by Hughes *et al.* (2007) is given by

$$\alpha_\infty(\theta) = 1 + \left[\frac{(1-\phi)\rho_s}{\langle\rho\rangle}\right]\cot^2\theta, \tag{3}$$

where $\langle\rho\rangle = \phi\rho_f + (1-\phi)\rho_s$, ρ_f and ρ_s being the mass densities of the fluid and solid, respectively, and ϕ is the porosity. This idealized angle-dependence implies infinite tortuosity for $\theta = 0°$ when sound travels normal to the parallel plates in the assumed parallel plate microstructure and a value depending on the relative densities of solid and fluid for $0° < \theta < 90°$. The tortuosity defined by Eq. (3) would be unity for propagation parallel to the plates if the plates are rigid; i.e., Equation (3) has an angle-dependence similar to that of the tortuosity in an idealized microstructure of parallel cylindrical pores in a rigid frame. In such a medium, the tortuosity

TABLE I. Properties and r [Eq. (4)] values for bone replicas (Attenborough *et al.*, 2005).

Replica type	Porosity	r
Iliac crest	0.8386	0.888
Femoral head	0.7426	0.591
Lumbar spine (LS2)	0.9173	0.521
Calcaneus	0.8822	0.816
Lumbar spine (LS4)	0.9121	0.259

would be given by $1/\sin^2(\theta)=\mathrm{cosec}^2(\theta)=1+\cot^2(\theta)$ where $\theta=0°$ is normal to the pore direction.

Cancellous bone microstructure departs significantly from either parallel plate or parallel pore idealizations. There is no evidence of values of tortuosity higher than 2.64 in the bone (see Table I in Hughes *et al.*, 2007). So the function given by Eq. (3) is least likely to be reliable for low angles, precisely where Hughes *et al.* (2007) found the biggest discrepancies between SB theory and data. According to the geometrical interpretation of tortuosity, it is determined entirely by the pore structure, is independent of the saturating fluid, and is independent of scaling. Consequently, extreme values of the angle-dependence of tortuosity may be derived empirically by referring to the average measured tortuosity values deduced from audio-frequency measurements on five air-filled stereo-lithographical cancellous (human) bone replicas at 13 times actual scale (Attenborough *et al.*, 2005). These data show that cancellous bone microstructure has orthotropic anisotropy. It is assumed that the dependence of tortuosity on porosity is given by (Berryman, 1980):

$$\alpha_\infty = 1 - r\left(1 - \frac{1}{\phi}\right), \tag{4}$$

where r is a variable calculated from a microscopic model of a frame moving in a fluid. The values of r required for consistency with the values of tortuosity for $\theta=0°$ deduced from the acoustical measurements on air-filled replica bones of known porosity (Attenborough *et al.*, 2005) are listed in Table I.

A heuristic form for porosity and angle dependent tortuosity may be written as

$$\alpha_\infty = 1 - r\left(1 - \frac{1}{\phi}\right) + k\cos^2(\theta), \tag{5}$$

where r and k can be considered adjustable. The assumed angle-dependence function is chosen arbitrarily but is simple and consistent with the expected variation in fast wave speed with angle. It should be noted that, if tortuosity has angle-dependency, as in Eq. (5), then so do the characteristic lengths and form factors [through Eq. (1)]. A range of possible values of r and k have been found by comparing predictions of Eq. (5) for $\theta=0°$ and 90°, respectively, with values deduced from air-filled replica bones (Attenborough *et al.*, 2005) of known porosity. Values of r and k are found by solving the resulting simultaneous equations. The angle dependent function representing the extremes of tortuosity measured in the bone replicas is

TABLE II. Default input parameters of the anisotropic Biot–Allard model for cancellous bone.

Parameters	Value
Density of solid bone, ρ_s	1960 kg/m^3
Density of fluid, ρ_f	1000 kg/m^3
Young's modulus of bone, E_s	20 GPa
Bulk modulus of fluid, K_f	2.2 GPa
Poisson's ratio of solid, υ_s	0.32
Poisson's ratio of frame, υ_b	0.32
Porosity, ϕ	0.65
Power index, n	$1.23\sin^2(\theta)+2.35\cos^2(\theta)$
Viscosity of fluid, η	0.001 Pa s
Permeability, k_0	5×10^{-9} m^3
Frequency, f	1 MHz
Fluid specific heat ratio, γ	1.0107
Prandtl number, B^2	7
Form factor, c	1
Form factor, c'	$c/2$

$$\alpha_\infty = 1.025 + 0.864\cos^2(\theta). \tag{6}$$

Williams (1992) suggested that the dependences of skeletal frame moduli (Young's modulus E_b, bulk modulus K_b, and rigidity modulus μ_b) in terms of bone volume fraction $(1-\phi)$ and Young's modulus of the solid material of the frame (E_s) are given by

$$E_b = E_s(1-\phi)^n, \tag{7a}$$

$$K_b = E_b/(1-2\upsilon_b), \tag{7b}$$

$$\mu_b = E_b/(1+2\upsilon_b), \tag{7c}$$

where the exponent n varies from 1 to 3 according to Gibson (1985), depending on the angle (θ) with respect to the dominant structural orientation (of the trabeculae, for example) according to

$$n = n_1\sin^2(\theta) + n_2\cos^2(\theta). \tag{8}$$

Values of $n_1=1.23$ and $n_2=2.35$ are chosen by Lee *et al.* (2007) to be consistent with the work of Williams (1992). Default values of the parameters required by the anisotropic Biot–Allard theory are listed in Table II. As remarked earlier, neither of the theoretical approaches used by Lee *et al.* (2007) includes an angle dependent tortuosity. They used a porosity dependent but angle independent tortuosity in two different formulations of Biot theory.

III. COMPARISONS WITH DATA

Figure 1 compares predictions of anisotropic Biot–Allard theory based on Eqs. (5) and (8) with data obtained on bovine femur by Hughes *et al.* (1999). The predictions include angle dependent tortuosity by allowing θ to vary in Eq. (5) but assume isotropic elasticity by setting $\theta=90°$ in Eq. (8). These predictions are similar to those of the stratified-Biot model in Hughes *et al.*, 2007 (see their Fig. 5) but assume a less extreme variation of tortuosity with angle.

Figure 2 compares predictions that allow angle-dependency in both tortuosity and elasticity with the same

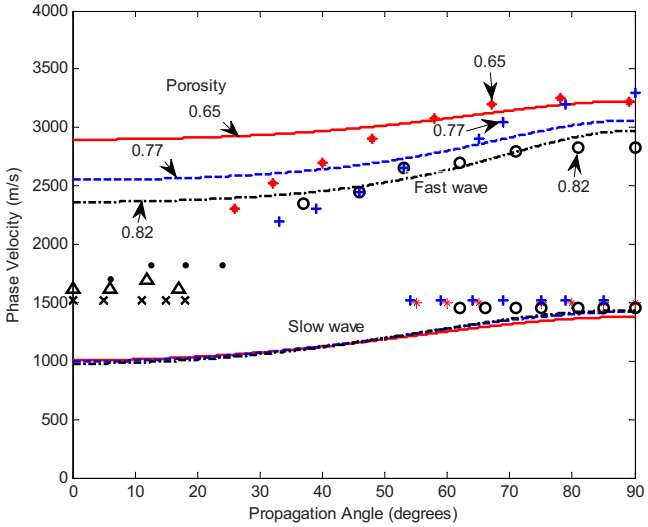


FIG. 1. (Color online) Hughes *et al.* (1999) data (symbols +, ○, and *) for three "parallel" samples on wave speeds as a function of angle (for porosities of 0.65, 0.77, and 0.82), and data (symbols ×, Δ, and ·) for three "perpendicular" samples compared with predictions (lines) assuming an angle and porosity dependent tortuosity function [Eq. (5) with $r=0.259$ and $k=0.864$] and isotropic elasticity [Eqs. (7) and (8) and Table II with $n_1=1.23$ and $\theta=90°$].

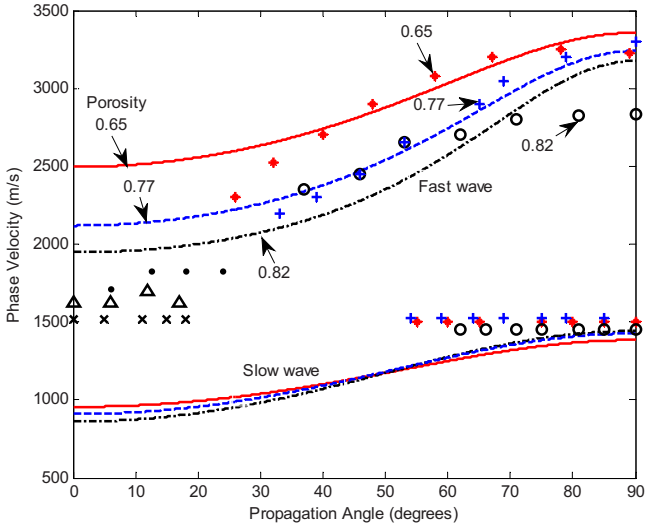


FIG. 3. (Color online) Data (symbols +, ○, and *; ×, Δ, and ·) corresponding to porosities of 0.65, 0.77, and 0.82, for wave speeds as a function of angle compared with predictions (lines) assuming an angle and porosity dependent tortuosity function [Eq. (5) with $r=0.259$ and $k=0.864$] and angle dependent elasticity [Eqs. (7) and (8) and Table II with $n_1=1.15$ and $n_2=1.6$].

data using Eqs. (5), (7), and (8) with $r=0.047$ and $k=0.864$. The resulting predictions are rather similar to those in Lee *et al.*, 2007 (see their Fig. 2). However, it should be noted that Lee *et al.* (2007) compared predictions for porosity of 0.65 with data for a porosity of 0.77. Although, as they asserted, the overall prediction of angle-dependence is improved through use of Eq. (8), it is at the cost of accuracy in the predicted porosity-dependence. In short, the predictions by Lee *et al.* (2007) of the influence of porosity on angle-dependence are not as good as they claim.

It should be noted that although the predictions of fast wave speeds in Fig. 2 are very similar to those in Fig. 2 of Lee *et al.* (2007), use of Eq. (5) rather than the fixed values

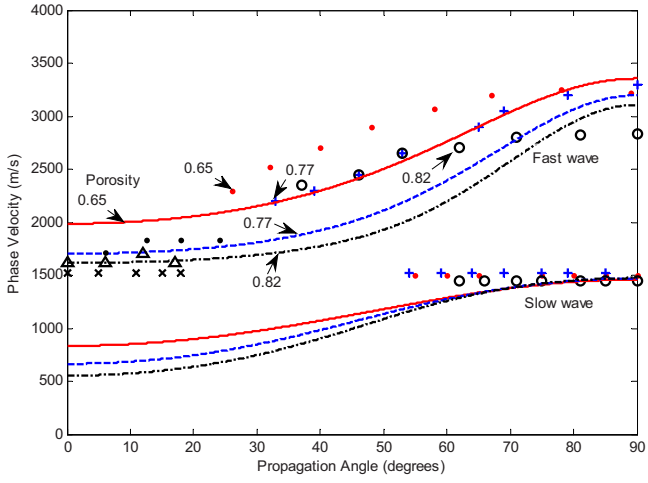


FIG. 2. (Color online) Data for three "parallel" (symbols +, ○, and *) samples and three "perpendicular" (symbols ×, Δ, and ·) samples (for porosities of 0.65, 0.77, and 0.82) on wave speeds as a function of angle compared with predictions (lines) assuming an angle and porosity dependent tortuosity function [Eq. (5) with $r=0.047$ and $k=0.864$] and angle dependent elasticity [Eqs. (7) and (8) and Table II].

of tortuosity used by Lee *et al.* (2007) means that the slow wave predictions at large angles are slightly improved compared with those in Lee *et al.*, 2007.

To obtain improved agreement between predicted and measured fast wave speeds over all angles when including both angle dependent tortuosity and elasticity in the predictions, the dependence on angle in Eq. (8) must be reduced. This means that the coefficient values n_1 and n_2 in Eq. (8) should be reduced. An example result, which confirms that, thereby, an improved prediction of porosity and angle-dependence can be obtained, is shown in Fig. 3. The values of the coefficients n_1 and n_2 have an important effect on the phase velocities of fast and slow waves, especially at low angles. Reducing the values of n_1 and n_2 increases the predicted phase speed of the fast wave particularly at low angles.

Lee *et al.* (2007) also compared predictions and data for wave speeds at 1 MHZ in directions perpendicular to and parallel with the dominant structural orientation. The corresponding predictions from Eqs. (5), (7), and (8) are shown in Fig. 4.

The value of r is predicted to have important influence on the fast wave speed variation with porosity perpendicular to the dominant structural orientation and on the slow wave speed variation with porosity parallel to the dominant structural orientation. Although not shown here, the value of n_1 is predicted to have an important influence on the fast wave speed variation with porosity, parallel to the dominant structural orientation. Other calculations suggest that an angle dependent viscous characteristic length has potentially important effects on the variation in slow wave speed with porosity for measurements close to the dominant structural direction.

IV. CONCLUSION

Previous work on the influence of anisotropic pore structure and elasticity in cancellous bone has been extended by

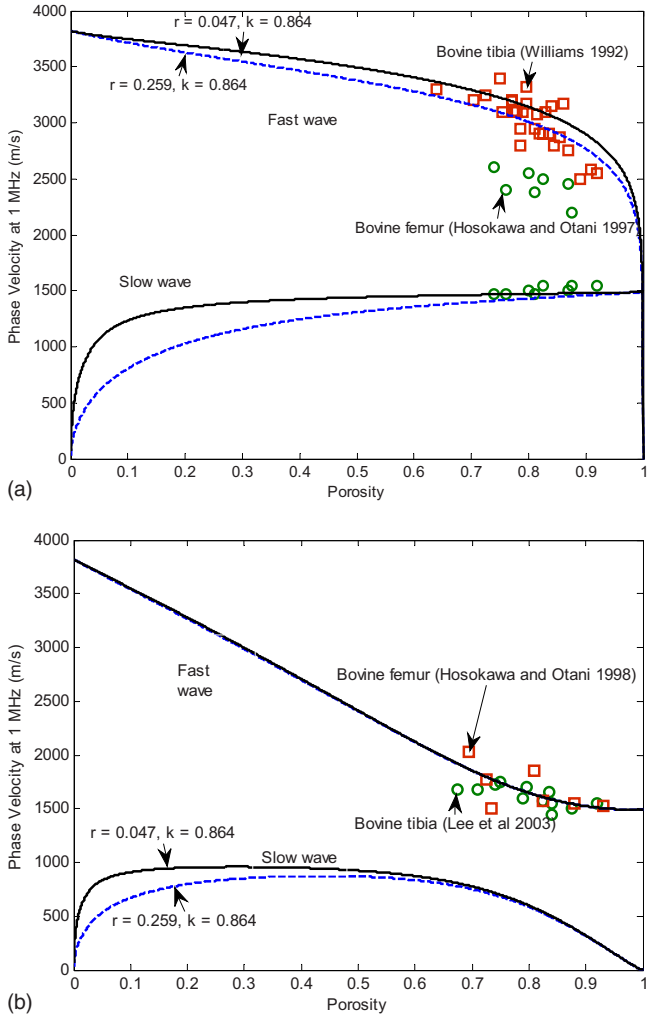


FIG. 4. (Color online) Predictions (lines) and data (symbols) for porosity-dependence of wave speeds (a) for propagation perpendicular to the dominant structural orientation direction assuming a porosity dependent tortuosity function [Eq. (5) with θ=90°, values of r and k as labeled] and anisotropic elasticity given by Eqs. (7) and (8) and Table II with θ=90°; (b) for propagation parallel to the dominant structural orientation assuming a porosity dependent tortuosity function [Eq. (5) with θ=0°, values of r and k as labeled] and elasticity given by Eqs. (6) and (7) with θ=0°, and parameter values in Table II.

developing an anisotropic Biot–Allard model allowing for angle dependent tortuosity and elasticity. The extreme angle-dependence of tortuosity corresponding to the parallel plate microstructure used by Hughes *et al.* (2007) has been replaced by angle dependent tortuosity values based on data for slow wave transmission through air-filled bone replicas. It has been shown that the good agreement claimed by Lee *et al.* (2007) using only angle dependent elasticity is misleading and that more complete predictions allowing for angle-dependency in both tortuosity and elasticity have greater validity. Although agreement with data even after adjustment of the parameter values for angle dependent elasticity used by Lee *et al.* (2007) is not particularly good, the anisotropic Biot–Allard model will be useful to give further insight into the factors that have the most important influence on the angle-dependency of wave speeds and attenuation in cancellous bone.

ACKNOWLEDGMENT

This work has been supported by Leverhulme Grant No. F/00 181/N.

Allard, J. F. (**1993**). *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Elsevier, London).

Attenborough, K., Qin, Q., Fagan, M. J., Shin, H.-C., and Langton, C. M. (**2005**). "Measurements of tortuosity in stereolithographical bone replicas using audio-frequency pulses," J. Acoust. Soc. Am. **118**, 2779–2782.

Berryman, J. G. (**1980**. "Confirmation of Biot's theory," Appl. Phys. Lett. **37**, 382–384.

Biot, M. A. (**1956a**). "Theory of propagation of elastic waves in a fluid saturated porous solid: I. Low frequency range," J. Acoust. Soc. Am. **28**, 168–178.

Biot, M. A. (**1956b**). "Theory of propagation of elastic waves in a fluid saturated porous solid: II. High frequency range," J. Acoust. Soc. Am. **28**, 179–191.

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

Gibson, L. J. (**1985**). "The mechanical behaviour of cancellous bone," J. Biomech. **18**, 317–328.

Hosoakawa, 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.

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

Lee, K. I., Hughes, E. R., Humphery, V. F., Leighton, T. G., and Choi, M. J. (**2007**). "Empirical angle dependent Biot and MBA models for acoustic anisotropy in cancellous bone," Phys. Med. Biol. **52**, 59–73.

McKelvie M. L., and Palmer S. B. _(**1991**). "The interaction of ultrasound with cancellous bone," Phys. Med. Biol. **36**, 1331–1340.

Sebaa, N., Fellah, Z., Fellah, M., Ogam, E., Wirgin, A., Mitri, F., Depollier, C., and Lauriks, W. (**2006**). "Ultrasonic characterisation of human cancellous bone using the Biot theory: Inverse problem," J. Acoust. Soc. Am. **120**, 1816–1824.

Williams, J. L. (**1992**). "Ultrasonic wave propagation in cancellous and cortical bone: predictions of some experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112.

Frequency dependence of average phase shift from human calcaneus *in vitro*

Keith A. Wear[a)]
Center for Devices and Radiological Health, U. S. Food and Drug Administration, Silver Spring, Maryland 20993

(Received 30 March 2009; revised 28 September 2009; accepted 2 October 2009)

If dispersion in a medium is weak and approximately linear with frequency (over the experimental band of frequencies), then it can be shown that the constant term in a polynomial representation of phase shift as a function of frequency can produce errors in measurements of phase-velocity differences in through-transmission, substitution experiments. A method for suppressing the effects of the constant phase shift in the context of the single-wave-model was tested on measurements from 30 cancellous human calcaneus samples *in vitro*. Without adjustment for constant phase shifts, the estimated phase velocity at 500 kHz was 1516 ± 6 m/s (mean ± standard error), and the estimated dispersion was -24 ± 4 m/s MHz (mean ± standard error). With adjustment for constant phase shifts, the estimated mean velocity decreased by 4–9 m/s, and the estimated magnitude of mean dispersion decreased by 50%–100%. The average correlation coefficient between the measured attenuation coefficient and frequency was 0.997 ± 0.0026 (mean ± standard deviation), suggesting that the signal for each sample was dominated by one wave. A single-wave, linearly dispersive model conformed to measured complex transfer functions from the 30 cancellous-bone samples with an average root-mean-square error of $1.9\% \pm 1.0\%$. [DOI: 10.1121/1.3257550]

PACS number(s): 43.80.Ev [TDM] Pages: 3291–3300

I. INTRODUCTION

Bone sonometry has emerged as a significant modality for the management of osteoporosis (Laugier, 2008; Krieg *et al.*, 2008). Despite the clinical utility of bone sonometry, the mechanisms underlying the interaction between ultrasound and cancellous bone are not well understood. An improved understanding of these mechanisms could result in the development of devices with improved diagnostic performance. Recent developments in device design include increased portability (Kaufman *et al.*, 2007), expansion to new sites such as the radius (Le Floch *et al.*, 2008), femur (Barkmann *et al.*, 2008) and tibia (Sarvazyan *et al.*, 2009), and a dual-frequency pulse-echo method for removing effects of soft tissues (Riekkinen *et al.*, 2006, 2008).

Several papers report negative dispersion (i.e., phase velocity decreasing with frequency) *in vitro* in human calcaneus (Stelitzki and Evans, 1996a, 1996b; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000a) and bovine cancellous bone (Hughes *et al.*, 1999; Kaczmarek *et al.*, 2002; Waters and Hoffmeister, 2005). This contrasts with the usual finding of positive dispersion in soft tissues (O'Donnell *et al.*, 1981). Dispersion is important because it can affect clinical measurements of ultrasound velocity in cancellous bone (Wear, 2000b; Haïat *et al.*, 2006, 2008a). Some authors proposed models to explain negative dispersion (Hughes *et al.*, 1999; Wear, 2001; Lin *et al.*, 2001; Kaczmarek *et al.*, 2002; Haïat *et al.*, 2008a).

Phase velocity in biologic tissues *in vitro* is often measured in a substitution experiment. Two through-transmission signals are acquired: one through a water-only-path and another through a water-sample-water path. By comparing frequency-dependent phases of the two signals, phase velocity can be measured provided that the speed of sound in water and the thickness of the sample are known.

The investigators in the aforementioned papers that report negative dispersion in cancellous bone measured frequency-dependent phase velocity from the difference in phases of signals with and without bone samples placed in a water path. Their method implicitly assumed that all measured phase shifts were attributable directly to phase-velocity differences. However, other factors (including transmission through boundary surfaces, refraction effects, diffraction effects, phase aberration, defocusing, multipath interference, reverberations, phase cancellation, finite sample effects, multiple scattering, experimental error, etc.) may also contribute to phase shifts. These effects may be insignificant for relatively homogeneous, fluid-like media (such as some soft tissues) but may be significant for relatively inhomogeneous, poroelastic media such as cancellous bone, which exhibit increased complexity of interaction with ultrasound (Fellah *et al.*, 2004, 2008; Sebaa *et al.*, 2008).

The present paper considers media with dispersion that is weak and approximately linear with frequency (over the experimental band of frequencies). The phase shift in a through-transmission substitution experiment is represented as a polynomial function of frequency. It is shown that, in the context of the single-wave-model (the model on which previous estimates of frequency-dependent phase velocity and dispersion in cancellous bone are based), the constant term in the polynomial is not necessarily directly attributable to phase-velocity differences. The effect of the constant phase shift on estimates of phase velocity and dispersion was mea-

a)Author to whom correspondence should be addressed. Electronic mail: keith.wear@fda.hhs.gov

sured in (1) a polycarbonate plate, (2) seven parallel-nylon-wire cancellous-bone-mimicking phantoms, and (3) 30 cancellous human calcaneus samples.

II. THEORY

Phase velocity and attenuation are often measured in through-transmission geometry, with one transmitting transducer and one receiving transducer. One measurement, $X(\omega)$, is performed with the ultrasound beam passing from the transmitter to the receiver through water only. A second measurement, $Y(\omega)$, is performed with a sample placed in the ultrasound beam path. Most reports of measurements of frequency-dependent phase velocities for bone samples in substitution experiments (Strelitzki and Evans, 1996a; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000a) are implicitly based on a sine-wave transfer function $H(\omega)=Y(\omega)/X(\omega)$ of the form

$$H(\omega)=A\ \exp[-\alpha(\omega)d]\exp\left[i\omega\frac{d}{v(\omega)}\right]\exp\left[-i\omega\frac{d}{c}\right], \quad (1)$$

where $\omega=2\pi f$, f is the ultrasound frequency, $\alpha(\omega)$ is attenuation coefficient$=\beta\omega/2\pi$, β is attenuation slope, $v(\omega)$ is phase velocity, d is sample thickness, c is speed of sound in water, and A is frequency-independent factor that includes transmission through boundaries.

Equation (1) may be regarded as a definition of effective bulk phase velocity $v(\omega)$ through a sample. Although Eq. (1) might be limited in its ability to completely represent the complex interactions between ultrasound in bone, it offers a simple, empirically supported, practical model without requiring extensive knowledge of material and structural properties of cancellous bone. Other models, including two-longitudinal-wave models (such as Biot theory), the Stratified model, and multiple scattering models are also valuable for increasing our intuition regarding the interaction between ultrasound and bone (Hughes, *et al.*, 1999; Fellah *et al.*, 2004; Haïat *et al.*, 2008a).

Sometimes the transfer function for a poroelastic medium may be approximated by the sum of two waves with a form similar to Eq. (1) (Marutyan *et al.*, 2006; Anderson *et al.*, 2008; Bauer *et al.*, 2008). (One small difference between the present model and previous models is that an exponential factor, $\exp[-i\omega d/c]$, has been included to explicitly account for the fact that, in a substitution experiment, the attenuating sample replaces an equivalent length of water in the acoustic beam path.) The two waves may correspond to the "fast" and "slow" waves predicted by Biot theory (Hosokawa and Otani, 1997, 1998; Cardoso *et al.*, 2003; Fellah *et al.*, 2004, 2008; Sebaa *et al.*, 2008; Cardoso *et al.*, 2008; Pakula *et al.*, 2008). The presence of two waves in cancellous bone has been observed mostly in cases when the ultrasound propagates *parallel* to the predominant trabecular orientation (i.e., the main load direction) (Hosokawa and Otani, 1997, 1998; Nicholson *et al.*, 1998; Mizuno *et al.*, 2009). For human calcaneus interrogated in the mediolateral orientation (as is the case with the four aforementioned studies and with clinical bone sonometers), however, the ultrasound propagates approximately *perpendicular* to the predominant trabecular orientation.

It is possible that what appears to be a single pulse is actually the superposition of two pulses that overlap in time (Anderson *et al.*, 2008). The likelihood of two overlapping pulses for typical human calcaneus samples interrogated in the mediolateral orientation *in vitro* may be investigated by comparing the difference in fast and slow wave transit times to the pulse duration. The difference in fast and slow wave transit times may be estimated as follows: Anderson *et al.* (2008) reported that for perpendicular interrogation of cancellous-bone samples with porosity of 0.85, Biot theory predicts a velocity difference between fast and slow waves of 865 m/s. This implies a fast-slow wave transit time difference of about 14 μs, assuming that $d=18$ mm, $v_{\text{fast}}=1565$ m/s, and $v_{\text{slow}}=700$ m/s (Anderson *et al.*, 2008, Fig. 1). The pulse duration may be estimated as follows: For a typical broadband transducer with a center frequency of 500 kHz, the cycle period is about 2 μs, and the pulse duration is about 2 cycles or 4 μs. Under these conditions, fast and slow waves, if they are both present, should be easily separated in time by $14/4=3.5$ pulse durations. The fact that only one pulse is usually observed for human calcaneus interrogated in the perpendicular orientation suggests that the signal is dominated by a single wave. However, this analysis is based on numerous assumptions, including an empirical extension of isotropic Biot theory to anisotropic media and is not necessarily definitive.

The two-wave model may have a comparative advantage for (1) propagation parallel to the predominant trabecular orientation and/or (2) bovine cancellous bone, which typically has lower porosity than human cancellous bone. Haïat *et al.* (2008b) suggested that the single-wave model may be deemed adequate when the correlation coefficient of a least-squares linear fit to the measured attenuation coefficient versus frequency exceeds a threshold (e.g., $r^2>0.97$). Regardless of whether or not the single-wave model is the best model for human calcaneus interrogated perpendicular to the predominant trabecular orientation, however, it is the model upon which previous published estimates of phase velocity and dispersion are implicitly based (Strelitzki and Evans, 1996a; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000a). Therefore, it is worthwhile to examine the previous data analysis method in the context of the single-wave model in order to see whether previous estimates of phase velocity and dispersion might be improved.

The phase of the single-wave transfer function is given by

$$\Delta\phi(\omega)=\arg[H(\omega)]=\arg[A]+\omega d\left[\frac{1}{v(\omega)}-\frac{1}{c}\right]$$
$$\equiv\Delta\phi_A+\Delta\phi_V(\omega), \quad (2)$$

where $\Delta\phi_A\equiv\arg[A]$ and $\Delta\phi_V(\omega)\equiv\omega d[1/v(\omega)-1/c]$. If the sample in the substitution experiment is a relatively homogeneous, fluid-like medium, then $\Delta\phi_A$ may be negligible and phase differences between water-only and water-sample-water measurements may be entirely attributable to phase-velocity differences between the sample and water, i.e.,

$\Delta\phi(\omega) \approx \Delta\phi_V(\omega)$. In this case, phase velocity may be computed as reported previously (Strelitzki and Evans, 1996a; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000a), using

$$v(\omega) = \frac{c}{1 + \frac{c\Delta\phi(\omega)}{\omega d}}. \tag{3}$$

This equation is often presented with a minus sign instead of a plus sign in the denominator. The ambiguity arises from ambiguity in the sign of $\Delta\phi(\omega)$, which depends on how the phase difference is computed.

If, however, the sample in the substitution experiment is an inhomogeneous, poroelastic medium such as cancellous bone, then $\Delta\phi_A$ may not necessarily be zero (Fellah *et al.*, 2004, 2008; Jocker and Smeulders, 2009). Authors (including the present author) who have previously reported measurements of frequency-dependent phase velocity in human calcaneus samples *in vitro* have used a form of Eq. (3) without explicitly accounting for $\Delta\phi_A$ (Strelitzki and Evans, 1996a; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000a). The remainder of this section develops a method for accounting for $\Delta\phi_A$ in phase-velocity measurements in weakly dispersive media.

Over a limited band of frequencies (e.g., the experimental frequency band), the phase velocity of the sample may be expanded in a polynomial

$$\begin{aligned} v(\omega) &= v_0 + v_1\omega + v_2\omega^2 + \cdots \\ &= v_0\left[1 + \frac{v_1}{v_0}\omega + \frac{v_2}{v_0}\omega^2 + \cdots\right]. \end{aligned} \tag{4}$$

Equation (4) has the form $v(\omega)=v_0(1+x)$ where $x=(v_1/v_0)\omega+(v_2/v_0)\omega^2+\ldots$ For a weakly dispersive medium, $|x| \ll 1$ so that $v(\omega)=v_0(1+x)\approx v_0/(1-x)$, and $\Delta\phi_V(\omega)$ can be rewritten (over the experimental frequency band) as

$$\Delta\phi_V(\omega) = \omega d\left[\frac{1}{v_0}\left(1 - \frac{v_1}{v_0}\omega - \frac{v_2}{v_0}\omega^2 - \cdots\right) - \frac{1}{c}\right]. \tag{5}$$

It may be seen in Eq. (5) that (1) if $v(\omega)=v_0$ (non-dispersive medium), then $\Delta\phi_V(\omega)$ depends *linearly* on frequency; (2) if $v(\omega)=v_0+v_1\omega$ (linearly dispersive medium), then $\Delta\phi_V(\omega)$ depends *quadratically* on frequency.

Now the total phase difference (over the experimental frequency band) becomes

$$\Delta\phi(\omega) = \Delta\phi_A + d\left(\frac{1}{v_0} - \frac{1}{c}\right)\omega - \frac{dv_1}{v_0^2}\omega^2 + \cdots. \tag{6}$$

From Eq. (5), it may be seen that, for a weakly dispersive medium, $\Delta\phi_V(\omega)$ does not contain a constant term (in the single-wave model). Therefore, for weakly dispersive media, the constant phase shift in the polynomial representation of $\Delta\phi$ is entirely attributable to $\Delta\phi_A$. Therefore, $\Delta\phi_A$ may be estimated by performing a least-squares polynomial fit to $\Delta\phi(\omega)$ over the experimental frequency band and extrapolating the fit to $\omega=0$. The effect of the constant phase shift term, $\Delta\phi_A$, may be suppressed with the following formula for phase velocity:

TABLE I. Phantom properties. Trabecular thickness (Tb.Th) corresponds to nylon-wire thickness. The variable s is the inter-wire spacing, which is equal to the sum of Tb.Th and Tb.Sp (trabecular separation). The VF is the fraction of volume occupied by nylon wire. Porosity=1−VF.

Tb.Th (μm)	s=Tb.Th+Tb.Sp (μm)	Volume fraction
152	1000	0.018
152	900	0.022
152	800	0.028
152	700	0.037
203	800	0.051
254	800	0.079
305	800	0.114

$$v(\omega) = \frac{c}{1 + \frac{c[\Delta\phi(\omega) - \Delta\phi_A]}{\omega d}}. \tag{7}$$

III. EXPERIMENTAL METHODS

A. Polycarbonate plate

Frequency-dependent phase-velocity measurements were performed on a 25.8-mm-thick polycarbonate sample. Measurements were compared with previously published measurements in polycarbonate (Droin *et al.*, 1998; Wear, 2000a).

B. Parallel-nylon-wire cancellous-bone-mimicking phantoms

Frequency-dependent phase-velocity measurements were performed on seven phantoms consisting of parallel nylon wires (simulating trabeculae) in two-dimensional rectangular grid arrays (custom-built by Computerized Imaging Reference Systems, Norfolk, VA). The nylon-wire diameter corresponded to trabecular thickness, often denoted by Tb.Th (Parfitt *et al.*, 1987). As shown in Table I, four values for Tb.Th were used: 152, 203, 254, and 305 μm. The mean value for Tb.Th for human calcaneus is 127 μm (Ulrich *et al.*, 1999). Trabecular spacing is given by s=Tb.Sp+Tb.Th, where Tb.Sp is trabecular separation. Four values for s were used: 700, 800, 900, and 1000 μm. The mean value for Tb.Sp in human calcaneus is 684 μm (Ulrich *et al.*, 1999), which corresponds to a mean value for s equal to 684 μm +127 μm=811 μm. The volume fraction (VF) occupied by wire (trabeculae) is given by $\mathrm{VF}=\pi(\mathrm{Tb.Th}/2)^2/s^2$. Porosity is given by $\beta=1-\mathrm{VF}$. The range of VF spanned by the seven phantoms (1.8%–11.4%) roughly corresponds to the range reported for human calcaneus, 2%–14% (Wear, 2005).

C. Bone samples

Thirty human calcaneus samples (genders, ages unknown) were defatted using a trichloro-ethylene solution. Removal of fat decreases the risk of microorganism contamination of samples. Some studies suggest that marrow removal has a small impact on attenuation and sound speed (Langton *et al.*, 1996; Hoffmeister *et al.*, 2002). Other stud-

ies suggest that marrow removal has a significant impact but that correlations between ultrasound measurements in marrow-filled and water-filled cancellous bone are high: $r^2=0.94$ for phase velocity and $r^2=0.94$ for attenuation (Alves *et al.*, 1996) and $r^2=0.77$ for phase velocity and $r^2=0.79$ for broadband ultrasound attenuation (Nicholson and Bouxsein, 2002). The cortical lateral layers were sliced off leaving two parallel surfaces with direct access to cancellous bone. The thicknesses of the samples varied from 16 to 21 mm. The other dimensions were approximately 35×70 mm^2.

In order to remove air bubbles, the samples were vacuum degassed underwater in a desiccator. After vacuum, samples were allowed to thermally equilibrate to room temperature prior to ultrasonic interrogation. Ultrasonic measurements were performed in distilled water at room temperature. The temperature was measured with a digital thermometer for each experiment and ranged between 19.1 and 21.2 °C. The relative orientation between the ultrasound beam and the bone samples was the same as with *in vivo* measurements performed with commercial bone sonometers, in which sound propagates in the mediolateral direction.

D. Ultrasonic data acquisition and analysis

A Panametrics (Waltham, MA) 5077PR pulser/receiver was used. Samples were interrogated in a water tank using a pair of coaxially aligned, Panametrics broadband, 1 in. (2.5 cm) diameter, 1.5 in. (3.8 cm) focal-length transducers. The polycarbonate plate was interrogated using transducers with center frequency equal to 1 MHz. The phantoms and bone samples were interrogated with transducers with center frequency equal to 500 kHz. The propagation path between transducers was twice the focal length. Received radio-frequency signals were digitized (8 bits, 10 MHz) using a LeCroy (Chestnut Ridge, NY) 9310C Dual 400 MHz oscilloscope and stored on computer (via general purpose interface bus) for off-line analysis.

The single-wave transfer function, $H(\omega)$, was computed from the ratio of the fast Fourier transforms of the signals acquired with and without the bone sample in the water path. The attenuation coefficient was computed using a log spectral difference technique as described previously (Wear, 1999). The adequacy of a single-wave model [Eq. (1)] was assessed with the method of Haïat *et al.* (2008b), which entails measuring the correlation coefficient of a linear fit to the measured attenuation coefficient versus frequency and determining whether it exceeds a threshold ($r^2>0.97$). In addition, the root-mean-square-error (RMSE) between a model fit [Eq. (1)] and the data were computed.

The phase difference, $\Delta\phi(\omega)$, was computed as the inverse tangent of the ratio of the imaginary to real parts of $H(\omega)$. Since the inverse tangent function yields principal values between $-\pi$ and π, the phase had to be unwrapped by adding an integer multiple of 2π to all frequencies above each frequency where a discontinuity appeared.

Frequency-dependent phase-difference data were least-squares-error fit to linear and quadratic polynomials over the range from 500 kHz–1.2 MHz (polycarbonate plate), 300–700 kHz (phantoms), or 300–600 kHz (bone samples). Frequency-dependent phase velocity, $v(\omega)$, was computed using two or more of three methods:

(1) with Eq. (3), i.e., no correction for $\Delta\phi_A$ (polycarbonate, phantoms, and bones);
(2) with Eq. (7), subtracting the constant phase shift term, $\Delta\phi_A$, estimated using a least-squares *linear* fit for phase difference vs frequency (bones); and
(3) with Eq. (7), subtracting the constant phase shift term, $\Delta\phi_A$, estimated using a least-squares *quadratic* fit for phase difference vs frequency (polycarbonate, phantoms, and bones).

The temperature-dependent speed of sound in distilled water (Kaye and Laby, 1973) was used in Eqs. (3) and (7).

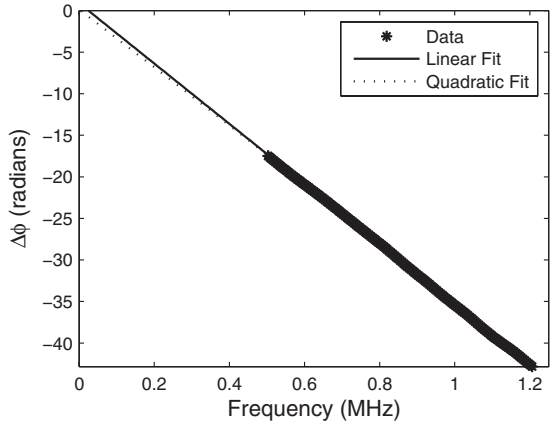


FIG. 1. Measurements of phase difference, $\Delta\phi(\omega)$, vs frequency $(\omega/2\pi)$ for the polycarbonate plate. Linear and quadratic fits to the data are also shown. Error bars (standard errors) are plotted but may be too small to discern.

IV. RESULTS

Figure 1 shows measurements, a linear fit, and a quadratic fit for average phase difference, $\Delta\phi(\omega)$, between water-polycarbonate-water and water-only signals as functions of frequency $(\omega/2\pi)$. The phase difference depends nearly linearly on frequency, suggesting little or no disper-

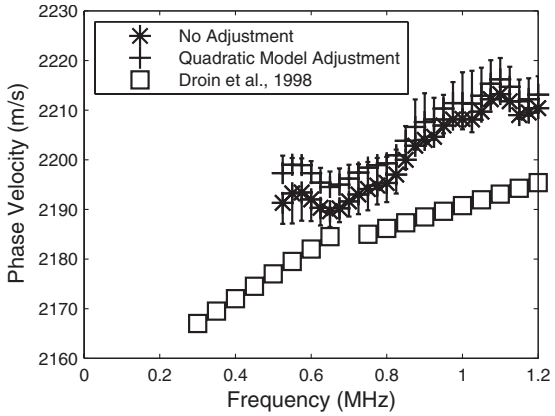


FIG. 2. The effect of subtracting the y-intercept from a least-squares quadratic fit to the phase-velocity vs frequency data in order to suppress boundary effects in the polycarbonate plate. Previously published measurements by Droin *et al.* (1998) are also shown.

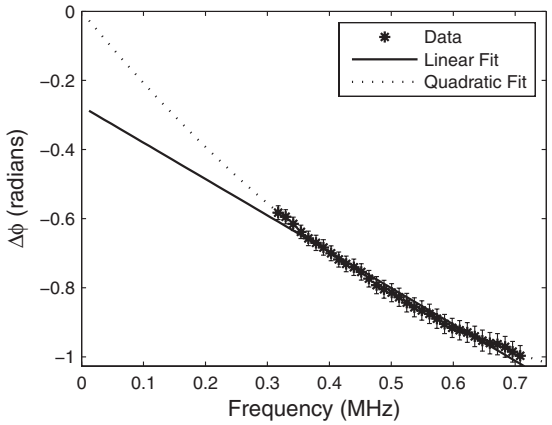


FIG. 3. Measurements of phase difference, $\Delta\phi(\omega)$, vs frequency $(\omega/2\pi)$ for the parallel-nylon-wire cancellous-bone-mimicking phantom with Tb.Th =305 μm and spacing=800 μm. Linear and quadratic least-squares fits to the data are also shown. Error bars are standard errors.

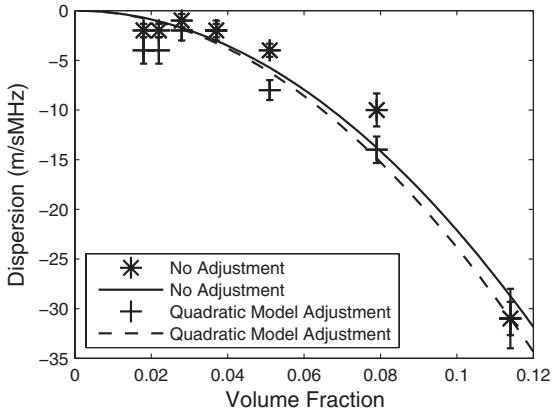


FIG. 5. Estimates of dispersion with and without adjustment for y-intercept $(\Delta\phi_A)$ in seven parallel-nylon-wire cancellous-bone-mimicking phantoms. Parabolic fits to the data are also shown.

sion. Figure 1 is very similar in appearance to results reported by Kaufman *et al.* (1995) [Fig. 4(d)] for the phase of the transfer function for a Lucite block. In both cases, phase (1) has a positive y-intercept, and phase (2) decreases linearly with frequency.

Figure 2 shows previously published measurements of frequency-dependent phase velocity for polycarbonate (Droin *et al.*, 1998—squares; Wear, 2000a—asterisks). The dispersion rate for polycarbonate in this frequency range is approximately 20–30 m/s MHz. Figure 2 also shows the effect of subtracting the estimate of $\Delta\phi_A$ from a quadratic fit to phase-velocity vs frequency data. The resulting change in phase-velocity values is slight, presumably because $\Delta\phi_A$ (0.11 rad) is small relative to the frequency-dependent phase shifts in the experimental frequency band (500 kHz–1.2 MHz), which range from −20 to −45 rad (as may be seen in Fig. 1). The frequency-dependencies for the uncorrected and $\Delta\phi_A$-subtracted data are similar to that reported previously by Droin *et al.* (1998). Small differences may be attributable to differences in composition between the polycarbonate plate studied here and the one studied by Droin *et al.*

Figure 3 shows measurements, a linear fit, and a quadratic fit for average phase difference, $\Delta\phi(\omega)$, between water-phantom-water and water-only signals as functions of frequency $(\omega/2\pi)$ for the parallel-nylon-wire cancellous-bone-mimicking phantom with Tb.Th=305 μm and spacing=800 μm. The correlation coefficient (r) between the measurements and the least-squares linear fit was 0.994 (95% confidence interval: 0.987–0.997). The correlation coefficient between the measurements and the least-squares quadratic fit was 1.000 (95% CI: 0.999–1.000), suggesting that the quadratic fit may be regarded as superior to the linear fit. Figures 4 and 5 show measurements of phase velocity and dispersion with and without subtraction of $\Delta\phi_A$ (based on the quadratic fit to phase difference vs frequency). Since the quadratic fits produced estimates of $\Delta\phi_A$ very close to zero, the subtraction of $\Delta\phi_A$ did not have much impact on the estimates of phase velocity and dispersion. This reinforces a recent theoretical model for the prediction of frequency-dependent phase velocity in parallel-nylon-wire cancellous-bone-mimicking phantoms (Haïat *et al.*, 2008a) that attributes negative dispersion to the effects of multiple scattering.

Figures 6 and 7 show experimental measurements of single-wave linearly dispersive transfer functions, $H(\omega)$, and model fits for two cancellous-bone samples. The average cor-

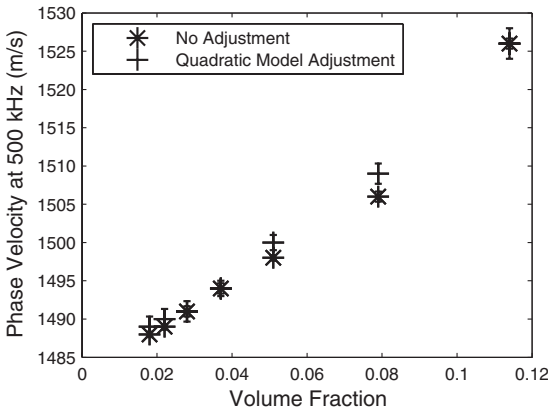


FIG. 4. Estimates of phase velocity at 500 kHz with and without adjustment for y-intercept $(\Delta\phi_A)$ in seven parallel-nylon-wire cancellous-bone-mimicking phantoms.

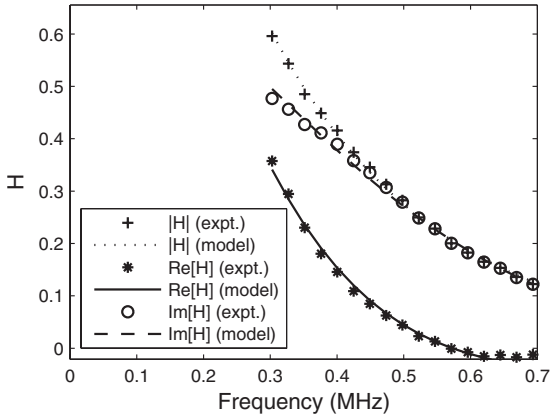


FIG. 6. Experimental measurements of complex transfer function, $H(\omega)$, and single-wave linearly dispersive transfer function fits [Eq. (1)] for a single bone specimen with attenuation coefficient slope of 19.6 dB/cm MHz and a phase velocity at 500 kHz of 1525 m/s.

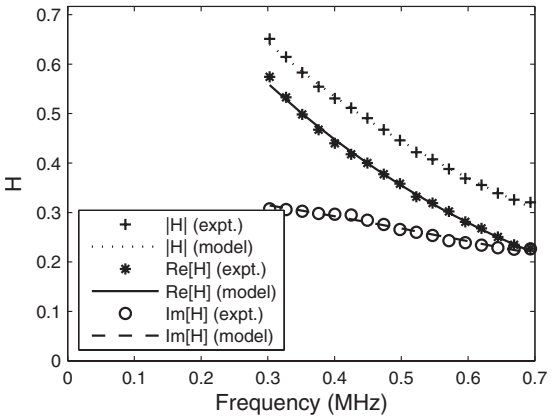


FIG. 7. Experimental measurements of complex transfer function, $H(\omega)$, and single-wave linearly dispersive transfer function fits [Eq. (1)] for a single bone specimen with attenuation coefficient slope of 8.0 dB/cm MHz and a phase velocity at 500 kHz of 1497 m/s.

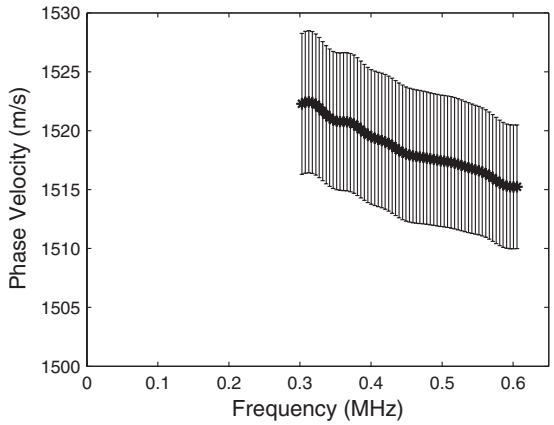


FIG. 9. Phase velocity computed from phase-difference data with no adjustment for boundary effects. Error bars are standard errors.

relation coefficient between the experimental attenuation coefficients and least-squares linear fits for the 30 bone samples was 0.997 ± 0.0026 (mean $\pm$ standard deviation). The average 95% confidence interval for the correlation coefficient was $(0.994 \pm 0.0043, 0.998 \pm 0.0016)$. The criterion of Haïat *et al.* (2008b) of $r^2 > 0.97$ (i.e., $r > 0.985$) supported the single-wave model. Additional support for the single-wave model comes the fact that the average RMSE between the five-parameter ($|A|$, $\Delta\phi_A$, β, v_0, and v_1), single-wave, *linearly dispersive* model and the data for the 30 cancellous-bone samples was only $1.9\% \pm 1.0\%$. The RMSE between the four-parameter ($|A|$, $\Delta\phi_A$, β, and v_0), single-wave, *nondispersive* model for the 30 cancellous-bone samples was only slightly higher, $2.1\% \pm 1.0\%$, suggesting little or no value to adding the fifth parameter (v_1).

Figure 8 shows measurements, a linear fit, and a quadratic fit for average phase difference, $\Delta\phi(\omega)$, between water-bone-water and water-only signals as functions of frequency $(\omega/2\pi)$ for the 30 bone samples. The correlation coefficient (r) between the measurements of $\Delta\phi(\omega)$ and the least-squares linear fit was 0.998 (95% CI: 0.997–0.999). The correlation coefficient between the measurements and the least-squares quadratic fit was 0.999 (95% CI: 0.998–0.999). Because of substantial overlap of the 95% confidence intervals, it is unclear in this case whether the quadratic fit should be preferred to the linear fit. The quasi-linear behavior of $\Delta\phi(\omega)$ over the experimental bandwidth supports the polynomial form in Eqs. (4)–(6).

Figure 9 shows the estimate of average phase velocity vs frequency for the 30 bone samples computed using Eq. (3), i.e., without subtraction of $\Delta\phi_A$, as has been done by previous investigators (Stelitzki and Evans, 1996a, 1996b; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000a). Figure 9 implies negative dispersion as has been reported previously in human cancellous-bone samples by these investigators.

Figure 10 shows the estimate of average phase velocity vs frequency for the 30 bone samples computed using Eq. (7) with $\Delta\phi_A$ estimated from a least-squares linear fit to measurements of $\Delta\phi(\omega)$. As suggested by Eqs. (6) and (7), a linear model for $\Delta\phi(\omega)$ results in an estimated phase velocity with essentially no dispersion.

Figure 11 shows the estimate of average phase velocity vs frequency for the 30 bone samples computed from Eq. (7) with $\Delta\phi_A$ estimated from a least-squares quadratic fit to measurements of $\Delta\phi(\omega)$. The quadratic model resulted in an estimated phase velocity with negative dispersion but at a more gradual rate than that shown in Fig. 9. The quasi-linear (or

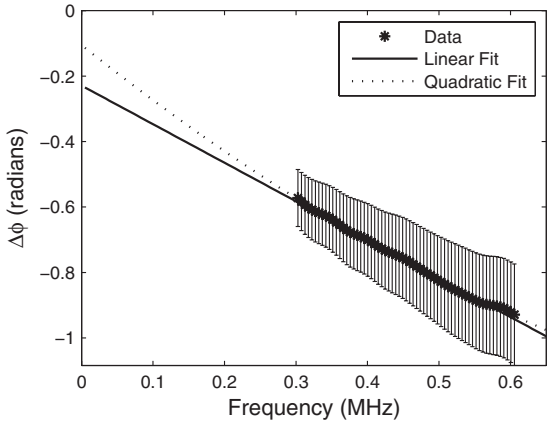


FIG. 8. Measurements of phase difference, $\Delta\phi(\omega)$, vs frequency $(\omega/2\pi)$ for 30 bone samples. Linear and quadratic least-squares fits to the data are also shown. Error bars are standard errors.

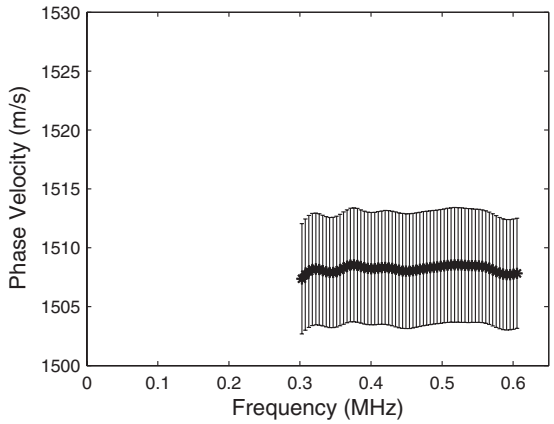


FIG. 10. Phase velocity computed from phase-difference data with an adjustment for boundary effects based on a linear model for phase difference, $\Delta\phi(\omega)$, vs frequency $(\omega/2\pi)$. Error bars are standard errors.

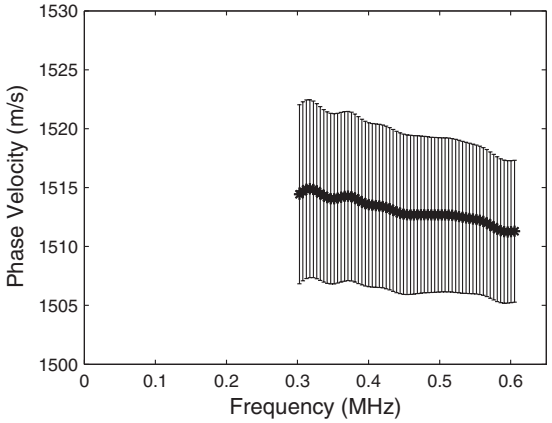


FIG. 11. Phase velocity computed from phase-difference data with an adjustment for boundary effects based on a quadratic model for phase difference, $\Delta\phi(\omega)$, vs frequency $(\omega/2\pi)$. Error bars are standard errors.

perhaps mildly quadratic) behavior of average phase velocity over the experimental bandwidth in Figs. 10 and 11 is consistent with the polynomial in Eq. (4).

Table II lists results in cancellous bone. The rightmost column shows that the weak dispersion assumption, $|(v_1/v_0)\omega| \ll 1$, was satisfied by all three models at 500 kHz, the center frequency.

V. DISCUSSION

Many investigators have reported measurements of frequency-dependent phase velocity in human calcaneus *in vitro*. These investigators (1) implicitly assumed a single-wave model for wave propagation, (2) estimated frequency-dependent phase velocity from through-transmission measurements of phase differences of signals with and without bone samples placed in a water path, and (3) attributed all measured phase shifts to phase velocity differences between bone samples and water. In the present paper, however, it has been shown that if the dispersion in the medium is weak and approximately linear with frequency (over the experimental band of frequencies), then the constant term in a polynomial representation of phase shift as a function of frequency can lead to errors in measurements of phase velocity differences. A method for suppressing the constant phase shift in the context of the single-wave-model was tested on measurements from (1) a polycarbonate plate, (2) seven cancellous-bone-mimicking phantoms, and (3) 30 cancellous human calcaneus samples *in vitro*. Constant phase shifts were not significant for the polycarbonate plate or the phantoms but were significant for the 30 human calcaneus samples. Without adjustment for constant phase shifts in the 30 cancellous-bone samples, the estimated phase velocity at 500 kHz was 1516 ± 6 m/s (mean $\pm$ standard error), and the estimated dispersion was -24 ± 4 m/s MHz (mean $\pm$ standard error). With adjustment for constant phase shifts, the estimate for mean velocity decreased by 4–9 m/s, and the estimate of the magnitude of mean dispersion decreased by 50%–100%. Constant phase shifts may have resulted in overestimation of the magnitude of dispersion in other studies (Strelitzki and Evans, 1996a; Nicholson *et al.*, 1996; Droin *et al.*, 1998).

Figure 8 shows that the measurements of phase difference between water-bone-water and water-only signals was very nearly linear, supporting the use of low-order polynomial approximate forms of Eqs. (4)–(6). In fact, the linear fit to the frequency-dependent phase-difference measurements was so close to the measurements that the possibility that the measurements reveal no dispersion whatsoever could not be completely ruled out. The quadratic fit to the phase-difference measurements, which may have been more accurate than the linear fit (but this is not conclusively shown here), does suggest negative dispersion, however.

The approach taken in the present paper was based on a single-wave model. The linearly dispersive form is based on a relatively small number (5) of parameters ($|A|$, $\Delta\phi_A$, β, v_0, and v_1) and conformed to measured transfer functions with an average RMSE of $1.9\% \pm 1.0\%$. The non-dispersive form is based on four parameters ($|A|$, $\Delta\phi_A$, β, and v_0) and conformed to measured transfer functions with an average RMSE of $2.1\% \pm 1.0\%$, suggesting little or no value to adding the fifth parameter (v_1). Regardless of whether or not the single-wave model is the best model for human calcaneus interrogated perpendicular to the predominant trabecular orientation, it is the model upon which previous estimates of phase velocity and dispersion are based (Strelitzki and Evans, 1996a; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000a). The present paper offers a method by which previous estimates of phase velocity and dispersion for weakly linearly dispersive media might be improved in the context of the single-wave model. Other models including two-longitudinal-wave models (such as Biot theory), the Stratified model, and multiple scattering models are also valuable for increasing our intuition regarding the interaction

TABLE II. The effect of boundary effects adjustment on estimates phase velocity at 500 kHz and dispersion in human calcaneus samples *in vitro*. Values are expressed as mean $\pm$ standard error of estimate of the mean. The weak dispersion assumption is valid if $|(v_1/v_0)\omega| \ll 1$.

Adjustment for boundary effects	Correlation coefficient (r) (95% CI)	Average $\Delta\phi_A$ (rad)	Phase velocity at 500 kHz (m/s)	Dispersion (m/s MHz)	$\lvert(v_1/v_0)\omega\rvert$ at 500 kHz
None			1516 ± 6	-24 ± 4	0.05
Linear phase model	0.998 (0.997, 0.999)	-0.23 ± 0.03	1507 ± 5	0 ± 1	0
Quadratic phase model	0.999 (0.998, 0.999)	-0.11 ± 0.04	1512 ± 6	-12 ± 6	0.025

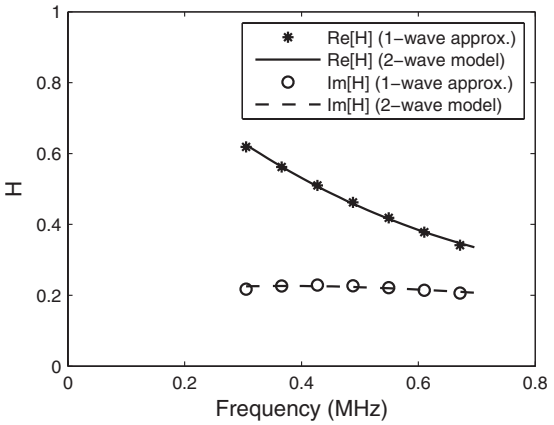


FIG. 12. Real and imaginary parts (continuous curves) of the two-wave transfer function, $H=H_{\text{fast}}+H_{\text{slow}}$, for a set of parameters that Anderson *et al.* (2008) showed to result in an attenuation coefficient that appears to vary approximately linearly with frequency: $A_{\text{fast}}=0.3$, $A_{\text{slow}}=0.7$, $\beta_{\text{fast}}=20$ dB/cm MHz, $\beta_{\text{slow}}=6.9$ dB/cm MHz, $v_{\text{fast}}(\omega_0)=1600$ m/s, $v_{\text{slow}}(\omega_0)=1500$ m/s, and $\omega_0/2\pi=300$ kHz. Also shown are real and imaginary parts (asterisks and circles) of a single-wave linearly dispersive model (with frequency-independent form for A) fit to H, which conform very closely to the two-wave model over the experimental frequency band.

between ultrasound and bone (Hughes, *et al.*, 1999; Fellah *et al.*, 2004; Haïat *et al.*, 2008a). However, if they require more parameters than the single-wave models, then it may be difficult to provide adequate statistical support for them based on typical data sets from human cancellous-bone samples *in vitro* because data set size is limited by the difficulties involved in (1) obtaining human tissue samples, (2) preparing samples for measurement, and (3) performing measurements.

Following Haïat *et al.* (2008b), the present investigation supported the single-wave model by showing that the correlation coefficient of a least-squares linear fit to the measured attenuation coefficient versus frequency was sufficiently high ($r^2>0.97$). There may be limitations to this approach. For example, Anderson *et al.* (2008) and Bauer *et al.* (2008) showed that the attenuation coefficient can appear to vary approximately linearly with frequency even under a two-wave model. However, for parameter values relevant to fast and slow waves in cancellous bone, this tends to happen primarily when the phase velocity difference between fast and slow waves is relatively small (on the order of 100 m/s), as illustrated in Fig. 3 of Anderson *et al.* (2008). It is instructive to investigate the difference between the two-wave and single-wave models in this limit of low phase-velocity differential. Anderson *et al.* (2008) utilized a two-wave transfer function that was the sum of two terms of a form similar to Eq. (1). Figure 12 shows the real and imaginary parts (continuous curves) of the two-wave transfer function, $H(\omega)=H_{\text{fast}}(\omega)+H_{\text{slow}}(\omega)$, for a set of parameters that Anderson *et al.* (2008) showed to result in an attenuation coefficient that appears to vary approximately linearly with frequency: $d=1$ cm, $A_{\text{fast}}=0.3$, $A_{\text{slow}}=0.7$, $\beta_{\text{fast}}=20$ dB/cm MHz, $\beta_{\text{slow}}=6.9$ dB/cm MHz, $v_{\text{fast}}(\omega_0)=1600$ m/s, $v_{\text{slow}}(\omega_0)=1500$ m/s, and $\omega_0/2\pi=300$ kHz. (Note that in this example, the fast-slow-wave velocity differential, 100 m/s, is much less than the previously mentioned value estimated by Anderson *et al.* (2008), for the case of perpendicular incidence, 865 m/s.) Figure 12 also shows real and imaginary parts (asterisks and circles) of a single-wave linearly dispersive model approximately fit to $H(\omega)$, which conforms very closely to the two-wave model. Therefore, even when the signal truly contains two waves, the single-wave model may serve as an adequate approximation over a typical experimental band of frequencies. Moreover, in the presence of typical levels of noise encountered in experiments on bone samples, it might be difficult to reject the single-wave model in favor of the two-wave model based on the underlying functions shown in Fig. 12. In fact, a single-wave linearly dispersive model might even be preferable from a statistical standpoint based on parsimony considerations.

The average correlation coefficient between the measured attenuation coefficient in the 30 cancellous-bone samples and frequency was 0.997 ± 0.0026 (mean $\pm$ standard deviation) over the range from 300 to 600 kHz. This high correlation is consistent with other reports. Strelitzki and Evans (1996b) reported an average correlation coefficient in ten human calcaneus samples *in vitro* of 0.987 ± 0.012 over the range from 400 and 800 kHz. The slight reduction in correlation coefficient (compared with the present study) may be partially attributable to the shifted, broader bandwidth used by Strelitzki and Evans. Chaffaï *et al.* (2000) reported correlation coefficients for power law fits for attenuation coefficient, $\alpha(f)=\alpha_0+\alpha_1 f^n$. Their average correlation coefficient for 11 human calcaneus samples *in vitro* was 0.990 ± 0.005, over the range from 500 kHz and 1.3 MHz. Their values for n (1.09 ± 0.3) were close to 1, suggesting that they may have achieved high correlations with linear fits (i.e., $n=1$), especially over a much smaller bandwidth such as that used in the present study (300–600 kHz). Substantial nonlinearity has been reported for frequencies above 600 kHz in young women *in vivo*, however, (Langton *et al.*, 1984). Therefore, the conclusions drawn from the present study may be more appropriate for relatively low-density calcaneus samples.

Fellah *et al.* (2004) reported a theoretical prediction for a transfer function that suggested approximately exponential decay with frequency (as in the present study) but modulated by a factor that oscillated with frequency. This modulating oscillation was not seen in the present study, however. [$|H(\omega)|$ is closely modeled by a decaying exponential in Figs. 6 and 7.] However, Fellah *et al.* (2004) considered bone samples with parameter values different that those in the present paper: thickness=7 mm (Fellah) vs 16–21 mm (present paper) and porosity=0.83 (Fellah) vs 0.86–0.98 (present paper) (Wear, 2005). The empirical results reported in the present paper suggest that over the range of frequencies investigated for the present set of human calcaneus samples, signals transmitted through bone were dominated by a single wave.

The single-wave, linearly dispersive model was convenient for the present application because it allowed mathematical separation of the constant phase shift term, $\Delta\phi_A$, from other sources of phase shift. Unfortunately, it may not be as easy to separate these phase shifts factors for a two-wave model, which may be more appropriate for bovine samples or for human samples when propagation is parallel

to the predominant trabecular orientation. However, even for situations where the two-wave model may be more appropriate than the single-wave linearly dispersive model, the analysis in the present paper suggests that the relationship between measured phase shifts and frequency-dependent phase velocity may be more complicated than previously thought, and therefore estimates of dispersion should be interpreted with caution.

The present analysis assumed that the multiplicative constant, A, in Eq. (1) was frequency-independent. It is instructive to consider the case of a frequency-dependent multiplicative factor, $A \rightarrow A(\omega)$. A could depend on ω, for example, if the efficiency of transmission through the boundaries of the sample depended on the ratio of the wavelength to the average pore size. In this case, the polynomial representation for $\arg[A(\omega)]$ might contain not just a constant ($\Delta\phi_A$) but also linear, quadratic, and all higher order terms in ω. Therefore, all terms in the polynomial representation of measurements of total phase shift, $\Delta\phi$, would contain contributions from $\arg[A(\omega)]$ as well as from $\Delta\phi_V$. In this case, it would be impossible to separate phase shifts due to $A(\omega)$ from those directly attributable to bulk phase-velocity differences. However, the fact that the variation in attenuation coefficient with frequency was measured to be so nearly linear suggests that the magnitude of $A(\omega)$ either (1) was approximately frequency-independent or (2) obeyed an exponential form, $|A(\omega)| = A_0 e^{-q\omega}$ so that its frequency-dependent effects masqueraded as linear attenuation. Since the latter possibility would be quite coincidental, it is plausible that $|A(\omega)|$ may have been approximately frequency-independent for human cancellous calcaneus samples over the range from 300 to 600 kHz. If the phase of $A(\omega)$ was also frequency-independent, then measurements of $\Delta\phi(\omega)$, once adjusted for the effects of $\Delta\phi_A$, should be relatively free from contamination due to a hypothetical frequency dependence of A.

Differences between the cancellous-bone-mimicking phantoms interrogated here and cancellous bone should be acknowledged. First, the longitudinal sound speed in nylon (2600 m/s) is somewhat lower than that for mineralized bone material (2800–4000 m/s, near 500 kHz) (Duck, 1990) (but still far greater than that for water or marrow—near 1500 m/s). Second, the parallel-nylon-wire phantoms lacked cross-links that can connect nearby trabeculae in cancellous bone. Third, wires in the phantom were aligned more parallel than trabeculae in cancellous bone. Despite these differences, these parallel-nylon-wire phantoms have previously been shown to exhibit similar dependences of scattering (Wear, 1999, 2004) and phase velocity (Wear, 2005; Haïat *et al.*, 2008a, 2008b) on frequency and scatterer thickness as cancellous bone. Fourth, nylon wires may have lower absorption than trabeculae. However, this may not be a serious limitation because simulations that ignore absorption have been shown to reproduce many experimental results for cancellous bone (Bossy *et al.*, 2005).

Phase velocity and attenuation measurements are more challenging *in vivo* than *in vitro* because of the presence of soft tissue and cortical bone layers surrounding the cancellous bone and because of the unknown thickness of the calcaneus. In particular, since the speed of sound in cortical bone is much faster than the speed of sound in cancellous bone, the presence of cortical plates would likely increase effects due to transmission through boundaries, refraction, diffraction, phase aberration, defocusing, multipath interference, reverberations, phase cancellation, finite sample effects, and reverberation. Cortical plates also affect attenuation (Xia *et al.*, 2005).

VI. CONCLUSION

Like previous reports of frequency-dependent phase velocity and dispersion in cancellous bone, the present paper was based on a single-wave model for propagation of ultrasound. It was shown that for a medium with dispersion that is weak and approximately linear with frequency, the constant term in a polynomial representation of phase shift as a function of frequency can lead to errors in measurements of phase-velocity differences. Without adjustment for constant phase shifts in 30 human cancellous-bone samples, the estimated phase velocity at 500 kHz was 1516 ± 6 m/s (mean $\pm$ standard error), and the estimated dispersion was -24 ± 4 m/s MHz (mean $\pm$ standard error). With adjustment for constant phase shifts, the estimated mean velocity decreased by 4–9 m/s, and the estimated magnitude of mean dispersion decreased by 50%–100%.

ACKNOWLEDGMENTS

The author is grateful to the U.S. Food and Drug Administration Office of Women's Health for funding, the Heather Pierce of Computerized Imaging Reference Systems, Norfolk, VA, for assistance in phantom design and construction, and Reviewer 1 of this manuscript for excellent suggestions.

Alves, J. M., Ryaby, J. T., Kaufman, J. J., Magee, F. P., and Siffert, R. S. (**1996**). "Influence of marrow on ultrasonic velocity and attenuation in bovine trabecular bone," Calcif. Tissue Int. **58**, 362–367.

Anderson, C. C., Marutyan, K. R., Holland, M. R., Wear, K. A., and Miller, J. G. (**2008**). "Interference between wave modes may contribute to the apparent negative dispersion observed in cancellous bone," J. Acoust. Soc. Am. **124**, 1781–1789.

Barkmann, R., Laugier, P., Moser, U., Dencks, S., Klausner, M., Padilla, F., Haïat, G., Heller, M., and Glüer, C. (**2008**). "In vivo measurements of ultrasound transmission through the human proximal femur," Ultrasound Med. Biol. **34**, 1186–1190.

Bauer, A. Q., Marutyan, K. R., Holland, M. R., and Miller, J. G. (**2008**). "Negative dispersion in bone: The tole of interference in measurements of the apparent phase velocity of two temporally overlapping signals," J. Acoust. Soc. Am. **123**, 2407–2414.

Bossy, E., Padilla, F., Peyrin, F., and Laugier, P. (**2005**). "Three-dimensional simulation of ultrasound propagation through trabecular bone structures measured by synchrotron micro-tomography," Phys. Med. Biol. **50**, 5545–5556.

Cardoso, L., Meunier, A., and Oddou, C. (**2008**). "In vitro acoustic wave propagation in human and bovine cancellous bone as predicted by Biot's theory," J. Mech. Med. Biol. **8**, 183–201.

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.

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

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

Control **45**, 581–592.
Duck, F. A. (**1990**). *Physical Properties of Tissue* (Cambridge University Press, Cambridge, UK).
Fellah, Z. E. A., Chapelon, J. Y., Berger, S., Lauriks, W., and Depollier, C. (**2004**). "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73.
Fellah, Z. E. A., Sebaa, N., Fellah, M., Mitri, F. G., Ogam, E., Lauriks, W., and Depollier, C. (**2008**). "Application of the Biot model to ultrasound in bone: Direct problem," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1508–1515.
Haïat, G., Lhémery, A., Renaud, F., Padilla, F., Laugier, P., and Naili, S. (**2008a**). "Velocity dispersion in trabecular bone: Influence of multiple scattering and of absorption," J. Acoust. Soc. Am. **124**, 4047–4058.
Haïat, G., Padilla, F., Cleveland, R. O., and Laugier, P. (**2006**). "Effects of frequency-dependent attenuation and velocity dispersion on in vitro ultrasound velocity measurements in intact human femur specimens," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 39–51.
Haïat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2008b**). "Fast wave ultrasonic propagation in trabecular bone: Numerical study of the influence of porosity and structural anisotropy," J. Acoust. Soc. Am. **123**, 1694–1705.
Hoffmeister, B. K., Auwarter, J. A., and Rho, J. Y. (**2002**). "Effect of marrow on the high frequency ultrasonic properties of cancellous bone," Phys. Med. Biol. **47**, 3419–3427.
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.
Jocker, J., and Smeulders, D. (**2009**). "Ultrasonic measurements on poroelastic slabs: Determination of reflection and transmission coefficients and processing for Biot input parameters," Ultrasonics **49**, 319–330.
Kaczmarek, M., Kubik, J., and Pakula, M. (**2002**). "Short ultrasonic waves in cancellous bone," Ultrasonics **40**, 95–100.
Kaufman, J. J., Luo, G., and Siffert, R. S. (**2007**). "A portable real-time ultrasonic bone densitometer," Ultrasound Med. Biol. **33**, 1445–1452.
Kaufman, J. J., Xu, W., Chiabrera, A. E., and Siffert, R. S. (**1995**). "Diffraction effects in insertion mode estimation of ultrasonic group velocity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 232–242.
Kaye, G. W. C., and Laby, T. H. (**1973**). *Table of Physical and Chemical Constants* (Longman, London, UK).
Krieg, M. A., Barkmann, R., Gonnelli, S., Stewart, A., Bauer, D. C., Del Rio Barquero, L., Kaufman, J. J., Lorenc, R., Miller, P. D., Olszynski, W. P., Poiana, C., Schott, A., Lewiecki, E. M., and Hans, D. (**2008**). "Quantitative ultrasound in the management of osteoporosis: The 2007 ISCD official positions," J. Clin. Densitom. **11**, 163–187.
Langton, C. M., Njeh, C. F., Hodgskinson, R., and Carrey, J. D. (**1996**). "Prediction of mechanical properties of the human calcaneus by broadband ultrasonic attenuation," Bone **18**, 495–503.
Langton, C. M., Palmer, S. B., and Porter, R. W. (**1984**). "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91.
Laugier, P. (**2008**). "Instrumentation for in vivo ultrasonic characterization of bone strength," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1179–1196.
Le Floch, V., McMahon, D. J., Luo, G., Coeh, A., Kaufman, J. J., Shane, E., and Siffert, R. S. (**2008**). "Ultrasound simulation in the distal radius using clinical high-resolution peripheral CT images," Ultrasound Med. Biol. **34**, 1317–1326.
Lin, W., Qin, Y., and Rubin, C. (**2001**). "Ultrasonic wave propagation in trabecular bone predicted by the stratified model," Ann. Biomed. Eng. **29**, 781–790
Marutyan, K. R., Holland, M. R., and Miller, J. G. (**2006**). "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL55–EL61.
Mizuno, K., Matsukawa, M., Otani, T., Laugier, P., and Padilla, F. (**2009**). "Propagation of two longitudinal waves in human cancellous bone: An *in vitro* study," J. Acoust. Soc. Am. **125**, 3460–3466.
Nicholson, P. H. F., and Bouxsein, M. L. (**2002**). "Bone marrow influences quantitative ultrasound measurements in human cancellous bone," Ultrasound Med. Biol. **28**, 369–375.
Nicholson, P. H. F., Lowet, G., Langton, C. M., Dequeker, J., and Van der Perre, G. (**1996**). "Comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurements in trabecular bone," Phys. Med. Biol. **41**, 2421–2435.
Nicholson, P. H. F., Müller, R., Lowet, G., Cheng, X. G., Hildebrand, T., Rüegsegger, P., Van der Perre, G., Dequeker, J., and Boonen, S. (**1998**). "Do quantitative ultrasound measurements reflect structure independently of density in human vertebral cancellous bone," Bone **23**, 425–431.
O'Donnell, M., Jaynes, E. T., and Miller, J. G. (**1981**). "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701.
Pakula, M., Padilla, F., Laugier, P., and Kaczmarek, M. (**2008**). "Application of Biot's theory to ultrasonic characterization of human cancellous bone: Determination of structural, material, and mechanical properties," J. Acoust. Soc. Am. **123**, 2415–2423.
Parfitt, A. M., Drezner, M. K., Glorieux, F. H., Kanis, J. A., Malluche, H., Meunier, P. J., Ott, S. M., and Recker, R. R. (**1987**). "Bone histomorphometry: Standardization of nomenclature, symbols, and units," J. Bone Miner. Res. **2**, 595–609.
Riekkinen, O., Hakulinen, M. A., Timonen, M., Töyräs, J., and Jurvelin, J. S. (**2006**). "Influence of overlying soft tissues on trabecular bone acoustic measurement at various ultrasound frequencies," Ultrasound Med. Biol. **32**, 1073–1083.
Riekkinen, O., Hakulinen, M. A., Töyräs, J., and Jurvelin, J. S. (**2008**). "Dual-frequency ultrasound—New pulse-echo technique for bone densitometry," Ultrasound Med. Biol. **34**, 1703–1708.
Sarvazyan, A., Tatarinov, A., Egorov, V., Airapetian, S., Kurtenok, V., and Gatt, C. J. (**2009**). "Application of the dual-frequency ultrasonometer for osteoporosis detection," Ultrasonics **49**, 331–337.
Sebaa, N., Fellah, Z. E. 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**, 1516–1523.
Strelitzki, R., and Evans, J. A. (**1996a**). "On the measurement of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213.
Strelitzki, R., and Evans, J. A. (**1996b**). "An investigation of the measurement of broadband ultrasonic attenuation in trabecular bone," Ultrasonics **34**, 785–791.
Ulrich, D., van Rietbergen, B., Laib, A., and Ruegsegger, P. (**1999**). "The ability of three-dimensional structural indices to reflect mechanical aspects of trabecular bone," Bone **25**, 55–60.
Waters, K. R., and Hoffmeister, B. K. (**2005**). "Kramers-Kronig analysis of attenuation and dispersion in trabecular bone," J. Acoust. Soc. Am. **118**, 3912–3920.
Wear, K. A. (**1999**). "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664.
Wear, K. A. (**2000a**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.
Wear, K. A. (**2000b**). "The effects of frequency-dependent attenuation and dispersion on sound speed measurements: Applications in human trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 265–273.
Wear, K. A. (**2001**). "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control. **48**, 1079–1083.
Wear, K. A. (**2004**). "Measurement of frequency dependence of scattering from cylinders using focused transducers—With applications in trabecular bone," J. Acoust. Soc. Am. **115**, 66–72.
Wear, K. A. (**2005**). "The dependencies of phase velocity and dispersion on trabecular thickness and spacing in trabecular bone-mimicking phantoms," J. Acoust. Soc. Am. **118**, 1186–1192.
Xia, Y., Lin, W., and Qin, Y. (**2005**). "The influence of cortical end-plate on broadband ultrasound attenuation measurements at the human calcaneus using scanning confocal ultrasound," J. Acoust. Soc. Am. **118**, 1801–1807.

Influence of the filling fluid on frequency-dependent velocity and attenuation in cancellous bones between 0.35 and 2.5 MHz

Michal Pakula[a)]
Institute of Mechanics and Applied Computer Science, Kazimierz Wielki University, 85-064 Bydgoszcz, Poland and Laboratoire d'Imagerie Paramétrique, Université Pierre et Marie Curie–Paris 6, F-75005 Paris, France

Frederic Padilla and Pascal Laugier
Laboratoire d'Imagerie Paramétrique CNRS, UMR 7623, Université Pierre et Marie Curie–Paris 6, F-75005 Paris, France

(Received 21 May 2009; revised 29 September 2009; accepted 1 October 2009)

The paper is focused on experiments on human cancellous bones filled with different fluids with the goal of evaluating their contribution to velocity dispersion, absorption, and scattering mechanisms. The specimens were measured first filled with marrow and subsequently, after marrow removal, with water and alcohol. No significant influence of the fluids was evidenced on the attenuation coefficient. Given the absence of impact of viscosity of the saturating fluid, the authors hypothesized that the source of attenuation is associated with viscoelastic absorption in the solid trabeculae and with scattering. Alteration of scattering obtained by changing the acoustic impedance mismatch between the fluid (alcohol vs water) and the trabeculae was reflected neither in the attenuation nor in its slope. This led the authors to suggest that longitudinal-to-shear scattering together with absorption in the solid phase are candidates as main sources for the attenuation. The differences in velocity values indicate that the elastic properties of the fluid are main determinants of the phase velocity. This finding is particularly significant in the context of /*in vivo*/ measurements, because it demonstrates that the subject-dependent properties of marrow may partly explain the inter-subject variability of speed of sound values.
 [DOI: 10.1121/1.3257233]

PACS number(s): 43.80.Ev, 43.80.Qf, 43.80.Cs, 43.20.Jr [CCC] Pages: 3301–3310

I. INTRODUCTION

Quantitative ultrasound (QUS) techniques have received considerable attention during the past decade for their application to the characterization of cancellous bone, particularly in the context of non-invasive *in vivo* assessment of bone quality. Despite many clinical successes of bone QUS, the basic physics associated with wave propagation in cancellous bone are not fully elucidated yet. Understanding the mechanisms that relate QUS variables such as attenuation and speed of sound (SOS) to cancellous bone properties is needed to properly interpret *in vivo* measurements.

One of the reasons of moderate comprehension of the fundamental interaction between ultrasound and bone arises from the complex, heterogeneous, and highly porous (typically above 80% porosity) structure of cancellous bones. Ultrasonic waves propagating in bones are influenced by material and mechanical properties of the solid phase,[1–3] its anisotropy,[4–10] macroscopic structural parameters such as porosity,[3,11–14] permeability or tortuosity,[15,16] as well as by physical properties of the saturating fluid.[17–20]

Considering *in vivo* ultrasonic studies of cancellous bones, the pores are filled with a fluid-like mixture of red (hematopoietic) and yellow (fatty) marrow,[21] while in *in vitro* experiments usually marrow is removed and replaced with water.[19] *In vitro* studies of water-saturated bone specimens are motivated by practical reasons such as preservation, specimen manipulation, and ease of experimentation in water. In addition, because marrow composition is known to vary with age, skeletal site, and health condition,[21] replacing marrow with water is beneficial for reducing the variability of the results and for accurately determining the relationships of QUS variables to structural and material bone properties. On the other hand, a few questions arise when bone marrow is removed and replaced with water.

(1) It remains unclear to what extent the inherent material properties of the saturating fluid significantly affect the measured QUS properties, by changing the absorption and/or scattering conditions. Mass density and bulk compression modulus are similar for water and marrow, which leaves scattering conditions almost unchanged in water or marrow-saturated specimens. In contrast, marrow and water differ by their viscosity, which may influence significantly internal dissipation in the fluid phase or the dissipation caused by friction at the fluid/trabeculae interface. The impact of both mechanisms on total loss depends on their relative contribution and may change with frequency.[22]

[a)]Author to whom correspondence should be addressed. Electronic mail: michalp@ukw.edu.pl

Comparison of sound velocity and attenuation coefficient measurements for marrow- and water-saturated bones has been investigated, but inconsistent results have been reported. For example, the presence of water in the pores instead of marrow has been reported to have a significant impact on SOS ($m\,s^{-1}$) and the slope of the frequency-dependent attenuation [the so-called normalized broadband ultrasonic attenuation or nBUA, $dB/(MHz\,cm)^{-1}$] experimentally in human femur[17] and human calcaneus specimens[19] and numerically in finite difference time domain simulations of wave propagation.[23] In contrast, other groups reported no significant difference in nBUA as well as in the phase velocity when marrow was replaced with water in human calcaneus specimens both experimentally[24,25] and numerically.[18] Reasons for these inconsistencies have not been investigated in detail. There may be several reasons, such as experimental imprecision, differences in structure of investigated bone samples (e.g., bone volume fraction and trabecular orientation), or in signal processing procedures.

(2) Similarly the effect of inherent individual variability in marrow on the measured QUS properties and on relationship of QUS to bone properties is unclear. To our knowledge, only one study has specifically looked at this effect on the correlation between QUS measurements and bone mineral density (BMD).[19] They found that QUS measurements in marrow-saturated specimens correlated less strongly with BMD than did corresponding measurements in water-saturated specimens. The authors justified this result by the inter-specimen marrow heterogeneity and concluded that the potential impact of marrow should be considered when interpreting clinical QUS measurements. Such effects cannot be mimicked when marrow is replaced with water.

New data are needed to enlighten the role of the saturating fluid in determining attenuation and speed of sound of ultrasonic waves transmitted through cancellous bone. This paper describes the influence of the physical properties of the interstitial fluid on the phase velocity and attenuation coefficient using new ultrasonic measurements of human cancellous bone specimens saturated with three different fluids. Two different fluids, water and alcohol, were used to replace marrow with the specific aim of varying the absorption and scattering conditions. Water and marrow have similar bulk modulus but huge difference in their viscosity [viscosity in marrow is several orders of magnitude higher than in water,[17,26,27] while water and alcohol have similar viscosity but differ by their bulk modulus (bulk modulus of water and marrow is almost twice higher than that of alcohol)]. The influence of changing the density is not addressed here, as density is approximately the same for marrow, water, and alcohol. Our aim is to provide new experimental data to test in future work the accuracy of models of ultrasound transmission through cancellous bone. An ancillary objective is to evaluate whether water-saturated cancellous bone specimens reproduce the same acoustical properties as marrow-saturated ones and thus can be considered as a good experimental model to investigate the properties of marrow-saturated bone. Toward these goals, cancellous bone plates were prepared from fresh human condyles. Ultrasonic measurements were performed at using three pairs of transducers having center frequencies of 0.5, 1, and 2 MHz on marrow-saturated specimens and subsequently (after marrow removal from the pores) on water-saturated specimens. The third series of experiments was performed in alcohol-saturated specimens, using the 0.5 and 1 MHz center frequency transducers.

II. MATERIAL AND METHOD

A. Specimen preparation

26 human cancellous bone specimens were obtained from the distal end of human femurs (13 from left and 13 from right condyle) belonging to 13 individuals (skeletal status of the donors was not known). Ethics approval for collection of samples was granted by the Human Ethics Committee of the Institute of Anatomy at the University René Descartes (Paris, France). The donors or their legal guardians provided informed written consent to provide their tissues for investigation in accordance with legal clauses stated in the French Code of Public Health (Code de la Santé Publique Français).

The femoral head and the diaphysis were first removed, leaving the distal part of the femur (condyles). Slices of cancellous bone with parallel faces and thickness of approximately 15 mm in thickness were cut from the femoral condyles frozen at −20 °C. The specimens were cut perpendicularly to the long femur axis using a circular precision diamond saw (Isomet 4000, Buehler, Lake Bluff, IL). In order to protect the specimens from local heating during cutting, fluid cooling along with slow cutting rotational velocity (4000 r/min) and linear velocity (7 mm/min) was applied. The final size of the specimens depended on the anatomical size of each femur, but typically dimensions did not exceed $65\times65\times15$ mm^3 (Fig. 1). Each sample thickness was measured using an electronic calliper.

After the initial measurements were performed on marrow-saturated specimens, all the specimens were defatted by supercritical carbon dioxide followed by inactivation using sequential exposure to hydrogen peroxide, sodium hydroxide, monosodium dihydrogen phosphate, and ethanol. Technical details concerning the procedure are not presented within the paper, because in our studies intertrabecular fluid removal was performed accordingly to the standards published elsewhere,[28–30] by the Biobank company (Biobank S.A., Presle-en-Brie, France). The treatment resulted in a decrease in the apparent mass of the specimens from 25.8 to 50 g (average value of 37.8 g).

B. Ultrasonic measurements

1. Data acquisition

Ultrasonic measurements were performed in immersion following the insertion method described in detail elsewhere.[22,31,14,32] The wave parameters (phase velocity and attenuation coefficient as a function of frequency) were computed from the ratio of the spectrum of the broadband pulse transmitted through the specimen to the spectrum of the reference pulse recorded when there is no specimen between the transducer and receiver. Experiments were done using three pairs of custom-made 10-mm-diameter unfocused ul-

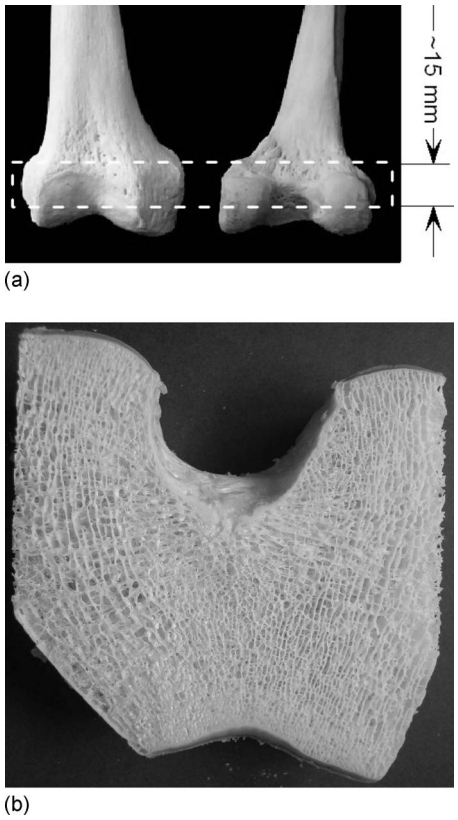


FIG. 1. Orientation of cut slice (a) and picture of the slice of bone specimen (b).

trasonic transducers having center frequencies of 0.5, 1, and 2 MHz (Optel, Wroclaw, Poland). During all experiments, the distance between transducers was maintained constant (about 8 cm), while the specimen was positioned at mid-distance between the transmitter and the receiver. Transmitted radio-frequency (rf) signals were recorded along two-dimensional (2D) scans in steps of 2 mm. The size of the scans (70×70 mm^2) was chosen to fit the sample size. During the scan, the transducers were maintained in a fixed position while the bone specimens were automatically moved in X and Y directions, using high precision stepping motors mounted on a crossed slide assembly (NewPort Motion Controller MM4006, Irvine, CA). The radio-frequency signals were generated and acquired with 60 MHz sampling frequency, by an 8-bit ultrasonic programmable PC board SFT 4001H PCI (Sofratest, Ecquevilly, France). In order to increase signal-to-noise ratio, the amplitude of rf signals was automatically adjusted to the dynamic range of the analog/digital converter and were time averaged (20 times). The scheme of the experimental setup is presented in Fig. 2.

Ultrasonic measurements were performed in three steps: (1) intact or marrow-saturated bone specimens (with marrow kept inside the pores), (2) defatted specimens refilled with water, and (3) defatted specimens refilled with alcohol (ethanol 95%–96% pure Technisolv, VWR Prolabo, France). The relative orientation between the ultrasound beam and the trabeculae was assumed as not known. In all cases the specimens were measured following an identical experimental protocol described above. Marrow-saturated specimens (step 1) were immersed in a water bath. In order to avoid presence

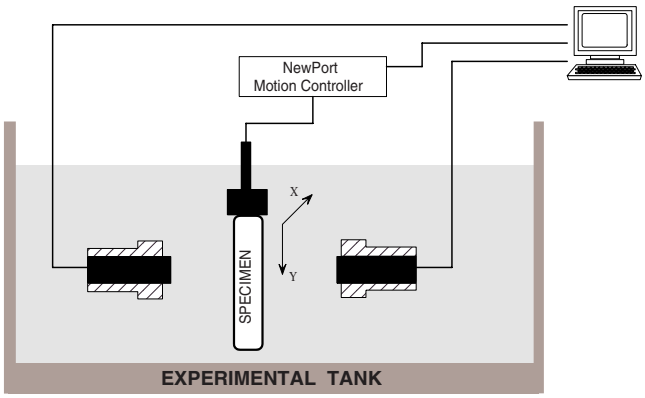


FIG. 2. (Color online) Sketch of ultrasonic measurement system.

of air bubbles, each specimen was degassed at room temperature (system composed by vacuum oil pump with refrigerated vapor trap, Savant, Thermo Scientific, Waltham, MA) and kept in the water/alcohol under the vacuum for 24 h before the measurements. Defatted specimens were also degassed and water (step 2) or alcohol (step 3) vacuum saturated. The bone specimens were immersed in a water (step 2) or alcohol (step 3) bath. The speed of sound in the coupling fluid (water or alcohol) was measured using a pulse echo technique. The experiments for marrow-filled specimens were performed within the temperature range 21–24 °C, and for each sample the temperature values were recorded. Then, the studies for water-filled state and alcohol-saturated specimens were done at the same temperatures as experiments for marrow-saturated samples controlled by thermostat (Polystat II, Bioblock Scientific, France) with an accuracy of 0.1 °C.

The physical properties of marrow water and alcohol are indicated in Table I.

2. Data processing

Figure 3 presents ultrasonic images of one representative specimen. The ultrasonic images (obtained by 1 MHz transducer) represent the ratio of amplitude of the signal transmitted through the bone to the amplitude of the reference signal transmitted through the coupling fluid, and expressed in dB. Such ultrasonic images were calculated for marrow-, water-, and alcohol-saturated cancellous bones at three different frequencies (0.5, 1, and 2 MHz) and were treated as the first step of signal processing. It is worth noticing that the shape of the specimen was well represented.

Because it was difficult to accurately reproduce the positioning of each specimen for all pairs of transducers and for different saturating fluids, a critical step of data preprocessing consisted in defining site-matched regions of interest (ROIs) for all ultrasonic scans obtained at different frequencies and with different filling fluids.

While a two-dimensional cross-correlation algorithm applied to the 2D signal magnitude images was successful to detect and correct for the translational or rotational shift between 1 and 2 MHz images, the same algorithm failed for the 0.5 MHz images, presumably because the spatial resolution achieved at 0.5 MHz was too low. We therefore decided to analyze square ROIs centered on the scan point where the

TABLE I. Physical and mechanical properties of the saturating fluids used in the paper.

Fluid	Density (kg m^{-3})	SOS (m s^{-1})	Attenuation (dB cm^{-1}) at 1 MHz	Bulkmodulus (GPa)	Viscosity (mPa s)
Marrow (bovine 20 °C)	950	1445[a]	⋯	⋯	1500[b]
Water (25 °C) [c,d]	998	1497	2.2×10^{-3}	2.25	0.89
Alcohol (20 °C) [e]	806[e]	1220[f]	4.2×10^{-3} [c,d]	1.19[f]	1.2[e]

[a]Reference 58.
[b]References 6 and 27.
[c]Reference 59.
[d]Reference 60.
[e]Producer specification. Ethanol 95%–96% vol. Technisolv C_2H_6O, VWR Prolabo, Paris, France.
[f]Own experiments.

transmitted signal had the highest magnitude. Such a region, illustrated in Fig. 3, could consistently be defined in the center of all the specimens and all frequencies. We assumed that for a given specimen, the region of maximum amplitude remains located at a fixed position whatever the filling fluid or the measuring frequency. Following this assumption, the coordinates (x,y) of the scan point of maximum amplitude could automatically be determined and the ROI subsequently centered on this scan point.

Subsequently, the transmitted signals of the ROI were used to obtain the frequency-dependent phase velocity and attenuation coefficient. Briefly, as an initial step, a Tukey window function was applied in order to extract the pulse of highest amplitude from the rest of the transmitted or reference signal. The center of the Tukey window was positioned at the maximum of the pulse, while the window width was chosen based on the pulse width of the reference pulse (thus the window width was frequency dependent). Then, the ratio of the Fourier transform of the windowed signal transmitted through the sample to that of a windowed reference signal transmitted through the coupling fluid (water for the marrow- and water-saturated specimens, and alcohol for the alcohol-saturated specimens) was computed. The useful frequency bandwidth for each transducer was set at the −6 dB level of the maximum of the amplitude spectrum of the transmitted signal through bone, resulting in frequency bandwidths of approximately 0.35–0.6, 0.6–1.2, and 1.8–2.5 MHz for 0.5, 1, and 2 MHz transducers, respectively. The log-magnitude of the spectra ratio normalized to the specimen thickness gave the frequency-dependent attenuation coefficient while the phase of the ratio yielded the phase velocity. The slope of a linear regression fit to the attenuation coefficient gave the slope of the attenuation coefficient (nBUA). The phase velocity was obtained following the method proposed by Sachse[33] and extensively described by Droin *et al.*[34] in the case of bone. SOS values were obtained as the value of the phase velocity at center frequencies of 0.5, 1, and 2 MHz, respectively. Square ROIs of 9 pixels (13×13 mm^2) were analyzed by averaging the frequency-dependent phase velocity and attenuation coefficient of the 9 pixels of the ROI. Figures 3(a)–3(c) illustrate the ROI for one of representative specimens for one frequency (1 MHz) and each saturating fluid.

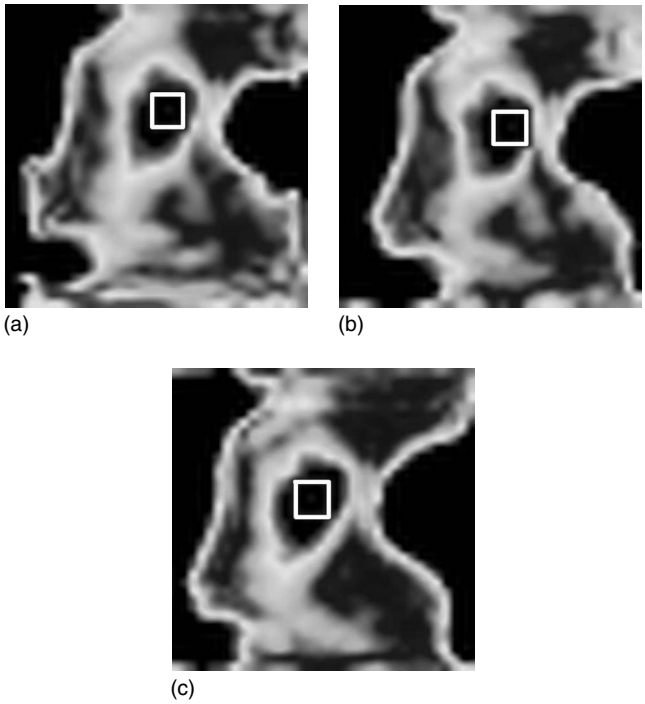


FIG. 3. Ultrasonic images (obtained by 1 MHz transducer) for one cancellous bone specimen with marked ROI (square in the center part of each image) including nine transmitted signals for marrow- (a), water- (b), and alcohol-saturated (c) specimens. Amplitude ratios of the signal transmitted through the bone and the reference signal transmitted through the coupling fluid are represented in arbitrary units.

3. Reproducibility

In order to assess the experimental reproducibility and to estimate the potential effect of alcohol filling on the solid trabeculae, repeated measurements were reproduced on a subset of four representative water-saturated specimens (the least attenuating, the most attenuating, and two intermediate attenuating specimens) after an interval of 2 months following the initial measurements in the water-filled state. In the interval, the specimens were measured and saturated in alcohol-filled state. The difference between the two measurements was assessed using the paired-sample Wilcoxon signed rank test. Variation in the attenuation coefficient was found to vary from 0.2 to 1 dB cm^{-1} for the least and the most attenuating specimens, respectively, while the maximum difference in the phase velocity did not exceed 20 m s^{-1}. None of the difference was found to be statistically significant. This suggests that alcohol filling had no

observable effect on the trabecular microstructure. The reproducibility was better than 10% and 2% for the attenuation coefficient and the phase velocity, respectively.

III. RESULTS

The heterogeneity of the bone structure is reflected in the spatial distribution of the amplitude of the transmitted signal as illustrated in Fig. 3. Typically, the within-specimen variability of the amplitude of the transmitted signal is 60 dB. Moreover, in some regions, the transmitted signal contains two overlapping pulses, which presumably correspond to the fast and slow waves already reported by others experimentally[6,15,20,35–38] or in computer models[39,40] and predicted by Biot's theory.[41] Such pulse overlapping, depending on the time separation and amplitude ratio of the two waves, may be responsible for an apparent non-linear frequency-dependent attenuation and "anomalous" negative phase velocity dispersion.[42–44] It precludes any reliable estimate of the ultrasonic parameters without former wave separation. Mixed waveforms composed of overlapping pulses were mostly observed in highly attenuating regions where low amplitude signals were transmitted.

Due to the difficulty with subsequent interpretation of the results obtained from mixed waveforms consisting of overlapped pulses, we decided to restrict the analysis to low attenuating regions where the rf signals contain presumably a single apparent waveform. Fortunately, such ROIs were consistently located in the center of the specimens where the porosity (assessed by microcomputed tomography, results not shown in this paper) is maximal and where the transmitted pulses had the highest amplitude (see marked ROIs in the central part of each ultrasonic image in Fig. 3). Figure 4(a) illustrates such waveform transmitted through the rather low attenuating center ROI using 1 MHz transducers. In this case, the attenuation coefficient frequency variation is quasi-linear and the dispersion is weak $[-16.6\ \mathrm{m/(s\ MHz)^{-1}}]$. Finally, 25 specimens could be selected (from the set of 30 specimens).

For these specimens, we assumed transmission of a single waveform, and wave parameters (attenuation coefficient and phase velocity) as a function of frequency were calculated based on the transmitted signals through bone and the reference signal transmitted through the coupling fluid following the procedure described above.

Figure 5 and Table II summarize the average values and standard deviations of SOS, ΔSOS, α, and nBUA for all measured specimens as a function of frequency. Fluid effect on measured parameters (at 0.5 and 1 MHz) was determined by one-way analysis of variance (ANOVA) followed by *post hoc* multiple comparison Tukey–Kramer tests. When only two fluids were measured (at 2 MHz), the paired *t*-test for differences between water and marrow was performed. Results of the independent one-sample *t*-test testing the null hypothesis for velocity dispersion that the population mean is equal to 0 are also shown. The main conclusions are summarized as follows.

(1) One-way ANOVA showed a significant effect of the filling fluid for phase velocity at each center frequency. All differences between the three groups (alcohol vs marrow, alcohol vs water, and marrow vs water) were statistically significant. The phase velocity of marrow-saturated specimens was approximately 20 $\mathrm{m\ s^{-1}}$ lower than that of water-saturated specimens. The difference was statistically significant ($p<0.0001$) at all frequencies. The velocity was significantly lower (about 300 $\mathrm{m\ s^{-1}}$) in alcohol-saturated specimens compared to water-saturated specimens with a significant level $p<0.0001$.

(2) A strong negative dispersion was observed for marrow-saturated $[-108\pm71\ \mathrm{m/(s\ MHz)^{-1}}]$ and water-saturated $[-98\pm60\ \mathrm{m/(s\ MHz)^{-1}}]$ specimens measured at low frequencies (0.5 MHz). The magnitude of dispersion decreases as the frequency increases until absence of dispersion at 2 MHz. One-way ANOVA showed a significant effect of the filling fluid on velocity dispersion. Dispersion in alcohol-filled state was significantly different from dispersion in marrow- or water-saturated states ($p<0.0001$). Velocity dispersion showed no significant difference between the marrow- and water-saturated states. The same group of specimens filled with alcohol exhibits on average a slightly but significant negative dispersion $[-8\pm13\ \mathrm{m/(s\ MHz)^{-1}}]$ at 0.5 MHz. Practically, no dispersion $[1\ \mathrm{m/(s\ MHz)^{-1}}]$ was measured at 1 MHz for alcohol-saturated specimens.

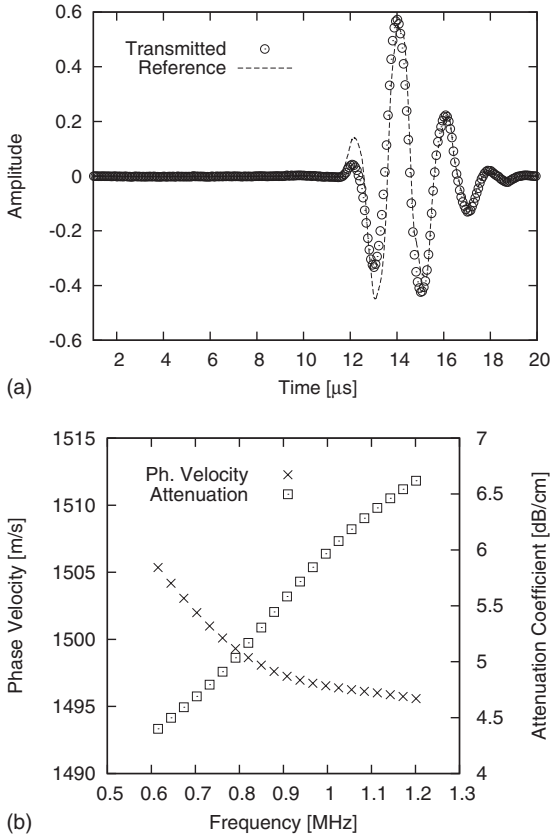


FIG. 4. (a) Typical waveform (at 1 MHz frequency) transmitted through a water-saturated bone specimen (circle) compared to the reference signal (continuous line) and the corresponding attenuation coefficient (square) and phase velocity (cross) as a function of frequency (b). The rf transmitted and reference signals (a) are normalized by their maximum amplitude (real values can be obtained by multiplying the reference signal by a factor 83.5, and the transmitted signal by a factor 45.7).

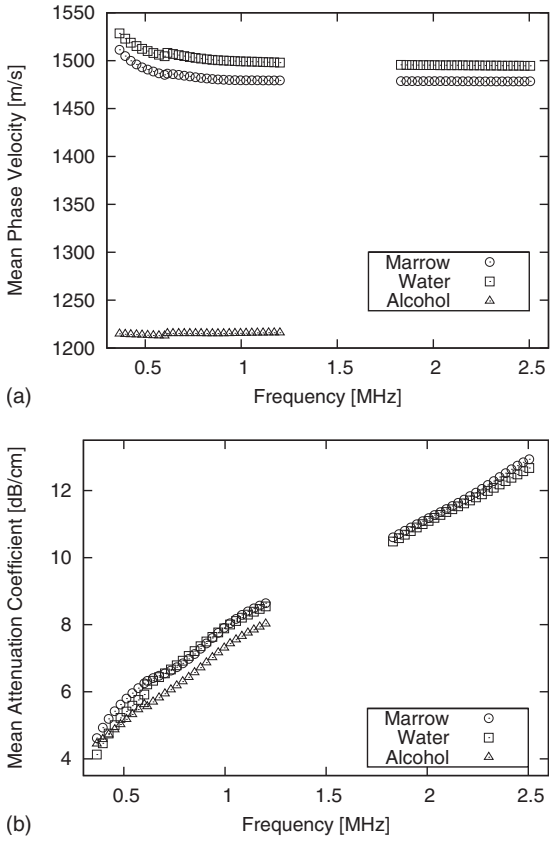


FIG. 5. Mean phase velocities (a) and mean attenuation coefficients (b) vs frequency for marrow-saturated (circle), water-saturated (square), and alcohol-saturated (triangle) specimens ($N=25$).

(3) One-way ANOVA did not show any significant effect of the filling fluid for the attenuation coefficient nor for nBUA at any frequency. For nBUA at 0.5 MHz, there was a trend for lower values in the alcohol-saturated state $[5.1 \pm 3.2\ \mathrm{dB/(cm\ MHz)^{-1}}]$ compared to water-saturated specimens $[7.3 \pm 4.0\ \mathrm{dB/(cm\ MHz)^{-1}}]$; however, this result was not confirmed by one-way ANOVA. For the three saturating fluids, the attenuation coefficient increases quasilinearly with frequency, though a slightly lower than linear frequency dependence of attenuation is obtained at 0.5 MHz when ultrasound propagates in marrow- and water-saturated specimens. This non-linear behavior seems to be less pronounced in the case of alcohol-filled state. A similar behavior is found for marrow- and water-filled specimens.

IV. DISCUSSION

The primary focus of this study was to investigate experimentally the effect of the saturating fluid on the phase velocity and attenuation coefficient of ultrasonic waves transmitted through fluid-filled cancellous bone specimens. Such experiments have already been reported in bovine cancellous bones saturated with a variety of silicone oils having different viscosities[45] or with water and alcohol.[20] In other experiments, human cancellous bones have been compared in the marrow-filled or water-filled states.[17,23,25] Our study is the first report in which QUS measurements are achieved as a function of frequency in human cancellous bone specimens saturated with three different fluids selected to study the specific impact of fluid bulk compression modulus and viscosity on the measured QUS parameters.

In experiments on bovine bone, two longitudinal waves known as the fast and slow waves can generally be observed and their properties can be assessed separately.[6,36] Interestingly, Camus *et al.*[45] concluded that the properties of the pore-filling fluid were affecting the velocity of the slow wave only. In contrast, in experiments in human bone, the ultrasound parameters have generally been computed as if the signal observed in transmission through the marrow- or water-filled specimens contained only one wave. This approach was also adopted here by restricting our analysis to weakly attenuating regions in the center of the specimens and to their corresponding signals apparently consisting of a single waveform. The bone regions measured here were the least attenuating part of the bone and were characterized by a high porosity. Porosity values assessed by micro-computed tomography were found to vary from 90% to 98% (data not shown here).

Within these ROIs, the amplitude of the slow wave is expected to be much larger than the amplitude of the fast wave. Assumption of a single waveform gives the advantage of using conventional spectroscopy techniques to compute the attenuation coefficient and phase velocity. On the other hand, the absence of visible interference on rf signals in the

TABLE II. Wave parameters for $N=25$ representative marrow-saturated, water-saturated, and alcohol-saturated specimens obtained at center frequencies 0.5, 1, and 2 MHz and for the whole measured frequency range: SOS (m s^{-1}), phase velocity; ΔSOS m/(s MHz)$^{-1}$, dispersion of phase velocity; α, attenuation coefficient (dB cm^{-1}); nBUA, slope of attenuation coefficient dB/(cm MHz)$^{-1}$; and SD, standard deviation.

	0.5 MHz				1 MHz				2 MHz			
	SOS (±SD)	ΔSOS (±SD)	α (±SD)	nBUA (±SD)	SOS (±SD)	ΔSOS (±SD)	α (±SD)	nBUA (±SD)	SOS (±SD)	ΔSOS (±SD)	α (±SD)	nBUA (±SD)
Marrow	1493[a] (±19)	−108[b] (±71)	5.7 (±2.8)	6.6 (±4.4)	1480[a] (±8)	−12[b] (±11)	7.9 (±3.5)	4.4 (±2.1)	1478[a] (±4)	0 (±1)	11.2 (±4.8)	3.5 (±1.4)
Water	1512 (±16)	−98[b] (±60)	5.3 (±2.5)	7.3 (±4)	1500 (±7)	−17[b] (±11)	7.9 (±3.5)	4.2 (±2.2)	1495 (±4)	−1 (±1)	11.1 (±5)	3.2 (±1.4)
Alcohol	1214[c] (±8)	−8[c,b] (±13)	5.1 (±2.6)	5.1 (±3.2)	1215[c] (±4)	1[c] (±4)	7.3 (±3.4)	4.4 (±2)	…	…	…	…

[a] $p<0.05$, *post hoc* multiple comparison Tukey–Kramer test for differences between water and marrow.
[b] $p<0.05$, *t*-test for independent one-sample *t*-test in testing the null hypothesis that the population mean is equal to 0.
[c] $p<0.05$, *post hoc* multiple comparison Tukey–Kramer test for differences between water and alcohol.

time domain does not guarantee the absence of such interferences between two overlapping pulses,[44] especially if the difference in arrival time (determined by disparities of velocities) of both waves is small compared to the temporal extent (determined by frequency) or if the amplitude ratio between the fast and slow waves is small (depending on the porosity and orientation of the trabeculae).

Inconsistent results were reported by investigators who compared marrow- and water-saturated human specimens.[17,19,23,25] However, the tremendous heterogeneity of cancellous bone must be kept in mind when performing comparative measurements on the same specimens saturated with different fluids or when using different transducers to explore a wide frequency bandwidth. The strength of our study relies on the comparison of measurements performed in a single set of human cancellous bone samples using a strict site-matching control based on automatic data processing ensuring that exactly the same regions of our bone samples were assessed when they were saturated with different fluids or measured at different frequencies. Such a protocol was designed to minimize the influence of heterogeneity and decrease the measurement errors. The precision of our experiments is illustrated by the continuity of the frequency-dependent attenuation coefficient and phase velocity recorded for different saturating fluids in various bandwidths.

A consistent trend emerged for the velocity results. The presence of marrow decreased the speed of sound by approximately 20 $m\ s^{-1}$ compared to the water-filled state, in general agreement with the results reported in Refs. 17 and 19. Velocity values for marrow itself are not well known, but the fact that the values recorded for marrow-saturated specimens are slightly lower than for the water-saturated specimens is probably be due to the high proportion of fat (known to have a lower sound speed than water) contained in cancellous bone marrow. The difference in velocity values recorded in water and alcohol-saturated specimens is close to the difference in velocity for pure water and alcohol. These results reflect the strong impact on speed of sound values of the saturating fluid, which predominates in the highly porous regions of cancellous bone investigated in the present study.

Like other investigators, we observed a negative dispersion for the marrow- and water-filled states in the low frequency range 0.35–0.6 MHz.[34,46,47] For the first time, velocity dispersion is also reported at higher frequencies until 2.5 MHz. Additional interesting features revealed by our measurements are as follows:

(1) less negative dispersion values, as the frequency becomes higher;
(2) virtually no dispersion in the alcohol-filled state except for a weak negative dispersion around 0.5 MHz; and
(3) velocity dispersion is similar in the marrow- and water-filled states.

The sources of velocity dispersion in cancellous bone are complicated by the fact that two waves can be transmitted through the bone porous network and subsequently overlap in time domain. As shown by Anderson *et al.*,[44] velocity dispersion depends on a variety of factors including relative amplitudes and velocities of the fast and slow waves. In particular, the authors showed that the dispersion becomes more negative, as the fast and slow waves become closer in magnitude. Attenuation, which increases as a function of frequency, is a potential mechanism by which velocity dispersion can change when the frequency increases. Assuming that the attenuation coefficient increases with frequency more for the fast wave (a wave component that propagates in the solid network) than for the slow wave (a wave component that propagates in the fluid phase), the amplitudes of the fast and slow waves become more disparate when frequency increases, resulting in less impact of the fast wave in the interference and producing less negative dispersion. Other sources of negative dispersion have been recognized such as multiple scattering[46] or phase cancellation[43] but they are not considered in the present discussion.

In alcohol-filled specimens negative dispersion is negligible in comparison with water- or marrow-filled state. A possible explanation of such behavior is lower (about 300 $m\ s^{-1}$) wave velocity in alcohol than in water or marrow. It can result in a better separation in time domain of the two wave modes in the alcohol-filled state compared to the water- or marrow-filled states, and as a consequence it is more likely that the one of the wave modes (if two existed) was gated out of the analysis window. Accordingly to the results of Marutyan *et al.*[42] an almost complete time separation of the two wave components presumably causes less or even the absence of negative dispersion.

The overlap between the two wave modes in the time domain does not depend only on the velocity disparities between the solid and fluid parts. It is further complicated by the influence of the frequency, which determines the length of the pulses in time. A small overlap in the time window of analysis for long pulses at low frequency may explain the small negative dispersion in the alcohol-filled state at 0.5 MHz. Increasing the frequency decreases the degree of overlap between the two pulses as a result decreases the negative dispersion.

Although individual fast and slow waves exhibit positive dispersions in accordance with the Kramers–Kronig relations for attenuation and dispersion, no significant positive dispersion could be observed here, either because a pure mode was never obtained (even in high frequencies or in the alcohol-filled state) or because the dispersion of the slow wave in highly porous specimens is too weak and remains beyond the measurement precision. Future studies are required to elucidate the exact role of the two wave modes and their individual characteristics (amplitudes and velocities) in determining the measured apparent acoustic properties. Methods such as Bayesian approaches currently developed by other groups that unambiguously can recover the ultrasonic characteristics of the individual wave modes may be valuable to address these issues.[48]

The slight non-linear frequency dependence of attenuation may be explained to some extent by the overlap and interference of two waves. Indeed, it has been shown that attenuation begins to exhibit a non-linear-with-frequency behavior when multiple pulses overlap.[44] Other factors such as diffraction could also affect the frequency dependence of the attenuation coefficient. But these factors were not considered

in this study. A better characterization of diffraction effect would require further work. The more linear-with-frequency behavior of attenuation in alcohol-saturated compared to water-saturated specimens is consistent with a better time separation of the fast and slow waves and less negative dispersion.

In cancellous bone, several mechanisms can potentially be a source of attenuation. The mechanisms include the following:

(1) intrinsic absorption in the fluid,
(2) loss related to the viscous friction at the fluid/solid interface,
(3) intrinsic absorption (of either longitudinal or shear components) in the solid trabeculae, and
(4) scattering resulting from the acoustic impedance mismatch between the solid and fluid phases; scattering consists of longitudinal-to-longitudinal (LL) scattering and longitudinal-to-shear (LS) conversion (with subsequent scattering or absorption of mode-converted shear waves).

The estimation of the role of intrinsic absorption in the fluid and loss by friction at the solid/fluid interfaces was done by comparison of the attenuation coefficients for marrow- and water-saturated specimens (two fluids having similar elasticity and significantly different viscosity—see Table I). While there are three orders of magnitude between the viscosities in marrow and water/alcohol, no significant difference of attenuation coefficients and their slope (nBUA) could be found between the marrow- and water-saturated specimens. Therefore we conclude that neither intrinsic loss in the fluid nor viscous loss due to friction at the fluid/trabeculae interface play a significant role in attenuation mechanisms for all measured frequencies. This result therefore suggests that the source of attenuation in trabecular bones is due either to scattering or to absorption in the solid phase. This result might influence future modeling of ultrasound wave propagation through trabecular bones. It is, for example, in apparent contradiction with the Biot model, from which an increase in the saturating fluid viscosity would predict significant increase in attenuation for both waves.

The role of scattering was estimated by comparison of the values of attenuation coefficient for water- and alcohol-filled specimens. Given the different bulk velocities in the two fluids, a slight change in the propagating wavelengths was expected in the frequency range 0.5–2 MHz. The wavelength is in the range 0.75–3 mm for water and from 0.6 to 2.4 mm for alcohol. Scattering is expected to be slightly strengthened when the frequency increases and when the wavelength becomes closer to the characteristic size of the scattering trabeculae (i.e., in the alcohol-filled state). However, major changes in scattering were expected from replacement of water with alcohol in the pores. A 50% increase in acoustical impedance mismatch at the fluid/solid and solid/fluid boundaries results from the replacement of water with alcohol. Scattering (and subsequently attenuation losses) may be expected to be higher in alcohol-saturated specimens than in water-saturated specimens. However, experimentally, there was no significant difference of attenuation values measured in alcohol- and water-saturated specimens.

If scattering is a major component of total loss, the frequency dependence of attenuation should approach the frequency dependence of scattering. Models and experiments have shown a nearly cubic frequency dependence of LL scattering and backscatter coefficients[49–55] in cancellous bone. A nearly linear frequency dependence of attenuation has been reported in the majority of studies,[31,56] including the present one, particularly with alcohol-saturated specimens. Wear, noting the discrepancies between the frequency dependence of attenuation and scattering, hypothesized that LL scattering should not be a strong contributor to total loss.[52,53] Our experimental findings further reinforce this hypothesis. These results lead to the conclusion that LS mode conversion and intrinsic absorption in the solid should be considered in further experiments and models as potential significant components of total attenuation. In experimental studies performed on cancellous-bone-mimicking phantoms,[53] Wear concluded that this kind of LS mode conversion is a significant source of attenuation in the diagnostic frequency range. The Wear[53] findings are supported by the results of three-dimensional (3D) simulation study, done by Bossy *et al.*[22,39] who pointed the potential importance of LS mode conversion to explain the attenuation of ultrasonic waves in cancellous bone.

Insignificant variations of attenuation were obtained when the filling fluid changes, which contrasts with the strong variations of attenuation usually observed when the bone volume fraction changes by a few percent[1,11,14] or when the orientation of the trabecular network changes in the ultrasonic beam.[4,10] We therefore hypothesize that the source of wave attenuation can be associated with viscoelastic absorption in the solid phase and to longitudinal-shear mode conversion by the solid trabeculae. Moreover, shear waves, if present, are likely to be attenuated rapidly either by absorption or scattering.[57] The linear frequency dependence of attenuation coefficient of trabecular bone suggests that scattered shear wave energy, if present, is ultimately absorbed.

V. CONCLUSION

Overall, the fluid does not influence the attenuation coefficient of cancellous bone between 0.35 and 1.2 MHz. Moreover, a change in marrow by water did not induce changes in the attenuation coefficient for higher frequencies in the bandwidth 1.8–2.5 MHz.

For nBUA at 0.5 MHz, there was a trend for lower values in the alcohol-saturated state compared to water-saturated specimens. We suspect that two wave modes overlap at low frequencies in water-/marrow-saturated specimens, and that the difference in attenuation slope comes from processing together these overlapping pulses rather than from specific physical phenomenon associated with the elastic wave propagation in the bone structure. These results point to the difficulty of an accurate experimental determination of ultrasonic parameters in cancellous bone when two wave modes overlap.

Significant alteration of scattering conditions obtained by changing the acoustic impedance mismatch between the saturating fluid and the trabeculae was reflected neither in the attenuation nor in its slope. The attenuation remains nearly linearly frequency dependent in a broad frequency range. Our study also clearly demonstrated that the viscosity of the saturating fluid did not influence the amplitude of the attenuation coefficient neither its frequency dependence. These led us to suggest that attenuation is associated with the trabecular network and that LS scattering together with absorption in the solid phase are candidates as main sources for the attenuation.

It was found that the elastic properties of the fluid are main determinants of the phase velocity. These findings are particularly significant in the context of *in vivo* measurements, because they demonstrate that the subject-dependent properties of marrow may partly explain the inter-subject variability of SOS values. For *in vitro* investigations of the influence of bone material or microstructural properties, our suggestion is that marrow can be advantageously replaced with water, because attenuation will not be affected by the change, and the variability of SOS measurements will be reduced to the impact of the bone properties only.

Future studies will focus on the estimation of microstructural parameters of the trabecular network toward a better understanding of the velocity dispersion and attenuation mechanisms. Moreover, the analysis of wave parameters in regions where two longitudinal waves are observed will also be the subject of future work. Approaches developed to recover the individual properties of the fast and slow waves may prove to be useful for studying separately the characteristics of the fast and slow waves and their relation to the solid and fluid characteristics.

ACKNOWLEDGMENTS

Authors would like to thank the group of the Department of Physics, University of Jyvaskyla, Finland (particularly Dr. Petro Moilanen) for micro-tomographic studies of cancellous bones. Moreover, one of the authors (M.P.) acknowledges support from the European Commission (Marie Curie Fellowship).

[1]G. Haïat, F. Padilla, F. Peyrin, and P. Laugier, "Variation of ultrasonic parameters with microstructure and material properties of trabecular bone: A 3D model simulation," J. Bone Miner. Res. **22**, 665–674 (2007).
[2]B. K. Hoffmeister, S. A. Whitten, and J. Y. Rho, "Low-megahertz ultrasonic properties of bovine cancellous bone," Bone **26**, 635–642 (2000).
[3]O. Riekkinen, M. A. Hakulinen, M. J. Lammi, J. S. Jurvelin, A. Kallioniemi, and J. Töyräs, "Acoustic properties of trabecular bone—Relationships to tissue composition," Ultrasound Med. Biol. **33**, 1438–1444 (2007).
[4]C. Glüer, C. Wu, M. Jergas, S. Goldstein, and H. Genant, "Three quantitative ultrasound parameters reflect bone structure," Calcif. Tissue Int. **55**, 46–52 (1994).
[5]S. M. Han and J. Y. Rho, "Dependence of broadband ultrasound attenuation on the elastic anisotropy of trabecular bone," Proc. Inst. Mech. Eng., Part H: J. Eng. Med. **212**, 223–227 (1998).
[6]A. Hosokawa and T. Otani, "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722 (1998).
[7]P. H. F. Nicholson, M. J. Haddaway, and M. W. J. Davie, "The dependence of ultrasonic properties on orientation in human vertebral bone," Phys. Med. Biol. **39**, 1013–1024 (1994).
[8]C. F. Njeh, R. Hodgskinson, J. D. Currey, and C. M. Langton, "Orthogonal relationships between ultrasonic velocity and material properties of bovine cancellous bone," Med. Eng. Phys. **18**, 373–381 (1996).
[9]C. F. Njeh, C. W. Kuo, C. M. Langton, H. I. Atrah, and C. M. Boivin, "Prediction of human femoral bone strength using ultrasound velocity and BMD: An in vitro study," Osteoporosis Int. **7**, 471–477 (1997).
[10]K. A. Wear, "Anisotropy of ultrasonic backscatter and attenuation from human calcaneus: Implications for relative roles of absorption and scattering in determining attenuation," J. Acoust. Soc. Am. **107**, 3474–3479 (2000).
[11]S. Chaffai, F. Peyrin, S. Nuzzo, R. Porcher, G. Berger, and P. Laugier, "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microstructure," Bone **30**, 229–237 (2002).
[12]D. Hans, C. Wu, C. F. Njeh, S. Zhao, P. Augat, D. Newitt, T. Link, Y. Lu, S. Majumdar, and H. K. Genant, Calcif. Tissue Int. **64**, 18–23 (1999).
[13]P. H. Nicholson, R. Müller, G. Lowet, X. G. Cheng, T. Hildebrand, P. Rüegseger, G. van der Perre, J. Dequeker, and S. Boonen, "Do quantitative ultrasound measurement reflect structure independently of density in human vertebral cancellous bone," Bone **23**, 425–431 (1998).
[14]F. Padilla, F. Jenson, V. Bousson, F. Peyrin, and P. Laugier, "Relationships of trabecular bone structure with quantitative ultrasound parameters: In vitro study on human proximal femur using transmission and backscatter measurements," Bone **42**, 1193–1202 (2008).
[15]Z. E. Fellah, J. Y. Chapelon, S. Berger, W. Lauriks, and C. Depollier, "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73 (2004).
[16]M. Pakula, F. Padilla, M. Kaczmarek, and P. Laugier, "Application of Biot's theory to ultrasonic characterization of human cancellous bones: Determination of structural, material, and mechanical properties," J. Acoust. Soc. Am. **123**, 2415–2423 (2008).
[17]J. M. Alves, J. T. Ryaby, J. J. Kaufman, F. P. Magee, and R. S. Siffert, "Influence of marrow on ultrasonic velocity and attenuation in bovine trabecular bone," Calcif. Tissue Int. **58**, 362–367 (1996).
[18]J. J. Kaufman, G. Luo, and R. S. Siffert, "On the relative contributions of absorption and scattering to ultrasound attenuation in trabecular bone: A simulation study," presented at the 2003 Proceedings of the IEEE Ultrasonics Symposium (2003).
[19]P. H. Nicholson and M. L. Bouxsein, "Bone marrow influences quantitative ultrasound measurements in human cancellous bone," Ultrasound Med. Biol. **28**, 369–375 (2002).
[20]M. Kaczmarek, J. Kubik, and M. Pakula, "Short ultrasonic waves in cancellous bone," Ultrasonics **40**, 95–100 (2002).
[21]G. P. Liney, C. P. Bernard, D. J. Manton, L. W. Turnbull, and C. M. Langton, "Age, gender, and skeletal variation in bone marrow composition: A preliminary study at 3.0 Tesla," J. Magn. Reson Imaging **26**, 787–793 (2007).
[22]E. Bossy, P. Laugier, F. Peyrin, and F. Padilla, "Attenuation in trabecular bone: A comparison between numerical simulation and experimental results in human femur," J. Acoust. Soc. Am. **122**, 2469–2475 (2007).
[23]A. S. Aula, J. Töyräs, M. A. Hakulinen, and J. S. Jurvelin, "Effect of bone marrow on acoustic properties of trabecular bone—3D finite difference modeling study," Ultrasound Med. Biol. **35**, 308–318 (2009).
[24]C. M. Langton, C. F. Njeh, R. Hodgskinson, and J. D. Currey, "Prediction of mechanical properties of the human calcaneus by broadband ultrasonic attenuation," Bone **18**, 495–503 (1996).
[25]C. F. Njeh and C. M. Langton, "The effect of cortical endplates on ultrasound velocity through the calcaneus: An in vitro study," Br. J. Radiol. **70**, 504–510 (1997).
[26]J. D. Bryant, "On the mechanical function of marrow in long bones," Eng. Med. **17**, 55–58 (1988).
[27]J. D. Bryant, T. David, P. H. Gaskell, S. King, and G. Lond, "Rheology of bovine bone marrow," Proc. Inst. Mech. Eng., Part H: J. Eng. Med. **203**, 71–75 (1989).
[28]J. Fages, A. Marty, C. Delga, J. S. Condoret, D. Combes, and P. Frayssinet, "Use of supercritical CO_2 for bone delipidation," Biomaterials **15**, 650–656 (1994).
[29]L. Vastel, C. Masse, E. Crozier, F. Padilla, P. Laugier, D. Mitton, R. Bardonnet, and J. P. Courpied, "Effects of gamma irradiation on mechanical properties of defatted trabecular bone allografts assessed by speed-of-sound measurement," Cell Tissue Bank. **8**, 205–210 (2007).
[30]L. Vastel, C. Masse, P. Mesnil, E. Crozier, F. Padilla, P. Laugier, D. Mitton, and J. P. Courpied, "Comparative ultrasound evaluation of human trabecular bone graft properties after treatment with different sterilization

procedures," J. Biomed. Mater. Res., Part B: Appl. Biomater. **90**, 430–437 (2009).
[31]S. Chaffai, F. Padilla, G. Berger, and P. Laugier, "In vitro measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz," J. Acoust. Soc. Am. **108**, 1281–1289 (2000).
[32]P. Laugier, P. Droin, A. M. Laval-Jeantet, and G. Berger, "In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone **20**, 157–165 (1997).
[33]W. Sachse and Y. H. Pao, "On the determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **49**, 4320–4327 (1978).
[34]P. Droin, G. Berger, and P. Laugier, "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans Ultrason Ferroelectr. Freq Control. **45**, 581–592 (1998).
[35]N. Sebaa, Z. E. Fellah, M. Fellah, E. Ogam, A. Wirgin, F. G. Mitri, C. Depollier, and W. Lauriks, "Ultrasonic characterization of human cancellous bone using the Biot theory: Inverse problem," J. Acoust. Soc. Am. **120**, 1816–1824 (2006).
[36]A. Hosokawa and T. Otani, "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562 (1997).
[37]L. Cardoso, F. Teboul, L. Sedel, C. Oddou, and A. Meunier, "In vitro acoustic waves propagation in human and bovine cancellous bone," J. Bone Miner. Res. **18**, 1803–1812 (2003).
[38]K. Mizuno, M. Matsukawa, T. Otani, P. Laugier, and F. Padilla, "Propagation of two longitudinal waves in human cancellous bone: An in vitro study," J. Acoust. Soc. Am. **125**, 3460–3466 (2009).
[39]E. Bossy, F. Padilla, F. Peyrin, and P. Laugier, "Three-dimensional simulation of ultrasound propagation through trabecular bone structures measured by synchrotron microtomography," Phys. Med. Biol. **50**, 5545–5556 (2005).
[40]G. Haïat, F. Padilla, F. Peyrin, and P. Laugier, "Fast wave ultrasonic propagation in trabecular bone: Numerical study of the influence of porosity and structural anisotropy," J. Acoust. Soc. Am. **123**, 1694–1705 (2008).
[41]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).
[42]K. R. Marutyan, M. R. Holland, and J. G. Miller, "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL55–61 (2006).
[43]A. Q. Bauer, K. R. Marutyan, M. R. Holland, and J. G. Miller, "Negative dispersion in bone: The role of interference in measurements of the apparent phase velocity of two temporally overlapping signals," J. Acoust. Soc. Am. **123**, 2407–2414 (2008).
[44]C. C. Anderson, K. R. Marutyan, M. R. Holland, K. A. Wear, and J. G. Miller, "Interference between wave modes may contribute to the apparent negative dispersion observed in cancellous bone," J. Acoust. Soc. Am. **124**, 1781–1789 (2008).
[45]E. Camus, K. Mukuda, T. Otani, and Y. Watanabe, "Influence of fluid viscosity and temperature on the propagation of ultrasonic waves in bovine cancellous bone," Technical Report Nos. US97-85 and EA97-81, IEICE, 1998.
[46]G. Haïat, A. Lhémery, F. Renaud, F. Padilla, P. Laugier, and S. Naili, "Velocity dispersion in trabecular bone: Influence of multiple scattering and of absorption," J. Acoust. Soc. Am. **124**, 4047–4058 (2008).
[47]K. A. Wear, "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646 (2000).
[48]K. R. Marutyan, G. L. Bretthorst, and J. G. Miller, "Bayesian estimation of the underlying bone properties from mixed fast and slow mode ultrasonic signals," J. Acoust. Soc. Am. **121**, EL8–15 (2007).
[49]S. Chaffaï, V. Roberjot, F. Peyrin, G. Berger, and P. Laugier, "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results," J. Acoust. Soc. Am. **108**, 2403–2411 (2000).
[50]K. A. Wear, "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664 (1999).
[51]K. A. Wear, "Measurement of dependence of backscatter coefficient from cylinders on frequency and diameter using focused transducers—With applications in trabecular bone," J. Acoust. Soc. Am. **115**, 66–72 (2004).
[52]K. A. Wear, "Ultrasonic scattering from cancellous bone: A review," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1432–1441 (2008).
[53]K. A. Wear, "Mechanisms for attenuation in cancellous-bone-mimicking phantoms," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 2418–2425 (2008).
[54]K. A. Wear, F. Padilla, and P. Laugier, "Comparison of the Faran cylinder model and the weak scattering model for predicting the frequency dependence of backscatter from human cancellous femur in vitro," J. Acoust. Soc. Am. **124**, 1408–1410 (2008).
[55]K. A. Wear and G. R. Harris, "Frequency dependence of backscatter from thin, oblique, finite-length cylinders measured with a focused transducer—With applications in cancellous bone," J. Acoust. Soc. Am. **124**, 3309–3314 (2008).
[56]K. A. Wear, "Ultrasonic attenuation in human calcaneus from 0.2 to 1.7 MHz," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 602–608 (2001).
[57]A. Hosokawa, "Simulation of ultrasound propagation through bovine cancellous bone using elastic and Biot's finite-difference time-domain methods," J. Acoust. Soc. Am. **118**, 1782–1789 (2005).
[58]A. A. El-Sariti, J. A. Evans, and J. G. Truscott, "The temperature dependence of the speed of sound in bovine bone marrow at 750 kHz," Ultrasound Med. Biol. **32**, 985–989 (2006).
[59]D. R. Lide, *Handbook of Chemistry and Physics*, 74th ed. (CRC, Boca Raton, FL, 1994).
[60]R. C. Weast, *Handbook of Chemistry and Physics*, 45th ed. (The Chemical Rubber Co., Cleveland, OH, 1962).

Pinna-rim skin folds narrow the sonar beam in the lesser false vampire bat (*Megaderma spasma*)

Xiaobin Wang and Rolf Müller[a)]
School of Physics, Shandong University, Hongjia Lou 5, 250100 Jinan, China

(Received 17 October 2008; revised 28 September 2009; accepted 29 September 2009)

False vampire bats (genus *Megaderma*) employ active as well as passive biosonars. In the present work, the acoustic impact of a conspicuous feature of the animals' ear morphology, skin folds of the pinna rim linking the two pinnae at the midline, has been studied using a numerical approach. Automated methods have been devised to measure the largest width of the beam patterns irrespective of beam orientation. A total of six pinna shapes from three individuals have been studied. For all these shapes, it was found that the reception biosonar beams had approximately elliptic cross-sections with the largest beamwidth being on average almost twice as large as the beamwidth in the orthogonal direction. The directions of the largest beamwidths were scattered around the azimuthal dimension. Removal of the skin folds resulted in significant widening of the beams along their widest dimensions with an increase in beamwidth of 9.2° (a 30% change) on average. The strength and repeatability of this effect across individuals suggest the hypothesis that skin folds are functionally relevant to the animals' biosonar system. It may be a morphological adaptation to biosonar tasks that benefit from a narrow beam such as the detection of faint sounds or precise localization. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3257210]

PACS number(s): 43.80.Ka [MCH] Pages: 3311–3318

I. INTRODUCTION

Bats are one of the most diverse living groups of mammals (Nowak, 1991). Among the more than 900 species of microbats (Simmons, 2005), an extensive adaptive radiation in the use of biosonar has taken place (Jones and Teeling, 2006). The biosonar systems that have evolved in this radiation can make use of echo returns elicited by self-emitted pulses (active sonar) as well as signals from other sources (passive sonar). The varied shapes of the animals' noseleaves and outer ears (pinnae) that contribute to the distinctive appearance of many species are another conspicuous manifestation of the biodiversity in bats. As baffle shapes surrounding the sites of the emission and reception of the animals' biosonar signals, these morphological features are in a position to affect the shape of the biosonar beam patterns through diffraction of the outgoing and incoming sound waves. It may hence be hypothesized that at least part of the diversity seen in these shapes has a significant effect on the function of the respective biosonar systems. In these cases, the baffle shapes could be evolutionary adaptations that have shaped the biosonar beam in a way that results in an enhanced performance in the specific sonar tasks that the animals face. Structures with such an adaptive value could also have technological potential as inspirations for parsimonious beamforming devices that are based on physical baffle shape alone (Müller and Kuc, 2007).

Widenings of the pinna rim are a feature seen in several groups of bats such as the genus of the noctule bats (*Nyctalus*), some genera in the free-tailed bats (e.g., *Tadarida*, *Molossus*, and *Otomops*, see Fig. 1), as well as the family of the old world leaf nosed bats (*Megadermatidae*). These widenings can take the form of either enlargements of the pinna rim proper or skin folds attached to it. The latter is the case in the species belonging to the family *Megadermatidae*, where conspicuous skin folds on the inner margins of the two pinnae connect them together over the midline of the head (see Fig. 2). These connecting skins hence create a fairly large, approximately flat surface between the openings of the pinnae and hence add to the already comparatively large pinna size in *Megaderma*.

The prominence of these skin folds, their close proximity to the pinnae, and the fact that they are part of a recurring pattern across different taxonomic groups of bats (see Figs. 1 and 2) all suggest the hypothesis that these structures are adaptations to an acoustic function in the biosonar system of these animals. Since the assemblies of the pinnae and folds are irregular three-dimensional shapes, it is not possible to make definite prediction regarding their acoustic effects without results from physical or numerical experiments. However, since the skin folds enlarge the aperture of the ear, they have the potential to narrow the biosonar beam, although this potential may not be realized by their particular shapes.

Of the two species in the genus, *Megaderma*, the greater false vampire bat, *Megaderma lyra*, has been studied extensively in terms of its sensory ecology (e.g., Marimuthu and Neuweiler, 1987; Rajan and Marimuthu, 1999; Schmidt *et al.*, 2000; Leippert *et al.*, 2002; Ratcliffe *et al.*, 2005) and the psychophysics of the biosonar system (e.g., Wiegrebe and Schmidt, 1996; Preisler and Schmidt, 1998; Krumbholz and Schmidt, 2001; Weißenbacher and Wiegrebe, 2003; Schuchmann and Wiegrebe, 2006). It is known to be a carnivorous

[a)]Author to whom correspondence should be addressed. Present address: Department of Mechanical Engineering, Virginia Tech and Institute for Advanced Learning and Research, 150 Slayton Avenue, Danville, VA 24540. Electronic mail: rolf.mueller@vt.edu.

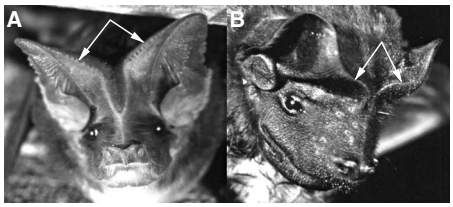


FIG. 1. Portrait photos showing different realizations of pinna rim folds in free-tailed bats: (A) large-eared free-tailed bat (*Otomops martiensseni*) and (B) black mastiff bat (*Molossus ater*). Photo: M. Brock Fenton, Department of Biology, University of Western Ontario; reproduced with permission.

predator and has been shown to use a combination of active and passive sonars (Rajan and Marimuthu, 1999; Schmidt *et al.*, 2000). Following the notion that a larger aperture is capable of forming a narrower beam, it has been hypothesized that the large pinnae of the animals are adaptations to detect faint, low-frequency sounds generated by moving prey (Obrist *et al.*, 1993). The other species in the genus, the lesser false vampire bat (*Megaderma spasma*), has been studied much less, but like the greater false vampire bat it is known to be a predator, although it appears to prefer grasshoppers and moths over vertebrate prey (Nowak, 1991).

In the present work, pinna samples from specimens of the lesser false vampire bat have been used to address the question whether the unique and conspicuous skin folds of the pinna that are characteristic for the family *Megadermatidae* (Koopman, 1994) have an effect on the receiver directivity pattern of the ears. Although much less is known about the sensory ecology and biosonar system of this species, the similarity in morphology should allow it to integrate the results obtained here with the large body of knowledge accumulated on the greater false vampire bat. In the research reported here, the pinna directivities have been predicted from digital models of the pinna shape with and without skin folds using numerical methods. This approach facilitates well-defined and reproducible manipulations of the pinna shapes and hence ensures that differences seen in the directivities are only due to the presence or absence of the skin folds and are not caused by accidental changes in other aspects of pinna shape or other experimental details. In carrying out this work, special attention has been paid to the problem of measuring beamwidth in beam patterns of arbitrary orientation and to identifying effects that are strong and consistent across individuals.

FIG. 2. Portrait photo of one of the lesser false vampire bat (*Megaderma spasma*) specimens studied (individual "A"). The arrows indicate the skin folds of the pinnae.

II. METHODS

A. Shape models

Ear shape samples were taken from three adult specimens, two males and one female, collected in Ninh Binh Province, Vietnam (one specimen) and in Kandal Province, Cambodia (two specimens). Ear shape in life was photographically documented for each animal prior to collection of the specimen. The two pinnae of each specimen were surgically removed from the carcass and from each other along the midline of the skin folds. Each of the six individual pinna shapes was remounted to approximate its configuration *in vivo* and scanned using x-ray microtomography (Skyscan 1172 desktop x-ray microtomograph). X-ray shadow images of each pinna were taken from directions spanning 180° with 0.9° angular resolution and had a gray-level depth of 16 bits. A cone beam reconstruction algorithm was used to obtain a three-dimensional digital tomographic reconstruction of the pinna shapes with an isotropic voxel resolution of ~36 μm and a gray-level depth of 8 bits. After pre-filtering with an isotropic Gaussian smoothing kernel, the voxels of the tomographic reconstruction were automatically classified as either "tissue" or "air" using a manually adjusted gray-level threshold. The voxels representing tissue in the skin flaps were then hand-labeled to permit the creation of a digital model of the unaltered natural pinna shape with the skin fold as well as a model of a modified with the fold removed. Thus, a total of 12 ear shapes, 6 natural shapes including the skin folds and the corresponding 6 modified shapes with the folds removed, were created (see Fig. 3). Since the ear canal of the specimens was not preserved in the pinna samples, an artificial ear canal was fitted to the original opening of the ear canal in the concha. The method used for this purpose has been described previously (Müller *et al.*, 2006). It finds the direction of the ear canal as the normal vector to the plane spanned by the two eigenvectors of the spatial distribution of the voxels marking the rim of the ear canal opening associated with the largest eigenvalues. Having found the direction for the artificial ear canal, the walls of the canal are created by adding additional tissue voxels to the voxels of the rim of the ear canal opening in that direction. The acoustic behavior of the artificial ear canals was also simulated in isolation (without a pinna being present) to test the assumption that the ear canal opening without pinna would act as an omni-directional sound source since its diameter was significantly shorter than even the smallest wavelengths employed.

B. Numerical methods

A finite-element method was used to predict the frequency-domain acoustic near-field elicited by an acoustic point source placed near the proximal end of the artificial ear

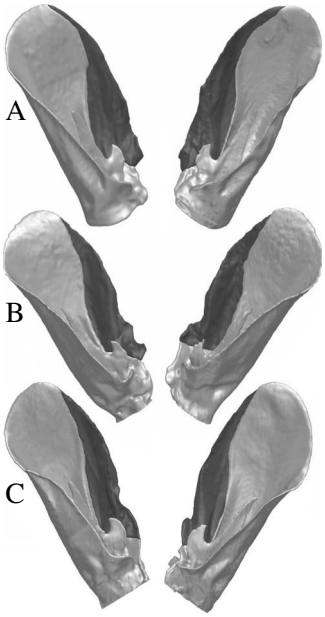


FIG. 3. Renderings of all pinna shapes from the three specimens (labeled "A," "B," and "C"). The skin flap is marked in dark gray. The average total height of the pinna shapes as shown is 4 cm.

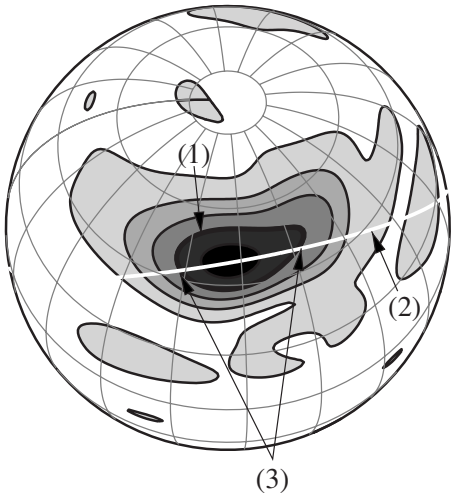


FIG. 4. Example illustrating the method used to measure beamwidth: (1) −3 dB contour of the directivity gain, (2) arc from the selected great circle through the gain maximum, and (3) angular distance between the intersection points of the great circle and the −3 dB gain contour used to measure beamwidth. The example beam pattern shown here is the directivity of the natural shape (with skin fold) of the left pinna of individual C at 66 kHz; gain values of the other contours shown are 0.3, 0.5, 0.7, and 0.9 of the normalized directivity; coordinate grid lines are spaced 20° apart in azimuth and in elevation.

canal. The complex near-field values on the boundary of the finite-element computational domain were then used to project the field outwards and estimate the far-field directivity. By reciprocity (Pierce, 1981), this far-field directivity for the pinna operated as an emitter baffle is identical to the receiving directivity of the baffle. A detailed description of the numerical methods used can be found in Zhuang and Müller, 2007. An isotropic cubic finite-element mesh was derived from the voxels of the binary digital shape representation. Each finite element spanned five voxels in each direction giving it an edge length of ~0.18 mm. Associated with each element was a set of eight linear shape functions (one for each corner of the element) used to approximate the solution to the Helmholtz differential equation. Boundary conditions on the pinna surface were chosen to represent perfect reflections. On the outer surface of the finite-element computational domain, reflection-free sound propagation was modeled by a layer of three-dimensional mapped wave-envelope infinite elements (Astley and Macaulay, 1998) with Jacobi polynomials of order 3 and with parameter values of $\alpha=2$ and $\beta=0$ serving as basis functions (Dreyer and von Estorff, 2003). The linear problem of finding the complex wavefield amplitudes was solved iteratively using a bi-conjugate gradient stabilized method (Barrett *et al.*, 1994; van der Vorst, 2003) with a successive over-relaxation (SOR) preconditioner (relaxation factor $\omega=1.2$). The routines used for solving the linear system were taken from the implementations in the portable, extensible toolkit for scientific computation (PETSC) software library (Balay *et al.*, 1997). Numerical modeling of diffraction-free outward propagation of the wavefield from the surface of the finite-element computational domain was accomplished using a Kirchhoff integral formulation (Jackson, 1999; Ramahi, 1997). The projection was used to obtain the field values at a set of points spaced 1° apart in azimuth and elevation on the surface of a sphere in the far-field (radius 10 m). For each frequency, the magnitude of the projected field values at these points was normalized (divided by its maximum value over all directions) to yield estimates of the real-valued, normalized directivity. For each of the natural and modified pinna shapes (12 shapes in total), directivity estimates were computed for 25 frequencies ranging from 40 to 88 kHz in steps of 2 kHz, i.e., a total of 300 directivity estimates were computed. The studied frequency range covers slightly more than the lower two-thirds of the spectral band known to be used by the species at least in part of its range (40–110 kHz, H. Raghuram, personal communication). All frequency values tested were converted into wave numbers assuming a sound speed of 343 m/s for the simulations.

C. Beamwidth measurement

The directivity gain is a scalar function of position on the surface of a sphere. In the present work, the center of the sphere was positioned at the location of the sound source and the tip of the pinna oriented toward the north pole. For the purpose of the present work, the mainlobe of the directivity pattern is defined as the peak in this function that includes its global maximum value. Since the position, shape, and orientation of the mainlobe can all vary considerably with the individual pinna as well as with frequency, an objective method is needed to determine the greatest beamwidth irrespective of where the beam is positioned and how it is oriented. The method should also be robust against shape irregularities, at least as long as they are small compared to the overall extent of the mainlobe. In the present work, the following *ad-hoc* method was used to achieve this goal (see Fig. 4). The beamwidth was measured along an orthodrome (i.e., an arc segment of a great circle) that intersects the direction of the maximum directivity gain and runs along the

principal extent of the mainlobe. The width of the lobe was then measured as the angular distance between the two points where the arc cuts the directivity-gain contour defined by the gain threshold at which the width was to be determined. All beamwidth data presented here were obtained with a threshold value of −3 dB below the maximum. The plane that cuts the sphere to generate the great circle was defined by two vectors: the vector from the origin of the sphere to the location of the directivity-gain maximum and a vector that describes the principal extent of the locations of above-threshold directivity gains. To obtain the latter vector, the entire cloud of sample points on the sphere surface with above-threshold directivity gains was transformed into Cartesian coordinate space. The eigenvector associated with the largest eigenvalue of this point cloud was then used as the second vector to define the cutting plane that yielded the great circle. Beamwidth was measured by the gain-threshold contour that was closest to the global gain maximum. This was accomplished by finding the group of neighboring directions with above-threshold gain values using a union-find algorithm (Sedgewick, 1998). Limiting the measurement to a contiguous region of directions with above-threshold gain values excluded sidelobes from the measured beamwidth.

III. RESULTS

The directivity pattern estimates obtained for the generated artificial ear canals alone (pinna removed) did not exhibit any significant directionality within the studied frequency range (see Fig. 5). The shorter wavelengths (higher frequencies) tested manifested themselves only in a greater small-scale variability in the estimated directivity gains (reflected in a rougher beam surface in Fig. 5), but did not lead to a noticeable overall narrowing of the beam. In contrast to this, all beam patterns estimated for the pinna shapes (excited through the artificial ear canals) were characterized by a strong directionality manifest in the presence of a single prominent mainlobe (see Fig. 6). Despite noticeable individual variations, the width of the mainlobe tended to decrease with increasing frequency for all studied shapes (see Fig. 8). For the width of the mainlobe averaged over all six natural ear shapes, the maximum value of ~41° was observed at the lowest studied frequency (40 kHz) and the minimum value of ~21° at the highest studied frequency (88 kHz). The beamwidth estimates obtained for the modified shapes (see below) showed the same overall dependence on frequency. In general, the beam patterns were characterized by few and relatively weak sidelobes. 33 out of the 300 beam patterns analyzed contained sidelobes that exceeded an amplitude threshold of −3 dB relative to the maximum gain and were separated from it by a trough of at least that depth. Among these beam patterns, 28 contained only a single sidelobe of above-threshold strength. It was observed qualitatively that sidelobes became more prominent at the higher frequencies studied, but they were almost always found to be much weaker overall than the mainlobe. Both trends, narrowing of the beam and rising prominence of the sidelobes, were seen in the natural shapes (with skin folds present) as well as in the modified shapes with the skin folds removed.

FIG. 5. Examples of three-dimensional directivity-gain isosurfaces for the directivity of an artificial ear canal alone at frequencies close to the lower end, center, and upper end of the studied frequency band.

A qualitative trend was also discernible in the shape of the mainlobe as a function of frequency, although there were many exceptions. Near the bottom of the studied frequency band, the beams tended to be uniformly broad in all directions (see Figs. 6 and 7). Likewise, beams tended to be uniformly narrow in all directions near the upper limit of the studied frequency band (see Fig. 7). In between, the isogain contour lines tended to have elliptic shapes indicating beams that were noticeably wider in one direction (referred to as the "major direction" below) than in the orthogonal direction (referred to as the "minor direction" below, see Fig. 7). The ratio between the beamwidths measured along the major and minor directions was found to be greater than unity in all but 2 of the 300 beam patterns analyzed (in these two exceptional cases, the measured beamwidths along the two directions were identical). On average (over all analyzed beam patterns), the ratio between the beamwidth along the major and the minor direction was 1.9 (±0.6 standard deviation); i.e., the beams were almost twice as wide as they were tall. Although variable, the major direction was found to be closely aligned with the azimuthal direction (latitude); the average angle between the two directions was 3° (±10° standard deviation).

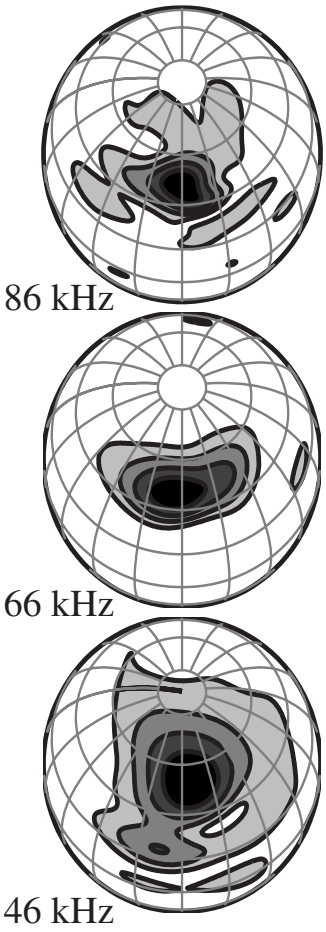


FIG. 6. Examples of pinna directivity patterns for frequencies close to the lower end, center, and upper end of the studied frequency band. The examples shown are directivities of the natural shape (with skin fold) of the left pinna of individual C. The gain contours are spaced linearly between values of 0.2 and 0.8 in steps of 0.2; the lines of the coordinate grid are spaced 20° apart in azimuth and in elevation.

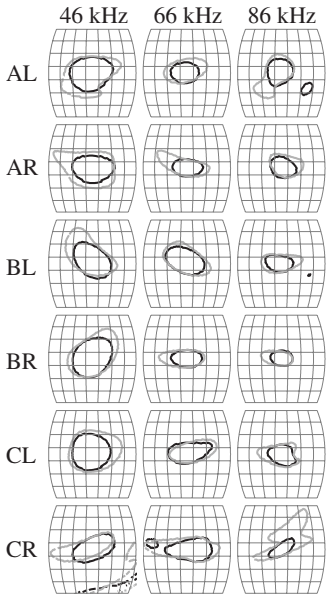


FIG. 7. Comparison of −3 dB reception directivity-gain isocontours obtained for pinna shapes with and without skin folds. The pinna shapes (rows) are labeled with a two-letter code in which the first letter denotes the individual (A, B, and C, see Fig. 3) and the second letter indicates whether the shape is a left ("L") or right ("R") pinna. Three frequencies (columns: 46, 66, and 86 kHz) near the bottom, center, and top of the studied band are shown. Gain contours for the natural ear shapes are rendered in black, while gain contours obtained for shapes with skin folds removed are rendered in gray. The lines of the spherical coordinate grid are spaced 10° apart in azimuth and in elevation.

Removal of the skin folds from the pinnae resulted in a consistent additional widening of the beam along the major direction that characterized the beam patterns of the natural shapes (see Fig. 7). In contrast to this, the beamwidths measured along the minor directions did not undergo any significant changes. The same was true for the orientation of the major direction, which only changed less than 1° on average. This effect was seen in all six studied pinnae and over the entire frequency range analyzed (see Fig. 8). Only in 6 out of the 150 combinations of individual pinna and frequencies studied was the estimated maximum beamwidth with skin folds larger than without skin folds. Averaged over all individuals and frequencies, the difference in beamwidth was 9.2° (±6.5° standard deviation). The relative increase in beamwidth due to removal of the folds was found to be ~31% on average (±8% standard deviation). The largest increase in beamwidth found upon removal of the fold was 31°; the largest relative increase was ~39%. For the individual frequencies studied, the mean absolute beamwidth difference between pinnae with and without skin folds averaged over all six pinnae ranged between 2.7° and 13.7° with standard deviations between 3° and 9.5° (see Fig. 9, top). The changes in relative beamwidth (reference: beamwidth in the presence skin folds) averaged over the individual pinnae ranged between 11.5% and 43.1% with standard deviations between 13.3% and 37.8% (see Fig. 9, bottom). The probability of falsely rejecting the null hypothesis of beamwidth differences being equal or less than zero was estimated to be less than 0.05 for all but 2 frequencies (42 and 80 kHz) out of the 25 examined frequencies (one-sample right-tailed

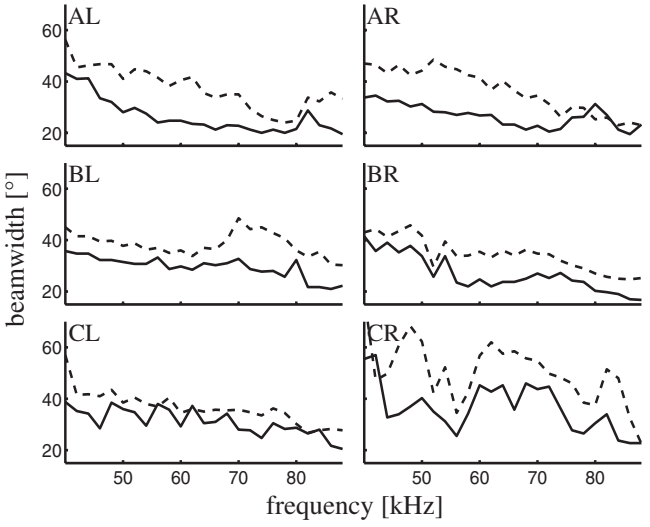


FIG. 8. Comparison of −3 dB beamwidths obtained for pinna shapes with and without skin folds as a function of frequency. Results for each individual pinna are presented in a separate graph. Pinnae are labeled using the same notation as in Fig. 7. The beamwidth of the natural pinna shapes (with skin folds) is represented by solid lines, and the beamwidth of modified shapes with the skin folds removed by dashed lines.

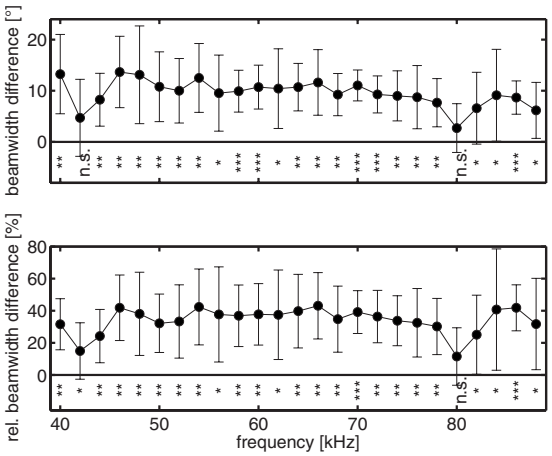


FIG. 9. Differences in −3 dB beamwidth across frequency averaged over individual pinnae (individual data shown in Fig. 8): Top: absolute differences in beamwidth between pinnae without and with skin folds. Bottom: differences in beamwidth relative to the beamwidth with skin folds. Filled circles represent averages (over all six studied pinnae) and error bars represent the standard deviations in the sample. Below each studied frequency, the significance level at which the null hypothesis (difference in beamwidth less or equal to zero) is rejected is indicated (∗: $p<0.05$, ∗∗: $p<0.01$, ∗∗∗: $p<0.001$, and n.s.: not significant).

T-test with five degrees of freedom, see Fig. 9). The probability of error was less than 0.01 for more than two-thirds of the studied frequencies (18 out of 25). When performing the hypothesis test over all pinnae and frequencies, the probability of error in rejecting the null hypothesis was less than 0.001 (149 degrees of freedom).

IV. DISCUSSION

The pinna shapes that have been studied here are not readily described in terms of simple geometries, and hence numerical models are the only feasible approach for predicting their acoustic behavior. Nevertheless, the major properties of the estimated beam patterns can be evaluated qualitatively against expectations derived from basic physical principles. In all these cases, the experimental results were found to be in accordance with expectations.

An example is the acoustic behavior of the artificial ear canals when tested without pinnae. For these shapes, the absence of any significant directionality is to be expected considering the small diameter of the ear canal (less than 1 mm) relative to all used wavelengths (3.9 mm and larger). Besides verifying that the highly directional beam patterns are entirely due to the pinna, this finding also provides further confirmation that the numerical results capture the physical behavior of these structures accurately.

The same is true for the observed decrease in beamwidth with frequency, where the general trend seen in the data conforms qualitatively with the approximately hyperbolic relationship between aperture width and beamwidth that is expected from, e.g., the Fourier-transform relationship between the two quantities that is obtained from the Fraunhofer approximation for planar apertures (Goodman, 2004).

Most importantly, the difference between the beamwidths predicted for shapes with and without the skin folds is likewise consistent with the expectation that a larger aperture is capable of producing a narrower beam. Noteworthy about the present results is the presence of the fold's effect on beamwidth in all studied pinnae as well as the magnitude of the effect. There were considerable variations between the beam patterns of the individual pinna samples (see Fig. 7). Such variations have been observed previously in bats (Aytekin *et al.*, 2004) as well as in other mammals (Xu and Middlebrooks, 2000). In particular, differences between the directivities of the left and right ears of the same individual have been reported previously for bats (Aytekin *et al.*, 2004). Sources of this variability are the natural variability in the pinna shapes, errors made in recreating life-like shapes when remounting the samples *post mortem*, and errors in the numerical predictions. Despite all the individual variabilities seen among the samples, the effect was clearly expressed in each of the six studied ears and hence in all three studied specimens (see Fig. 8). The magnitude of the effect of the skin folds on beamwidth is noteworthy because it reduced the beamwidth by about 30% on average and the observed beamwidths fell into a range where the values are likely neither too small nor too large for such a change in matter. While the *capability* of a larger aperture to produce a narrower beam is in accordance with expectations, the observed strength of the effect is neither a necessary nor a trivial consequence of a large aperture. A poorly designed three-dimensional baffle could fail to deliver any significant narrowing of the beam. The strong effect of the skin folds hence suggests the hypothesis that their shape and placement might be an outcome of evolutionary optimization for this function.

Beyond what has been observed in the numerical experiments reported here, it is conceivable that the effect of the skin folds on beamwidth is even larger *in vivo*. A likely reason why the numerical experiments reported here could have underestimated the skin folds' effect on beamwidth lies in the artificial separation of the skin folds of the left and right pinnae along the midline in the experiments. This separation amounts to a one-sided truncation of the aperture that the intact animal has. In life, the skin folds of the pinnae may have acoustic effects across the midline and hence the skin fold of the contralateral pinna could contribute to the size of the aperture that determines the beamwidth produced by the ipsilateral pinna. If the contralateral portion of the aperture had a significant effect on beamwidth, this effect would have been obliterated by the separation of the pinnae. Another factor that may influence the effect of the folds on beamwidth is the relative orientation of pinna and skin fold. While the arrangements of the pinnae and skin folds in the samples replicated configurations seen in life and also captured some variability in the geometry (see Fig. 3), it is conceivable that the animals can use muscular actuation to achieve configurations that fall outside of the range studied here. Such configurations could either enhance or reduce the narrowing of the beam caused by the skin folds. Because of these unexplored possibilities, the effect of the skin flaps demonstrated here is best regarded as a lower bound on the impact the skin folds could have *in vivo*. But even without relying on such additional factors, the magnitude of the observed effect and its presence across all samples suggest the hypothesis that the influence of the skin folds on the beamwidth is function-

ally relevant and that the narrowing of the beam through the skin folds contributes to the specific capabilities of the animals' biosonar system.

The pinna of *Megaderma* is significantly taller than it is wide (see Fig. 3). As is expected from such an aperture geometry, it produces a beam that is significantly (almost two times for the natural pinna shapes) narrower in elevation than in azimuth. Comparing these—monaural—beamwidths along the dimensions of azimuth and elevation does not, however, provide an adequate assessment of the beam patterns that can be formed at the level of the entire biosonar system. At the system's level, the animals could combine the signals from both ears to achieve a narrower beam in azimuth. Since this strategy would be limited to one angular dimension, the oval shape of the pinna could be seen as consistent with the hypothesis that the bat relies primarily on one mechanism of beamforming (a physical baffle) in one direction (elevation) and another mechanism (binaural array processing) in the other direction (azimuth). This would also conform to the general notion about the main mechanisms of auditory localization in humans (Blauert, 1996). The presence of the skin folds and their monaural beam-narrowing effect on *Megaderma* suggest the hypothesis that a narrow monaural beam with few sidelobes both in azimuth and in elevation is important to the specific uses of biosonar in these species. These monaural beam patterns could play a role in conjunction with binaural processing or in enhancing performance in monaural sensing tasks. Uses for narrow beam patterns could be found in the context of the active sonar as well as in the passive sonar that is known to be employed by the false vampire bats. Narrow beams concentrate most of the ear's sensitivity within a narrow set of directions. This aids the detection of low-amplitude signals, which could either originate from weak sources or have been weakened through propagation losses over a large distance. Dealing with faint prey-generated sounds as has been suggested previously (Obrist *et al.*, 1993) would fall into this category. In addition, narrow beams support precise localization through their steep flanks on which small changes in the source bearing yield large changes in the received signal amplitude (Müller *et al.*, 2008). The strong spatial filtering in direction due to the narrow beams restricts the locations and hence the number of scatters that can contribute significantly to the received echo. Together with the finding that false vampire bats do not automatically use echo suppression ("precedence effect") (Schuchmann and Wiegrebe, 2006), this could be an indication that these animals' biosonar tends to survey the environment with a long narrow window of attention. The downside of such a narrow spatial filter is that it does not allow the animals to be alerted by or otherwise evaluate faint sounds from any of the directions other than the beam aim. To make up for this, the bats would have to scan its beam patterns through movements of either the pinnae or the entire head.

Future work is needed to investigate how these monaural beam patterns are integrated with other aspects of the biosonar system such as binaural processing and behavioral components such as scanning behaviors or muscular actuation of the pinnae and their skin folds. Most importantly, the hypothesis that the observed features are of functional relevance to the animal needs to be tested. Experimental evaluations of the possible impact that manipulations of the skin folds may have on certain aspects of the animals' performance in biosonar tasks would be an appropriate way to address this. However, design of such experiments would have to ensure that the experimental subjects cannot compensate for a deterioration in certain aspects of system performance by solving the same task in a different way.

ACKNOWLEDGMENTS

This research was supported by the European Commission (CILIA project), the National Natural Science Foundation of China (Project No. 10774092), and Shandong University. The bat specimen from Vietnam was collected by the Institute of Ecology and Biological Resources, Vietnamese Academy of Science and Technology. H. Raghuram provided valuable comments.

Astley, R. J., and Macaulay, G. J. (**1998**). "Three-dimensional wave-envelope elements of variable order for acoustic radiation and scattering. Part I. Formulation in the frequency domain," J. Acoust. Soc. Am. **103**, 49–63.

Aytekin, M., Grassi, E., Sahota, M., and Moss, C. F. (**2004**). "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," J. Acoust. Soc. Am. **116**, 3594–3605.

Balay, S., Gropp, W. D., McInnes, L. C., and Smith, B. F. (**1997**). "Efficient management of parallelism in object oriented numerical software libraries," in *Modern Software Tools in Scientific Computing*, edited by E. Arge, A. M. Bruaset, and H. P. Langtangen (Birkhäuser, Boston, MA), pp. 163–202.

Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and der Vorst, H. V. (**1994**). *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA).

Blauert, J. (**1996**). *Spatial Hearing-Revised Edition: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Dreyer, D., and von Estorff, O. (**2003**). "Improved conditioning of infinite elements for exterior acoustics," Int. J. Numer. Methods Eng. **58**, 933–953.

Goodman, J. W. (**2004**). *Introduction to Fourier Optics*, 3rd ed. (Roberts and Company, Greenwood Village, CO).

Jackson, D. E. (**1999**). *Classical Electrodynamics*, 3rd ed. (Wiley, New York).

Jones, G., and Teeling, E. C. (**2006**). "The evolution of echolocation in bats," Trends Ecol. Evol. **21**, 149–156.

Koopman, K. F. (**1994**). *Chiroptera: Systematics* (Walter de Gruyter, Berlin, Germany).

Krumbholz, K., and Schmidt, S. (**2001**). "Evidence for an analytic perception of multiharmonic sounds in the bat, *Megaderma lyra*, and its possible role for echo spectral analysis," J. Acoust. Soc. Am. **109**, 1705–1716.

Leippert, D., Frank, E., Gabriel, P., Kutter, S., Scheidemann, K., von Stillfried, N., and Weller, F. (**2002**). "Prey-correlated spectral changes in echolocation sounds of the Indian false vampire *Megaderma lyra*," Ethology **108**, 139–156.

Marimuthu, G., and Neuweiler, G. (**1987**). "The use of acoustical cues for prey detection by the Indian false vampire bat, *Megaderma lyra*," J. Comp. Physiol. [A] **160**, 509–515.

Müller, R., and Kuc, R. (**2007**). "Biosonar-inspired technology: Goals, challenges, and insights," Bioinspir. Biomim. **2**, S146–S161.

Müller, R., Lu, H., and Buck, J. R. (**2008**). "Sound-diffracting flap in the ear of a bat generates spatial information," Phys. Rev. Lett. **100**, 108701.

Müller, R., Lu, H., Zhang, S., and Peremans, H. (**2006**). "A helical biosonar scanning pattern in the Chinese noctule, *Nyctalus plancyi*," J. Acoust. Soc. Am. **119**, 4083–4092.

Nowak, R. M. (**1991**). *Walker's Mammals of the World*, 5th ed. (Johns Hopkins University Press, Baltimore, MD).

Obrist, M., Fenton, M., Eger, J., and Schlegel, P. (**1993**). "What ears do for

bats: A comparative study of pinna sound pressure transformation in chiroptera," J. Exp. Biol. **180**, 119–152.
Pierce, A. D. (**1981**). *Acoustics* (McGraw-Hill, New York).
Preisler, A., and Schmidt, S. (**1998**). "Spontaneous classification of complex tones at high and ultrasonic frequencies in the bat, *Megaderma lyra*," J. Acoust. Soc. Am. **103**, 2595–2607.
Rajan, K., and Marimuthu, G. (**1999**). "Localization of prey by the Indian false vampire bat, *Megaderma lyra*," Mammalia **63**, 149–158.
Ramahi, O. M. (**1997**). "Near- and far-field calculations in FDTD simulations using Kirchhoff surface integral representation," IEEE Trans. Antennas Propag. **45**, 753–759.
Ratcliffe, J., Raghuram, H., Marimuthu, G., Fullard, J., and Fenton, M. (**2005**). "Hunting in unfamiliar space: Echolocation in the Indian false vampire bat, *Megaderma lyra*," Behav. Ecol. Sociobiol. **58**, 157–164.
Schmidt, S., Hanke, S., and Pillat, J. (**2000**). "The role of echolocation in hunting terrestrial prey-new evidence for an underestimated strategy in the gleaning bat, *Megaderma lyra*," J. Comp. Physiol. [A] **186**, 975–988.
Schuchmann, M. H., and Wiegrebe, L. (**2006**). "The absence of spatial echo suppression in the echolocating bats *Megaderma lyra* and *Phyllostomus discolor*," J. Exp. Biol. **209**, 152–157.
Sedgewick, R. (**1998**). *Algorithms in C* (Addison-Wesley Longman, Reading, MA).
Simmons, N. (**2005**). "Order chiroptera," in *Mammal Species of the World: A Taxonomic and Geographic Reference*, 3rd ed. (Johns Hopkins University Press, Baltimore, MD), Vol. **1**, pp. 312–529.
van der Vorst, H. A. (**2003**). *Iterative Krylov Methods for Large Linear Systems*, Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, Cambridge).
Weißenbacher, P., and Wiegrebe, L. (**2003**). "Classification of virtual objects in the echolocating bat, *Megaderma lyra*," Behav. Neurosci. **117**, 833–839.
Wiegrebe, L., and Schmidt, S. (**1996**). "Temporal integration in the echolocating bat, *Megaderma lyra*," Hear. Res. **102**, 35–42.
Xu, L., and Middlebrooks, J. C. (**2000**). "Individual differences in external-ear transfer functions of cats," J. Acoust. Soc. Am. **107**, 1451–1459.
Zhuang, Q., and Müller, R. (**2007**). "Numerical study of the effect of the noseleaf on biosonar beamforming in a horseshoe bat," Phys. Rev. E **76**, 051902.

Bowhead whale (*Balaena mysticetus*) songs in the Chukchi Sea between October 2007 and May 2008

Julien Delarue,[a] Marjo Laurinolli, and Bruce Martin
JASCO Applied Sciences, Suite 432, 1496 Lower Water Street, Halifax B3J 1R9, Canada

(Received 16 March 2009; revised 2 September 2009; accepted 25 September 2009)

This paper reports on the acoustic detection of bowhead whale (*Balaena mysticetus*) songs from the Bering-Chukchi-Beaufort stock, including the first recordings of songs in the fall and early winter. Bowhead whale songs were detected almost continuously in the Chukchi Sea between October 30, 2007 and January 1, 2008 and twice from April 16 to May 5, 2008 during a long-term deployment of five acoustic recorders moored off Point Lay and Wainwright, AK, between October 21, 2007 and August 3, 2008. Two complex and four simple songs were detected. The complex songs consisted of highly stereotyped sequences of four units. The simple songs were primarily made of sequences of two to three moan types whose repetition patterns were constant over short periods but more variable over time. Multiple song types were recorded simultaneously and there is evidence of synchronized song variation over time. The implications of the spatiotemporal distribution of song detection with respect to the migratory and mating behavior of western Arctic bowheads are discussed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3257201]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 3319–3328

I. INTRODUCTION

The range of bowhead whales (*Balaena mysticetus*) from the Bering-Chukchi-Beaufort (BCB) stock (International Whaling Commission, 1992) extends from the Bering Sea north into the Chukchi Sea and east to the Canadian Beaufort Sea (Moore and Reeves, 1993). Bowhead whales migrate northward from the Bering Sea into the Chukchi Sea in April and May. Most calving occurs in the Chukchi Sea (Koski *et al.*, 1993) as individuals follow leads in shorefast ice and stretches of open water called polynyas (Mel'nikov *et al.*, 1997) on their way to summer feeding grounds in the Canadian Beaufort Sea (Moore and Reeves, 1993). Bowhead whales' summer presence in the Chukchi Sea is thought to be scarce (Dahlheim *et al.*, 1980; Miller *et al.*, 1986) even though incidental sightings suggest that they may occur in the northeastern part of the area in summer more commonly than previously believed (Moore, 1992). The fall migration through the Beaufort Sea starts in late August and September. Most bowhead whales then pass through the Chukchi Sea from mid-September to late October (Moore and Reeves, 1993), even though their migration path in the area is not well understood (Mel'nikov *et al.*, 1997; Mate *et al.*, 2000). The majority of the population has typically entered the Bering Sea by the end of November (Moore and Reeves, 1993) where they spend the winter and breed.

The acoustic behavior of bowhead whales from the BCB stock is the best described of all five stocks (Ljungblad *et al.*, 1980, 1982; Clark and Johnson, 1984; Cummings and Holliday, 1987). Würsig and Clark (1993) summarized findings from over 5500 h of ice-based and sonobuoy recordings and divided sounds associated with bowheads into three categories: (1) blow and crunch sounds, gunshots, and slaps; (2) simple, low-frequency (LF), frequency-modulated (FM) calls and complex, high-frequency (HF), pulsive calls ordered randomly; and (3) songs, defined as patterned sequences of sounds repeated over time.

Of interest to this paper is the singing behavior of bowheads. Songs recorded in the spring off Alaska were first reported and analyzed in the early 1980s (Cummings *et al.*, 1983; Clark and Johnson, 1984; Cummings and Holliday, 1987). Recently, Stafford *et al.* (2008) provided the first detailed description of elaborate songs produced by bowheads in springtime. Songs are typically composed of one to three themes, themselves made up of repeated phrases containing one to five units (Würsig and Clark, 1993; Stafford *et al.*, 2008). Song units are usually longer and more broad band than simple and complex calls although simple songs characterized by a repetition of simple moans have been described (Ljungblad *et al.*, 1982). A single song lasts about 1 min but song bouts, i.e., continuous repetitions of identical songs, can last several hours (Würsig and Clark, 1993; Stafford *et al.*, 2008). Cummings and Holliday (1987) initially reported that a single singer sang at a given time, but Stafford *et al.* (2008) recently showed that multiple singers can sing simultaneously, as is the case in humpback whales (*Megaptera novaeangliae*).

Singing has so far only been recorded in spring, although Blackwell *et al.* (2007) reported sequences of identical calls repeated regularly and lasting between 14 s and 50 min in September. These sequences were not considered to be songs but exhibited a basic level of organization similar to songs. Since the peak mating season is believed to occur in late winter-early spring (Nerini *et al.*, 1984; Reese *et al.*, 2001), songs are thought to be involved in the breeding behavior of bowhead whales, and more particularly male breeding displays (Würsig and Clark, 1993; Stafford *et al.*, 2008). This is supported by similar evidence in related species such as humpback (Darling *et al.*, 2006; Smith *et al.*,

[a]Author to whom correspondence should be addressed. Electronic mail: julien.delarue@jasco.com

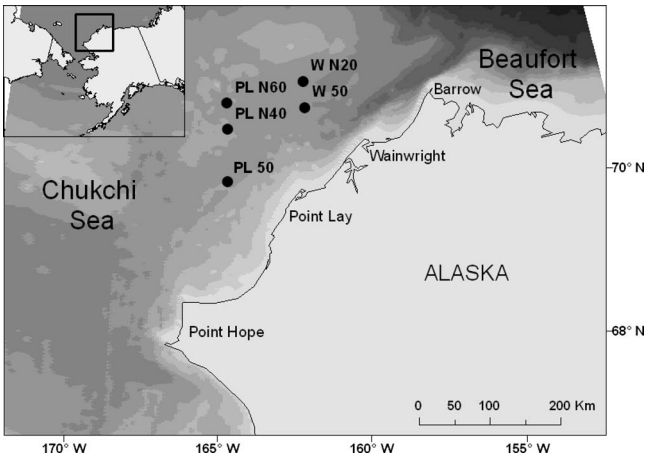


FIG. 1. Map of the Chukchi Sea and adjacent areas showing the recorders' location. Bathymetry is shown in shades of gray.

2008) and fin whales (*Baleanoptera physalus*) (Watkins *et al.*, 1987; Croll *et al.*, 2002). Bowhead songs previously recorded off northwestern Alaska in the spring are considered to be from the end of the mating season when males are potentially less sexually active, which may result in degraded or incomplete songs, as observed late in the humpback breeding season (Würsig and Clark, 1993).

Würsig and Clark (1993) stressed that virtually nothing is known about bowheads' life in late fall and winter and that recordings over these periods would likely contain an exceptionally rich assortment of sounds, including singing bouts. Here we report on recordings of bowhead songs obtained in the Chukchi Sea including the first recordings in late fall-early winter. These songs are described and compared to previous descriptions for this and other populations. The implications of these acoustic detections at a time when most of the population is presumed to be in the Bering Sea (Moore and Reeves, 1993) are discussed.

II. METHODS

A. Recording protocol

Acoustic data were obtained during a long-term deployment of five autonomous underwater recorders for acoustic listening (AURAL) (Multi-Electronique, Inc., Rimouski, Quebec, Canada) recorders deployed on the seabed in the Chukchi Sea (Fig. 1) to monitor marine mammal vocal activity during the Arctic winter. The recorders incorporated an HT196 hydrophone (1–10 000 Hz; −164 dB re 1 μPa), hard drive, and batteries packaged in pressurized fiber glass tubes mounted on a stainless steel frame. They were deployed between October 21 and 25, 2007 off Wainwright and Point Lay, AK, at a depth of about 45 m and recovered between July 20 and August 3, 2008 (Table I). Acoustic data were recorded during the first 48 min of every 4-h period and sampled at a rate of 16 384 Hz with 16-bit resolution and an AURAL recorder gain setting of +22 dB.

B. Data analysis

The recording schedule generated 727 Gbytes of data. To speed up the data analysis process, all 48-min sound files were scanned using a purpose-built computer program to search for energy peaks in specific frequency bands. These bands were defined to locate any marine mammal calls and were not specific to bowheads. Four frequency bands contained most of the bowhead detections: 90–220, 180–440, 360–880, and 720–1760 Hz. Detection was performed using a split window normalizer applied to each of the bands, with a notch of 30 time cells, a window of 80 time cells, and a threshold of 8. All the energy values of the normalized spectrogram above this threshold were considered as part of an acoustic event that may have been a marine mammal call. The fast Fourier transform (FFT) used was 4096 real points in a 8192-point FFT, with 2048-point overlap.

Sound files containing acoustic detections were then visualized as spectrograms (Hanning window, 8192-point FFT, 4096 real points, 50% overlap) and checked for the presence of bowhead sounds. Due to the large number of files containing acoustic events, we first checked a random sample representing about 15% of all files among those showing detection peaks. When a bowhead call was identified, we checked adjacent files in order to define the length of the calling period, since bowhead calls typically occur in clusters (Blackwell *et al.*, 2007). Verification was stopped when ten adjacent files, corresponding to about 48 h, did not yield any positive identification. Since singing occurred almost continuously during the main detection period (fall-early winter), all files between the first and last one containing bowhead calls were manually checked at each station.

As no visual data were collected simultaneously with acoustic data, bowhead calls were identified based on the resemblance to published descriptions (Würsig and Clark, 1993; Stafford *et al.*, 2008) as well as confirmation from experienced researchers. Time and frequency characteristics of song units were measured using SYRINX.[1]

TABLE I. Position, deployment duration, and perpendicular distance to shore of five acoustic recorders deployed in the Chukchi Sea. Stations are ordered by increasing distance to Point Barrow, AK.

Station	Latitude (°N)	Longitude (°W)	Depth (m)	Deployment	Recovery	Distance to shore (km)
WN20	7138.5618	16132.2499	47.5	23-Oct-07	3-Aug-08	116
W50	7118.6390	16132.2463	49.5	25-Oct-07	2-Aug-08	92
PLN60	7123.9337	16435.2472	44.0	22-Oct-07	30-Jul-08	154
PLN40	7104.0222	16434.4752	38.5	21-Oct-07	20-Jul-08	131
PL50	7023.9794	16433.9426	42.0	21-Oct-07	20-Jul-08	88

C. Song classification

The analysis was restricted to songs. This is not meant to imply that bowheads only produced songs during this study or that non-stereotyped sequences and isolated calls were not present in the recordings. As a matter of fact, they were more common than songs in the spring, but they were not the focus of this paper and are not discussed further. A song was defined as a set of phrases each containing one or more repetitions of the same unit or group of units. Continuous repetitions of songs are referred to as song bouts. Song complexity varied greatly across the recording period, ranging from simple, stereotyped moan sequences to elaborate series of complex sounds. The predictability of the arrangement of units into phrases and their repetition over time were two key factors in determining whether a call sequence was indeed a song. Different songs were labeled based on aural or spectrographic qualities of the predominant phrase or a noticeable feature. Two sound files per 24-h period were checked to quantify the daily song diversity at each station and to produce time series of song occurrence. We did not attempt to enumerate the number of whales singing simultaneously or the number of songs per file because of the uncertainty introduced by overlapping songs from different whales.

D. Ice concentration data

In order to evaluate the impact of ice on the presence of bowheads in the study area, we obtained daily ice concentration data derived from passive microwave telemetry from the National Snow and Ice Data Center (Cavalieri *et al.*, 2004). The data had a 12.5×12.5 km^2 spatial resolution and were obtained in .hdf format. Each file was converted to a georeferenced raster using the HDF-EOS to GeoTIFF Conversion Tool,[2] which was subsequently imported into a geographic information system (ARCVIEW 9.2, ESRI, 2006) to extract daily ice concentration values at each of our recorders. Ice concentration time series were then plotted against the song detection time series for each recorder.

III. RESULTS

The five recorders provided acoustic data for an average of 274.6 days on a duty cycle of 48 min every 4 h. Many species were detected including bowhead whales, walrus, belugas, bearded and ringed seals, and unidentified ice seals. We hereafter only focus on bowhead songs.

A. Song description

The recordings contained six bowhead whale song types that were divided into four simple and two complex songs, each containing two to four units (Table II). The simple songs were stereotyped sequences of narrow-band, FM moans, at times associated with a complex call. The complex songs consisted of diverse sequences of simple and complex calls. The structure of complex songs was highly preserved over time, except during the initial development of the song when different combinations of the same units were recorded until the final, stable versions described below emerged. On the other hand, simple songs were less constrained structurally. The songs are described hereafter in chronological order of appearance.

TABLE II. Duration and frequency parameters (with standard deviation) of songs and their units. *n* refers to the number of measured units.

Song	Duration	Max freq (Hz)	Min freq (Hz)	*n*
Shriek	73.8 (5.4)			
Unit 1	2.8 (0.2)	2899.9 (199.4)	236.6 (44.5)	52
Unit 2	6.5 (1.0)	1604.7 (131.9)	890.7 (116.6)	52
Unit 3	0.6 (0.1)	623.7 (75.6)	77.1 (49.5)	35
Unit 4	2.0 (0.3)	355.5 (14.2)	225.6 (21.6)	47
Whistle	83.5 (9.5)			
Unit 1	3.5 (0.9)	1421.3 (97.6)	1025.7 (112.8)	48
Unit 2	0.5 (0.06)	956.8 (129.9)	160.0 (39.8)	48
Unit 3	2.4 (0.4)	2777.7 (278.3)	543.1 (30.9)	43
Unit 4	1.5 (0.2)	1288.5 (57.5)	147.1 (25.8)	45
Upsweep	65.8 (12.0)			
Unit 1	6.5 (1.7)	780.2 (235.6)	355.6 (42.5)	47
Unit 2	1.6 (0.4)	312.1 (31.5)	181.1 (16.1)	72
Unit 3	2.0 (0.4)	347.6 (16.0)	268.6 (11.7)	61
Downsweep	49.8 (13.0)			
Unit 1	1.7 (0.2)	354.7 (10.0)	304.9 (14.0)	55
Unit 2	0.9 (0.2)	230.5 (8.3)	207.6 (8.9)	67
Unit 3	1.5 (0.2)	328.9 (16.8)	257.4 (16.6)	56
Mixed	62.2 (5.8)			
Unit 1	0.3 (0.1)	397.6 (18.0)	337.3 (14.5)	58
Unit 2	1.4 (0.3)	371.9 (17.6)	227.4 (14.2)	51
U				
Unit 1	1.1 (0.3)	586.0 (45.7)	425.76 (27.0)	53
Unit 2	1.4 (0.3)	1375.3 (207.3)	579.3 (104.1)	53

The first simple song recorded was the *upsweep* song. It was primarily composed of upsweeping units (Figs. 2 and 3) but also included a warble sound (unit 1), though this unit was not systematically present. A typical upsweep song started with a warble followed by a phrase containing 2–12 (absolute range) repetitions of unit 2, itself followed by another phrase characterized by 2–12 repeats of unit 3. Units 2 and 3 were both upsweeping moans, but unit 2 had lower minimum and maximum frequencies and a larger bandwidth (Fig. 2; Table II). Variations in this song were numerous but limited to changes in the number of units per phrase, the

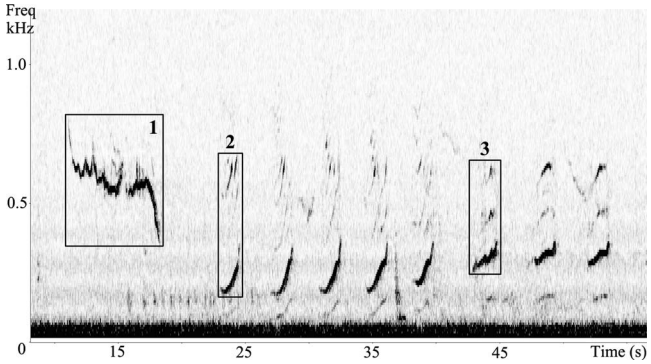


FIG. 2. Close-up of the *upsweep* song units. Each unit type is identified by a box. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

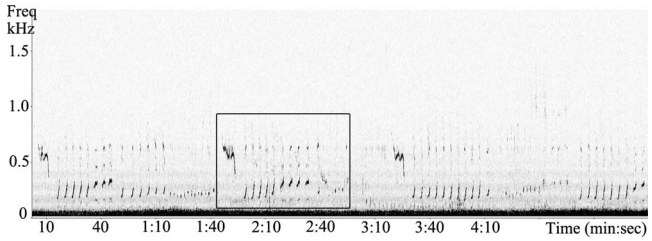


FIG. 3. Bout of the *upsweep* song. The box encompasses one song from the bout. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

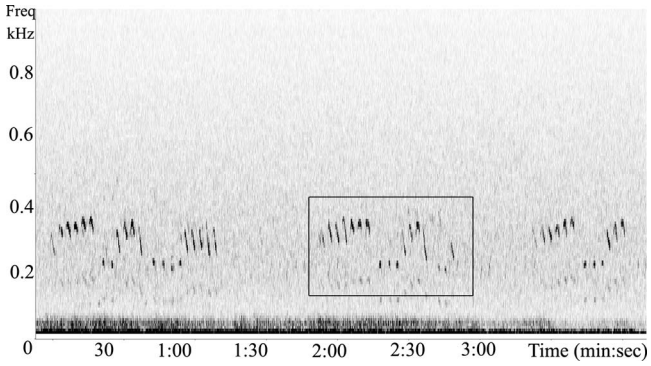


FIG. 5. Bout of the *downsweep* song. The box encompasses one song from the bout. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

presence of a phrase, or the order of phrases. The most common variant was the omission of unit 1. Songs were separated by a gap of 32.0 (±6.0) s, while units were separated by about 2–3 s. The mean duration of this song was 65.8 (±12.0) s. The upsweep song was the first song to be recorded, starting on November 8 at stations PL50, PLN40, and PLN60. It was detected on all five stations and stopped almost simultaneously on all of them on November 17. It was the only song detected during this early singing phase.

The second song recorded was the *downsweep* song, owing its name to the downswept nature of one of its units and to emphasize the contrast with the other simple songs (Figs. 4 and 5). This song was composed of three units (Fig. 4): a tonal up-down call (unit 1), a short constant or slightly upsweeping moan (unit 2), and a downswept moan (unit 3). The most common variant of this song consisted of a phrase containing 5–10 (absolute range) repetitions of unit 1 followed by about five repetitions of unit 2 and concluded by 2–5 repetitions of unit 3. As in the upsweep song, phrases contained a variable number of units but their order was fairly constant. This song almost invariably started with phrases of unit 1, but phrases of unit 3 were sometimes inserted before the middle phrase (unit 2) and omitted at the end. It also appeared that the relative proportion of unit 3 in comparison with units 1 and 2 increased over the period characterized by the downsweep song. Consecutive songs were separated by 38.6 (±9.5) s and the entire song lasted 49.8 (±13.0) s. The downsweep song was recorded in the middle of the detection period and was also transitory. It was first detected at PL50, PLN40, and PLN60 around November 22 and then started on November 30 and December 6 at W50 and WN20, respectively. It was last detected on December 21 at PL50.

The *shriek* song was the first of the two complex songs recorded in this study (Figs. 6 and 7). The first phrase consisted of 14–19 repetitions of unit 1 sung uninterruptedly and lasting 43.6 (±5.1) s. Unit 1 was a FM shriek made of upsweeping and downsweeping components. The first two to three units in this phrase generally did not include the upsweeping component, and the loudness of units increased toward the end of the phrase. The first phrase was followed within about 0.5 s by a second phrase formed by a warble (unit 2) and five to seven broad-band grunts (unit 3). Both units were produced simultaneously which supports previous findings suggesting that bowhead whales may possess several sound production organs (Würsig and Clark, 1993). The last phrase started about 5 s after the second phrase and consisted of 4–6 repetitions of unit 4 each 2 s apart. Unit 4 was a downswept moan with harmonic structure. The loudest moans contained up to ten harmonics, the second being dominant (and measured in Table II). A whole song lasted 73.8 (±5.4) s and was repeated after a 5–10 s gap.

All the sound files containing the shriek song that were checked visually contained uninterrupted singing, suggesting that the song bouts lasted for a minimum of 48 min. This song was first detected on December 3 at station PL50 and was only recorded at PL50 and PLN40. However, prior to that, primitive versions were detected as early as November

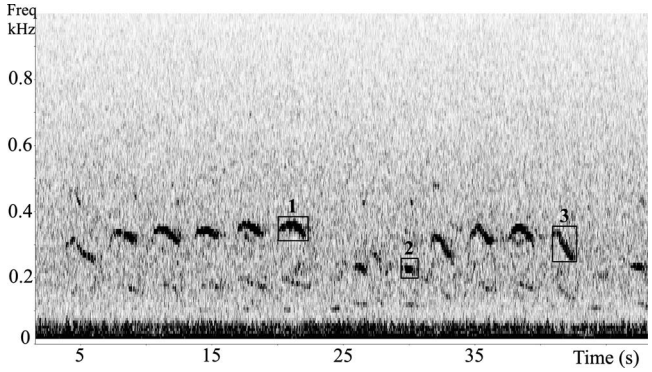


FIG. 4. Close-up of the *downsweep* song units. Each unit type is identified by a box. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

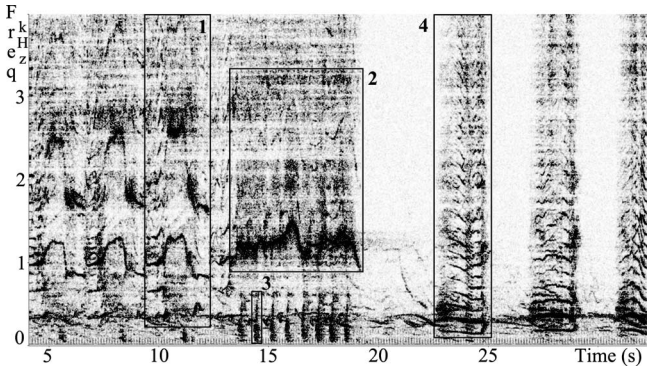


FIG. 6. Close-up of the *shriek* song units. Each unit type is identified by a box. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

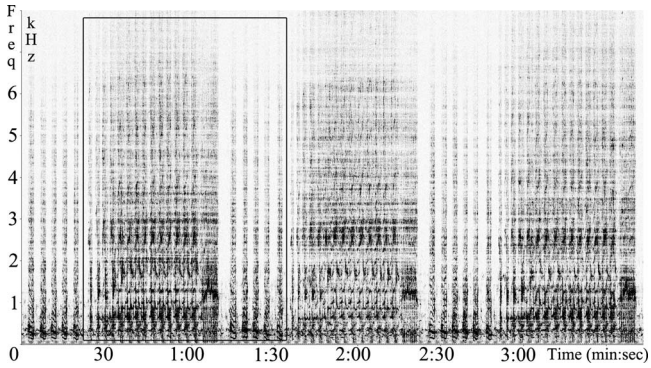


FIG. 7. *Shriek* song bout. The box encompasses one song from the bout. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

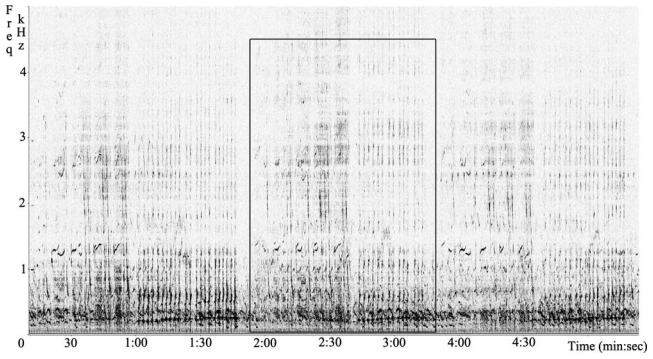


FIG. 9. *Whistle* song bout. The box encompasses one song from the bout. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

2 at PL50. The song recorded on that day started with a warble similar to unit 1 of the upsweep song or unit 2 of the shriek song and was followed by a phrase similar to the first phrase of the shriek song but with short gaps between units. The shriek song was last recorded on December 21 at PL50.

The second complex song recorded was the *whistle* song (Figs. 8 and 9). The first phrase was composed of repeats of a sound pair consisting of units 1 and 2 displayed simultaneously followed by one or two repeats of unit 3. The first phrase typically included five of these sound pairs. Unit 1 was a slightly upsweeping whistle evolving into a shriek with a sharp increase in frequency at the end; unit 2 was a broad-band grunt. Similar to units 2 and 3 of the shriek song, these were produced simultaneously. However, the broad-band grunts (unit 2) were only displayed with the last two to three repeats of unit 1 as the loudness of the phrase increased. Unit 3 was a downsweeping whistle-like call. This first phrase lasted 41.5 (±5.5) s. It was followed after a break ranging between 10 and 15 s by a second phrase consisting of 13–22 repetitions of unit 4. Unit 4 can be described as a LF constant tone leading to a "whoo-up" sound sharply increasing in frequency. The second phrase lasted 34.1 (±7.3) s and the overall song lasted 83.8 (±9.5) s. As with the shriek song, all files that were checked suggest that bouts of the whistle song last for a minimum of 48 min. This song was first recorded at PLN40 on December 8 and was only found there and at PL50. It was last recorded on December 19 at PL50.

The shriek and whistle songs were often recorded simultaneously with each other and with the downsweep and/or *mixed* (see below) song. Considering complex songs only, the shriek song was clearly dominant throughout the recording period, both in terms of the relative abundance of this song when both were present on a sound file and in terms of the number of files containing only one song type.

The fifth song type recorded was the mixed song. This was a simple song that included both upsweeping and downsweeping moans (Figs. 10 and 11). It consisted of a phrase composed of 12–22 repetitions of unit 1 followed by a second phrase composed of 3–6 repetitions of unit 2. Unit 1 was a short, narrow-band upsweep while unit 2 was a more broad-band FM call made up of a short upsweeping phase followed by a downsweep. Variations in this song were limited to changes in the number of units in each of the two phrases. A whole song lasted 62.2 (±5.8) s; consecutive songs were separated by 9.0 (±2.0) s. This song characterized the end of the acoustic detection period. It was first detected on December 5 at PL50 and subsequently at PLN40 and PLN60 on December 10. The mixed song overlapped with the downsweep song until this one disappeared and persisted after the whistle and shriek song vanished. It was last recorded on January 1, 2008 at PL50.

The last song recorded was the U song, a simple song

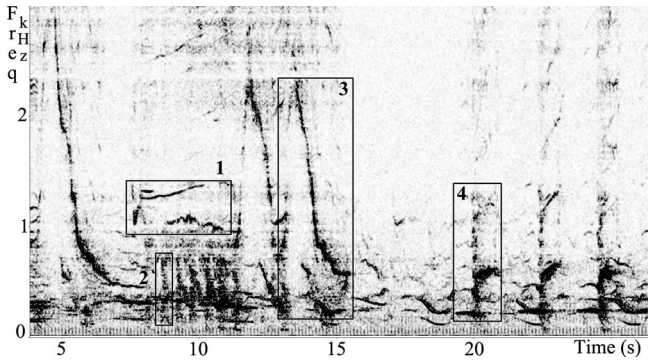


FIG. 8. Close-up of the *whistle* song units. Each unit type is identified by a box. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

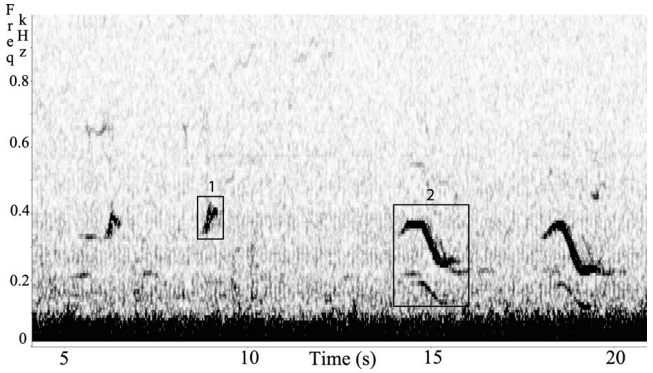


FIG. 10. Close-up of the units of the *mixed* song. Each unit type is identified by a box. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

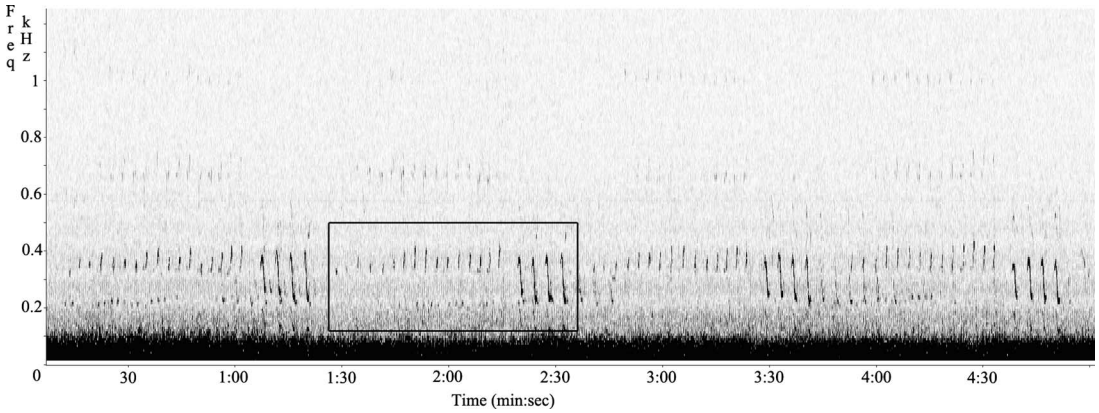


FIG. 11. Bout of the *mixed* song. The box encompasses one song from the bout. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

detected between April 16 and 18 and on May 4 and 5, 2008 (Figs. 12 and 13) and owing its name to the U-shape of its second unit. In the April version, it consisted of a repeated alternation of two units. The first unit was an up-down moan with a typically more prominent upsweeping phase. The second unit was a whistle-like call first decreasing then increasing in frequency. The end frequency was generally less than the start frequency but at times similar. Individual songs were made of a highly variable number of unit pairs and their duration could thus not be meaningfully measured. During the short May detection period, the second unit had essentially lost its initial downsweeping component, which was predominant in mid-April while the first unit remained almost unchanged. The May recordings contained a cacophony of bearded seal calls with several bowheads singing simultaneously. These features made it difficult to assess the structure of those songs precisely. However, these followed the pattern of the U song—a primarily upsweeping moan followed by a broad-band, FM whistle—although the structure of the second unit had evolved to some extent and the overall song structure appeared less constrained.

In fall and winter, many sound files contained songs that could not be clearly identified due to several whales singing simultaneously, and only the complex songs could be reliably labeled. This situation occurred for example at PL50 between the December 13 and 18. The unidentified songs—typically one or two of the simple songs described above—were grouped under the label "other" (Fig. 14).

B. Spatiotemporal distribution of songs

Songs were recorded as early as October 30, 2007 (Fig. 14), but these detections first occurred only on the southernmost station (PL50) and not at the closest station to Point Barrow, AK (WN20), where bowhead whales exit the Beaufort Sea. Except for these early detections, singing started almost simultaneously at all stations between November 8 and 10. Songs were recorded continuously between October 30 and January 1 on at least one hydrophone except during five periods: between October 31 and November 2, November 3 and 8, November 19 and 21, and December 22 and 27. Songs first stopped at WN20, the north-easternmost station, on December 17 and 2 days later at W50, PLN60, and PLN40. Songs were last recorded on January 1, 2008 at the southernmost station (PL50). The minimum distance to shore for all stations varied between 88 and 154 km (Table I). In spring, songs were only detected during 5 days at W50 only: between April 16 and 18 and between May 4 and 5, 2008.

C. Non-song sequences

A few seemingly stereotyped sequences of simple calls did not fit the organization patterns observed in the other six song types, thus one cannot rule out that other simple songs

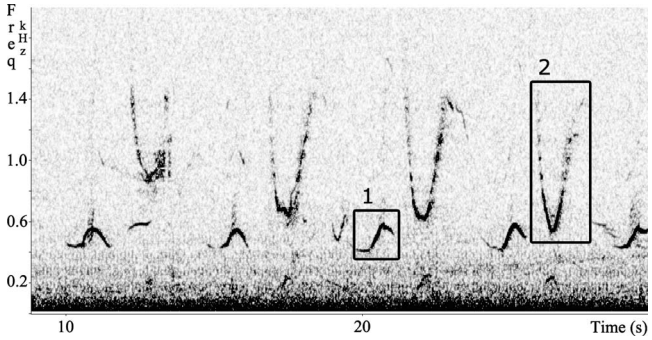


FIG. 12. Close-up of the units of the U song. Each unit type is identified by a box. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

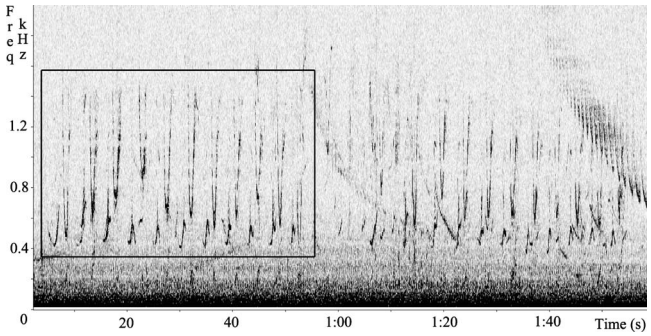


FIG. 13. Bout of the song. The box encompasses one song from the bout. Spectrogram generated using a Hanning window, 2048-point FFT, and no overlap.

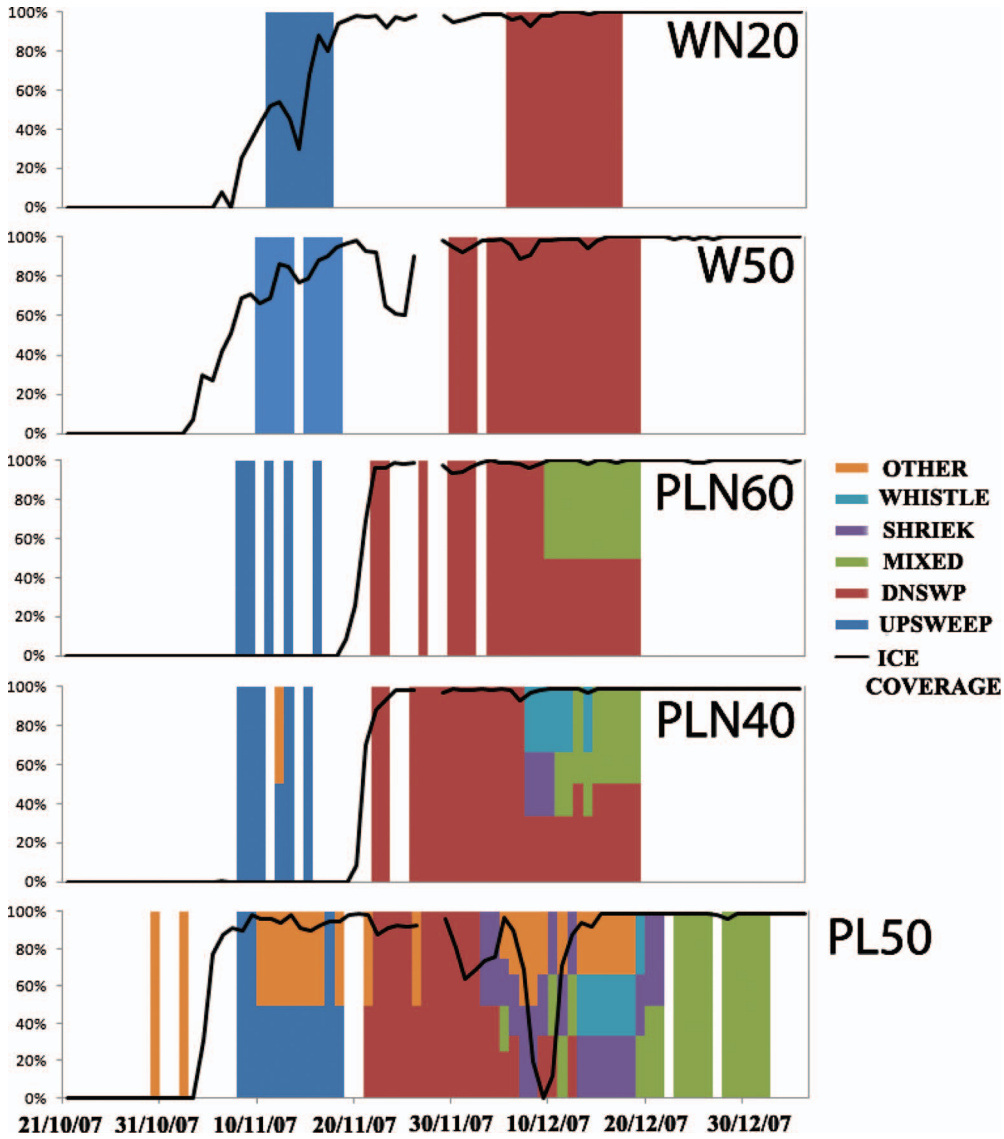


FIG. 14. Spatiotemporal distribution of bowhead whale song detection and ice concentration at five recording stations in the Chukchi Sea from October 21, 2007 until January 5, 2008. The vertical axis refers to ice concentration. The different colors in a vertical band represent daily detections of individual songs. The black line indicates the concentration of ice at each station.

exist. However, these unusual sequences were never detected on more than three consecutive files and they seem to be exempted. These sound files as well as those containing non-stereotyped sequences or isolated calls were not considered in evaluating the temporal distribution of songs (Fig. 14). Thus, the lack of detections between singing phase in fall or spring does not necessarily mean a complete absence of bowhead calls but rather an absence of songs.

D. Ice concentration data

Ice first appeared at stations PL50 and W50 closest to shore (Fig. 14) on November 3, and at WN20 5 days later. Ice did not form until November 20 at PLN40 and PLN60. Ice coverage increased gradually at W50 and WN20 and rapidly at the other three stations, reaching over 95% within 4 days. Once reaching 90% coverage or more, ice concentration remained at that level for most of the study period. Ice concentration at PL50 varied differently; it disappeared completely before increasing again to full coverage between December 6 and 13. Songs were recorded in heavy ice conditions, sometimes in large number, at all recorders though the first fall detections were in ice-free or light ice conditions. Ice noise was frequently overlapping with bowhead songs. In April and early May, the W50 station was fully covered in ice.

IV. DISCUSSION

A. Comparison with previous song descriptions

This study represents the first detailed description of complex songs for the BCB bowhead whale stock. Complex songs were previously recorded during the spring migration (Clark and Johnson, 1984; Würsig and Clark, 1993) but were not described in detail. Songs from different years are known to differ and singers within a season generally sing the same song, with limited intra- and inter-individual variations (Würsig and Clark, 1993). Stafford *et al.* (2008) recorded and described two complex and one simple bowhead songs

in springtime off Greenland. We recorded six different songs, including two complex ones, in the Chukchi Sea. Thus it appears that within a given season, more than one song can be heard in a given area.

Although it is not possible to compare the present songs to those recorded in the 1980s, it is worth noting that some phrases are very similar to those described in the past. For instance, the second phrase of the shriek song was highly reminiscent of the first phrase in a 1988 song (see Fig. 5.9 in Würsig and Clark, 1993) described as a series of broad-band grunts produced simultaneously with a high-frequency whistle. A similar sequence was also found in the whistle song. So although songs may differ between years, some patterns of association between certain types of units may be preserved over time.

The detailed description of bowhead song off Greenland in spring 2007 (Stafford *et al.*, 2008) provides an opportunity to compare songs from different areas. Aural and visual inspections of songs from Greenland and the Chukchi Sea revealed that they were all unique to the area where they were recorded. On the other hand, some phrases extracted from songs from both areas showed similar contours. For instance, the first phrase of the shriek song in this study resembled unit 2 of the *screech* song in Stafford *et al.*, 2008 [see Fig. 3(b)], and a warble-like sound was also part of one song in both areas. However, the resemblance was only structural and time and frequency parameters of these components differed between both areas. Overall, the different songs recorded in one area were as different from each other as from those in other areas. Since the Greenland and Chukchi Sea songs were recorded in different, though consecutive, breeding seasons and bowhead songs are known to change from 1 year to another, one can only speculate regarding the existence of geographic variations in bowhead songs. Geographic variations in songs have already been demonstrated for a number of marine mammal species, including humpback whales (Cerchio *et al.*, 2001), fin whales (Hatch and Clark, 2004), and bearded seals (*Erignathus barbatus*) (Risch *et al.*, 2007). Recordings obtained during the same breeding season in different areas should provide an opportunity to assess whether bowheads from different areas sing different songs.

B. Spatiotemporal song distribution and implications

Several songs were recorded during the study. Simple songs were individually relatively constrained in time (Fig. 14). The beginning of the detection period was characterized by the upsweep song. The apparition and disappearance of this song each occurred within 2 days at all stations and there was no overlap with the next simple song—the downsweep song—which was recorded in the middle of the fall detection period. The downsweep song was still recorded when the third simple song—the mixed song—first appeared. During the overlap between both songs, the downsweep song was first dominant but gradually faded away to completely disappear 10 days after the apparition of the mixed song. The latter was recorded on its own for another 2 weeks. The downsweep song appeared simultaneously at the three Point Lay stations and a few days later at the Wainwright stations. However, it disappeared within 36 h at all stations. Although no songs were recorded after that at all stations but PL50, the downsweep song was absent at PL50 later on, suggesting that its disappearance at the other stations was not only due to the whales leaving the area. The temporal patterns of complex songs at PL50 and PLN40 were less evident but they occurred within the same time frame. The simple song recorded in spring was different from the last one recorded in early winter. Both spring detections, separated by about 2 weeks, showed evidence of transformation for one of the two units.

One can propose two explanations for the relative temporal segregation of simple songs. Either different clusters of migrating whales passing through the area were characterized by different songs or new songs evolved due to individual novelties spreading through the population as is known to occur in humpback whales (Noad *et al.*, 2000; Eriksen *et al.*, 2005). In the latter case, one may expect to see gradual changes from one song to the next. Here, gradual changes, if any, are likely to have gone undetected in the fall due to the saturation of most sound files with moans from overlapping songs. However, there was evidence of such gradual changes between the two song detections in the spring.

Bowheads songs change from year to year (Würsig and Clark, 1993; Clark 1990) and even though more than one active song can be recorded at a given time (Stafford *et al.*, 2008; this study), all versions of a song produced by different singers in an area are identical. This, in addition to the present finding of songs appearing and disappearing almost simultaneously at stations up to 90 nautical miles apart, argues in favor of song cultural transmission in bowheads, as previously suggested by Rendell and Whitehead (2001).

The detections of bowhead whale sounds at distances between 88 and 154 km from shore confirms that at least a component of the population can be found relatively far offshore in the fall, presumably as they migrate toward the Bering Sea. Mel'nikov *et al.* (1997) suggested that after rounding Point Barrow, AK, part of the population heads for the northwestern portion of the Chukotka Peninsula, while the rest heads southwest toward the Bering Strait. Similarly, Moore and Clarke (1990) speculated on the existence of two migration routes, one heading southwest after passing Point Barrow and another one heading west along latitude 72°N. One satellite-tagged bowhead whale was found to follow the latter route (Mate *et al.*, 2000). Although the northernmost recorders could have detected whales using the northern route, the station PL50 (located at N70°24) is more likely to have recorded whales heading straight to the Bering Sea. It is also the station that recorded songs most consistently, in highest numbers and for the longest period. Considering the relatively limited geographical coverage of the array in light of the potential area used by bowheads while migrating, the current data cannot be used to locate migration routes or assess the relative significance of different routes. However, bowheads' intense vocal activity in fall and early winter offers interesting perspectives regarding the use of acoustics to study migration routes and distribution in these seasons. The

song recordings in the spring suggest that some bowheads can be found as far as 50 nautical miles from shore while migrating north.

The first song was recorded on October 30, 2007, when the migration is presumed to be already well underway (Moore and Reeves, 1993). Fall and winter detections occurred until the beginning of January. This suggests that singing by bowhead whales is not limited to the mating season, which peaks in March–April (Nerini *et al.*, 1984; Reese *et al.*, 2001). Similar observations were made for humpback whales whose songs were recorded on the feeding grounds (Mattila *et al.*, 1987) and during the migration toward the breeding grounds (Clapham and Mattila, 1990). In fin whales, song detection on a north-south array in the northwestern Atlantic brought evidence of a southern movement attributed to the fall migration to lower-latitude breeding grounds (Clark, 1995).

Stafford *et al.* (2008) speculated that male bowhead whales sing to mediate sexual competition and mate selection behaviors. This was based on such evidence in the related humpback whale (Darling *et al.*, 2006; Smith *et al.*, 2008) and the synchrony of songs with the mating season. Although mating is not believed to take place in the Chukchi Sea, the localized detection of two complex songs during 18 consecutive days produced by several singers indicates an aggregation of whales potentially engaged in courtship behaviors. If songs are indeed linked to mating, their timing in this study is earlier than the known peak mating time in March–April (Nerini *et al.*, 1984; Reese *et al.*, 2001). Koski *et al.* (1993) previously suggested that mating may start as early as January, so it is not unreasonable to believe that complex songs recorded in the second half of December in the Chukchi Sea represent the onset of the mating season.

The very limited number of song detections in spring may be due to the rather offshore (>50 nautical miles) location of the hydrophone array. Indeed, during the spring migration bowheads are known to follow an open-water lead that forms between Point Hope and Point Barrow along the Alaskan coast (Moore and DeMaster, 1998) and may thus be predominantly found close to shore.

These song detections indicate that bowhead whales were present in the Chukchi Sea until early January during the 2007–2008 winter and as early as mid-April in spring 2008. Mel'nikov *et al.* (1997) showed that in years with little ice, bowheads inhabit the Bering Strait and the southern portion of the Chukchi Sea in winter, which is consistent with the present findings. Ice concentration data indicated that by the beginning of December, all stations were almost 100% covered in ice, except for a week at PL50, but this did not appear to prevent bowheads from passing through or staying in the area, confirming the species' affinity for sea ice (Moore and Lairdre, 2006). This was also the case for the spring detections at W50. Although the coarse resolution of the ice maps analyzed (12.5×12.5 km^2) prevented us from detecting leads that bowheads likely used to migrate, ice conditions were still unarguably heavy. The lack of ice at PL50 during the second week of December could explain the greater number of songs and singers recorded there at that time. Overall, it is unknown whether the observed late residency of bowheads in the Chukchi Sea is an unusual phenomenon, but continuing overwinter acoustic monitoring should allow studying potential shifts in bowheads' winter distributions and habitat use in a currently warming Arctic characterized by receding sea ice (Stroeve *et al.*, 2007).

C. Further studies

The different level of complexity and consistency between simple and complex songs was striking. Simple songs appear to fit the definition of songs in that their units remained unchanged and their structural organization was fairly predictable even though a few variations were observed. On the other hand, complex songs were highly conserved once established. Further studies should investigate whether different age classes, genders, or clusters of migrating whales are involved in the production of different song type and whether simple and complex songs serve different behavioral functions. In addition, the temporal segregation of different songs and their development and spreading throughout the population deserve more attention. Finally, more work should be dedicated to the potential linkage between differences in songs and population structure. Although it was initially assumed that bowheads found in the Bering, Chukchi, and Beaufort Seas form a single stock (International Whaling Commission, 2003), several hypotheses are currently being explored to explain spatiotemporal differences observed among bowheads from these areas (International Whaling Commission, 2005), which potentially argue in favor of a multi-stock structure. In other species such as blue whale (McDonald *et al.*, 2006), fin whale (Delarue *et al.*, 2009) or killer whale (Ford *et al.*, 1982) geographic variations in songs or calls have been used or proposed to assess stock structure and similar findings in bowheads may improve the species' management.

An eight-recorder array was redeployed in October 2008 to survey the fall 08-spring 09 period. The area covered extends further north and southwest than in this study which should improve the current knowledge on fall and winter habitat use and migration routes of bowhead whales. This may become significant if the Chukchi Sea was found to be consistently used by bowheads in these periods as anthropogenic use of that area is likely to increase in the future.

ACKNOWLEDGMENTS

This work was funded by Shell Exploration and Production Co. The authors would like to thank Sue Moore, Kathleen Stafford, and Jim Darling for their assistance in confirming the identity of the species producing these songs. Christine Erbe, Dale Funk, and David Hannay and two anonymous reviewers provided useful comments on earlier drafts. We are also thankful for the support of Shell and to the staff of JASCO Research and crews of Norseman II for their assistance in deploying and recovering the recording units.

[1]http://www.syrinxpc.com/ (Last viewed 11/4/2009).
[2]http://newsroom.gsfc.nasa.gov/sdptoolkit/HEG/HEGHome.html (Last viewed 11/4/2009).

Blackwell, S. B., Richardson, W. J., Greene, C. R., Jr., and Streever, B. (**2007**). "Bowhead whale (*Balaena mysticetus*) migration and calling behavior in the Alaskan Beaufort Sea, autumn 2001-04: An acoustic localization study," Arctic **60**, 255–270.

Cavalieri, D., Markus, T., and Comiso, J. **2004**, AMSR-E/Aqua Daily L312.5 km Brightness Temperature, Sea Ice Concentration, and Snow Depth Polar Grids V002, Boulder, CO, National Snow and Ice Data Center, Digital Media.

Cerchio, S., Jacobsen, J. K., and Norris, T. F. (**2001**). "Temporal and geographical variation in songs of humpback whales, *Megaptera novaeangliae*: Synchronous change in Hawaiian and Mexican breeding assemblages," Anim. Behav. **62**, 313–329.

Clapham, P. J., and Mattila, D. K. (**1990**). "Humpback whale songs as indicator or migration routes," Marine Mammal Sci. **6**, 155–160.

Clark, C. W. (**1995**). "Application of US Navy underwater hydrophone arrays for scientific research on whales," Rep. Int. Whal. Comm. **45**, 210–212.

Clark, C. W. (**1990**). "Acoustic behavior in mysticete whales," in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York).

Clark, C. W., and Johnson, J. H. (**1984**). "The sounds of the bowhead whales, *Balaena mysticetus*, during the spring migrations of 1979 and 1980," Can. J. Zool. **62**, 1436–1441.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., and Urban, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809.

Cummings, W. C., and Holliday, D. V. (**1987**). "Sounds and source levels from bowhead whales off Pt. Barrow, Alaska," J. Acoust. Soc. Am. **82**, 814–821.

Cummings, W. C., and D. V. Holliday, W. T. Ellison, and B. J. Graham. (**1983**). "Technical feasibility of passive acoustic location of bowhead whales in population studies off Point Barrow, AK: A report to the North Slope Borough," Report No. T-83-06-002(U), Tracor Inc., San Diego, CA.

Dahlheim, M. E., Bray, T., and Braham, H. (**1980**). "Vessel survey for bowhead whales in the Bering and Chukchi Seas, June-July 1978," Mar. Fish. Rev. **42**, 51–57.

Darling, J. D., Jones, M. E., and Nicklin, C. F. (**2006**). "Humpack whale songs: Do they organize males on the breeding grounds?," Behaviour **143**, 1051–1101.

Delarue, J., Todd, S. K., Van Parijs, S. M., and Di Iorio, L. (**2009**). "Geographic variations in Northwest Atlantic fin whale (*Balaenoptera physalus*) song: Implications for stock structure assessment," J. Acoust. Soc. Am. **125**, 1774–1782.

Eriksen, N., Miller, L. A., Tougaard, J., and Helweg, D. A. (**2005**). "Cultural change in the songs of humpback whales (*Megaptera novaeangliae*) from Tonga," Behaviour **142**, 305–328.

ESRI (Environmental Systems Research Institute, Inc.). (**2006**). ArcGIS 9.2, Redlands, CA.

Ford, J. K. B., and Fisher, H. D. (**1982**). "Killer whale (*Orcinus orca*) dialects as an indicator of stocks in British Columbia," Rep. Int. Whal. Comm. **32**, 671–679.

Hatch, L. T., and Clark, C. W. (**2004**). "Acoustic differentiation between fin whales in both the North Atlantic and North Pacific Oceans, and integration with genetic estimates of divergence," presented at the IWC Scientific Committee, Sorrento, Italy, Paper No. SC/56/SD6.

International Whaling Commission (**1992**). "Report of the scientific committee, Annex I," Rep. Int. Whal. Comm. **42**, 121–138.

International Whaling Commission (**2003**). "Annex F. Report of the subcommittee on bowhead, right and gray whales," http://www.iwcoffice.org/_documents/sci_com/SCRepFiles2003/Annex%20F%20-%20BRG.pdf (Last viewed November 24, 2008).

International Whaling Commission (**2005**). "Report of the sub-committee on bowhead, right and gray whales," http://www.iwcoffice.org/_documents/sci_com/SCRepFiles2005/AnnexFsq.pdf (Last viewed January 16, 2009).

Koski, W. R., Davis, R. A., Miller, G. W., and Withrow, D. E. (**1993**). in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and J. C. Cowles (Society for Marine Mammalogy, Allen, Lawrence, KS), Special Publication No. 2, pp. 239–274.

Ljungblad, D. K., Leatherwood, S., and Dahlheim, M. E. (**1980**). "Sounds recorded in the presence of an adult and calf bowhead whales," Mar. Fish. Rev. **42**, 86–87.

Ljungblad, D. K., Thompson, P. O., and Moore, S. E. (**1982**). "Underwater sounds recorded from migrating bowhead whales, *Balaena mysticetus*, in 1979," J. Acoust. Soc. Am. **71**, 477–482.

Mate, B. R., Krutzikowsky, G. K., and Winsor, M. H. (**2000**). "Satellite-monitored movements of radio-tagged bowhead whales in the Beaufort and Chukchi Seas during the late-summer feeding season and fall migration," Can. J. Zool. **78**, 1168–1181.

Mattila, D. K., Guinee, L. N., and Mayo, C. A. (**1987**). "Humpback whale songs on a North Atlantic feeding ground," J. Mammal. **68**, 880–883.

McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. (**2006**). "Biogeographic characterization of blue whale song worldwide: Using song to identify populations," J. Cetacean Res. Manage. **8**, 55–65.

Mel'nikov, V. V., Zelensky, M. A., and Ainana, L. I. (**1997**). "Observations on distribution and migration of bowhead whales (Balaena mysticetus) in the Bering and Chukchi Seas," Scientific Report No. SC/50/AS3, International Whaling Commission, Department of Wildlife Management, North Slope Borough, Barrow, AK.

Miller, R. V., Rugh, D. J., and Johnson, J. H. (**1986**). "The distribution of bowhead whales, *Balaena mysticetus*, in the Chukchi Sea," Marine Mammal Sci. **2**, 214–222.

Moore, S. E. (**1992**). "Summer records of bowhead whales in the northeastern Chukchi Sea," Arctic **454**, 398–400.

Moore, S. E., and Clarke, J. T. (**1990**). "Distribution, abundance and behavior of endangered whales in the Alaskan Chukchi and Western Beaufort Sea, 1989. Report by SEACO/Science Applications International Corporation to the U.S. Minerals Management Service, Alaska Outer Continental Shelf Region," NTIS Report No. PB91-105494, National Technical Information Service, Springfield, VA.

Moore, S. E., and DeMaster, D. P. (**1998**). "Cetacean habitats in the Alaskan Arctic," J. Northwest Atl. Fish. Sci. **22**, 55–69.

Moore, S. E., and Lairdre, K. L. (**2006**). "Trends in sea ice cover within habitat used by bowhead whales in the western Arctic," Ecol. Appl. **16**, 932–944.

Moore, S. E., and Reeves, R. R. (**1993**). in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and J. C. Cowles (Society for Marine Mammalogy, Allen, Lawrence, KS), Special Publication No. 2, pp. 313–386.

Nerini, M. D., Braham, H. W., Marquette, W. M., and Rugh, D. J. (**1984**). "Life History of the bowhead whale, *Balaena mysticetus* (Mammalia: Cetacea)," J. Zool. **204**, 443–468.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., and Jenner, K. C. S. (**2000**). "Cultural revolution in whale songs," Nature (London) **408**, 537.

Reese, C., Calvin, J. A., George, J. C., and Tarpley, R. J. (**2001**). "Estimation of fetal growth and gestation in bowhead whales," J. Am. Stat. Assoc. **96**, 915–938.

Rendell, L., and Whitehead, H. (**2001**). "Culture in whales and dolphins," Behav. Brain Sci. **24**, 309–382.

Risch, D., Clark, C. W., Corkeron, P. J., Elepfandt, A., Kovacs, K. M., Lydersen, C., Stirling, I., and Van Parijs, S. M. (**2007**). "Vocalizations of male bearded seals, *Erignathus barbatus*: Classification and geographical variation," Anim. Behav. **73**, 747–762.

Smith, N. S., Goldizen, A. W., Dunlop, R. A., and Noad, M. J. (**2008**). "Song of make humpback whales, *Megaptera novaeangliae*, are involved in intersexual interactions," Anim. Behav. **76**, 467–477.

Stafford, K. M., Moore, S. E., Lairdre, K. L., and Heide-Jørgensen, M. P. (**2008**). "Bowhead whale springtime sing off West Greenland," J. Acoust. Soc. Am. **124**, 3315–3323.

Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M. (**2007**). "Arctic sea ice decline: Faster than forecast," Geophys. Res. Lett. **34**, L09501.

Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). "The 20-Hz signals of finback whales (*Balaenoptera physalus*)," J. Acoust. Soc. Am. **82**, 1901–1912.

Würsig, B., and Clark, C. W. (**1993**). *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and J. C. Cowles (Society for Marine Mammalogy, Allen, Lawrence, KS), Special Publication No. 2, pp. 157–199.

Acoustic analysis of primate air sacs and their effect on vocalization

Bart de Boer[a)]
Universiteit van Amsterdam, Spuistraat 210, 1012 Vermont Amsterdam, The Netherlands,

(Received 31 July 2008; revised 2 October 2009; accepted 2 October 2009)

This paper presents an analysis of the acoustic impedance of primate air sacs and their interaction with the vocal tract. A lumped element model is derived and it is found that the inertance of the neck and the volume of the air sac are relevant, as well as the mass and stiffness of the walls (depending on the tissue). It is also shown that at low frequencies, radiation from the air sac can be non-negligible, even if the mouth is open. It is furthermore shown that an air sac can add one or two low resonances to the resonances of the oral tract, and that it shifts up the oral tract's resonances below approximately 2000 Hz, and shifts them closer together. The theory was verified by acoustic measurements and applied to the red howler monkey (*Alouatta seniculus*) and the siamang (*Symphalangus syndactylus*). The theory describes the physical models and the siamang calls correctly, but appears incomplete for the howler monkey vocalizations. The relation between air sacs and the evolution of speech is discussed briefly, and it is proposed that an air sac would reduce the ability to produce distinctive speech, but would enhance the impression of size of the vocalizer.
 [DOI: 10.1121/1.3257544]

PACS number(s): 43.80.Ka, 43.70.Bk, 43.70.Aj, 43.20.Ks [DOS] Pages: 3329–3343

I. INTRODUCTION

Air sacs are a feature of many primate (and indeed many mammal) vocal tracts. They are air-filled cavities that are connected to the vocal tract. Hewitt *et al.* (2002) and Riede *et al.* (2008) described four main types of primate air sacs. The first type consists of lateral ventricular air sacs, which either exist in pairs or merge ventrally into a single air sac. They connect laterally to the larynx through the ventricles (which are located above the vocal folds) and extend ventrally or laterally. The second type consists of subhyoidal air sacs, connected to the larynx just above the vocal folds, and extending under the hyoid. Subhyoidal air sacs can either be contained completely in an extension of the hyoid (the bulla), or be only partly contained in the hyoid. The third type consists of infraglottal air sacs, connected to the larynx at the level of or below the vocal folds, and extending ventrally. The fourth type consists of dorsal air sacs, which connect between the larynx and the trachea (below the vocal folds) and extend dorsally. Frey *et al.* (2007), who discussed air sacs in other mammals than primates as well, distinguished a fifth type of air sac, the ventral laryngeal air sac, which exists inside a thyroid bulla, connects to the vocal tract above the vocal folds, and extends ventrally in the laryngeal cavity. Most air sac types have soft walls (and therefore can change volume), but the subhyoidal air sacs that are contained completely in the hyoid bulla have hard walls and fixed volume. For more details, the reader is referred to Hewitt *et al.*, 2002, Riede *et al.*, 2008, Frey *et al.*, 2007, and references therein.

This paper investigates the effect of air sacs on the resonance properties of the primate vocal tract, and is therefore concerned primarily with air sacs that connect to the vocal tract *above* the glottis: the lateral ventricular and subhyoidal air sacs. These are illustrated in Fig. 1. Subhyoidal air sacs are found among others in howler monkeys (genus *Alouatta*), which are known for their extremely loud and low-frequency calls. Lateral ventricular air sacs are found in many species of primates, but most relevantly in apes. All apes have such air sacs: chimpanzees (e.g., Gautier, 1971; Nishimura *et al.*, 2007), orangutans (e.g., Avril, 1963; Brandes, 1932; Camper, 1779), and gorillas (e.g., Kleinschmidt, 1938), while Avril (1963) presented a comparison. However, in humans, only the ventricle remains as a vestige of the air sac. There is also some evidence, in the shape of fossil hyoid bones, that *Homo neanderthalensis* (Arensburg *et al.*, 1989) and *Homo heidelbergensis* (Martínez *et al.*, 2008) did not have air sacs, while *Australopithecus afarensis* (Alemseged *et al.*, 2006) did (most likely lateral ventricular sacs, as these are the ones found in chimpanzees and gorillas). It is therefore tempting to speculate that the loss of air sacs in humans has something to do with the evolution of speech (e.g., Fitch, 2000; Nishimura, 2008), and this may partly explain a recent increase in interest in the study of air sacs in primates.

The acoustic function of these air sacs, however, is not clear. Authors in the biological literature write about their function as resonators (e.g., Avril, 1963; Schön, 1971), as impedance matchers (e.g., Fitch and Hauser, 1995), or about their potential to suppress resonances (Haimoff, 1983). There is even controversy about whether air sacs have an acoustic function at all. Hilloowalla and Lass (1978) surgically removed air sacs in rhesus macaques (*Macaca mulatta*, which has partly subhyoidal air sacs) and did not find a difference in the acoustics of their calls. Gautier (1971) on the other hand found that puncturing an air sac in Brazza's monkey (*Cercopithecus neglectus*, which has lateral ventricular air sacs) leads to a decrease in volume of the produced calls. It has also been proposed that air sacs are an accidental byprod-

[a)]Electronic mail: b.g.deboer@uva.nl

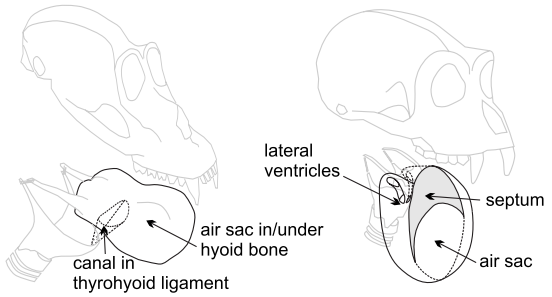


FIG. 1. Illustration of subhyoidal air sac (left) and lateral ventricular air sac (right) and roughly representing the anatomy of the red howler monkey and siamang, respectively. Note that the howler monkey in reality also has two lateral ventricular sacs, but these are omitted to better illustrate the subhyoidal air sac. Also note that in the siamang the two lateral ventricular air sacs have merged to one sac (the septum is a vestige of the wall between the two sacs).

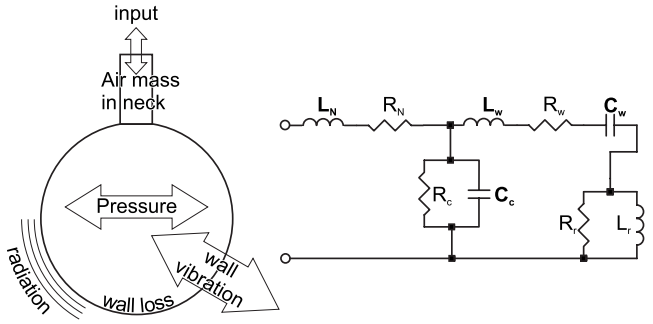


FIG. 2. Schematic representation of air sac, with modeled factors on the left and the lumped element electrical circuit on the right. Elements that are non-negligible are indicated in boldface. Values for elements of the circuit are given in the text.

uct of forced expiration (Brandes, 1932), that they serve for rebreathing air (Negus, 1949), and recently a case has been made that they can serve to prevent hyperventilation (Hewitt *et al.*, 2002). It is likely that air sacs play a range of different roles, and it is unlikely that they do not have any acoustic role whatsoever, especially in species that both have loud calls and large air sacs, such as siamangs (Haimoff, 1981, 1983) and howler monkeys (Schön, 1971; Schön Ybarra, 1986; Whitehead, 1995).

In order to make the discussion of the acoustic role of air sacs more concrete, an accurate theoretical model is needed. The purpose of this paper is to develop such a model. This model will analytically describe the acoustic effects of air sacs connected to a vocal tract. It will make it possible to predict (for low frequencies) the main acoustic effects for any primate air sac on the basis of anatomical measurements, without the need for complex computational simulations. However, the correspondence is by necessity approximate. As a first effort to understand the effect of air sacs, the work is therefore more comparable to early work on side branches (such as reviewed in, e.g., Fant, 1960; Flanagan, 1965) rather than more quantitatively precise work that takes anatomical detail into account (e.g., Dang and Honda, 1997; Dang *et al.*, 1994). The theory is also used to develop a computer model for higher frequencies.

The work presented here and elsewhere (de Boer, 2008a, 2008b) is closely related to the work presented by Riede *et al.* (2008). However, the aim and scope of the present paper are different. Riede *et al.*'s (2008) paper is mostly experimental in nature, and their computational model serves to validate their experiments. Another difference is that Riede *et al.*'s (2008) paper investigates the interaction between the vocal folds and the vocal tract including the air sac. Although this interaction is likely to be extremely important in ape vocalizations (e.g., Titze, 2008) it falls outside the scope of the present paper.

The paper consists of two parts. In Section II, the theoretical model is developed, it is shown which properties of the air sac are important and an analysis is made of how an air sac influences the resonances of the vocal tract to which it is connected. In Sec. III, the theory is put to the test by comparing its predictions with measurements on a physical model. The theory is also applied to animal data. Calls from two species that are known for their air sacs, siamangs (*Symphalangus syndactylus*) and red howler monkeys (*Alouatta seniculus*), are predicted on the basis of anatomical data, and the predictions are compared with the actual calls.

II. DEVELOPING A MODEL

A. A lumped element model

A lumped element model is a simplified representation of an acoustic system in which parts of the system are approximated by components that each represent a compliance, a resistance, or an inertance. In such a system each component represents a spatially extended part of the acoustic system. This means that lumped element models are only good approximations at low frequencies—corresponding to wavelengths that are long in comparison with the dimensions of the modeled parts. Given that ape air sacs have diameters of approximately 10 cm, and that the minimal allowed wavelength must be larger than two times this length, the maximal frequency for which the model developed here is valid is *lower* than $35\,000/20 \approx 1750$ Hz, and therefore a rounded down limit of 1500 Hz was assumed. However, as this approach simplifies the system to be modeled, analysis of a lumped element model can provide theoretical insights into the system's behavior. The elements that are relevant for the acoustics of an air sac are the volume of air that resonates in the air sac, the wall that might vibrate and radiate sound, and the tube (the neck) with which the air sac is connected to the rest of the vocal tract.

For the lumped element analysis, it will be assumed that air in the neck moves with uniform speed, that pressure everywhere in the cavity of the sac is the same, that the wall moves with the same phase and amplitude everywhere, and that the air sac radiates sound as a sphere. The factors that are modeled are illustrated in Fig. 2.

1. The neck

At low frequencies, the air in the neck of the air sac acts as a mass that moves when pressure is exerted on it. There will also be losses caused by viscosity of the air. Such a system can be modeled by a (frequency dependent) resis-

tance in series with an inertance. According to Flanagan (1965) (Sec. 3.21), the value per unit length of the inertance is given by

$$L = \frac{\rho_a}{A}, \tag{1}$$

where ρ_a is the density of air and A is the cross-sectional area of the tube. The total inertance is given by

$$L_n = \int_0^l \frac{\rho_a}{A(x)} dx, \tag{2}$$

where l is the length of the tube. For a neck with constant cross-sectional area, this reduces to

$$L_n = \frac{\rho_a l}{A}. \tag{3}$$

The inertance of an air sac neck of any shape can be expressed as that of an equivalent neck with constant cross-sectional area that has a length and area that are chosen to give the same value for inertance. Given the form of the expression for viscous losses, it is most convenient to choose the minimal cross-sectional area as the equivalent area, and calculate the equivalent length accordingly. This is because this form maximizes the viscous losses.

The value for the viscous loss per unit length (following Flanagan, 1965, Sec. 3.22) is given by

$$R = \frac{S}{A^2}\sqrt{\frac{\omega \rho_a \mu}{2}}, \tag{4}$$

where S is the circumference of the neck, ω the angular frequency, and μ is the viscosity coefficient of air. Integrating over the length of the neck gives the total resistance. It turns out that, given the form of Eq. (4), assuming constant cross-sectional area equal to the minimal area of the original tract, using the equivalent length giving the same inertance as the original tract, and using the maximal circumference of the original tract will always result in a tract *overestimating* the losses. In Sec. II A 6 it will be shown that even this upper bound on the losses can be ignored for primate air sacs.

The values for inertance and loss at low frequencies for the neck of an air sac are therefore given by

$$L_n = \frac{\rho_a l_e}{A_{\min}} \tag{5}$$

and

$$R_n = \frac{S_{\max} l_e}{{A_{\min}}^2}\sqrt{\frac{\omega \rho_a \mu}{2}}, \tag{6}$$

where l_e is the equivalent length, $A_{\min}$ is the minimal cross-sectional area of the original neck, and $S_{\max}$ is its maximal circumference.

For frequencies that are low enough for the lumped element model to be valid, this analysis also works for necks that contain branches or for air sacs with two necks (such as the air sac of the siamang). In this case, it suffices to add the volume of the two necks, take the minimal area as the equivalent area, and calculate the equivalent length by dividing the total volume of the necks by their minimal area.

TABLE I. Constants and parameters used in the lumped element model. Constants taken from Flanagan, 1965 (Sec. 3.25).

	Description	Value
c	Sound speed	350 m s^{-1}
ρ_a	Air density	1.14 kg m^{-3}
c_p	Air specific heat	1.00×10^3 J kg^{-1} K^{-1}
λ	Air heat conduction	0.023 W m^{-1} K^{-1}
k_2	Wall elasticity	9×10^3 soft tissue
		4.5×10^5 cartilage
		15×10^9 bone
η	Air adiabatic constant	1.4
$A_{\min}$	Neck area	$2.5\pi\times10^{-5}$ m^2
r	Sac radius	0.05 m
d	Wall thickness	5×10^{-3} m
μ	Air viscosity	18.6×10^{-6} Pa s
ρ_w	Wall density	10^3 kg m^{-3}
Q	Wall quality factor	1 soft tissue
		10 cartilage
		10 bone
l_e	Neck length	0.03 m
$S_{\max}$	Neck circumference	$10\pi\times10^{-3}$ m

2. The cavity

The air in the cavity acts as a spring that stores and releases energy when it is compressed and released and can therefore be modeled as a compliance. As compressed air is heated slightly, and as this heat can be transferred to the walls, some losses might occur. This can be modeled by a (frequency dependent) resistance.

The value of the compliance, C_c in acoustic terms, is given by

$$C_c = \frac{V}{\rho_a c^2}, \tag{7}$$

where V is the cavity volume, ρ_a is the density of air, and c is the speed of sound (Fletcher *et al.*, 2004).

The value R_c of the resistance due to losses at the wall is adopted from Flanagan (1965) (Sec. 3.24) and is given by

$$R_c = \frac{\rho_a c^2}{A(\eta - 1)}\sqrt{\frac{2 c_p \rho_a}{\lambda \omega}}, \tag{8}$$

where A is the cavity's area, η is the adiabatic constant, c_p is the specific heat of air at constant pressure, and λ is the coefficient of heat conduction. Values of all constants are given in Table I.

3. The wall

In the lumped element model the wall is assumed to vibrate with the same amplitude and phase everywhere. It is also assumed that it behaves as a damped oscillator, which is described by the following equation:

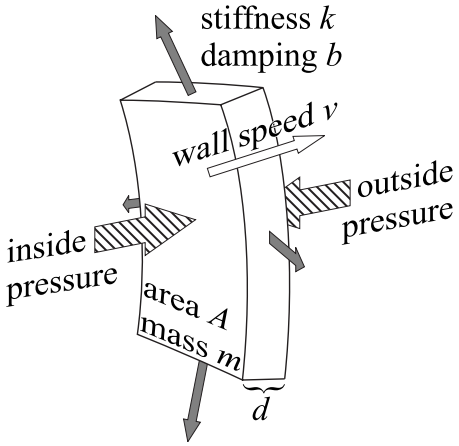


FIG. 3. Free body diagram for wall motion. Note that although damping b and stiffness k are generated inside the wall, only the component parallel to the wall movement (due to the wall's curvature) is considered.

$$p(t)A = m\frac{dv}{dt} + bv + k\int v dt, \tag{9}$$

where p is the differential pressure, v is the wall's velocity, m is its mass, b is its damping, and k is its stiffness. A free-body diagram of an element of the wall is given in Fig. 3. Differential pressure is the difference between the (vibrating) pressure in the air sac and the pressure of the air outside the air sac.

Acoustic impedance is the ratio between pressure p and volume velocity $u=vA$. Rewriting Eq. (9) to volume velocity gives

$$p(t) = \frac{m}{A^2}\frac{du}{dt} + \frac{b}{A^2}u + \frac{k}{A^2}\int u dt \tag{10}$$

(as wall motion is small with respect to the radius of the air sac, A is assumed to be constant to first order). Now if p and u are harmonic vibrations, with angular frequency ω, they can be written as

$$p(t) = Pe^{i\omega t}, \quad u(t) = Ue^{i\omega t}, \tag{11}$$

where P and U are the (complex) amplitudes of the vibrations. Substituting this in Eq. (10) and dividing out $e^{i\omega t}$ give

$$P = \left(\frac{m}{A^2}i\omega + \frac{b}{A^2} + \frac{k}{A^2}\frac{1}{i\omega}\right)U. \tag{12}$$

This is equivalent to an inertance, a resistance, and a compliance in series.

The wall's mass can be calculated from its area, its thickness, and its density. This results in the following value for wall inertance L_w:

$$L_w = \frac{m}{A^2} = \frac{\rho_w dA}{A^2} = \frac{\rho_w d}{A}, \tag{13}$$

where ρ_w is the density of wall tissue and d is the wall's thickness.

The wall's stiffness depends on Young's modulus of the tissue. However, as the deformation of the wall is the stretching of a two-dimensional membrane, and Young's modulus is

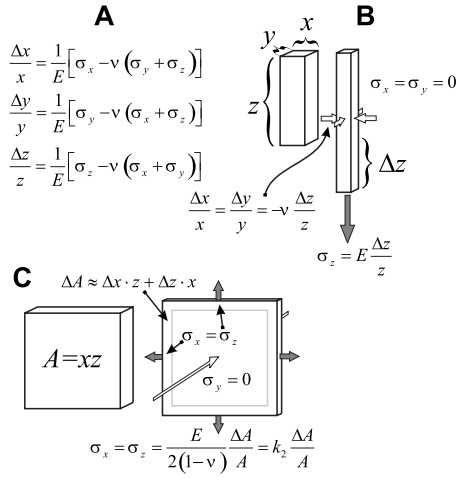


FIG. 4. Derivation of k_2 from generalized Hooke's law. (A) shows the equations for generalized Hooke's law, where $\sigma_{x,y,z}$ are the stresses (force per unit area) along the three axes. (B) shows its application to one-dimensional stretching, and obtains one-dimensional Hooke's law. (C) shows its application to two-dimensional stretching with constant strain in all stretching directions and obtains the relation between Young's modulus, Poisson's ratio, and k_2.

defined for stretching along one axis, Young's modulus E will have to be converted to a two-dimensional equivalent k_2 using Poisson's ratio ν (illustrated in Fig. 4)

$$k_2 = \frac{E}{2(1-\nu)}. \tag{14}$$

This quantity gives a ratio between stretching of the surface and strain as follows:

$$k_2 = \frac{S/d}{\delta A/A}, \tag{15}$$

where S is the surface tension, d is the thickness of the wall, and δA is the change in area. Now in order to find a relation between pressure and volume velocity (as is required for acoustic impedance) a relation between surface tension and pressure is needed, as well as a relation between change in area and volume velocity.

The relation between surface tension and pressure is given by the Young–Laplace equation

$$p = S\left(\frac{1}{r_1} + \frac{1}{r_2}\right), \tag{16}$$

where p is the differential pressure, and r_1 and r_2 are the principal radii of curvature. For a spherical air sac, where $r_1=r_2=r$, this simplifies to

$$p = \frac{2S}{r}. \tag{17}$$

In order to calculate the relation between change in area δA and volume velocity u, a relation between a radius of the air sac and its surface is needed. For all three dimensional shapes such a relation has the following form:

$$A = Cr^2, \tag{18}$$

where C is a constant that depends on the object's shape (4π for a sphere, for example). Now for a small change in area δA this becomes

$$A + \delta A = C(r + \delta r)^2 \approx Cr^2 + 2Cr\delta r = A + 2Cr\delta r. \tag{19}$$

For small values of δA and δr, such that second order terms in the small quantities can be ignored, the following relation therefore holds:

$$\delta A = 2Cr\delta r = 2A\frac{\delta r}{r} \tag{20}$$

or

$$\frac{\delta A}{A} = 2\frac{\delta r}{r}. \tag{21}$$

Finally, the total displacement of air (equal to the integral of volume velocity over time) is equal to A times δr. Substitution in Eq. (21) gives

$$\delta A = 2\frac{\int u dt}{r}. \tag{22}$$

Taking Eqs. (15), (17), and (22) together and solving for p finally give

$$\frac{4dk_2}{r^2 A}\int u dt = p. \tag{23}$$

From this it follows that the compliance of the wall in terms of properties of the tissue is

$$C_w = \frac{r^2 A}{4dk_2}. \tag{24}$$

Following Fletcher *et al.* (2004), the quality factor Q is used to model damping. The wall resistance is calculated from the quality factor and the inertance and compliance of the wall as follows:

$$R_w = \frac{1}{Q}\sqrt{\frac{L_w}{C_w}}. \tag{25}$$

4. Radiation

For simplicity's sake it will be assumed that the air sac is a spherical radiator. This is admittedly quite unlike reality, but the relevant quantity is the amount of acoustic energy radiated per unit area, and this is not too different for different radiators (Flanagan, 1965, Fig. 3.6). The spherical radiator results in the simplest equations. Furthermore, a hemispherical radiator in an infinite plane baffle (which is more like an air sac on an ape's chest) would result in the same radiation impedance per unit area.

The acoustical impedance Z_R of a spherical radiator is (e.g., Blackstock, 2000, Sec. 10D2b)

$$Z_r = \frac{\rho_a c}{A}\frac{1}{1 + \frac{1}{ikr}}, \tag{26}$$

where k is the wave number $2\pi f/c$. This is equivalent to a resistance

$$R_r = \frac{\rho_a c}{A}, \tag{27}$$

in parallel with an inertance

$$L_r = \frac{\rho_a r}{A}. \tag{28}$$

5. The complete circuit

Four different subcircuits determine the behavior of the air sac. The circuit of the neck consists of a resistance that represents viscous losses in series with the inertance describing the mass of air in the neck. This is a series circuit, as the loss is caused by the air flow. For the wall, the form of Eq. (12) indicates a circuit of elements in series. The circuit describing the behavior of air in the cavity consists of a resistance that represents compressional losses in parallel to the compliance that describes the compression of the air in the air sac. This is a parallel circuit, as losses and compression are due to the same pressure. The radiation circuit is a parallel circuit because of the form of Eq. (26).

The circuits describing the wall motion and the outside pressure must be in series, as the volume velocity outside of the wall is equal to that inside of the wall. This circuit must be in parallel to that describing the air in the cavity, as both are subject to the same pressure (that at the end of the neck). As the air flow coming out of the neck must be equal to that compressing the air in the sac and moving the walls, the parallel circuit of cavity and wall must be in series with the circuit describing the neck. The complete circuit is illustrated in Fig. 2.

6. Parameter values

For modeling primate air sacs, it is necessary to estimate a number of parameters. Some of these, such as the properties of air, are relatively well known, and need no discussion. Estimates of Young's modulus and Poisson's ratio for tissue can be found in literature. However, these values depend very much on the kind of tissue involved. For soft tissue (e.g., Hunter and Titze, 2007) a value of around 10 kPa appears to be realistic. Given that soft tissue consists for a large part of fluids, and is therefore flexible but incompressible, Poisson's ratio must be 0.5 or slightly lower. This results in k_2 being approximately equal to, but slightly lower than Young's modulus, and therefore a value of 9000 can be used. For cartilaginous tissue, Young's modulus is much higher. For human tracheal rings, values between 2.5 and 7.7 MPa are given by Lambert *et al.* (1991), while Jin and Lewis (2003) gave values between 0.5 and 2 MPa for articular cartilage. Poisson's ratio of cartilage is assumed to be around 0.45 (Jin and Lewis, 2003, for articular cartilage) resulting in values for k_2 in the same range as that of Young's modulus.

Here, a value of 0.45 MPa was used (the lowest value in the range, selected because it best illustrates the effects of intermediate elasticity). As for bony tissue, Martin *et al.* (1998) (Table 4.1) gave a value of 17.4 GPa for Young's modulus of human cortical bone and 0.39 for Poisson's ratio. This results in a value for k_2 around 15 GPa. The values have very different orders of magnitude, and as will be shown below, result in qualitatively different behaviors of the air sac.

A density of tissue of 10^3 kg m^{-3} was assumed. Values for damping are harder to find in literature, but in terms of the quality factor they range from 0.5 (which appears to be implicit in Maeda's (1982) vocal tract model to 10 (Fletcher *et al.*, 2004). Probably for stiffer materials, the quality factor is higher. A quality factor of 1 has been used for soft tissue, assuming a relatively strong damping. For cartilage and bone a quality factor of 10 was used resulting in less damping. Finally, the radius of the air sac was chosen to be 5 cm, its shape to be spherical, and its walls to be 0.5 cm thick.

7. Air sac–oral tract interaction

The next step is to analyze the interaction between an oral tract and an air sac. Riede *et al.* (2008) also investigated this question, but did not appear to take power radiated from the air sac into account in their computational model. Also, in the present paper, a more analytical approach is chosen in order to explain the acoustic effects of air sacs.

Three different vocal tracts roughly corresponding to [a], [ə], and [y] were investigated. All were 16 cm long. The tract for [ə] consisted of a single cylinder of 1.5 cm radius. The tracts for [a] and [y] consisted of two cylinders, each of 8 cm length, and with radii of 2 and 1 cm. For [a] the wide tube was nearest to the mouth, and for [y] the narrow tube was nearest to the mouth.

The air sac is modeled as an impedance parallel to the oral tract. This is warranted, as the tube connecting the air sac to the supralaryngeal vocal tract often joins the vocal tract right above the glottis (see Fig. 1 and illustrations in, e.g., Hewitt *et al.*, 2002; Riede *et al.*, 2008; Schön Ybarra, 1986). The glottal source was assumed to be an ideal volume velocity source (which has infinite impedance). Although this is an idealization, it makes analysis much easier and is generally considered acceptable as a first approximation (Fant, 1960; Flanagan, 1965; Stevens, 1999). A better approximation would have to take into account that for maximum acoustic output efficiency, the glottal source, the subglottal system, and the supraglottal system (including air sacs) must be matched to each other (Titze, 2002, 2008).

8. Higher frequencies

At higher frequencies, the impedance of the walls increases. Therefore, the motion of the walls becomes negligible, and the difference between the different wall types disappears. It also means that radiation from the air sac becomes negligible. At higher frequencies, the air sac can therefore be analyzed as a system of rigid-walled tubes. As pressure and volume velocity can no longer be assumed constant over lengths comparable to the dimensions of the air sac, analysis in terms of a lumped element model is no longer possible. A transmission line model (such as the model used in Riede *et al.*, 2008) can be used to gain qualitative insights, but because the cross-sectional dimensions of an air sac are large compared to the wavelengths involved, the assumption of one-dimensional wave motion is not valid (the limit of 1500 Hz for the lumped element model then applies to the transmission line model) and therefore no quantitative predictions can be made.

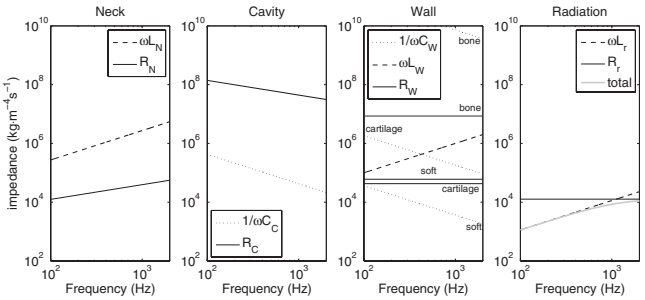


FIG. 5. Impedances for frequencies between 100 and 2000 Hz for the different elements of the lumped element model. The values of C_w (dotted line) and R_w (solid line) in the 3rd graph from left are given (from bottom to top) for soft, cartilaginous, and bony tissues.

To investigate the resonances and antiresonances at higher frequencies, finite element models were used. An axisymmetrical geometry was assumed, allowing the use of a two-dimensional mesh. All factors that were identified in the lumped element model, except viscous losses, were taken into account. Viscous losses were not modeled, as this would have considerably complicated the finite element model and their effect is small for the frequencies investigated. The finite element model approximated the Helmholtz equation in the specified geometry, implemented the walls as interface boundaries, and implemented radiation using a Bayliss–Turkel first order boundary condition (taken from Givoli, 1991). It was implemented in the GMSH/GETDP software environment (Dular *et al.*, 1998) using the Galerkin method, and was run to convergence (a relative residual value of 10^{-10} was used). For the computational verification with the finite element model presented in Sec. II B 3 a triangular mesh of 4667 elements was used.

B. Results

1. Dominant elements of the model

Given the complete circuit, it is possible to determine which elements dominate the circuit's behavior. The values of the different elements of the circuit were calculated for frequencies between 100 and 2000 Hz. The results are shown in Fig. 5. Keeping in mind that parallel impedances are dominated by the lowest impedance, and series impedances by the highest impedance, it is clear that a good estimate of the total impedance of an air sac for the relevant frequency range can be obtained by only looking at the inertance of the neck, the compliance of the cavity, and the inertance (mass) and the compliance (stiffness) of the walls. For walls of soft tissue, stiffness can be ignored, while for walls of bony tissue, mass can be ignored. For walls of cartilage, both mass and stiffness play a role. In none of the cases is damping crucial for the frequencies of the resonances (but in the case

of cartilage it is relevant for the bandwidth of the wall resonance). The case of soft tissue is in complete agreement with what Fletcher *et al.* (2004) found when studying the inflated esophagus of the ring dove. All air sac dimensions within the range that is plausible for primates give this same qualitative result.

The behavior of the air sac in the low-frequency approximation is thus found to be equivalent to that of a circuit consisting of an inertance (representing the air in the neck) in series with a parallel circuit made up of a compliance (representing the cavity air) in parallel with an inertance (representing the wall mass) and a compliance (representing wall stiffness) in series (the elements are indicated in bold in Fig. 2).

However, for low frequencies and soft walls, the amount of sound that is radiated from the air sac is non-negligible with respect to the amount of sound radiated from the mouth, and therefore radiation has to be taken into account in that case as well.

2. Air sac qualitative behavior

The whole impedance Z_{sac} of the circuit consisting of the air sac and the connecting tube is

$$Z_{\text{sac}} = i\frac{1-\omega^2 L_w C_w - \omega^2 C_w L_n - \omega^2 C_c L_n + \omega^4 L_w C_w C_c L_n}{\omega^3 L_w C_w C_c - \omega C_w - \omega C_c}. \tag{29}$$

From this equation it can be derived that the impedance has a pole (maximum) at angular frequency

$$\omega_p = \sqrt{\frac{C_w + C_c}{L_w C_c C_w}}, \tag{30}$$

which corresponds to the frequency at which the air in the air sac and the walls resonate. Equation (29) also has a pole at frequency 0, which corresponds to the fact that the air sac cannot be inflated without limits.

The air sac impedance has zeroes[1] at two angular frequencies:

$$\omega_z = \sqrt{\frac{\left(L_w C_w + C_w L_n + L_n C_c \pm \sqrt{(L_w C_w + C_w L_n + L_n C_c)^2 - 4 L_w C_w L_n C_c}\right)}{2 L_w C_w L_n C_c}}. \tag{31}$$

This is a rather complex expression, so it is instructive to first look at the limits of soft and hard walls. For soft-walled tissue (low stiffness and, therefore, infinite C_w) Equation (31) reduces to

$$\omega_z = \sqrt{\frac{L_n + L_w}{L_n L_w C_c}}. \tag{32}$$

Under the same conditions, the frequency of the pole becomes

$$\omega_p = \sqrt{\frac{1}{L_w C_c}}. \tag{33}$$

For hard-walled tissue (C_w approximately zero) Eq. (31) reduces to

$$\omega_z = \sqrt{\frac{1}{L_n C_c}} = c\sqrt{\frac{A_{\text{neck}}}{V l_e}}, \tag{34}$$

which is equal to the Helmholtz resonance frequency ω_H (as was to be expected). For the hard-walled case, the pole at nonzero frequency disappears. Equations (32)–(34) suggest a relation between the Helmholtz frequency and the frequencies of the poles and zeros of the soft-walled case

$$\omega_z^2 = \omega_H^2 + \omega_p^2, \tag{35}$$

which is equivalent to an equation given by Fant (1972) in relation to resonance frequencies of the human vocal tract when taking wall vibration into account.

Similar qualitative observations can be made about the intermediate stiffness case. Comparing the values of Eqs. (30) and (31), it is found that one of the zeros has a lower frequency than the pole, and the other zero has a higher frequency. A similar comparison shows that the zero at the highest frequency has a frequency that is higher than that of the zero for the soft-walled case, given in Eq. (32). Comparing Eqs. (30) and (33) and using the fact that all inertances and compliances are positive, it is clear that the pole frequency of the intermediate case is also higher than the pole frequency of the soft-walled case. A similar comparison shows that the zero with the lowest frequency occurs at a frequency that is lower than that of the wall resonance

$$\omega_w = \sqrt{\frac{1}{L_w C_w}}, \tag{36}$$

and finally that for the lowest-frequency zero the following condition holds:

$$\omega_{z\ \text{low}}^2 < \omega_p^2 - \omega_w^2. \tag{37}$$

The qualitative behavior of the air sac impedance at low frequencies can be summarized as follows: In the hard-walled case, there is a single zero at the Helmholtz frequency. For the soft-walled case, there is a pole followed by a zero higher than the Helmholtz frequency. For the intermediate case, there is a zero, followed by a pole, followed by a final zero. The pole and the highest-frequency zero are higher than those in the soft-walled case. The lowest-frequency zero is lower than the resonance frequency of the wall.

3. Computational results

For soft, cartilaginous, and bony air sacs with dimensions from Table I, the impedance for frequencies ranging from 20 to 4000 Hz were calculated using the lumped element model. Its absolute value is shown in Fig. 6.

The absolute values of the impedances calculated from the finite element model are given in Fig. 7. The peaks and zeros in the range 20–500 Hz correspond to the ones in Fig. 6. Also, the impedances are essentially equal to the ones in

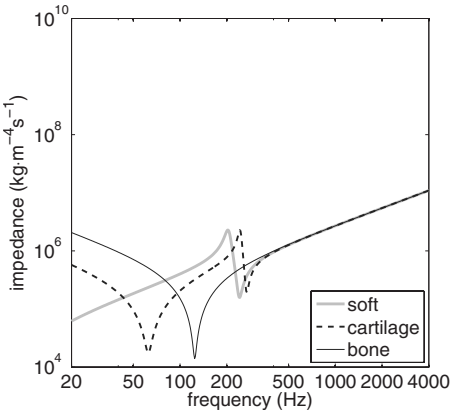


FIG. 6. Absolute value of the impedances of an example air sac of the dimensions given in Table I with soft, cartilaginous, and bony walls. Both the horizontal and vertical scales are logarithmic. Note the peaks and the zeroes.

Fig. 6 up to about 1500 Hz. At higher frequencies extra resonances and antiresonances appear. It should be noted that there is no significant difference between the different wall types for these higher resonances. The first anti-resonance/resonance pair (showing up as a peak followed by a valley around 2300 Hz) corresponds to the resonance of the cavity. The isolated antiresonance (the peak at 2700 Hz) corresponds to the first antiresonance of the neck, and the final resonance/anti-resonance pair (the valley followed by a peak around 3750 Hz) corresponds to the second resonance of the cavity. Note that due to the cavity's spherical shape, its resonances do not occur at the integer multiples expected from a straight tube.

Air sacs with the dimensions given in Table I and with the three different wall types (their impedances calculated with the finite element model) were connected to the three different vocal tracts. A comparison of the total radiated power of the tracts with and without air sacs is given in Fig. 8.

The major changes caused by the addition of an air sac to the vocal tract are that new peaks occur around the frequencies of the lower zeroes in the air sac response. The hard-walled and soft-walled sacs both add one low peak. The

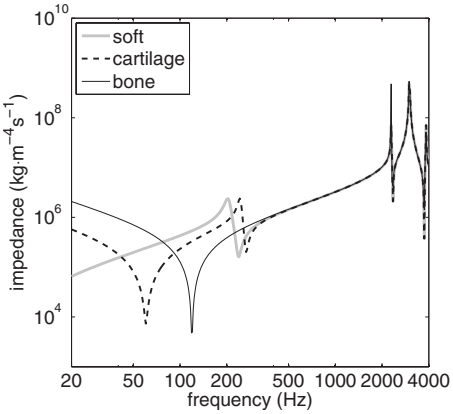


FIG. 7. Impedances of the finite element air sac models. Air sac dimensions are as given in Table I. Note the similarity for lower frequencies to Fig. 6. Also note that extra pairs of peaks and zeros, as well as extra isolated peaks, appear at higher frequencies in a very similar way for the three wall types.

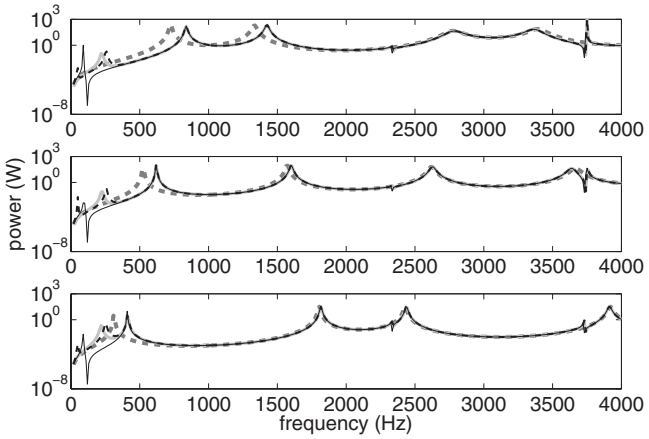


FIG. 8. Radiated power (given a constant volume velocity source, and including power radiated from the air sac and the mouth) of three different vocal tracts. From above to below, [a], [ə], and [y] are shown. Fat gray dotted lines indicate tracts without air sacs. Solid gray lines indicate tracts with soft-walled air sacs, dashed black lines tracts with cartilaginous air sacs, and solid thin black lines tracts with bony-walled air sacs. Exact dimensions are given in the text. Black lines represent tracts with air sacs, and gray lines represent tracts without air sacs.

cartilaginous-walled air sac adds two extra peaks. The frequency of these peaks is relatively constant. Peaks corresponding to the resonance frequencies of the vocal tracts without air sacs can also be found in the power spectra of the tracts with air sacs. However, these frequencies are shifted up and closer together for all tracts investigated here. Exact values for the frequencies of the peaks in the radiated power of the different tracts can be found in Table II.

Finally, peaks and zeros corresponding to those of the zeros in the air sac impedance at higher frequencies can also be found in the power spectrum of the combined tracts. The position of these peaks is relatively insensitive to the shape of the oral tracts. All the observations made here correspond qualitatively with the findings of Riede *et al.* (2008).

A major difference between the results presented here and the computational model of Riede *et al.* (2008) is in the importance of power radiated through the air sac wall. It turns out that most of the acoustic energy that is responsible for the low-frequency peaks is radiated from the air sac in the case of soft and cartilaginous walls. This is the result of the fact that at low frequencies, the mass of the air sac wall presents less of an obstacle for the acoustic vibrations, as well as from the fact that the air sac has a larger area than the opening of the mouth, and therefore functions as a much more efficient radiator at low frequencies. The ratio of power radiated through the mouth and power radiated from the air sac for the examples investigated here is presented in Fig. 9.

C. Discussion

A theoretical understanding of the observed changes in the vocal tract response can be gained by inspecting the impedances of the vocal tract without an air sac and that of the air sac. As the glottis is assumed to be a perfect volume velocity source (which has infinite impedance), the maximum energy theorem (e.g., Titze, 2002) states that most en-

TABLE II. Frequencies (in hertz) of the lower peaks of tracts with air sacs, and of the first and second resonances of tracts without and with air sacs, as well as the shifts in frequency, caused by the soft-walled air sac. Peaks caused by the air sac are indicated with F_{sac}, peaks caused by the vocal tract with F.

	F_{sac} 1	F_{sac} 2	F_1	Shift	F_2	Shift
[a]	⋯	⋯	724	⋯	1334	⋯
$[a]_{soft}$	⋯	223	838	114	1424	90
$[a]_{cart}$	47	256	838	114	1424	90
$[a]_{bone}$	89	⋯	837	113	1424	90
[ə]	⋯	⋯	524	⋯	1572	⋯
$[ə]_{soft}$	⋯	225	620	96	1601	29
$[ə]_{cart}$	50	257	620	96	1601	29
$[ə]_{bone}$	95	⋯	620	96	1601	29
[y]	⋯	⋯	312	⋯	1805	⋯
$[y]_{soft}$	⋯	220	412	100	1816	11
$[y]_{cart}$	47	252	413	101	1816	11
$[y]_{bone}$	88	⋯	410	98	1816	11

ergy is transferred when the vocal tract also has infinite impedance. When dealing with infinite impedances, it is easier to work with impedance's reciprocal, the *admittance*. Infinite impedance means zero admittance. This is achieved when the sum of the oral tract's admittance and the air sac's admittance is zero. The absence of an air sac is equivalent to zero admittance, and hence the resonances occur at the frequencies where the oral tract admittance is zero, which are just the resonances of an open/closed tube.

When the air sac has nonzero size, its impedance (and its admittance) can be calculated by Eq. (29) (ignoring radiation and losses). Because all losses are ignored, the real part of the admittance is zero. Therefore, only the imaginary part of admittance, called *susceptance,* needs to be studied. The susceptances of air sacs with the dimensions from Table I are given in Fig. 10. Also shown (gray line) is the negative of the susceptance of a straight tube [ə] with the dimensions as used above. It can be seen that the susceptance of the air sac consists of monotonically increasing curves with vertical asymptotes at the points of zero impedance. The susceptance at higher frequencies is negative, and approximately equal for all types of air sacs. The negative susceptance of the tube consists of monotonically decreasing lines, separated by vertical asymptotes where the impedance is zero. A qualitatively similar curve is obtained for differently shaped tracts, but with differently spaced asymptotes. A minimum for admittance (and therefore a maximum for impedance) for the parallel circuit is achieved where the tract curves and the air sac curves intersect. For reference, without an air sac, resonances occur where the oral tract curve intersects the zero susceptance line (dotted horizontal line in Fig. 10).

From Fig. 10 it becomes clear why the air sac changes the oral tract's response the way it does. The resonances at low frequencies occur near the intersection of the oral tract curve and the vertical asymptotes of the air sac susceptance. As the intersections' frequency is mostly determined by the position of the vertical asymptotes, it is therefore expected to be relatively independent of the oral tract shape. As for the other intersections, they are expected to occur at higher frequencies than those of the oral tract without an air sac. This is because the air sac susceptance is negative, and the slope of the oral tract susceptance curve is also negative. Because the air sac susceptance curve approaches zero for higher frequencies, it is also expected that higher-frequency resonances are shifted less than lower ones, therefore pushing them closer together.

The above analysis is only valid for low frequencies and for situations where the dimensions of the oral tract and the

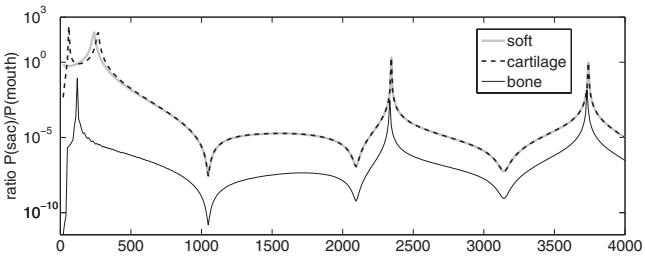


FIG. 9. Ratio of power radiated through the air sac to power radiated through the mouth for the three different wall types and for air sacs connected to the straight vocal tract. Note that at most frequencies radiation through the mouth dominates, but that at the low-frequency peaks, radiation through the air sac dominates for soft and cartilaginous walls.

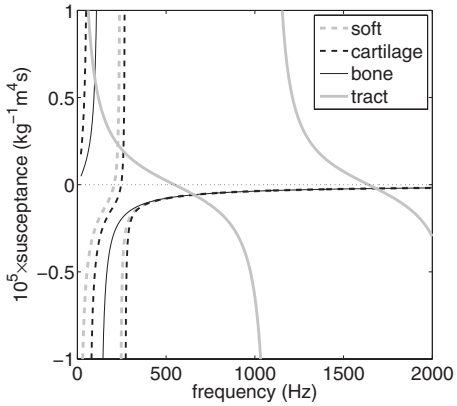


FIG. 10. Nomogram to calculate the resonances of an oral tract (a straight tube) with an air sac. The narrow gray line represents the negative of the vocal tract susceptance, the broad gray line the soft tissue air sac susceptance, the dashed black line the cartilaginous air sac susceptance, and the solid black line the bony air sac susceptance. The dotted line represents zero susceptance (absence of a parallel impedance). Resonances occur where the lines intersect. Tract and air sac dimensions are given in the text.

air sac are not such that the interactions occur at qualitatively different places. However, for all dimensions that are realistic for primates, the analysis appears to remain valid. As for higher frequencies, it can be observed in Fig. 8 that extra peaks at higher frequencies occur in the combined spectrum. These peaks correspond to resonances of the air sac, and are therefore insensitive to vocal tract shape.

III. VALIDATING THE MODEL

The simplifications made in the lumped element model appear to be valid in a mathematical sense, as it gives the same result as the finite element model. However, two things remain to be investigated: Is the model physically plausible and is it able to explain properties of primate vocalizations? For this reason two short experimental studies were conducted.

For testing the accuracy of the physics, artificial models made out of Perspex tubes were used. This was done because the dimensions and physical properties of these models could be measured exactly, and therefore the only uncertainty in these experiments was whether the modeled physics correspond to reality, and therefore a good quantitative correspondence is expected.

For testing the applicability of the theory to primate vocalizations, a first attempt was made to model two primates' vocal tracts and air sacs based on anatomical data, and to compare the calculated resonance frequencies with those measured in actual animal calls. The two species selected were the red howler monkey (*Alouatta seniculus*) and the siamang (*Symphalangus syndactylus*). Both have prominent air sacs as well as calls with prominent low-frequency components. The red howler monkey has a complex air sac system with both paired lateral ventricular and subhyoidal air sacs, and most likely has the largest hollow hyoid bone of all primates. The subhyoidal air sac is situated almost completely in the hyoid bulla, while the lateral air sacs are situated in the thyrocuneiform space. The lateral ventricular and subhyoidal air sacs appear each to be connected to the larynx separately, and therefore form three separate air sacs. The subhyoid air sac has bony walls, while the lateral ventricular air sacs has partly cartilaginous walls. The whole system is described in considerable detail by Schön (1971), who proposed that the lateral air sacs do not have an acoustic function, but serve to keep the channel into the subhyoid air sac open. The siamang has a single air sac, comprised of the fusion of the paired lateral ventricular air sacs. The diagrams in Fig. 1 are simplified representations of the red howler (subhyoidal air sac only) and siamang anatomy. As there was considerable uncertainty in the anatomical data (only data from literature were used) only a qualitative correspondence can be expected.

TABLE III. Dimensions of the Perspex models without air sacs, used for the measurements. Sections are numbered from the glottis.

Model	Length 1 (cm)	Area 1 (cm^2)	Length 2 (cm)	Area 2 (cm^2)
[a]	7.7	2.0	7.8	10.2
[ə]	15.8	5.1	⋯	⋯
[y]	7.8	10.2	7.7	2.0

A. Materials and methods

1. The Perspex models

Physical models of vocal tracts with air sacs were constructed from acrylic glass (Perspex) tubes with a wall thickness of 2 mm. Two sets of three models were constructed: one set with and one without air sacs for two-tube models corresponding roughly to the vowels [a], [ə], and [y] and having dimensions given in Tables III and IV. The density of Perspex is 1.2 kg m^{-3}, Young's modulus is approximately 3 GPa, and Poisson's ratio is 0.35. The value for k_2 was therefore estimated to be 2.3 GPa. The models are illustrated in Fig. 11(A).

The resonance frequencies of the Perspex models were measured by exciting them with a single pulse (generated by tapping the closed end). The sound was recorded with an Altec Lansing AHS30 electret microphone at 5 cm from the model's mouth and sampled at 44.1 kHz using a Soundmax Integrated Digital HD Audio PC card. A power spectrum was calculated with the PRAAT software package version 5.0.16 (Boersma and Weenink, 2008) using a 0.05 s Hann window. The position of the resonance peaks in the spectrum was calculated as the average of measurements on three different pulses.

The air sac was modeled with the complete lumped element model and with the finite element model (using a mesh of 4708 triangular elements) described in Sec. II A 8, while the vocal tracts were modeled with Flanagan's (1965) transmission line model. Radiation at the lips was also modeled after Flanagan (1965). The radiated power was calculated for frequencies between 1 and 4000 Hz, and the peaks were recorded. For these calculations, the speed of sound at

TABLE IV. Dimensions of the Perspex models with air sac attachment. Tubes are numbered from the glottis. The column labeled branch indicates the distance between the glottis and the center of the tube connecting the air sac. Due to the connection between the air sac and the vocal tracts, the neck consists of two tubes with numbers 0 and 1. The air sac body is tube number 2.

Model	Length 0 (cm)	Area 0 (cm^2)	Branch (cm)	Length 1 (cm)	Area 1 (cm^2)	Length 2 (cm)	Area 2 (cm^2)
[a]	⋯	⋯	1.2	7.8	2.0	7.8	10.2
[ə]	⋯	⋯	1.4	15.8	5.1	⋯	⋯
[y]	⋯	⋯	1.8	7.4	10.2	7.8	2.0
Air sac	1.9	1.1	⋯	2	2.0	8.7	10.2

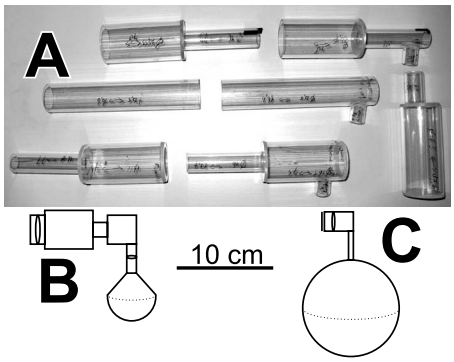


FIG. 11. Models used in the validation experiment. (A) Perspex models. The vertically displayed model is the air sac. The horizontally displayed models are the vocal tracts. Top row are the [a] models, middle the [ə] models, and bottom the [y] models. Models without air sac are on the left, and models with air sac attachment on the right. (B) Red howler monkey model (the air sac is not spherical because the ratio between area and volume does not correspond to a sphere; at the frequencies involved, exact shape does not matter, only the area and the volume are relevant). (C) Siamang model. All models are to scale. Vocal tracts are depicted horizontally and air sacs vertically. For all models, the glottis is on the right, and the mouth is on the left. Note that the siamang model is closed at the mouth.

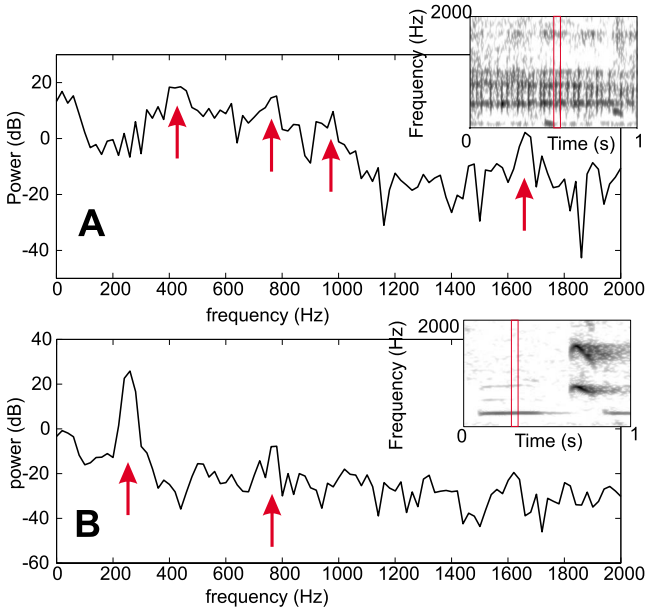


FIG. 12. (Color online) Spectra and spectrograms of (A) the red howler monkey and (B) the siamang. Resonances are indicated with arrows. It is indicated in the spectrograms from which part of the signal the spectrum was calculated. Note that in the paper, only the siamang boom is investigated, while in the spectrogram, also the high pitched part of the call (the "shriek," not analyzed here) is shown (to the right). Also note that there is considerable variation in the howler monkey call (especially in the presence or absence of the resonance around 1700 Hz), while only one part of its call is modeled.

room temperature (345 m s^{-1}) was used rather than the speed of sound at body temperature as given in Table I.

2. The animal models

The model of the red howler monkey was constructed using anatomical data (Kelemen and Sade, 1960; Schön, 1971; Schön Ybarra, 1986) and an estimated vocal tract length of 11 cm. The model consisted of a four-tube vocal tract with an attached simple air sac. The air sac does not model the complex anatomy of the howler monkey air sac system, but consists of a hard-walled resonator of 5 cm diameter, 40 cm^3 volume, and 40 cm^2 (radiating) surface area, connected to the glottis with a 3 cm tube of 1 cm diameter. The howler monkey's lateral ventricular sacs were not modeled in this first approximation, as Schön, 1971 proposed that the subhyoidal air sac is the main resonator (Schön, 1971, in the discussion). The oral tract consists of a 3 cm long tube of 3 cm diameter modeling the chamber above the glottis, a 1.5 cm long tube of 2 cm diameter modeling the constriction near the laryngeal introitus (No. 7 in Schön Ybarra 1986, Fig. 5), a 5 cm long tube of 4 cm diameter modeling the oral cavity, and a 1.5 cm long tube of 3 cm diameter modeling the rounded lips. This model is based on rough estimates, but it is meant to show a qualitative, rather than a complete quantitative resemblance to the howler monkey's bark. The model is illustrated in Fig. 11(B).

Schön Ybarra (1986) presented an analysis of calls from the red howler monkey. In that study, different types of calls (called barks and roars) occur that have resonances at around 350–400, 900–1100, and 1800–2200 Hz, although not all resonances appear in all calls. Some calls lack either or both peaks around 1000 Hz and 1800 Hz, while others have all three. Whitehead (1995) presented resonance patterns for roars that have been derived from a "multi-animal chorus" (multiple animals vocalizing during the same time period) with comparable values as the ones found by Schön Ybarra (1986), and Whitehead (1995) also mentioned that he did not consistently find the third peak around 2000 Hz. An example of a howler monkey's spectrum and spectrogram is illustrated in Fig. 12(A) (based on Davis, 1981, digitized at 44.1 kHz); the spectrum was calculated from a 2205-sample Hann window. This figure shows the same qualitative picture as found by Schön Ybarra (1986), and Whitehead (1995), except that it appears that there are *two* resonances in the range 700–1000 Hz. In Whitehead's (1995) spectrogram (Fig. 7), the peak appears wide, but one cannot clearly distinguish two peaks. However, an interpretation in terms of two resonances seems possible.

A second model was constructed to model the air sac of the siamang (*Symphalangus syndactylus*). In this case a larger air sac, of the dimensions given in Table I, was used. Given the near absence of higher harmonics in the boom call, the oral part of the vocal tract was assumed to be closed (no radiation). However, a short laryngeal-pharyngeal section of 3 cm length and 2 cm diameter was assumed (lacking precise anatomical data about the siamang, these dimensions were based on the *Hylobates lar* larynx shown in Nishimura, 2003). Assuming a closed vocal tract also means that the air sac inflates or deflates during vocalization (depending on whether it occurs on exhalation or inhalation) and that its acoustic properties change. Therefore the spectrum presented here should be seen as a single time slice of the vocalization. The model is illustrated in Fig. 11(C).

Siamang booms (the low-frequency parts of the call, produced with the air sac inflated, Haimoff, 1981) have a clear peak around 250–300 Hz (Haimoff, 1981; Riede *et al.*, 2008, Fig. 10B), illustrated in Fig. 12(B) (based on Priori, 1973, digitized at 44.1 kHz); the spectrum was calculated

TABLE V. Values for resonance frequencies of the Perspex models as calculated by the lumped element model, the finite element model, and as measured from the real models. For models without air sacs, only the transmission line model is used, so only one calculated value and the measured value are given. For models with air sacs, the values are given in the order lumped element model/finite element model/measured value.

	F_{sac}	F_1	F_2
[a] no sac		788/–/740	1366/–/1307
[a] with sac	222/214/203	888/866/891	1630/1478/1429
[ə] no sac	⋯	525/–/522	1577/–/1580
[ə] with sac	233/224/218	729/681/690	1579/1536/1601
[y] no sac	⋯	281/–/284	1833/–/1804
[y] with sac	179/176/170	567/511/499	1834/1815/1833

from a 2205-sample Hann window. In this recording, the frequency of the boom increases from about 245 Hz at the beginning to 260 Hz at the end, indicating that the air sac deflates during vocalization. The vocal folds probably vibrate close to the frequency of this first resonance, as second and third harmonics are visible in the spectrum and spectrogram. As the third harmonic is more prominent than the second, there must be a supraglottal resonance at around 800–900 Hz.

The vocal tract was modeled with Flanagan's (1965) transmission line model, and the air sac was modeled with the complete lumped element model. The radiated power was calculated for frequencies between 1 and 4000 Hz, and the peaks were recorded.

B. Results

1. The Perspex models

The measurements and the values predicted by the computer model are given in Table V and illustrated in Fig. 13. It can be observed that the differences between calculated and measured resonances are small and that the lumped element model gives qualitatively similar, but slightly less accurate predictions than the finite element model. The differences between the finite element model and the measured values are on average 2.9% with a maximum of 5.3% while for the lumped element model, the average is 6.3% and the maximum 14.1%. Variation between different measurements of the recorded signal was approximately 2%.

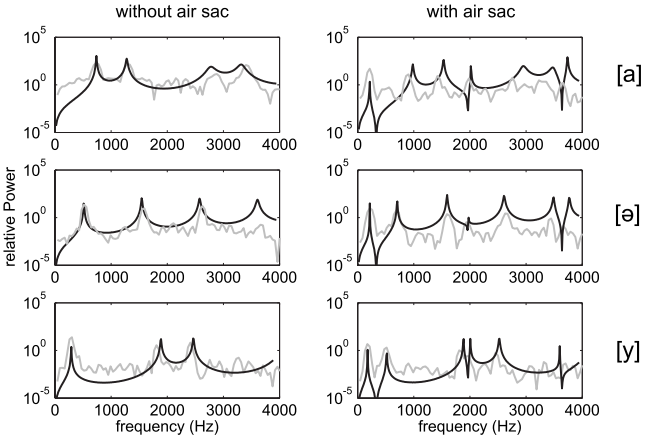


FIG. 13. Comparison of calculated (solid lines) and measured (gray lines) spectra of the Perspex models without (left column) and with air sacs (right column).

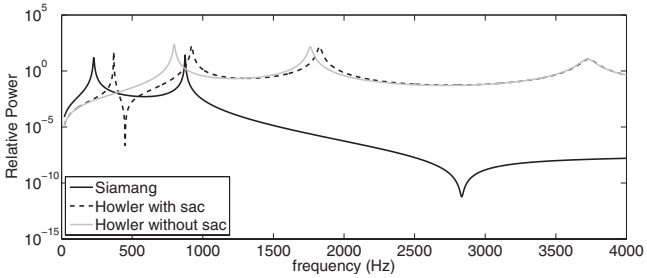


FIG. 14. Radiated power of the "howler monkey" (dashed line) and the "siamang" (solid line) models. For reference, the radiated power of a howler vocal tract without an air sac is also given (gray line). As in the siamang radiation was exclusively through the air sac, only radiated power from the vocal tract with air sac can be depicted. Details about the models can be found in the text.

The results conform to the theory: A new resonance at approximately the resonance frequency of the air sac appears and the frequency of this resonance is insensitive to the shape of the vocal tract. Simultaneously, the first two resonance frequencies of the original vocal tract are shifted up and closer together.

2. The animal models

The calculated spectrum of the howler monkey model with air sac is shown in Fig. 14. Resonances occur at 369, 921, and 1827 Hz. For reference, the spectrum of the Howler monkey vocal tract without an air sac is also given, and its resonances occur at 798 and 1761 Hz. It can be observed that there are no prominent higher resonances below 3000 Hz, just as is the case in Schön Ybarra's sonogram (Schön Ybarra, 1986, Fig. 1). There is also a reasonable qualitative similarity between the curve in Fig. 14 and the spectrogram shown in Whitehead, 1995 and in Fig. 12(A). With respect to the data shown in Fig. 12(A), it is important to note that in the real howler monkey call there are two peaks in the range 700–1000 Hz instead of one.

The calculated radiated power of the siamang model is given as the solid line in Fig. 14. Two prominent peaks can be observed at 228 and 875 Hz, which correspond qualitatively to the siamang's boom, although in the siamang [Fig. 12(B)] the resonance near 800 Hz is much less prominent. The first resonance is due to the air sac, and the second to the laryngopharyngeal cavity. The small peak around 500 Hz (best visible in the spectrogram) corresponds to the second harmonic of vocal fold vibration (which if tuned at the resonance of the air sac would be around 228 Hz). As the model does not take source characteristics into account, no corresponding peak is expected in the modeled spectrum.

It should be noted that all acoustic energies in the siamang model are radiated from the air sac. For the low frequencies involved, the air sac is a much better radiator than the open mouth, as it has a larger surface area. Its function as a radiator might explain the absence of hair on the skin covering the siamang's air sac.

C. Discussion

In the best controlled case, that of the Perspex models, the correspondence between the resonance frequencies calculated by the finite element model and those observed, is within 5.3%, which is comparable with the uncertainties with which the resonances and the dimensions of the models could be measured. This confirms that the underlying physical analysis of air sac behavior is sound. The results also confirm the qualitative analysis of the lumped element model: An extra peak appears at a frequency that is close to the resonance frequency of the air sac itself, while the original resonances of the vocal tract to which the air sac is attached are shifted up and closer together.

As for the animal models, the uncertainties here are much bigger, as only limited anatomical data were available. The correspondence between the howler monkey's call and the calculated spectrum is incomplete. The three predicted resonances occur at approximately the right frequencies, but it appears as if the real howler monkey spectrum has an extra resonance in the 700–1000 Hz range. It is unlikely that this is due to the shape of the vocal tract that was modeled. A different shape of the vocal tract could result in a different pattern of resonances, but the shape of the vocal tract investigated here already results in resonances that have low frequencies for a vocal tract with a length of 11 cm. Shapes with as many low resonances as observed would require a vocal tract area function with unrealistically sharp transitions. It is more likely that the extra resonance is caused by the lateral air sacs of the howler monkey. These structures are not taken into account in the present model, but might produce extra resonances in the correct frequency range [given a very rough size estimate derived from Schön (1971) Figure 6]. It therefore appears that Schön Ybarra's (Schön, 1971, in discussion) remark that the "…bulla hyoidea can act as the main resonator..." is mostly correct, but that the lateral sacs may cause an extra resonance. This should be investigated with a more complete model, but that requires accurate anatomical data about the howler monkey's lateral air sacs. Using accurate anatomical data is critical, as the match between the animal's resonance spectrum and the model's resonances can always be improved by tuning of parameters, such as the dimensions of the vocal tract and the air sac. Such parameter tuning should be avoided, however, as one can easily fool oneself by constructing a model that shows similar resonances as the animal model, but for totally different acoustic reasons. A more anatomically correct model should also be able to explain how the variations between the different howler calls can be made, potentially by attaching and detaching the different air sacs, and by opening and closing the vocal tract.

The qualitative correspondence between the predicted and observed siamang spectra is good. The only important difference is that in the real siamang spectrum, the second resonance is much less prominent. This could be due to coupling between the source of vocal energy and the air sac, which causes the source to contain much less energy at the second resonance, but this should be investigated further with more accurate models.

IV. CONCLUSION

The acoustic effect of air sacs with (to first order) constant volume has been investigated in this paper. It was shown that only the inertance of the air in the neck of the air sac, the volume of air in the sac itself, and the mass and stiffness of the wall of the air sac have significant influence on the resonance frequencies. Whether mass or stiffness dominates depends on the air sac's wall tissue. For soft tissue, mass dominates. For bony tissue, stiffness dominates. For cartilaginous tissue both mass and stiffness need to be taken into account. It was also shown that for large air sacs with flexible walls, radiation of acoustic power through the walls makes a significant contribution to the total sound produced at low frequencies.

As for the spectrum of the sound produced by a vocal tract with an attached air sac, it has been shown that the presence of a large air sac results in extra resonances at the approximate resonance frequencies of the isolated air sac. Attaching an air sac to a given oral tract also results in a shifting up of the original resonances (at least for resonances lower than the second resonance frequency of the air sac). As lower resonances are shifted up more than higher resonances, this process results in resonances moving closer together. These findings correspond to the experimental and modeling results of Riede *et al.* (2008).

The theoretical model was able to predict the resonances of a Perspex model accurately. Moreover, models based on the anatomy of a red howler monkey and of a siamang resulted in resonance patterns that were qualitatively (and to some degree quantitatively) similar to those of certain vocalizations produced by these species. The major resonances of the siamang "boom" could be modeled correctly, whereas three of the four major resonances of the red howler monkey could be modeled correctly by only modeling its bony-walled subhyoidal air sac. It was proposed that for a complete howler monkey model, the lateral air sacs also need to be taken into account. The model developed in this paper can be used for the case of multiple air sacs, by adding for each air sac a lumped element circuit as an extra parallel impedance. The results support the acoustical model developed in this paper, as well as the hypothesis that air sacs have an acoustic function in howler monkeys and siamangs.

Only the properties of the air sac as an acoustic filter have been investigated, while it was assumed that the source of vocal energy is a perfect volume velocity source. This simplification was made in order to enable theoretical analysis. However, interaction between the vocal folds and the vocal tract is probably very important in primate calls. Riede *et al.* (2008) found such interactions between the air sac and the vocal folds in an artificial model. Even in the absence of source-filter interaction, it would make a large difference in radiated power at what exact frequency the vocal folds vibrate. Tuning the fundamental frequency to fall near an antiresonance of the air sac might, on the contrary, help to suppress its first harmonic. This would concentrate acoustic output at higher frequencies, something which has been proposed for the high-frequency calls of siamangs by Haimoff (1983).

Finally, it is tempting to speculate about the influence that air sacs would have on speech, and their possible relevance for the study of the evolution of speech. The presence of an air sac results in a resonance at low frequencies, whose frequency is insensitive to the shape of the oral tract. In addition, higher resonances get shifted up and closer together. As distinctions between speech sounds are perceived primarily on the basis of the differences between their resonance patterns, this means that a smaller range of distinctive speech sounds could be produced when an air sac is present. This could be investigated by determining the maximal vowel space (Boë *et al.*, 1989) with and without an air sac, in a way similar to what has been done by Boë *et al.* (2002) and de Boer (2008b).

In this context, it would be interesting to investigate the perceptual effect of connecting an air sac to a vocal tract. This could be done with psychoperceptual experiments, in which the ability to distinguish between different vocal tract shapes with and without air sacs is investigated. Such investigations, however, fall outside the scope of this paper. A first impression of the Perspex models with and without air sacs is that of a clear decrease in perceived pitch when an air sac is attached to a vocal tract, while the different vocal tracts with an air sac sound more similar than those without air sacs.

A potential effect of the presence of the lower resonance is that it may help to exaggerate size. Without air sacs, the resonance patterns produced by the models with air sacs could only be produced by much longer vocal tracts. This can be an evolutionary advantage (Fitch, 2000; Fitch and Hauser, 1995, 2002). However, when being able to produce a wide range of distinctive resonance patterns becomes evolutionarily important, such as must have happened in the evolution of speech, air sacs become a disadvantage. Even if it is possible that air sacs can be disconnected from the vocal tract (which is likely) they can nevertheless be a source of infection, called airsacculitis (Lawson *et al.*, 2006, and references therein). It therefore makes sense that they would disappear if they are no longer useful in vocalization.

Although the model has been developed for understanding the function of air sacs in primates and ancestral hominins, there is no reason that the model could not be applied to other animals with air sacs, as long as their structure consists of a neck (a tube) connected to a cavity. This appears to be the case in many other animals. Frey *et al.* (2007) studied air sacs in reindeer and gave an overview of many mammalian species, while, for example, Reidenberg and Laitman (2008) gave an overview of cetacean air sacs and Weissengruber *et al.* (2001) gave an overview of epipharyngeal pouches (equivalent to air sacs) in bears. The model should be applicable to understanding these animals' vocalizations.

In any case a better understanding of the acoustic function of air sacs, through analysis of simplified models, through extension of the theory to include source-filter interactions and through experiments such as presented here and in Riede *et al.*, 2008 will help to advance the understanding of the evolutionary and behavioral functions of these fascinating anatomical structures.

ACKNOWLEDGMENTS

This work is part of the NWO vidi project "Modeling the evolution of speech" Grant No. 016.074.324. The author thanks Didier Demolin for discussion on the function and anatomy of air sacs, Rob van Son for comments on the manuscript, and Wendy van Bohemen for letting the author use the anatomical collection of the Amsterdam Zoological Museum.

[1]In the context of impedance "a zero" is a frequency at which the impedance is zero. It should not be confused with the *number* zero.

Alemseged, Z., Spoor, F., Kimbel, W. H., Bobe, R., Geraads, D., Reed, D., and Wynn, J. G., (**2006**). "A juvenile early hominin skeleton from Dikika, Ethiopia," Nature (London) **443**, 296–301..

Arensburg, B., Tillier, A. M., Vandermeersch, B., Duday, H., Schepartz, L. A., and Rak, Y. (**1989**). "A middle Palaeolithic human hyoid bone," Nature (London) **338**, 758–760..

Avril, C. (**1963**). "Kehlkopf und kehlsack des schimpansen, *Pan troglodythes* (Blumenbach 1799) (mamalia, primates, pongidae) (Larynx and air sac of the chimpanzee, *Pan troglodytes* (Blumenbach 1799) (mamalia, primates, pongidae))," Gegenbaurs Morphol. Jahrb **105**, 75–129.

Blackstock, D. T. (**2000**). *Fundamentals of Physical Acoustics* (Wiley, New York).

Boë, L.-J., Heim, J.-L., Honda, K., and Maeda, S. (**2002**). "The potential Neandertal vowel space was as large as that of modern humans," J. Phonetics **30**, 465–484..

Boë, L.-J., Perrier, P., Guerin, B., and Schwartz, J.-L. (**1989**). "Maximal vowel space," paper presented at the Eurospeech, Paris, France.

Boersma, P., and Weenink, D. (**2008**). PRAAT: Doing phonetics by computer (version 5.0.16), Amsterdam.

Brandes, R. (**1932**). "Über den kehlkopf des orang-utan in verschiedenen altersstadien mit berücksichtigung des kehlsackfrage (About the larynx of the orangutan in different stages of development, with discussion of the matter of air sacs)," Gegenbaurs Morphol. Jahrb **69**, 1–61.

Camper, P. (**1779**). "Account of the ogans of speech of the orang," Philos. Trans. R. Soc. London **69**, 139–159..

Dang, J., and Honda, K. (**1997**). "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. **101**, 456–465..

Dang, J., Honda, K., and Suzuki, H. (**1994**). "Morphological and acoustical analysis of the nasal and paranasal cavities," J. Acoust. Soc. Am. **96**, 2088–2100..

Davis, T. H. (**1981**). Macaulay Library Recording No. 25513, Suriname.

de Boer, B. (**2008a**). "The acoustic role of supralaryngeal air sacs," J. Acoust. Soc. Am. **123**, 3732–3733..

de Boer, B. (**2008b**). "The joy of sacs," *The Evolution of Language*, edited by A. D. M. Smith, K. Smith, and R. Ferrer i Cancho (World Scientific, Singapore), pp. 415–416.

Dular, P., Geuzaine, C., Henrotte, F., and Legros, W. (**1998**). "A general environment for the treatment of discrete problems and its application to the finite element method," IEEE Trans. Magn. **34**, 3395–3398..

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, 's-Gravenhage).

Fant, G. (**1972**). "Vocal tract wall effects, losses, and resonance bandwidths," Speech Transm. Lab. Q. Prog. Status Rep. **13**, 28–52.

Fitch, W. T. (**2000**). "The evolution of speech: A comparative review," Trends Cogn. Sci. **4**, 258–267..

Fitch, W. T., and Hauser, M. D. (**1995**). "Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on 'honest' advertisement," Am. J. Primatol **37**, 191–219..

Fitch, W. T., and Hauser, M. D. (**2002**). "Unpacking 'honesty': Vertebrate vocal production and the evolution of acoustic signals," in *Acoustic Communication*, edited by A. M. Simmons, R. R. Fay, and A. N. Popper (Springer, New York), pp. 65–137.

Flanagan, J. L. (**1965**). *Speech Analysis, Synthesis and Perception* (Springer, Berlin).

Fletcher, N. H., Riede, T., Beckers, G. J. L., and Suthers, R. A. (**2004**). "Vocal tract filtering and the 'coo' of doves," J. Acoust. Soc. Am. **116**, 3750–3756..

Frey, R., Gebler, A., Fritsch, G., Nygrén, K., and Weissengruber, G. E. (**2007**). "Nordic rattle: The hoarse vocalization and the inflatable laryngeal

air sac of reindeer (*Rangifer tarandus*)," J. Anat. **210**, 131–159..
Gautier, J.-P. (**1971**). "Etude morphologique et fonctionelle des annexes extra-laryngées des Cercopithecinae; liaison avec les cris d'espacement (Morphological and functional study of the extralaryngeal sacs of the Cercopithecines; relation to separation calls)," Biologia Gabonica **7**, 229–267.
Givoli, D. (**1991**). "Non-reflecting boundary conditions," J. Comput. Phys. **94**, 1–29..
Haimoff, E. F. (**1981**). "Video analysis of siamang (*Hylobates syndactylus*) songs," Behaviour **76**, 128–151..
Haimoff, E. F. (**1983**). "Occurrence of antiresonance in the song of siamang (*Hylobates syndactylus*)," Am. J. Primatol **5**, 249–256..
Hewitt, G. P., MacLarnon, A., and Jones, K. E. (**2002**). "The functions of laryngeal air sacs in primates: A new hypothesis," Folia Primatol. **73**, 70–94..
Hilloowala, R. A., and Lass, N. J. (**1978**). "Spectrographic analysis of laryngeal air sac resonance in rhesus monkey," Am. J. Phys. Anthropol. **49**, 129–132..
Hunter, E. J., and Titze, I. R. (**2007**). "Refinements in modeling the passive properties of laryngeal soft tissue," J. Appl. Physiol. **103**, 206–219..
Jin, H., and Lewis, J. L. (**2003**). "Determination of Poisson's ratio of articular cartilage in indentation test using different sized indenters," paper presented at the Summer Bioengineering Conference, Key Biscayne, FL.
Kelemen, G., and Sade, J. (**1960**). "The vocal organ of the howling monkey (*Alouatta palliata*)," J. Morphol. **107**, 123–140..
Kleinschmidt, A. (**1938**). "Die schlund und kehlorgane des gorillas "Bobby" unter besonderer berücksichtigung der gleichen organe von mensch und orang. Ein beitrag zur vergleichenden anatomie des kehlkopfes (The pharyngeal and laryngeal organs of the gorilla "Bobby" with special discussion of the equivalent organs of humans and orangs. A contribution to the comparative anatomy of the larynx.)," Gegenbaurs Morphol. Jahrb **81**, 78–157.
Lambert, R. K., Baile, E. M., Moreno, R., Bert, J., and Paré, P. D. (**1991**). "A method for estimating the Young's modulus of complete tracheal cartilage rings," J. Appl. Physiol. **70**, 1152–1159.
Lawson, B., Garriga, R., and Galdikas, B. M. F. (**2006**). "Airsacculitis in fourteen juvenile southern bornean orangutans (*Pongo pygmaeus wurmbii*)," J. Med. Primatol **35**, 149–154..
Maeda, S. (**1982**). "A digital simulation method of the vocal-tract system," Speech Commun. **1**, 199–229..
Martin, R. B., Burr, D. B., and Sharkey, N. A. (**1998**). *Skeletal Tissue Mechanics* (Springer-Verlag, New York).
Martínez, I., Arsuaga, J.-L., Quam, R., Carretero, J.-M., Gracia, A., and Rodríguez, L. (**2008**). "Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain)," J. Hum. Evol. **54**, 118–124..
Negus, V. E. (**1949**). *The Comparative Anatomy and Physiology of the Larynx* (William Heinemann Medical Books Ltd., London).
Nishimura, T. (**2003**). "Comparative morphology of the hyo-laryngeal complex in anthropoids: Two steps in the evolution of the descent of the larynx," Primates Med. **44**, 41–49.
Nishimura, T. (**2008**). "Origin of human speech an primate vocalizations: Paleoanthropology and bioacoustics," Anthropol. Sci. (Japanese Series) **116**, 1–14..
Nishimura, T., Mikami, A., Suzuki, J., and Matsuzawa, T. (**2007**). "Development of the laryngeal air sac in chimpanzees," Int. J. Primatol. **28**, 483–492..
Priori, A. L. (**1973**). Macaulay Library Recording No. 126331, Washington Zoo.
Reidenberg, J. S., and Laitman, J. T. (**2008**). "Sisters of the sinuses: Cetacean air sacs," Anat. Rec. **291**, 1389–1396..
Riede, T., Tokuda, I. T., Munger, J. B., and Thomson, S. L. (**2008**). "Mammalian laryngseal air sacs add variability to the vocal tract impedance: Physical and computational modeling," J. Acoust. Soc. Am. **124**, 634–647..
Schön, M. A. (**1971**). "The anatomy of the resonating mechanism in howling monkeys," Folia Primatol. **15**, 117–132..
Schön Ybarra, M. A. (**1986**). "Loud calls of adult male red howling monkeys (*Alouatta seniculus*)," Folia Primatol. **47**, 204–216..
Stevens, K. N. (**1999**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Titze, I. R. (**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. R. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749..
Weissengruber, G. E., Forstenpointner, G., Kübber-Heiss, A., Riedelberger, K., Schwammer, H., and Ganzberger, K. (**2001**). "Occurrence and structure of epipharyngeal pouches in bears (Ursidae)," J. Anat. **198**, 309–314..
Whitehead, J. M. (**1995**). "Vox alouattinae: A preliminary survey of the acoustic characteristics of long-distance calls of howling monkeys," Int. J. Primatol. **16**, 121–144..

Effect of ultrasound contrast agent dose on the duration of focused-ultrasound-induced blood-brain barrier disruption

Feng-Yi Yang[a)]
Department of Biomedical Imaging and Radiological Sciences, School of Biomedical Science and Engineering, National Yang-Ming University, 11221 Taipei, Taiwan and Division of Medical Engineering Research, National Health Research Institute, 35053 Zhunan, Miaoli, Taiwan

Shing-Hwa Liu
Institute of Toxicology, College of Medicine, National Taiwan University, 10051 Taipei, Taiwan

Feng-Ming Ho
Department of Internal Medicine, Tao-Yuan General Hospital, 33004 Taoyuan, Taiwan

Chi-Hong Chang
Department of Biomedical Imaging and Radiological Sciences, School of Biomedical Science and Engineering, National Yang-Ming University, 11221 Taipei, Taiwan

(Received 2 June 2009; revised 9 September 2009; accepted 9 September 2009)

It has been shown that focused ultrasound (FUS) is capable of noninvasive and reversible disruption of the blood-brain barrier (BBB) at target regions when applied in the presence of ultrasound contrast agent (UCA). The purpose of this study was to investigate the effects of UCA dose on the reversibility of BBB disruption induced by the same acoustical power of FUS. Sonications were applied at an ultrasound frequency of 1 MHz with a 5% duty cycle and a repetition frequency of 1 Hz. The brains of 66 male Sprague-Dawley rats were subjected to sonications with three doses of UCA. BBB integrity was evaluated via femoral vein injection of Evans Blue (EB) while the rats were anesthetized. The relationship between UCA dose and the region of EB extravasation was evaluated at an acoustic power of 1.43 W. The BBB disruption, as indexed by the amount of EB extravasation, was the largest immediately after the sonications. The quantity of Evans blue extravasation decreased as a function of time at various UCA doses. This study demonstrates that the appropriate dose of UCA not only enhance the BBB opening but also effectively aid in controlling the duration of BBB disruption. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3242376]

PACS number(s): 43.80.Sh [CCC] Pages: 3344–3349

I. INTRODUCTION

The central nervous system (CNS) is protected from the entry of foreign substances by the almost impenetrable blood-brain barrier (BBB). The BBB is a specialized vascular system consisting of endothelial cell tight junctions, basal lamina, and glial processes.[1] These structures characterize the BBB with low permeability to ionized water-soluble molecules with a molecular mass greater than 180 Da, hence chemotherapeutic agents with molecular masses between 200 and 1200 Da or antibody-based drugs with a molecular size of 150 kDa are easily impeded by the BBB.[2,3]

If drugs could be delivered to the CNS directly, the interaction would be highly target-specific, and the therapeutic effect could be dramatically improved. Many methods have been developed to overcome BBB impermeability when delivering drugs, such as increasing their lipid solubility, making them water-soluble with high affinities for carriers at the BBB, and by using vectors such as amino acids and peptide carriers.[4–6] The drawback of these techniques is that they promote the permeability of drugs throughout the brain.

Magnetic resonance imaging (MRI)-guided focused ultrasound (FUS) has been used to locally and reversibly increase the permeability of the BBB, and the changes in BBB permeability could be appropriately quantified by contrast-enhanced MRI.[7–10] Previously, Raymond *et al.*[11] showed ultrasound/mircobubble-induced vasospasm at low-power amplitudes in the brain. One study demonstrated that ultrasound can enhance the delivery of molecular imaging agents and large biological immunotherapeutics in transgenic mouse models of Alzheimer's disease.[12] The FUS-induced BBB disruption was affected by several acoustic parameters, including the applied pressure amplitude, operating frequency, burst length, and duty cycle of ultrasound transducer, the concentration, and bubble size of ultrasound contrast agent (UCA).

Some experimental evidences have shown that the normalized signal intensity of MRI in the focal volume is proportional to the applied pressure amplitude in the presence of MRI contrast agent.[10] Furthermore, the enhancement of MRI signal intensity due to BBB disruption is the greatest immediately after the sonications and decreased as a function of

[a)]Author to whom correspondence should be addressed. Electronic mail: fyyang@ym.edu.tw

time. 3 h following the sonication, the signal intensity increase was only approximately 10%–20% of that measured initially.[7,10] Another study[13] demonstrated that BBB integrity appeared to be re-established by 96 h because femoral vein injection of Evans blue (EB) at this time led to no tissue staining. Previously, we have indicated that injecting an appropriate quantity of UCA effectively increases and localizes the FUS-induced BBB disruption.[14,15] The UCA doses, duty cycle, and sonication time were all greater in the transcranial study than in the craniotomy study so as to compensate for the attenuation of acoustic propagating waves at the skull due to scattering and absorption. However, the effects of acoustic parameters and UCA dose on the duration of FUS-induced BBB disruption are still unknown. In fact, BBB disruption is still considered somewhat risky, particularly for therapies required regularly over a long term. Therefore, the duration of BBB disruption induced by FUS is one major concern if this technology is practical tool for clinical application.

The purpose of this study was to investigate the effects of UCA dose on the duration of BBB disruption following the ultrasound sonication. The reversibility of BBB disruption with UCA at three doses were quantitatively analyzed according to EB extravasation and evaluated by MRI at the FUS-targeted sites.

II. MATERIALS AND METHODS

A. Animals

A total of 66 male Sprague-Dawley rats weighing from 280 to 350 g were used in the experiments. All the procedures of the animal experiment adhered to the Guidelines for Care and Use of Experimental Animals by our institutional animal committee. Before the FUS sonications and MR imaging, each rat was anesthetized intraperitoneally with chloral hydrate (400 mg/kg), and the body temperature was maintained at 37 °C using a heating blanket. The scalp overlying the skull was incised to facilitate using the bregma of the rat skull as an anatomic landmark for targeting purposes. In the sonication experiments, FUS was applied to all rats in the presence of UCA at three doses (150, 300, and 450 μl/kg).

B. Ultrasound equipment

The pulsed FUS was generated by a 1 MHz single-element focused transducer (A392S, Panametrics, Waltham, MA) with a diameter of 38 mm and a radius of curvature of 63.5 mm. The half-maximum of the pressure amplitude of the focal zone had a diameter and length of 3 and 26 mm, respectively. The acoustic-power output and the focal-pressure amplitude were measured as described earlier.[15] The transducer was mounted on a removable cone filled with de-ionized and degassed water whose tip was capped by a polyurethane membrane, and the center of the focal spot was at approximately 5.7 mm below the cone tip. The FUS was precisely targeted using a stereotaxic apparatus (Stoelting, Wood Dale, IL) that utilized the bregma of the skull as the anatomical landmark. A function generator (33220A, Agilent Inc., Palo Alto, CA) was connected to a power amplifier (500-009, Advanced Surgical Systems, Tucson, AZ) to drive the FUS transducer and a power meter/sensor module (Bird 4421, Cleveland, OH) was used to measure the input electrical power. The animal positioning for the sonication arrangement was similar to our previous studies.[14,15] The rat was laid prone beneath the cone tip and ultrasound transmission gel (Pharmaceutical Innovations, Newark, NJ) was used to maximize the transmission of ultrasound between the transducer and the rat's brain.

C. Sonications

The rat's head was mounted on the stereotaxic apparatus with the nose bar positioned 3.3 mm below the interaural line. Sonications were pulsed with a burst length of 50 ms at a 5% duty cycle and a repetition frequency of 1 Hz. Each sonication protocol lasted 60 s. The FUS was delivered to one location in the right hemisphere brain at the position of 3.5 mm posterior and 2.5 mm lateral to the bregma, and 5.7 mm below the skull surface. The UCA (SonoVue, Bracco International, Amsterdam, The Netherlands) was injected into the femoral vein of the rats about 15 s before each sonication. It contains phospholipid-coated microbubbles with mean diameter = 2.5 μm and concentration = $(1–5)\times10^8$ bubbles/ml. In *in vivo* experiments, the rats were sonicated with UCA at three doses (150, 300, and 450 μl/kg) at an acoustic power of 1.43 W.

D. Assessment of blood-brain barrier disruption

It has been shown that BBB disruption could be quantified based on the extravasation of EB, which binds with albumin for the evaluation of vascular permeability.[14,15] Therefore, we tried to use EB to evaluate the relationship between duration of BBB disruption and UCA dose. The rats were injected intravenously by EB (Sigma, St. Louis, MO) with concentration of 100 mg/kg at desired time points (0, 15, 30, and 45 min or 1, 2, and 4 h) after FUS application. The animals were sacrificed approximately 4 h after the EB injection. Rats were perfused with saline via the left ventricle until colorless perfusion fluid appeared from the right atrium. After perfusion and brain removal, the brain was sectioned into six slices from 6 mm anterior to 6 mm posterior to the bregma and mounted on glass slides. The EB extravasation was visible to the unaided eye. The coronal sections were then divided into right and left hemispheres before measuring the amount of EB extravasated. The unsonicated left hemispheres were the control groups. Samples were weighed and then soaked in 50% trichloroacetic acid solution. After homogenization and centrifugation, the extracted dye was diluted with ethanol (1:3), and its amount was measured using a spectrophotometer (PowerWave 340, BioTek, Winooski, VT) at 620 nm.[16] The content of tissue's EB was quantified by a linear regression standard curve derived from seven concentrations of the dye and was denoted per gram of tissue. Results are typically expressed as means ± SEM. Differences among EB's concentrations were analyzed with a *t* test. Statistical significance was defined as p values $\leqslant 0.05$.

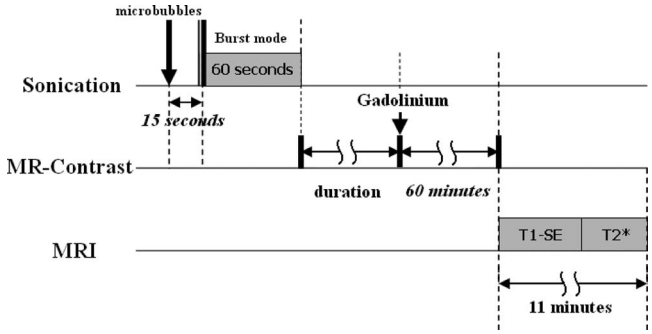


FIG. 1. Diagram of experimental time line for BBB opening duration. FUS sonication was applied 15 s after UCA injection. Gadolinium was injected intraperitoneally several minutes after sonication. All rats were scanned by MRI 60 min after gadolinium administration.

E. MR imaging acquisition

MR imaging was then performed after FUS sonication. The rats were anesthetized with 1.5% isoflurane mixed with O_2 and maintained at 1% isoflurane during the imaging procedures. The experiment was performed on a 3T MRI system (TRIO 3-T MRI, Siemens MAGNETOM, Germany). A loop coil (loop flex coil, approximately 4 cm in diameter) for rf reception was used. A multi-slice spin echo sequence was performed to obtain 20 slices of T1-weighted MR image covering the whole brain for depicting BBB disruption, TR/TE=435/12 ms; in-plane resolution=195×390 μm^2; slice thickness=1.5 mm. A multi-slice spoiled gradient-echo sequence (FLASH) was performed to obtain 20 slices of T2*-weighted MR image covering the whole brain for hemorrhage detection, TR/TE=400/10 ms; flip angle=25°. The in-plane resolution and slice thickness were 260 μm and 1.5 mm. The imaging plane was located across the focus (perpendicular to the beam axis) at the focal depth. MRI contrast agent (Omniscan, GE Healthcare, Cork, Ireland) gadolinium was injected (1 mmol/kg) intravenously several minutes (0, 20, and 60) after the sonication. Six rats were scanned by MRI 60 min after gadolinium administration (Fig. 1). These animals were not part of the EB extravasation study.

III. RESULTS

The BBB disruption was observed in the focal zone of the ultrasound beam with EB extravasation. Figures 2(a) and 2(b) illustrate the relationship between the injected dose of UCA and the degree of EB staining for EB injected intravenously immediately and 1 h after sonication, respectively. Both the size and degree of EB staining in the right brain with injection of 300 or 450 μl/kg UCA were significantly greater than that of 150 μl/kg UCA. In addition, there was no obvious EB staining 1 h after sonication for three doses of UCA. Figure 3 shows the average extravasation of EB (in μg/g of tissue) in the six coronal slices that were injected intravenously with EB immediately and 1 h after sonication at an acoustic power of 1.43 W in the presence of UCA at three doses. The EB extravasation of brain slices with injection of 300 or 450 μl/kg UCA was significantly greater than that of 150 μl/kg UCA. Furthermore, the amount of EB extravasation at 1 h was less than that at 0 h for the same dose of UCA.

Figure 4 shows the relationship between the time and the average extravasation of EB per unit mass (in μg/g of tissue) for the six coronal slices at the ultrasound-sonicated site after injecting UCA at doses of 150, 300, and 450 μl/kg. The EB extravasation across all slices of an individual brain hemisphere was greatest immediately after the sonications and decreased with the time of EB injection following the sonication, with only a small amount of EB extravasation being detected after more than 1 h following the sonication.

Figure 5 reveals that the degree of EB staining at 150 μl/kg at different intervals up to 60 min after sonication. BBB integrity appeared to be re-established after 1 h

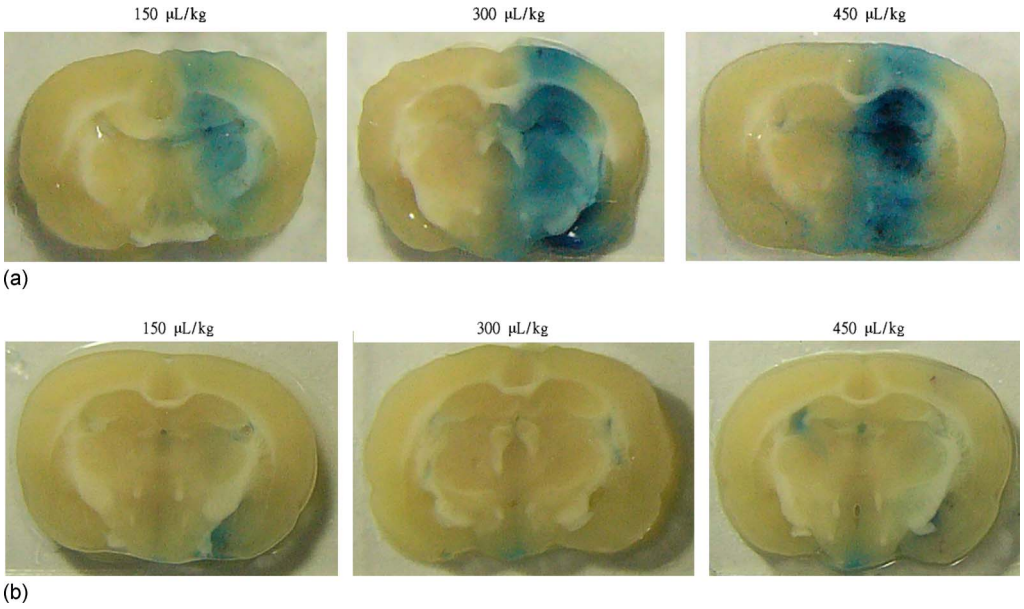
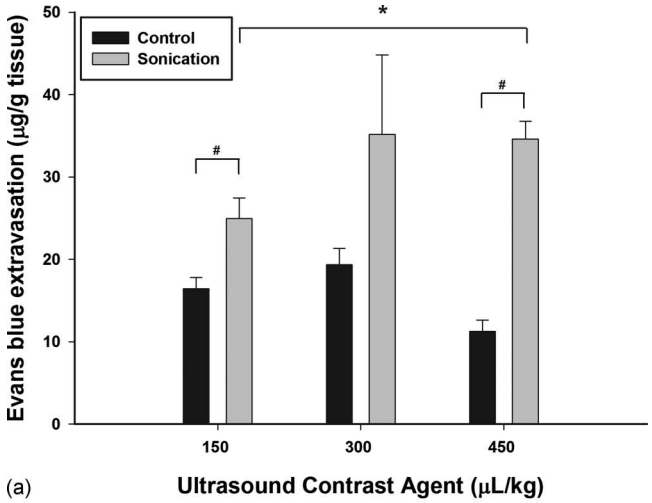


FIG. 2. (Color online) Distribution of BBB disruption for three doses of UCA as evaluated by the EB staining in the focal region at 0 h (a) and 1 h (b) after FUS application. Right brain: 1.43 W acoustic-power sonication. Left brain: no sonication.



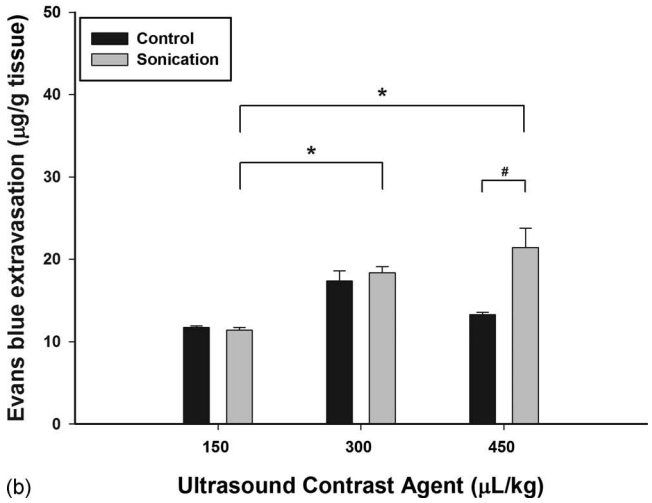


FIG. 3. Graphs showing the relationship between EB extravasation and the injected dose of UCA in the right brain hemisphere at 0 h (a) and 1 h (b) after sonication. The values plotted are the between-rat means of data averaged over six slices within a rat. Results are indicated as mean and SEM values. * and ♯ denote significant differences compared with 150 μl/kg and the contralateral untreated hemisphere, respectively (* and ♯, $p<0.05$, $n=4$).

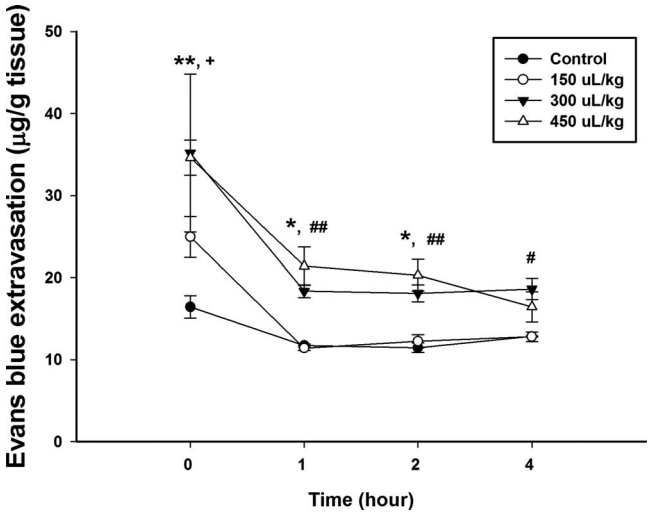


FIG. 4. Graph showing the amount of EB extravasation 4 h after EB administration as a function of time after sonication at three doses of UCA. The values plotted correspond to the between-rat means of data averaged over all six slices within each rat. EB extravasation was largest immediately after the sonications and rapidly decreased within 1 h. Statistical significance compared with the rats injected 150 μl/kg UCA without sonication. *, ♯, and + denote significantly different from 450, 300, and 150 μl/kg, respectively (* and ♯, $p<0.05$; ♯♯, $p<0.01$, $n=4$).

because injection of EB at this time led to no tissue staining. Moreover, Fig. 6 indicates the average extravasation of EB in the six coronal slices of one sonicated location and likewise in the same anatomical location on the opposite side of the brain. The EB extravasation at the sonicated hemispheres was correlated with the time. The BBB disruption upon sonication was significantly greater than that of control location at the time point 0 and 15 min. The EB extravasation showed continuous but declining disruption of the BBB and was approximately the same as the control location at the time point of 60 min after sonication.

Contrast enhancement on the T1-weighted spin echo MR images was used to detect the BBB disruption. Figures 7(a)–7(c) show MR images from the first set of experiments of three rats with doses of UCA at 150, 300, and 450 μl/kg, respectively. The area of contrast enhancement in the right brain with injection of 300 or 450 μl/kg UCA was significantly greater than that of 150 μl/kg UCA. This is consistent with the results of EB staining in Fig. 2(a). The hemorrhage could be observed in T2*-weighted MR images of brain with higher doses of UCA at 300 and 450 μl/kg, but no significant differentiable hemorrhage was observed at 150 μl/kg [Figs. 7(d)–7(f)]. In the second set of experiments, three rats were separately intravenously injected with MRI contrast agent at 0, 20, and 60 min after sonication with a 150-μl/kg dose of UCA. Figure 8 shows the area of contrast enhancement decreased with the time. Approximately 60 min after sonication, the MR image indicates the BBB closure because no contrast enhancement was seen at this time point.

IV. DISCUSSION

This study has revealed the relationship between UCA dose and the duration of BBB disruption induced by a single-element FUS transducer for sonication through the intact skull at a constant acoustic power (Fig. 4). Figures 2 and 3 show that both the area of EB staining and EB extravasation decrease with duration following the sonication at three

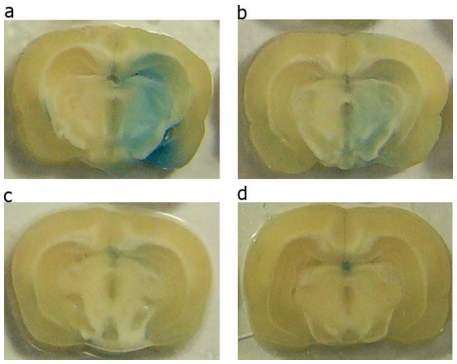


FIG. 5. (Color online) Distribution of BBB disruption for 150 μl/kg UCA as evaluated by the EB staining in the focal region of right brain for four different times after sonication: (a) 0 min, (b) 15 min, (c) 30 min, and (d) 60 min.

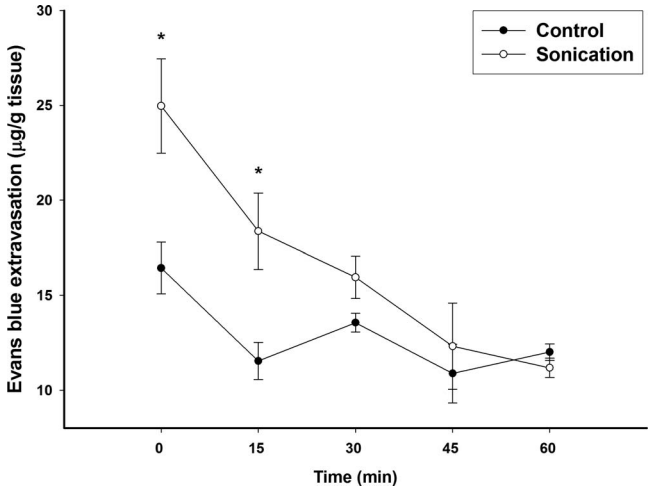


FIG. 6. Graph shows the EB extravasation of 150 μl/kg UCA within 1 h after sonication in right and left (control group) brain hemispheres with and without sonication at an acoustic power of 1.43 W (*, $p<0.05$, $n=4$).

doses of UCA. A higher dose of UCA means that there are more microbubbles in the vasculature to serve as nuclei for cavitation, which increases cavitation,[14,17] extravasation, and diffusion and convection of EB and gadolinium, and produces darker EB staining and brighter MR enhancement. Several studies suggest that the mechanism underlying BBB disruption is partially related to a physiological response to the sonication instead of a direct physical effect.[11,18,19] Interactions between the bubbles and the vessel walls—which include oscillatory forces, radiation force along the direction of the ultrasound beam, and shear stress related to streaming of fluid around the bubbles—are likely to trigger various physiological mechanisms.[20,21] Our previous study[14] found that the BBB disruption increased linearly with the bubble dose, which is consistent with Treat *et al.*,[22] demonstrating that drug delivery increased with the bubble concentration. In contrast, McDannold *et al.*[9] found that the microbubble dose had no significant effect on BBB disruption. Hynynen[20] inferred that BBB disruption should increase with the number of bubbles. These discrepant results suggest that the behavior varies markedly between experiments.

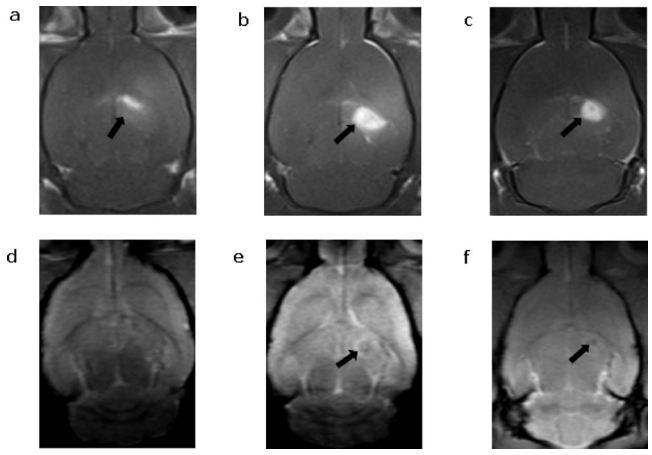


FIG. 7. T1-weighted MR images depict the BBB disruption (arrow) with UCA doses of 150 μl/kg (a), 300 μl/kg (b), and 450 μl/kg (c). T2*-weighted MR image detect brain hemorrhage induced by FUS. Hemorrhage can be detected at higher UCA doses of 300 μl/kg (e) and 450 μl/kg (f), but no visible hemorrhage can be found at 150 μl/kg (d).

Our rats were injected intravenously with EB at several time points after FUS application to evaluate the amount of time needed to repair the BBB disruption. As shown in Fig. 4, the EB extravasation decreased rapidly at 1 h following sonication for all three doses of UCA. At the 1-h time point, the EB extravasation for 150 μl/kg UCA was almost the same as that in the control region, but it differed significantly for both 300 and 450 μl/kg. This indicates that higher doses of UCA induced more long-lasting breakdown of the BBB. Therefore, the reversibility of BBB disruption depended on the dose of UCA for a constant acoustic power. An impermeable BBB is vital to maintaining normal brain physiology, and this also makes pharmacological treatment of the brain difficult. There may be safety concerns when the duration of BBB disruption is too long, and hence the duration of BBB opening in relation to the ultrasound exposure parameters needs to be explored. Further analysis of the reversibility of BBB disruption needs to be investigated so as to determine the safety, effectiveness, and most appropriate dose of UCA with different acoustic parameters.

In addition to tracking the pattern of gadolinium deposition, the hemorrhage condition was also monitored. Cerebral microbleeds are hemosiderin deposits in the brain that are caused by leakage of red blood cells from small blood vessels. These hemosiderin deposits appear in T2*-weighted gradient-echo MRI sequences as small foci of decreased signal intensity.[23] T2*-weighted sequences have been used widely for detecting hemorrhage in several diseases.[24] Figure 7 shows the pattern of contrast enhancement and the corresponding hemorrhage condition. The MRI images indicated that the effective BBB opening was larger at UCA doses of 300 and 450 μl/kg, but hemorrhage was also observed. Thus, a lower UCA dose, such as 150 μl/kg, is sufficient for BBB disruption at the focal zone; a higher UCA dose may induce more erythrocyte extravasations. Figure 4 shows a small (but statistically significant) amount of EB extravasation, which remained elevated even out to 4 h at the higher doses of UCA, indicating a reduced degree of reversibility. A future study should investigate whether more erythrocyte extravasations are responsible for this low level of leakage. Moreover, the area of the contrast enhancement is similar for

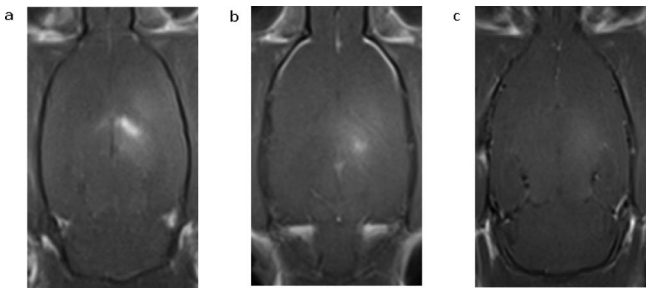


FIG. 8. T1-weighted MR images detect the duration of the FUS-induced BBB disruption. The deposition of gadolinium was monitored at the time points of 0 min (a), 20 min (b), and 60 min (c). The FUS-induced BBB disruption was thus reversed after less than 1 h.

300 and 450 μl/kg UCA, which is in good agreement with the degree of EB extravasation indicated in Figs. 2(a) and 3(a).

Figures 5 and 8 show images obtained several minutes after sonication with injection of EB and gadolinium, respectively. The EB and gadolinium first appear in the target location, and then slowly decrease from the same region, ultimately vanishing from the brain. Thus, BBB disruption was reversed within 1 h after sonication. Figure 6 shows the average extravasation of EB in the sonicated hemisphere within 1 h after sonication at an acoustic power of 1.43 W in the presence of UCA at 150 μl/kg. There was less EB extravasation in the control location due probably to the EB having diffused through some parts of brain lacking an intact BBB, such as the posterior pituitary and the pineal body. Transcranial ultrasound was used in this study to disrupt the BBB, and hence refraction and scattering waves would have affected neighboring regions. This might also explain why the EB level was high in the control region at the first time point, when no time had passed to allow recovery from a slightly disrupted BBB.

This study evaluated the relationship between UCA dose and the reversibility of BBB disruption based on the amount of EB extravasation. The results demonstrate that UCA not only enhances BBB opening but also can be used to control the duration BBB disruption. Future studies should attempt to optimize the ultrasound and UCA parameters so as to obtain appropriate opening and closure of the BBB for use in practical applications.

ACKNOWLEDGMENTS

This study was supported by grants from the National Science Council of Taiwan (No. NSC 97-2320-B-010-004) and was supported in part by the 3T MRI Core Facility in NYMU.

[1]L. L. Rubin and J. M. Staddon, "The cell biology of the blood-brain barrier," Annu. Rev. Neurosci. **22**, 11–28 (1999).
[2]R. A. Kroll and E. A. Neuwelt, "Outwitting the blood-brain barrier for therapeutic purposes: osmotic opening and other means," Neurosurgery **42**, 1083–1099 (1998).
[3]M. Kinoshita, N. McDannold, F. A. Jolesz, and K. Hynynen, "Noninvasive localized delivery of Herceptin to the mouse brain by MRI-guided focused ultrasound-induced blood-brain barrier disruption," Proc. Natl. Acad. Sci. U.S.A. **103**, 11719–11723 (2006).
[4]W. M. Pardridge, "Blood-brain barrier genomics and the use of endogenous transporters to cause drug penetration into the brain," Curr. Opin. Drug Discovery Dev. **6**, 683–691 (2003).
[5]W. M. Pardridge, "Drug and gene delivery to the brain: The vascular route," Neuron **36**, 555–558 (2002).
[6]W. M. Pardridge, "Drug and gene targeting to the brain with molecular Trojan horses," Nat. Rev. Drug Discovery **1**, 131–139 (2002).
[7]K. Hynynen, N. McDannold, N. A. Shcikov, F. A. Jolesz, and N. Vykhodtseva, "Local and reversible blood-brain barrier disruption by noninvasive focused ultrasound at frequencies suitable for trans-skull sonications," Neuroimage **24**, 12–20 (2005).
[8]J. J. Choi, M. Pernot, T. R. Brown, S. A. Small, and E. E. Konofagou, "Spatio-temporal analysis of molecular delivery through the blood-brain barrier using focused ultrasound," Phys. Med. Biol. **52**, 5509–5530 (2007).
[9]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).
[10]K. Hynynen, N. McDannold, N. Vykhodtseva, and F. A. Jolesz, "Noninvasive MR imaging-guided focal opening of the blood-brain barrier in rabbits," Radiology **220**, 640–646 (2001).
[11]S. B. Raymond, J. Skoch, K. Hynynen, and B. J. Bacskai, "Multiphoton imaging of ultrasound/optison mediated cerebrovascular effects in vivo," J. Cereb. Blood Flow Metab. **27**, 393–403 (2007).
[12]S. B. Raymond, L. H. Treat, J. D. Dewey, N. J. McDannold, K. Hynynen, and B. J. Bacskai, "Ultrasound enhanced delivery of molecular imaging and therapeutic agents in Alzheimer's disease mouse models," PLoS ONE **3**, e2175 (2008).
[13]A. H. Mesiwala, L. Farrell, H. J. Wenzel, D. L. Silbergeld, L. A. Crum, H. R. Winn, and P. D. Mourad, "High-intensity focused ultrasound selectively disrupts the blood-brain barrier in vivo," Ultrasound Med. Biol. **28**, 389–400 (2002).
[14]F. Y. Yang, W. M. Fu, W. S. Chen, W. L. Yeh, and W. L. Lin, "Quantitative evaluation of the use of microbubbles with transcranial focused ultrasound on blood-brain-barrier disruption," Ultrason. Sonochem. **15**, 636–643 (2008).
[15]F. Y. Yang, W. M. Fu, R. S. Yang, H. C. Liou, K. H. Kang, and W. L. Lin, "Quantitative evaluation of focused ultrasound with a contrast agent on blood-brain barrier disruption," Ultrasound Med. Biol. **33**, 1421–1427 (2007).
[16]H. S. Sharma and P. K. Dey, "Influence of long-term immobilization stress on regional blood-brain barrier permeability, cerebral blood flow and 5-HT level in conscious normotensive young rats," J. Neurol. Sci. **72**, 61–76 (1986).
[17]D. L. Miller, "Overview of experimental studies of biological effects of medical ultrasound caused by gas body activation and inertial cavitation," Prog. Biophys. Mol. Biol. **93**, 314–330 (2007).
[18]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).
[19]N. Sheikov, N. McDannold, N. Vykhodtseva, F. Jolesz, and K. Hynynen, "Cellular mechanisms of the blood-brain barrier opening induced by ultrasound in presence of microbubbles," Ultrasound Med. Biol. **30**, 979–989 (2004).
[20]K. Hynynen, "Ultrasound for drug and gene delivery to the brain," Adv. Drug Delivery Rev. **60**, 1209–1217 (2008).
[21]N. Vykhodtseva, N. McDannold, and K. Hynynen, "Progress and problems in the application of focused ultrasound for blood-brain barrier disruption," Ultrasonics **48**, 279–296 (2008).
[22]L. H. Treat, N. McDannold, N. Vykhodtseva, Y. Zhang, K. Tam, and K. Hynynen, "Targeted delivery of doxorubicin to the rat brain at therapeutic levels using MRI-guided focused ultrasound," Int. J. Cancer **121**, 901–907 (2007).
[23]F. Fazekas, R. Kleinert, G. Roob, G. Kleinert, P. Kapeller, R. Schmidt, and H. P. Hartung, "Histopathologic analysis of foci of signal loss on gradient-echo T2*-weighted MR images in patients with spontaneous intracerebral hemorrhage: evidence of microangiopathy-related microbleeds," AJNR Am. J. Neuroradiol. **20**, 637–642 (1999).
[24]V. Sehgal, Z. Delproposto, D. Haddar, E. M. Haacke, A. E. Sloan, L. J. Zamorano, G. Barger, J. Hu, Y. Xu, K. P. Prabhakaran, I. R. Elangovan, J. Neelavalli, and J. R. Reichenbach, "Susceptibility-weighted imaging to visualize blood products and improve tumor contrast in the study of brain masses," J. Magn. Reson Imaging **24**, 41–51 (2006).

Ambient pressure sensitivity of microbubbles investigated through a parameter study

Klaus Scheldrup Andersen and Jørgen Arendt Jensen
Department of Electrical Engineering, Center for Fast Ultrasound Imaging, Technical University of Denmark, Building 348, 2800 Kgs. Lyngby, Denmark

(Received 5 February 2009; revised 31 August 2009; accepted 8 September 2009)

Measurements on microbubbles clearly indicate a relation between the ambient pressure and the acoustic behavior of the bubble. The purpose of this study was to optimize the sensitivity of ambient pressure measurements, using the subharmonic component, through microbubble response simulations. The behavior of two microbubbles corresponding to two different contrast agents was investigated as a function of driving pulse and ambient overpressure p_{ov}. Simulations of Levovist using a rectangular driving pulse show an almost linear reduction in the subharmonic component as p_{ov} is increased. For a 20 cycle driving pulse, a reduction of 4.6 dB is observed when changing p_{ov} from 0 to 25 kPa. Increasing the pulse duration makes the reduction even clearer. For a pulse with 64 cycles, the reduction is 9.9 dB. This simulation is in good correspondence with measurement results presented in the literature. Further simulations of Levovist show that also the shape and the acoustic pressure of the driving pulse are very important factors. The best pressure sensitivity of Levovist was found to be 0.88 dB/kPa. For Sonazoid, a sensitivity of 1.14 dB/kPa has been found, although the reduction is not completely linear as a function of the ambient pressure.
 [DOI: 10.1121/1.3242359]

PACS number(s): 43.80.Vj, 43.25.Yw [CCC] Pages: 3350–3358

I. INTRODUCTION

Local blood pressure measurements provide important information on the state of health of the organs in the human body and can be used to diagnose severe heart, lung, and kidney diseases. The pressure is currently measured locally in arteries and organs by means of a pressure catheter. As this is an invasive technique, it is inconvenient to the patient, there is a risk of infection, and the catheter will inevitably introduce changes to the blood flow and thereby the pressure. Therefore, many attempts to find a noninvasive procedure have been made. When evaluating new approaches for noninvasive local blood pressure measurements, the sensitivity is a crucial factor. Although the human blood pressure varies between 0 and approximately 25 kPa (1 kPa=7.5 mm Hg), it should still be possible to distinguish pressure differences as low as 1–3 kPa to measure the blood pressure in the small veins and arteries. One noninvasive approach that has been suggested is to perform Doppler echocardiography using a simplified modification of the Bernoulli equation.[1,2] This method was, however, concluded not to yield reproducible or reliable results by Strauss *et al.*[3] Another type of approach, which is still being investigated, is to combine microbubbles injected into the blood and diagnostic ultrasound.

Bubbles in a fluid can be used for measurement of pressure gradients due to their size dependent oscillations.[4–6] Since the introduction of ultrasound contrast agents (UCAs), many approaches on how to exploit their ambient pressure sensitivity have been presented. One of the first to propose noninvasive measurement of cardiac pressure using an UCA were Fairbank and Scully[4] in 1977. They claimed that the acoustic properties of the microbubbles change when the size of the bubbles changes. To measure these changes, they suggested the use of resonance excitation. However, due to the large size distribution of the first generation UCAs containing free bubbles, their results were inconclusive. Other suggestions from that time are by Hok[5] in 1981 and Shankar *et al.*[6] and Newhouse and Shankar[7] in 1986. Newhouse and Shankar showed theoretically and experimentally that accurate bubble size measurements are possible using a double frequency technique for determination of the sum and difference frequencies. The rapid dissolution time after injection of the free air bubbles prevented, however, any practical implementation.

With the introduction of the more stable second generation UCAs, the circulation system can now be used to transport the encapsulated microbubbles to the region of interest. This has initiated new attempts to exploit the ambient pressure dependent acoustic properties that the high compressible air and gas bubbles possess. Bouakaz *et al.*[8] presented in 1999 an approach for measuring the disappearance time of free bubbles, which were generated at the region of interest by rupturing the contrast agent microbubbles using a low-frequency high acoustic amplitude pulse. From *in vitro* experiments they concluded the approach to have a resolution of 6.7 kPa (50 mm Hg). Later, they suggested that the resolution could be improved by using larger bubbles or by using wavelet processing or a combination of this.[9] No *in vivo* results or further investigations have, however, been presented using this approach yet. Around the same time, Shi *et al.*[10] observed from experiments that the subharmonic component of Levovist is highly sensitive to ambient pressure changes compared to the fundamental and the second harmonic component. They reported a 9.9 dB linear decrease in the peak amplitude of the subharmonic component when increasing the ambient hydrostatic pressure from 0 to 24.8 kPa

(186 mm Hg). Furthermore, they found that the ambient pressure-induced reduction was highest when the acoustic excitation pressure was around the growth stage of the subharmonic, which occurs when the acoustic driving pressure causes the subharmonic component to increase rapidly from background noise level to be clearly visible in the spectrum. Recently, the same group has presented similar results for Sonazoid, which was found to have an average decrease of 13.3 dB.[11] Furthermore, in 2005 the same group presented *in vivo* results for proof of concept of the capabilities of the subharmonic response.[12] As the measurements were performed directly on the abdominal cavity and the aorta by incision of two dogs, this can hardly be characterized as noninvasive. However, the results still showed that the subharmonic component decreased as the ambient pressure increased and thereby indicated that the subharmonic response of UCAs can be used for ambient pressure measurements. Also in 2005, Adam *et al.*[13] investigated microbubbles' response to cyclic ambient pressure changes by mimicking left ventricular pressure changes. They found that the subharmonic response correlated best with the cyclic changes compared to the fundamental and second harmonic, but also observed a transient delay before this correlation occurred. In 2008, Andersen and Jensen[14] presented a new experimental setup, which more realistically resembles a clinical setting using a single array transducer. The setup was used to measure the ambient pressure sensitivity of SonoVue and confirmed the previous findings, revealing a pressure sensitivity of 0.42 dB/kPa. The same group has also investigated the dependence on the acoustic driving pressure experimentally.[15] However, the driving pressure was selected too high causing bubble destruction, and the investigation was therefore inconclusive.

Modeling the acoustics of bubbles in a fluid is a still ongoing investigation, which was initiated by Rayleigh[16] in 1917 who studied damages to ship propellers due to bubble cavitation. In 1933, Minnaert[17] explained the characteristic resonance frequency of free bubbles when he did a theoretical and experimental study of bubbles' emission of sound. Since then, several modifications to the existing models and new theoretical models on how to predict the behavior of an oscillating bubble have been presented. Most models are based on modifications of the Rayleigh–Plesset[18] (RP) equation and are capable of handling shell encapsulating bubbles. This includes the models used by de Jong and Hoff[19] and Church.[20] Other models are based on the modified Herring equation to describe the radial motion (e.g., Morgan *et al.*[21]). Recently, Vos *et al.*[22] proposed a novel approach for investigation of full populations of microbubbles' behavior in acoustic fields, also based on the modified Herring equation. The study furthermore included a new method to estimate the viscoelastic shell properties of UCAs. Existing methods for this based on the Rayleigh–Plesset equation have been suggested by de Jong and Hoff[19] and de Jong *et al.*[23] and described further by Hoff.[24]

Despite the growing number of experiments within hydrostatic pressure measurements, no real parameter study investigating the response of microbubbles in respect to ambient pressure changes has been performed until now. The purpose of this study is to optimize the sensitivity of pressure measurements through bubble response simulations investigating the complex mechanisms for subharmonic generation. This is carried out by an extensive number of simulations of two commercial UCAs. Since the study focused on the effect of the driving pulse, the parameters of the microbubbles are fixed in all simulations, whereas several different settings regarding the excitation pulse were varied. Some part of this work has been presented at the 2008 SPIE Medical Imaging Symposium.[25]

This paper is organized as follows. Section II presents the choice of simulation model and parameters used for the investigation. Furthermore, it also describes the processing of the simulated response. The achieved results are presented and discussed in Sec. III. Finally, the investigation is summarized by a conclusion in Sec. IV.

II. THEORY AND METHOD

The investigation has been performed using the MATLAB (The Math Works Inc., Natick, MA) environment. To carry out the simulations, the free simulation program BUBBLESIM by Hoff[24] is used. BUBBLESIM is a toolbox that calculates the oscillation and scattered echo for a specified contrast agent microbubble and excitation pulse. It numerically solves a second order ordinary differential equation (ODE) that has been combined from a set of equations, each equation modeling different parts (bubble, shell, and surrounding liquid) of the system that makes up a contrast agent microbubble. In BUBBLESIM, the following four different models are implemented: The RP[18] model, the Trilling[26] model, the Keller–Miksis[27] model, and a modified version of the RP model, which is an intermediate model of the RP on one hand and the Trilling and Keller–Miksis models on the other hand. The largest disadvantage of the RP model is that it does not include radiation damping, which is energy loss caused by radiation of sound. This is accounted for in the Trilling and the Keller–Miksis models, which both include a finite but constant speed of sound in the liquid. However, both the Trilling and Keller–Miksis models have a risk of becoming numerically unstable when the bubble wall velocity becomes comparable to the speed of sound (acoustic Mach numbers, $M=\dot{R}/c$, around unity). This happens for high oscillation amplitudes and causes the models to have an unphysical negative inertia. Instead, Hilgenfeldt and Lohse[28] suggested a modified version of the RP model that includes the radiation damping term from the Trilling and Keller–Miksis models. This is also used in the modified version implemented in BUBBLESIM, which is the model selected for the parameter investigation. It was chosen because of its numerical stability, which is important when doing many simulations spanning a wide range of variable changes. The model is based on the theoretical description of microbubbles as air-filled particles with surface layers of elastic solids presented by Church[20] and later derived by Hoff[24] to describe a polymeric microbubble:

$$\rho_L a\ddot{a} + \frac{3}{2}\rho_L(\dot{a})^2 = p_{Ge}\left(\frac{a_e}{a}\right)^{3\kappa} - p_{ov} - p_{ac}(t) - 4\mu_L\frac{\dot{a}}{a} - 12\mu_S\frac{d_{Se}a_e^2}{a^3}\frac{\dot{a}}{a} - 12G_S\frac{d_{Se}a_e^2}{a^3}\left(1 - \frac{a_e}{a}\right). \quad (1)$$

In Eq. (1), ρ_L is the surrounding liquid density, $a(t)$ and a_e are the instantaneous and equilibrium outer bubble radii, respectively, $\dot{a}=da/dt$ denotes derivation with respect to time t, p_{Ge} is the pressure in the gas inside the bubble at equilibrium, κ is the polytropic exponent of the gas, p_{ov} is the hydrostatic pressure in the surrounding liquid, $p_{ac}(t)$ is the driving acoustic pressure, μ_L and μ_S are the shear viscosities in the liquid and shell, respectively, d_{Se} is the shell thickness, and G_S is the shell shear modulus. Rewriting Eq. (1) in a simplified form gives

$$\rho_L a\ddot{a} + \frac{3}{2}\rho_L(\dot{a})^2 = p_L - p_{ov} - p_{ac}(t), \quad (2)$$

where

$$p_L = p_{Ge}\left(\frac{a_e}{a}\right)^{3\kappa} - 4\mu_L\frac{\dot{a}}{a} - \Delta T_S \quad (3)$$

describes the pressure at the bubble surface and

$$\Delta T_S = 12\frac{d_{Se}a_e^2}{a^3}\left(G_S\left(1 - \frac{a_e}{a}\right) + \mu_S\frac{\dot{a}}{a}\right) \quad (4)$$

denotes the stress across the shell. The left hand side of Eq. (2) represents the inertia of the liquid due to the oscillating bubble while the right hand side denotes the damping viscous and restoring stiffness forces. The pressure in the gas at equilibrium p_{Ge} is equal to the hydrostatic pressure in the liquid and a term caused by surface tension.[20] However, to prevent immediate bubble shrinkage upon formation, the shell strain is assumed to counteract the latter part; hence $p_{Ge}=p_{ov}$. Finally, from Eq. (4) it is seen that the model requires the viscous and elastic properties of the shell described by μ_S and G_S. These must, however, be estimated experimentally as described below.

Any numerical solver can be used to solve the ODE in Eq. (1). In this study, the solver of variable order from 1 to 5 has been selected as it should be more reliable and stable for solving situations where the differential equation becomes stiff.[24] This occurs, for example, when the bubble radius changes slowly during the expansion phase but goes through very fast changes in radius and velocity under compression. The choices on simulation model and numerical solver, as well as other general setup parameters, for the simulations in this study are summarized in Table I.

In this study, a batch mode for BUBBLESIM has been created as this gives a bit more control and, more importantly, it makes it possible to perform multiple simulations automatically, which is essential in a parameter study like this. Furthermore, one modification has been made to BUBBLESIM. In its original form, it is not possible to change the ambient overpressure parameter denoted by p_{ov} in Eq. (1). Since this is crucial when investigating microbubbles' sensitivity to ambient pressure changes, this feature has been enabled by small modifications to the source code.

TABLE I. List of simulation parameters regarding the general setup of BUBBLESIM.

Parameter	Designation
ODE solver	ODE15s
Simulation model	Modified Rayleigh–Plesset
Thermal damping	Isothermal
Liquid	Water

While the bubble size distribution can be determined with a multisizer, it is somewhat more difficult to specify the elastic and viscous shell parameters. One way to do this is to perform a combination of experiments and model fitting as described by Hoff and co-workers.[19,23,24] This will, however, only give an estimate of proper designations, and usually an interval for some of the parameters is given. The procedure has been used by Yu *et al.*[29] and Hoff[24] to estimate suitable parameters for the commercial contrast agents Levovist (Schering AG, Berlin, Germany) and Sonazoid (GE Healthcare, Oslo, Norway), respectively. These values used in the investigation were fixed for all simulations and are listed in Table II. Before a simulation can be carried out, a driving pulse must be selected. Since the emphasis of this study was to optimize the subharmonic sensitivity to ambient pressure changes as a function of the excitation pulse, a large number of different driving pulses were examined. The driving pulse was generated based on four different characteristics being the center frequency f_c, the number of pulse cycles N_c, the maximum acoustic pressure p_{ac}, and the shape of the pulse. The possible designations used for the investigation are listed in the upper part of Table III. The center frequency was selected based on a preliminary study optimizing the energy of the subharmonic component to the fundamental, as shown for Sonazoid in Fig. 1. Figure 1(a) shows the energy of the subharmonic component as a function of the driving frequency. As can be seen, the energy peaks and is approximately the same for all driving frequencies displayed. Looking at the ratio of the subharmonic energy to the energy of the fundamental component in Fig. 1(b), however, shows that the subharmonic component is greater than the fundamental at two of the displayed frequencies. Therefore, f_c =2.46 MHz was selected as the driving frequency when simulating the Sonazoid microbubble. As can be seen from Table III, 30 different settings for the acoustic pressure are used. This was decided to ensure determination of the growth stage of the subharmonic component with a reasonable precision. Although an acoustic pressure of 950 kPa will prob-

TABLE II. List of the parameters from Yu *et al.* (Ref. 29) and Hoff (Ref. 24) used to describe the two different types of bubbles for the simulations in BUBBLESIM.

Contrast agent	Bubble radius (μm)	Shell thickness (nm)	Shear modulus (MPa)	Shear viscosity (Pa s)
Levovist	3.0	6.0	80	1.3
Sonazoid	3.2	4.0	52	0.99

TABLE III. List of parameters used in combination with the contrast agents listed in Table II. Combining all settings gives 3600 simulations in total for each agent.

Parameter	Designation										Unit
f_c	2.06	2.46									MHz
N_c	1	2	5	10	20	32	48	64	128	256	cycles
p_{ac}	100	150	200	250	275	300	325	350	375	400	kPa
	425	450	475	500	550	575	600	650	675	700	
	725	750	775	800	825	850	875	900	925	950	
Pulse shape	"Rectangular"	"Hanning"									
p_{ov}	0	5	10	15	20	25					kPa

ably destroy the microbubbles in real measurements, the high values were selected to cover the entire range of subharmonic growth and saturation. The bottom row of Table III lists the designations of the ambient overpressures which were used in the simulations. As can be seen, the range covers the interval between 0 and 25 kPa in steps of 5 kPa. In this way, the most common human blood pressure values are covered. Combining all the parameters in Table III gives a total of 3600 different simulations for each contrast agent.

When BUBBLESIM has completed a simulation, the simulated scattered pressure calculated 1 m from the center of the bubble is returned and the Fourier transformation is applied. Next, a search for the fundamental (f_0), the first subharmonic ($\frac{1}{2}f_0$), and the second harmonic ($2f_0$) component is performed, and the energy of each component is calculated. The center frequencies of the harmonic bands were selected as multiples of the emitted center frequency f_c. It should, however, be noted that initial simulations show that the frequency of the subharmonic component shifts slightly as the acoustic driving pressure is increased. The energy has been chosen over the peak amplitude since this is a more robust measure. The bandwidth to calculate the energy within was selected as the −10 dB bandwidth of the excitation pulse.

III. RESULTS AND DISCUSSION

This section presents the results obtained through the simulation study. First, the fundamental, subharmonic, and second harmonic dependences on acoustic pressure will be presented. This is a natural step for two reasons. First of all, generation of the subharmonic component must be ensured before looking into the ambient pressure dependency. Another reason is to see at which acoustic pressures the growth stage of the subharmonic occurs for the two types of microbubbles. Along with this investigation, the scattered responses and spectra have been examined to ensure useful responses and proper selection of the bandwidth intervals to calculate the energy of the respective frequency components within. Since these results are rather trivial and take up a lot of space, only a few selected examples are presented in this section. In Sec. III B, the influence of ambient overpressure will be examined.

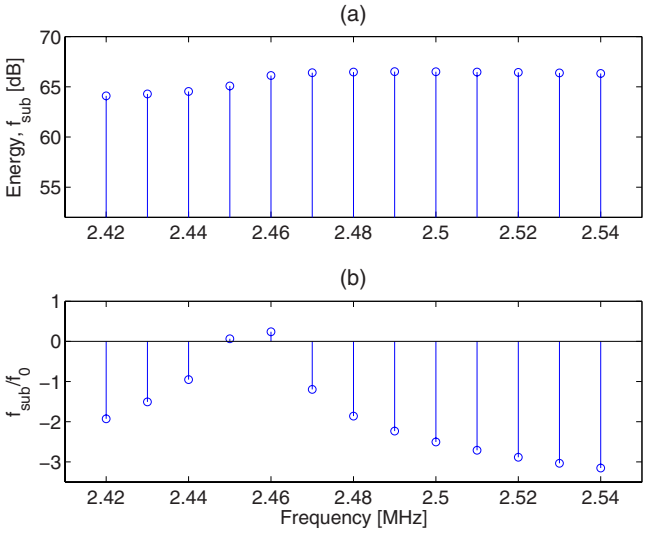


FIG. 1. (Color online) (a) shows the energy of the subharmonic component as a function of emitted frequency, while (b) shows the relation of the subharmonic to the fundamental component.

A. Dependence on acoustic pressure

Figure 2 shows the energy of the subharmonic, fundamental, and second harmonic components of Sonazoid as a function of acoustic pressure when a rectangular driving pulse for a different number of cycles is used. Each curve has been normalized by 88 dB, which corresponds to the maximum energy observed among all simulations for both agents. Examining the subharmonic component, three characteristic

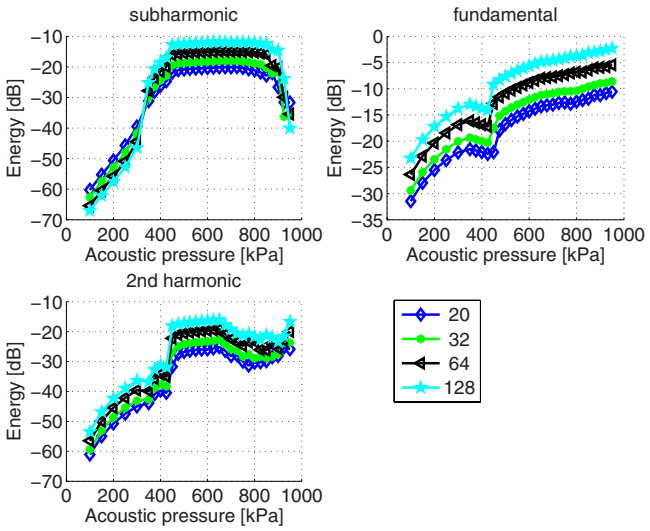


FIG. 2. (Color online) Energy as a function of acoustic pressure for Sonazoid. The driving pulse is a rectangular shaped sinusoid. Upper left graph shows the subharmonic behavior, upper right shows the first harmonic, and lower left presents the behavior of the second harmonic component. Each curve represents a different number of cycles in the driving pulse, as displayed in the legend to the lower right in the figure.

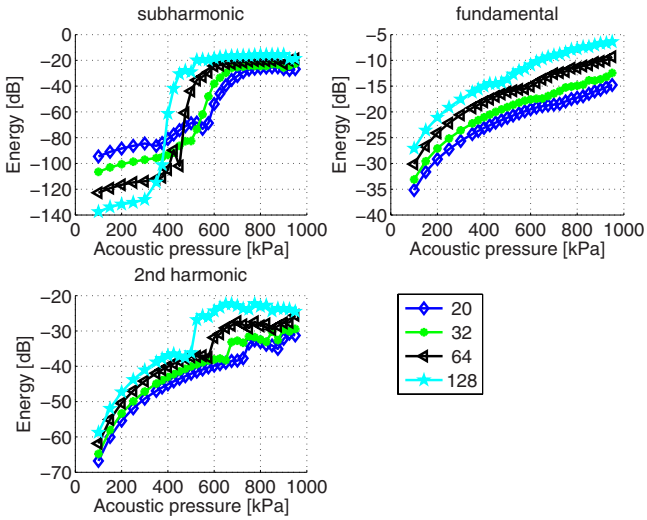


FIG. 3. (Color online) Energy as a function of acoustic pressure for Sonazoid. The driving pulse is a Hanning shaped sinusoid. Upper left graph shows the subharmonic behavior, upper right shows the first harmonic, and lower left presents the behavior of the second harmonic component. Each curve represents a different number of cycles in the driving pulse, as displayed in the legend to the lower right in the figure.

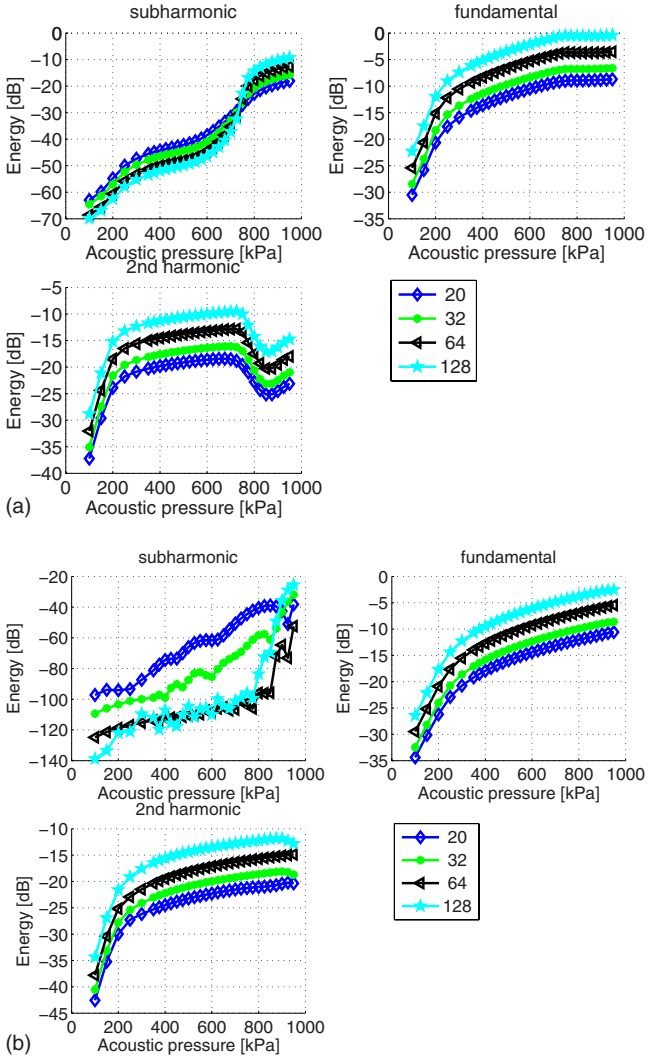


FIG. 4. (Color online) Energy as a function of acoustic pressure for Levovist. The driving pulse is a rectangular shaped (a) and a Hanning shaped (b) sinusoid, respectively. Upper left graph shows the subharmonic behavior, upper right shows the first harmonic, and lower left presents the behavior of the second harmonic component. Each curve represents a different number of cycles in the driving pulse, as displayed in the legend to the lower right in the figure.

stages are clearly observed. In the occurrence stage for acoustic pressures below 300 kPa, the subharmonic is weak compared to the other components. For acoustic pressures in the interval between 300 and 425 kPa, the subharmonic increases rapidly, and this part can be characterized as the growth stage. This interval is in good agreement with measurements by Sarkar *et al.*[30] who investigated the scattered response of Sonazoid at different driving pressures and frequencies. When increasing the acoustic pressure further, the growth eases off and can be compared to the saturation stage observed in measurements. Finally, when the acoustic pressure exceeds 875 kPa the energy decreases again. At these levels, the corresponding spectra look noisy, indicating a stage of chaos. The pattern is the same for pulses of other lengths than displayed here, although the subharmonic component cannot be distinguished from the fundamental for driving pulses smaller than 5 cycles. The chaotic behavior at high acoustic pressure levels predicted in the simulations is actually in good correspondence with experimental results of free bubbles achieved by Lauterborn and Cramer.[31] Looking at the fundamental, it increases almost linearly as expected. However, a slight drop is seen in the pressure interval corresponding to the growth stage of the subharmonic.

When the shape of the driving pulse is changed by applying a Hanning window, especially the subharmonic and second harmonic change behavior, as can be seen in Fig. 3. Regarding the subharmonic component, the three stage pattern is the same as observed for the rectangular driving pulse, although the interval of the growth period seems to have increased. This makes sense since less energy is transmitted using a Hanning shaped driving pulse compared to a rectangular signal of the same acoustic strength. Another interesting observation is that the acoustic pressure interval of the growth stage now is more dependent on the length of the driving pulse. The same pulse length dependent behavior is also seen for the second harmonic component. The fundamental, on the other hand, does not seem to be affected much although the small drop in energy observed for the rectangular driving pulse is hardly visible anymore.

The results for the simulations of Levovist as a function of acoustic pressure using a rectangular driving pulse are shown in Fig. 4(a). Once again, the three stage behavior of the subharmonic component is observed. However, now the growth stage occurs in the interval from 600 to 900 kPa. Although the increase in energy is the same, the interval is at much higher acoustic levels than experimental results achieved by Shi *et al.*,[10] who observed it to be between 300 and 600 kPa for a custom designed (*in vitro* batch) suspension of Levovist using a 64 cycle rectangular driving pulse with a center frequency of 2.0 MHz. Possible reasons for this deviation can be the selection of the shell parameters itself, the difference in UCA suspension when determining the shell parameters, or the different excitation frequencies used. The fundamental and second harmonics more closely resemble

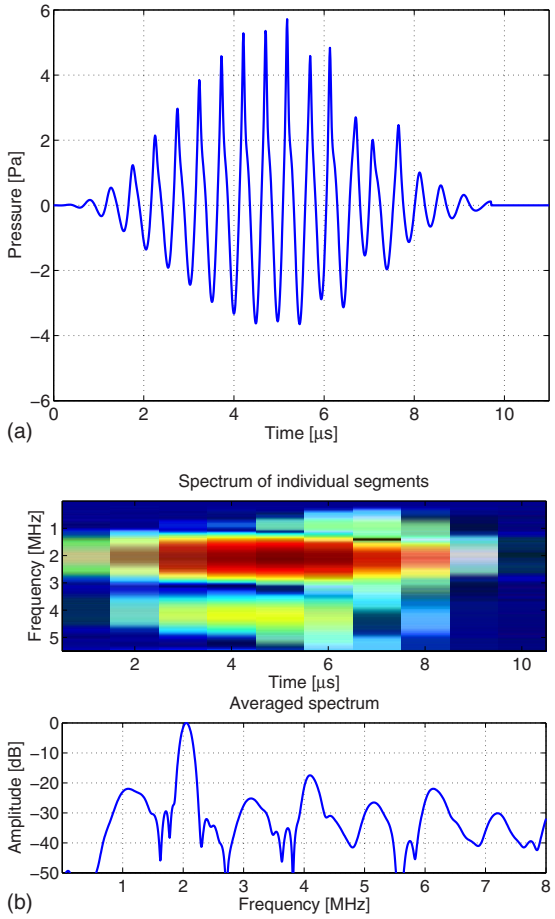


FIG. 5. (Color online) Example of scattered pressure (a) and its corresponding spectrum (b) when using a Hanning shaped driving pulse. Sonazoid is used and the excitation is a 20 cycle Hanning shaped signal with a center frequency of 2 MHz and an acoustic pressure of 525 kPa.

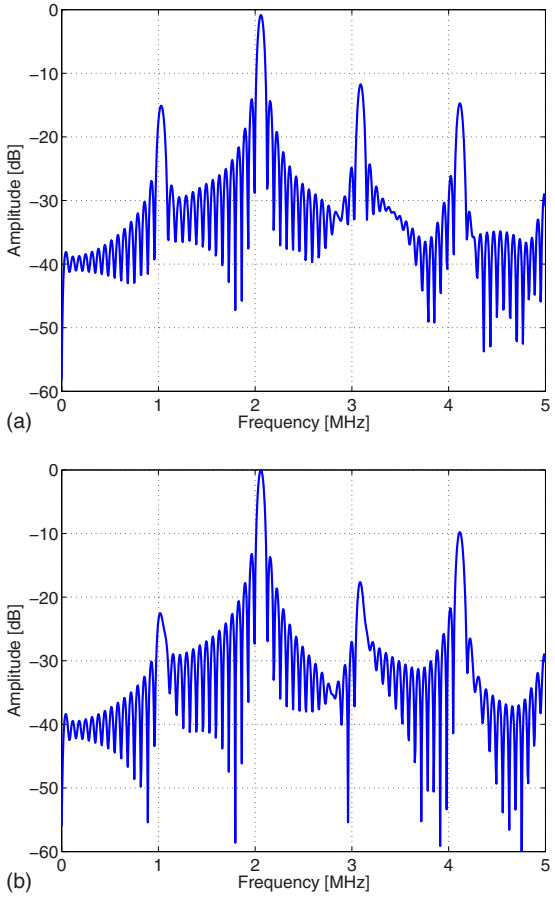


FIG. 6. (Color online) Example of spectrum of scattered response from excitation of microbubble corresponding to Levovist. The driving pulse is a 32 cycle rectangular shaped signal with a center frequency of f_c =2.06 MHz and an acoustic pressure of P_{ac}=800 kPa. (a) is when no overpressure is applied and (b) shows the response when an overpressure of 25 kPa is applied.

the obtained measurement results. The simulations of Levovist using a Hanning shaped excitation pulse indicate that it is very hard to generate the subharmonic component for this type of driving pulse, see Fig. 4(b). In fact, the subharmonic component is hardly visible in any of the spectra, not even at the very high driving pressures. Regarding the fundamental and second harmonic components, they are similar to what was observed using the rectangular driving pulse.

Except, possibly, for the last setup, common for all the simulations is that the subharmonic component has a threshold and is present only above a certain acoustic pressure. This observation was also reported by Prosperetti[32] who examined this experimentally on free bubbles and, as mentioned, by Shi *et al.*[10] In contrast to the threshold behavior of the subharmonic, the higher harmonics seem to be present to various degrees for all driving pressures.

Finally, one interesting observation regarding the scattered pressure, when using the Hanning shaped driving pulse for excitation of Sonazoid, should be noted. When the driving pressure is increased to a level where the subharmonic is generated, the scattered response suddenly changes characteristics halfway in the pulse, as shown in Fig. 5(a). In the first half, the traditional harmonic distortion is clearly observed but no subharmonics. Halfway in the signal, the characteristic oscillation at twice the driving period is seen and continues for the rest of the scattered response. As can be seen from the corresponding spectrum in Fig. 5(b), this is what gives rise to the sub- and ultraharmonics.

B. Dependence of overpressure

In this section, the simulation results achieved when changing the ambient overpressure will be shown. Figure 6 shows an example of how the scattered spectrum changes when the ambient pressure is the only parameter that is changed from one simulation to another. The example is for Levovist when driven by a rectangular pulse with 32 cycles and an acoustic pressure of 800 kPa. In Fig. 6(a), the scattered spectrum is shown when no pressure is seen and Fig. 6(b) shows the spectrum when an overpressure of 25 kPa is applied. Comparing the two spectra, a clear reduction in the subharmonic component at 1 MHz is observed. No theoretical explanation for the subharmonic sensitivity to the surrounding pressure has yet been presented. However, it is likely to believe that increased damping and, perhaps, microbubble stiffness are the main mechanisms for this property as they control how the bubble oscillates.[32]

Figure 7 shows the effect on the subharmonic component when the pulse length is varied. It displays the energy of the subharmonic component when using the same setup as

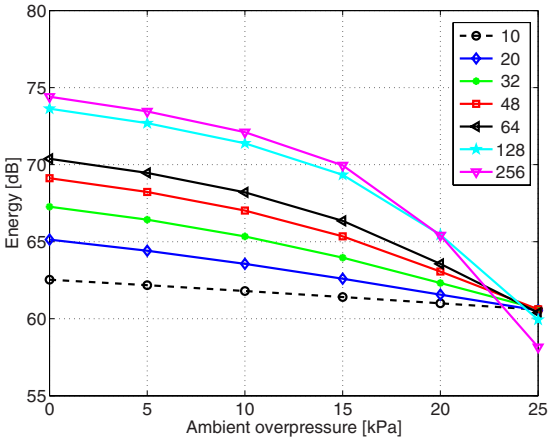


FIG. 7. (Color online) Energy of the subharmonic component scattered by Levovist when using a rectangular shaped driving pulse with an acoustic pressure of 800 kPa. The energy is displayed as a function of ambient pressure, and each curve in the plot represents a different number of cycles in the driving pulse, as indicated by the legend.

used to create Fig. 6. There is a clear tendency for all pulse lengths that the energy decreases as the overpressure is increased. Furthermore, the total decrease in energy also seems to be dependent on the number of cycles in the driving pulse. An explanation for this is probably found having Figs. 2–4, which displayed the subharmonic response simulated at 0 Pa, in mind. These indicated that a longer driving pulse generates more subharmonic energy. However, when the ambient pressure is increased to 25 kPa, the bubble is compressed to a degree that hardly allows any subharmonic oscillations. This theory is corroborated when looking at the last ambient pressure setting in Fig. 7, indicating almost the same amount of subharmonic energy disregarding the pulse length. One final thing that should be noted from Fig. 7 is that it also indicates that the decrease is not completely linear in all cases. For easy comparison of the change in energy for the different simulation setups, Fig. 8 shows the energy of the three frequency components as function of ambient overpressure when each simulation has been normalized to their respective maximum. Looking at the results for the fundamental, it is seen that this component is not affected by ambient pressure changes. The second harmonic seems to be affected and increases about 5 dB, slightly dependent on the pulse length. This increase is quite in contradiction to the experiments by Shi *et al.*,[10] who excited Levovist in the growth stage using a 64 cycle rectangular pulse. They found that the second harmonic decreases by 1.8 dB over the same ambient pressure interval. The explanation for this discrepancy is not clear as the increase in second harmonic energy is a clear trend from all the simulations performed in this study. When examining the subharmonic in Fig. 8, a highly pulse length dependent decrease is observed. As the number of pulse cycles is increased, the reduction in energy also increases. However, the decrease becomes less linear as the pulse length increases. For the driving pulse with 64 cycles, a decrease of 9.9 dB is found. This amount of reduction is in good agreement with similar experimental results by Shi *et al.*,[10] who measured a reduction of 9.6 dB using a center frequency of 2 MHz and a driving pressure of 0.39 MPa corresponding to the subharmonic growth interval of the current batch.

Figure 9 shows the ambient pressure sensitivity of the subharmonic component when the ambient pressure is increased from 0 to 25 kPa. The sensitivity corresponds to the absolute reduction divided by 25 and is shown as a function of the acoustic pressure and number of pulse cycles. Figure 9 thereby summarizes 252 of the most promising simulations of Levovist. Furthermore, to get a measure of the linearity between the energy of the subharmonic component and the overpressure, a straight line has been fitted using linear re-

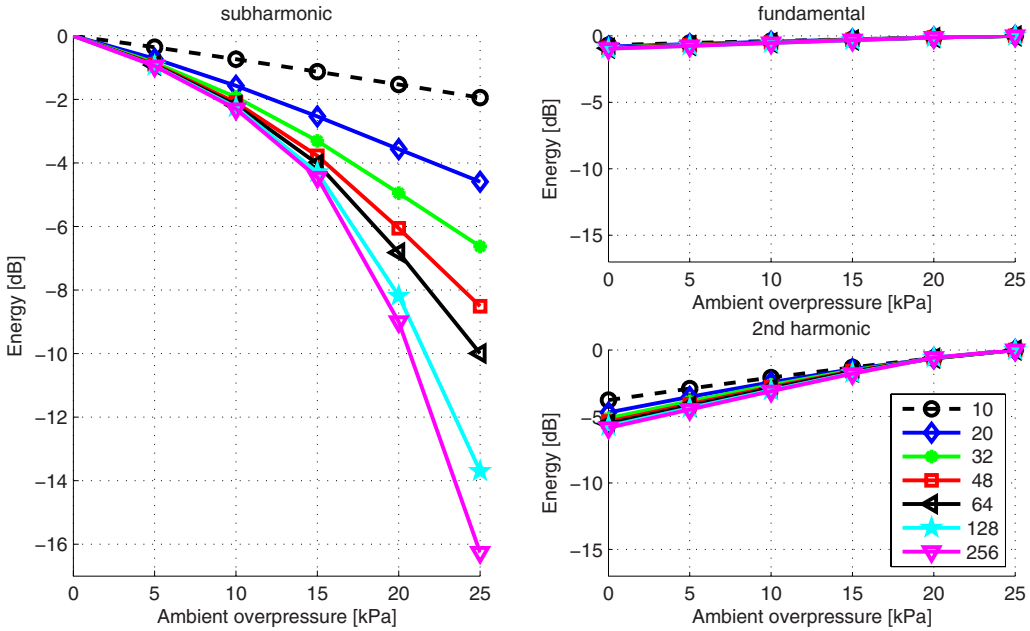
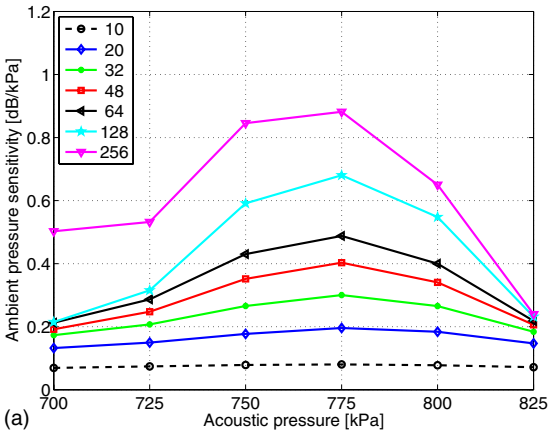


FIG. 8. (Color online) Decrease in energy of respective frequency components scattered by Levovist when using a rectangular shaped driving pulse with an acoustic pressure of 800 kPa. The energy is displayed as a function of ambient pressure, and each curve in the plots represents a different number of cycles in the driving pulse, as indicated by the legend.



(a)

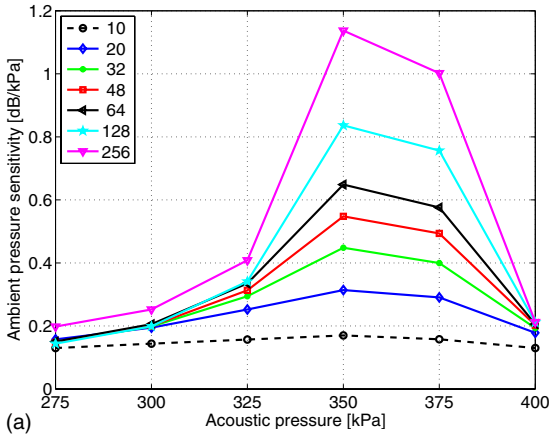
N_c \ P_{ac}	700	725	750	775	800	825
10	0.07	0.07	0.08	0.08	0.08	0.07
20	0.13	0.16	0.18	0.20	0.18	0.16
32	0.17	0.21	0.27	0.30	0.27	0.18
48	0.19	0.26	0.36	0.40	0.34	0.21
64	0.21	0.29	0.43	0.49	0.40	0.22
128	0.21	0.32	0.59	0.68	0.55	0.23
256	0.50	0.63	0.86	0.88	0.68	0.24

Pressure sensitivity [dB/kPa]

N_c \ P_{ac}	700	725	750	775	800	825
10	1.00	1.00	1.00	1.00	1.00	1.00
20	1.00	1.00	1.00	1.00	1.00	0.99
32	1.00	1.00	0.99	1.00	0.99	0.99
48	1.00	1.00	0.99	1.00	0.98	0.98
64	1.00	0.99	0.99	1.00	0.97	0.98
128	0.99	0.99	0.97	0.99	0.96	0.97
256	1.00	1.00	0.97	0.99	0.93	0.97

(b) Correlation coefficients

FIG. 9. (Color online) Ambient pressure sensitivity of the subharmonic component for Levovist when the ambient pressure is increased from 0 to 25 kPa. The sensitivity is shown as a function of acoustic pressure and number of cycles in the rectangular driving pressure. To the right, the respective correlation coefficients, when using a linear regression model, are shown.



(a)

N_c \ P_{ac}	275	300	325	350	375	400
10	0.13	0.14	0.16	0.17	0.16	0.13
20	0.16	0.19	0.26	0.31	0.29	0.18
32	0.18	0.20	0.29	0.45	0.40	0.19
48	0.18	0.20	0.31	0.65	0.49	0.20
64	0.18	0.21	0.33	0.65	0.58	0.21
128	0.14	0.20	0.34	0.84	0.76	0.21
256	0.20	0.26	0.41	1.14	1.00	0.21

Pressure sensitivity [dB/kPa]

N_c \ P_{ac}	275	300	325	350	375	400
10	1.00	1.00	1.00	1.00	1.00	1.00
20	0.99	1.00	1.00	1.00	0.99	0.99
32	1.00	0.99	0.99	0.99	0.98	0.99
48	1.00	0.99	0.99	0.99	0.97	0.99
64	1.00	0.99	0.98	0.99	0.98	0.98
128	1.00	0.99	0.98	0.97	0.94	0.98
256	1.00	1.00	0.98	0.98	0.90	0.98

(b) Correlation coefficients

FIG. 10. (Color online) Ambient pressure sensitivity of the subharmonic component for Sonazoid when the ambient pressure is increased from 0 to 25 kPa. It is shown as a function of acoustic pressure and number of cycles in the rectangular driving pressure.

gression for each simulation setup when only the overpressure is changed. Next, the correlation coefficient r has been calculated to see how well a linear relationship between subharmonic energy and ambient overpressure can be assumed. The respective correlation coefficients are shown to the right in Fig. 9. Figure 9 shows very clearly two characteristics: The optimal driving pressure is 775 kPa, which is in the upper end of the growth stage of the subharmonic component. Lower acoustic pressures do not generate as much subharmonic energy while higher driving pressures cause poorer sensitivity due to saturation. Furthermore, the sensitivity is increased as the driving pulse length is increased. This indicates, unfortunately, that a compromise between axial resolution and pressure sensitivity exists. The correlation coefficients to the right in Fig. 9 indicate a very good linearity. In fact, it can be seen that the two lowest coefficients are actually for the two simulations in Fig. 8 with 256 and 128 cycles, respectively. The rest of the coefficients are all equal to or above $r=0.97$. The maximum sensitivity for Levovist was achieved using a rectangular pulse of 256 cycles with a driving pressure of 775 kPa. Using this setting, a reduction in the subharmonic energy was simulated to be 22.0 dB ($r=0.99$), giving a pressure sensitivity of 0.88 dB/kPa. For a shorter driving pulse with 64 cycles, the best pressure sensitivity was found to be 0.49 dB/kPa ($r=1.0$).

Examining the results for Sonazoid gives the same indications as for Levovist, although the results are not as symmetric around a certain acoustic pressure. However, once again there is a clear tendency that a specific acoustic pressure in the upper end of the growth stage will optimize the ambient pressure sensitivity. Furthermore, the simulations also indicate the same relation between sensitivity and pulse length. The findings for Sonazoid are summarized in Fig. 10, which is the same as Fig. 9 for Levovist. Using a rectangular driving pulse with 256 cycles, a maximum pressure sensitivity of 1.14 dB/kPa ($r=0.96$) was found. For a driving pulse with 64 cycles, the best sensitivity was found to be 0.65 dB/kPa with a linear correlation coefficient of $r=0.99$.

IV. CONCLUSION

A simulation study consisting of 7200 simulations has been carried out to investigate and optimize the subharmonic response sensitivity to ambient pressure changes. Two different types of ultrasound contrast agents, corresponding to Levovist and Sonazoid, were simulated. While the param-

eters of the microbubbles were kept fixed, the parameters describing the driving pulse and ambient overpressure were changed in each simulation.

Simulations show that the subharmonic component is more easily generated using a rectangular shaped driving pulse compared to a Hanning shaped signal. For the case of Levovist, it was not possible to generate the subharmonic using the Hanning shaped excitation even for very high acoustic driving pressures. This dissimilarity in responses makes a study of the differences in shell properties of Levovist and Sonazoid interesting.

Investigations of the subharmonic energy as function of ambient overpressure showed two tendencies very clearly: The amount of reduction in energy of the subharmonic component is dependent on acoustic driving pressure and peaks when the acoustic pressure is in the upper end of the growth stage. Second, the investigations also showed a clear relation between the amount of energy reduction and length of the driving pulse. Simulations of Levovist indicate a linear change in energy of the subharmonic component as a function of ambient overpressure. Changing the overpressure from 0 to 25 kPa indicates pressure sensitivities of 0.49 and 0.88 dB/kPa for a rectangular driving pulse with 64 and 256 cycles, respectively. For Sonazoid, the sensitivities were found to be 0.65 and 1.14 dB/kPa when using the same excitation pulses as for Levovist.

ACKNOWLEDGMENTS

The authors wish to thank Lars Hoff for making BUBBLESIM publicly available. This work was supported by Grant No. 26-04-0024 from the Danish Science Foundation, the Technical University of Denmark, and by B-K Medical Aps.

[1]D. H. Evans, W. N. McDicken, R. Skidmore, and J. P. Woodcock, *Doppler Ultrasound, Physics, Instrumentation, and Clinical Applications* (Wiley, New York, 1989).
[2]A. C. Burton, *Physiology and Biophysics of the Circulation*, 2nd ed. (Year Book Medical, Chicago, IL, 1972).
[3]A. L. Strauss, F. J. Roth, and H. Rieger, "Noninvasive assessment of pressure gradients across iliac artery stenoses: Duplex and catheter correlative study," J. Ultrasound Med. **12**(1), 17–22 (1993).
[4]W. M. Fairbank and M. O. Scully, "A new noninvasive technique for cardiac pressure measurements: Resonant scattering of ultrasound from bubbles," IEEE Trans. Biomed. Eng. **24**, 107–110 (1977).
[5]B. Hok, "A new approach to noninvasive manometry: Interaction between ultrasound and bubbles," Med. Biol. Eng. Comput. **19**, 35–39 (1981).
[6]P. M. Shankar, J. Y. Chapelon, and V. L. Newhouse, "Fluid pressure measurement using bubbles insonified by two frequencies," Ultrasonics **24**, 333–336 (1986).
[7]V. L. Newhouse and P. M. Shankar, "Bubble size measurements using the nonlinear mixing of two frequencies," J. Acoust. Soc. Am. **75**, 1473–1477 (1984).
[8]A. Bouakaz, P. J. Frinking, N. de Jong, and N. Bom, "Noninvasive measurement of the hydrostatic pressure in a fluid-filled cavity based on the disappearance time of micrometer-sized free gas bubbles," Ultrasound Med. Biol. **25**, 1407–1415 (1999).
[9]A. Bouakaz, P. J. Frinking, and N. de Jong, "Noninvasive pressure measurement using microbubble contrast agent and wavelet transforms," Proc.-IEEE Ultrason. Symp. **2**, 1907–1910 (2000).
[10]W. T. Shi, F. Forsberg, J. S. Raichlen, and L. Needleman, "Pressure dependence of subharmonic signals from contrast microbubbles," Ultrasound Med. Biol. **25**, 275–283 (1999).
[11]L. M. Leodore, F. Forsberg, and W. T. Shi, "In vitro pressure estimation obtained from subharmonic contrast microbubble signals," Proc.-IEEE Ultrason. Symp., 2207–2210 (2007).
[12]F. Forsberg, J.-B. Liu, W. T. Shi, J. Furuse, M. Shimizu, and B. B. Goldberg, "In vivo pressure estimation using subharmonic contrast microbubble signals: Proof of concept," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 581–583 (2005).
[13]D. Adam, M. Sapunar, and E. Burla, "On the relationship between encapsulated ultrasound contrast agent and pressure," Ultrasound Med. Biol. **31**, 673–686 (2005).
[14]K. S. Andersen and J. A. Jensen, "In vitro measurement of ambient pressure changes using a realistic clinical setup," Proc.-IEEE Ultrason. Symp., 1096–1099 (2008).
[15]K. S. Andersen and J. A. Jensen, "Non-invasive estimation of blood pressure using ultrasound contrast agents," in International Congress on Ultrasonics (2009).
[16]L. Rayleigh, "On the pressure developed in a liquid during the collapse of a spherical cavity," Philos. Mag. **34**, 94–98 (1917).
[17]M. Minnaert, "On musical air-bubbles and the sound of running water," Philos. Mag. **16**, 235–248 (1933).
[18]M. S. Plesset, "The dynamics of cavitation bubbles," J. Appl. Mech. **16**, 277–282 (1949).
[19]N. de Jong and L. Hoff, "Ultrasound scattering properties of albunex microspheres," Ultrasonics **31**, 175–181 (1993).
[20]C. C. Church, "The effects of an elastic solid surface layer on the pulsations of gas bubbles," J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[21]K. E. Morgan, J. S. Allen, P. A. Dayton, J. Chomas, A. Klibanov, and K. Ferrara, "Experimental and theoretical evaluation of microbubble behavior: Effect of transmitted phase and bubble size," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1494–1509 (2000).
[22]H. J. Vos, F. Guidi, E. Boni, and P. Tortoli, "Method for microbubble characterization using primary radiation force," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1333–1345 (2007).
[23]N. de Jong, L. Hoff, and N. Bom, "Absorption and scatter of encapsulated gas filled microspheres: Theoretical considerations and some measurements," Ultrasonics **30**, 95–103 (1992).
[24]L. Hoff, *Acoustic Characterization of Contrast Agents for Medical Ultrasound Imaging* (Kluwer Academic, Dordrecht, 2001).
[25]K. S. Andersen and J. A. Jensen, "Simulation of microbubble response to ambient pressure changes," Med. Imag. V Symp. **6920**, 692016 (2008).
[26]L. Trilling, "The collapse and rebound of a gas bubble," J. Appl. Phys. **23**, 14–17 (1952).
[27]J. B. Keller and M. Miksis, "Bubble oscillations of large amplitude," J. Acoust. Soc. Am. **68**, 628–633 (1980).
[28]S. Hilgenfeldt and D. Lohse, "The acoustics of diagnostic microbubbles: Dissipative effects and heat deposition," Ultrasonics **38**, 99–104 (2000).
[29]J.-F. Yu, D. Zhang, X.-F. Gong, Y.-J. Gong, Z.-M. Zhu, and X.-M. Liu, "Frequency dependences of sound attenuation and phase velocity in suspensions containing encapsulated microbubbles," Chin. Phys. Lett. **22**, 892–895 (2005).
[30]K. Sarkar, W. T. Shi, D. Chatterjee, and F. Forsberg, "Characterization of ultrasound contrast microbubbles using in vitro experiments and viscous and viscoelastic interface models for encapsulation," J. Acoust. Soc. Am. **118**, 539–550 (2005).
[31]W. Lauterborn and E. Cramer, "Subharmonic route to chaos observed in acoustics," Phys. Rev. Lett. **47**, 1445–1448 (1981).
[32]A. Prosperetti, "Application of the subharmonic threshold to the measurement of the damping of oscillating gas bubbles," J. Acoust. Soc. Am. **61**, 11–16 (1977).

Erratum: "Acoustic signal horizontal coherence variability: Relationship to internal tide and storm events" [J. Acoust. Soc. Am. 126(4), 2158 (2009)]

Marshall H. Orr
Acoustics Division, Naval Research Laboratory, 4555 Overlook Ave., S.W., Washington, DC 20375-5032

Peter C. Mignerey
Naval Research Laboratory, Washington, DC 20375-5032

David Walsh
Naval Research Laboratory, Stennis Space Center, Mississippi 39529-5004

(Received 12 October 2009; accepted 12 October 2009)
[DOI: 10.1121/1.3257650]

PACS number(s): 43.30.Re, 43.10.Vx

There is a publisher's error in the list of authors of abstract 1aAO6 in the published program for the San Antonio meeting. The correct author and affiliation list is as follows:

Marshall H. Orr (Acoust. Div., Naval Research Lab., 4555 Overlook Ave., S.W., Washington, DC 20375-5032, marshall.orr@nrl.navy.mil), Peter C. Mignerey (Naval Res. Lab., Washington, DC 20375-5032), and David Walsh (Naval Res. Lab., Stennis Space Ctr., MS 39529-5004).

 0001-4966/2009/126(6)/3359/1/$25.00

Erratum: "Improving mp3 capability of mobile phones by linking acoustics information with vibrations" [J. Acoust. Soc. Am. 126(4), 2185 (2009)]

Cheol Hong Kim
Chonnam National University, Department of Electronics and Computer Engineering, 300 Yongbong-Dong, Buk-Gu Gwangju 500-757, Korea

Young Jin Park
Chonnam National University, Buk-Gu Gwangju 500-757, Korea

Jong-Myon Kim
University of Ulsan, Korea

(Received 12 October 2009; accepted 12 October 2009)
[DOI: 10.1121/1.3257648]

PACS number(s): 43.60.Vx, 43.10.Vx

There is a publisher's error in the list of authors of abstract 1pSPb4 that appears in the published program of the San Antonio meeting. The correct author and affiliation list is as follows:

Cheol Hong Kim (Chonnam Natl. Univ., Dept. of Electron. and Comp. Eng., 300 Yongbong-Dong, Buk-Gu Gwangju 500-757, Korea), Young Jin Park (Chonnam Natl. Univ., Korea), and Jong-Myon Kim (Univ. of Ulsan, Korea).

ACOUSTICAL NEWS

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.*

Announcement of the 2010 Election

In accordance with the provisions of the bylaws, the following Nominating Committee was appointed to prepare a slate for the election to take place on 9 April 2010:

Gilles Daigle, Chair
Dani Byrd
Gary Elko
Ronald Freiheit
Jody Kreiman
Alexandra Tolstoy

The bylaws of the Society require that the Executive Director publish in the *Journal*, at least 90 days prior to the election date, an announcementof the election and the Nominating Committee's nominations for the offices to be filled. Additional candidates for these offices may be provided by any Member or Fellow in good standing by letter received by the Executive Director not less than 60 days prior to the election date, and the name of any eligible candidate so proposed by 50 Members or Fellows shall be entered on the ballot. Biographical information about the candidates and statements of objectives of the candidates for President-Elect and Vice President-Elect will be mailed with the ballots.

CHARLES E. SCHMID
Executive Director

The Nominating Committee has submitted the following slate:

For President-Elect

Mardi C. Hastings

Peggy B. Nelson

For Vice President-Elect

Fredericka Bell-Berti

Brigitte Schulte-Fortkamp

 0001-4966/2009/126(6)/3361/12/$25.00

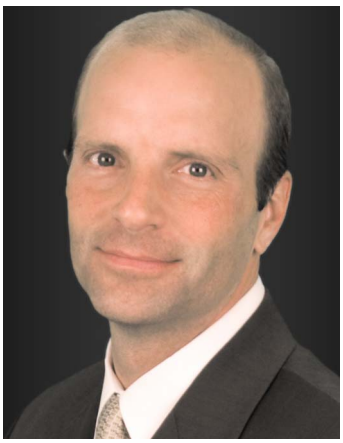
For Members of the Executive Council

Ann R. Bradlow

Michael J. Buckingham

Stan E. Dosso

Barbara G. Shinn-Cunningham



Brandon D. Tinianov

Peter F. Worcester

Preliminary Notice: Joint Meeting: 159th Meeting of the Acoustical Society of America and NOISE-CON 2010

The 159th Meeting of the Acoustical Society of America will be held Monday through Friday, 19–23 April 2010 at the Baltimore Marriott Waterfront Hotel, Baltimore, MD, USA. NOISE-CON 2010 will be held in conjunction with the meeting, Monday through Wednesday, 19–21 April. A block of rooms has been reserved at the Baltimore Marriott Waterfront Hotel at special ASA rates.

Information about the meeting also appears on the ASA Home Page at ⟨http://asa.aip.org/meetings.html⟩.

Charles E. Schmid
Executive Director

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 19–23 April 2010 for the ASA meeting and Monday through Wednesday, 19–21 April for NOISE-CON 2010.

Special Sessions

Acoustical Oceanography (AO)

Acoustics in polar environments
(Joint with Animal Bioacoustics and Underwater Acoustics)
Organized by: Jennifer L. Miksis-Olds, David K. Mellinger, Ann E. Bowles
Scattering, propagation, sound fields, and animal bioacoustics in polar environments
Impact of shallow water acoustic propagation by linear internal waves and neutrally buoyant intrusions
(Joint with Underwater Acoustics)
Organized by: Dajun J. Tang, David L. Bradley
Measurements, modeling and analysis of internal waves and neutrally buoyant intrusions and the consequent impact on acoustic signal behavior in shallow water

Animal Bioacoustics (AB)

Auditory attention, learning and memory: From neurons to behavior
(Joint with Psychological and Physiological Acoustics)
Organized by: Cynthia F. Moss
Recent research findings that shed light on the role of cognition in auditory processing and perception in a variety of animal species
Effects of anthropogenic noise on aquatic animals
Organized by: Carl R. Schilt

Studies of the specific and general effects of anthropogenic noise on all aquatic life, from invertebrates to marine mammals
Estimating spatial density of animal populations with passive acoustics
Organized by: David K. Mellinger
Methods, case studies, and results of passive acoustic estimation of the spatial density of animal populations
Incorporating context into acoustic automated detection and classification algorithms
(Joint with Noise and Signal Processing in Acoustics)
Organized by: Aaron M. Thode
Role of context on detection and logging and classification of animal calls
Signal processing for subtle and complex animal communications
(Joint with Signal Processing in Acoustics)
Organized by: Sean K. Lehman, Ann E. Bowles
Opportunity for signal processors to understand the issues with the analysis of animal communications to improve analysis and understanding, with the goal of matching signal processing solution with current animal bioacoustics research needs

Architectural Acoustics (AA)

Required accuracy of absorption, scattering, and diffusion coefficients
Organized by: Peter D'Antonio
Limitations, available libraries, new measurements, improved data, etc.
Acoustics of green buildings
(Joint with Noise and INCE)
Organized by: Brandon D. Tinianov, David M. Sykes
Acoustical issues in green buildings that result in low user ratings and performance issues
Case studies, applications, and integration of architectural acoustics in building information modeling three-dimensional modeling
(Joint with Noise and INCE)
Organized by: Norman H. Philipp
Architectural acoustics can have or might be integrated in the evolution of design in building information modeling
"Diamonds in the rough" of "hidden gems"
Organized by: Andrew N. Miller
Spaces that possess outstanding acoustical attributes but are not necessarily known about within the popular scheme of acoustical places
Physical acoustics in Boston Symphony Hall: A guide for the perplexed
(Joint with Physical Acoustics)
Organized by: James B. Lee
Physical acoustics, i.e., proximity effects on stage, in concert halls
Primary and secondary school special function spaces
(Joint with Noise)
Organized by: Robert C. Coffeen
Acoustical issues related to the multipurpose spaces in schools
Rooms for reproduced sound
(Joint with Engineering Acoustics)
Organized by: K. Anthony Hoover and Alexander U. Case
Interaction of reproduced sound and architecture in studios, cinemas, home theaters, schools, installations, and more
Speech intelligibility and privacy
(Joint with Speech Communication)
Organized by: Eric L. Reuter
Effects of architecture on speech communication

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Blast-induced traumatic brain injury: Mechanisms, assessment, therapy, and mitigation
(Joint with Animal Bioacoustics, Speech Communication, and Psychological and Physiological Acoustics)
Organized by: Steven G. Kargl, William C. Moss, Thomas J. Matula
Exposure to blast waves from improvised explosive devices has given rise to a significant increase in traumatic brain injury (TBI). Treatment of TBI requires a better understanding of the physics and biological mechanisms of injury, assessments of damage, therapies, and mitigation
Numerical modeling for medical ultrasound
(Joint with Physical Acoustics)
Organized by: Robert J. McGough
Linear and nonlinear models of ultrasound propagation in biological tissues such as models of frequency dependent attenuation and acoustic backscatter. Numerical approaches and software packages used to solve these problems will also be explored
Ultrasonic characterization of bone
Organized by: Keith A. Wear
Theoretical modeling and experimental measurements regarding attenuation, velocity, and scattering properties of cancellous and cortical bone and the use of ultrasound in the diagnosis and management of osteoporosis
Ultrasonically activated agents
(Joint with Physical Acoustics)
Organized by: Oliver D. Kripfgans
Review of fluid, gas, or other-phase based agents for therapeutic and diagnostic ultrasound. These have evolved to active agents that can be activated with spatial and temporal control
Ultrasound induced cellular bioeffects
Organized by: E. Carr Everbach, Stuart B. Mitchell
Effects of ultrasound on mesoscale biological systems through the action of heating, cavitation, or other processes

Education in Acoustics (ED)

Diversity issues in education in acoustics
(Joint with ASA Committee on Diversity)
Organized by: Preston S. Wilson, Juan I. Arvelo, Jr.
Diversity issues in acoustics education followed by a panel discussion to identify actions that the Acoustical Society of America and American Institute of Physics may take to foster diversity in acoustics
Listen up and get involved
(Joint with ASA Women in Acoustics Committee)
Organized by: Marcia J. Iskason, James M. Sabatier
Acoustic demonstrations for middle- and high-school age Girl Scouts

Engineering Acoustics (EA)

Acoustic impedance of the ear
(Joint with Psychological and Physiological Acoustics)
Organized by: Daniel M. Warren, Susan E. Voss
Modeling and measuring the acoustic impedance of the ear, including wideband reflectance (WBR) measures, and their utility for fitting hearing aids or identifying pathologies. Implications of open-canal versus occluded-canal fittings of hearing aids as well as other techniques for modifying the impedance of the ear (for example, active occlusion reduction) will also be considered
Computer modeling for complex acoustic environments
(Joint with Acoustical Oceanography and Underwater Acoustics)
Organized by: Kenneth M. Walsh
Methods for modeling complex acoustic environments involving stratified media such as shallow water, sub-bottom, or deep sub-bottom for oil survey
Directional Microphone Arrays
(Joint with Signal Processing in Acoustics)
Organized by: James E. West, Joshua D. Atkins
Microphone arrays and the application of digital signal processing for feature enhancement
Electret condenser microphones
Organized by: Allan J. Zuckerwar, Quamar Shams
Fabrication, properties, calibration, applications, historical development of electrets

Musical Acoustics (MU)

Extended instrument techniques
Organized by: Jonas Braasch
The acoustics of instruments when played with techniques that expand beyond the traditional repertoire of the instrument
Homemade musical instruments for teaching acoustics
Organized by: Thomas D. Rossing, Preston S. Wilson, Andrew Morrison
Students often build simple musical instruments in high-school physics classes to learn physical principles. Students and others are invited to describe and demonstrate these instruments and explain their physical principles

Measurement and modeling of the acoustic properties of the banjo
(Joint with Signal Processing in Acoustics and Engineering Acoustics)
Organized by: Grace A. Clark
Recent studies as applied to the acoustic properties of the banjo
String instruments
Organized by: Paul A. Wheeler
Research on the acoustics of string instruments including modeling, history, and performance
The contemporary traditional violin
Organized by: George A. Bissinger
The acoustics of the traditional violin as used in contemporary music (music performed today)

Noise (NS)

Acoustics and public policy
(Joint with Animal Bioacoustics and INCE)
Organized by: Nancy S. Timmerman
Exploring the effectiveness of public policies in acoustics
Effect of noise on humans and non-human animals
(Joint with Animal Bioacoustics and INCE)
Organized by: Brigitte Schulte-Fortkamp, Ann E. Bowles
Techniques, measurements, and further approaches in evaluation
Healthcare acoustics/noise and occupant perception and performance
(Joint with Architectural Acoustics)
Organized by: Kenneth P. Roy, Erica E. Ryherd, Mandy Kuchur
Combination of architectural design and noise sources within the hospital/clinic will affect both patients and health care professionals in terms of intelligibility, privacy, annoyance, and disturbance
Low frequency noise
(Joint with Architectural Acoustics and INCE)
Organized by: Natalia V. Sizov, Detlef Krause
Need for measurement techniques
Military noise environments—Continuous and impulsive
(Joint with Physical Acoustics and INCE)
Organized by: Richard L. McKinley
Descriptions of intense military noise environments, nearfield, farfield, and community and effects on personnel and structures
Noise and its control in complex and urban environments
(Joint with INCE)
Organized by: Siu-Kit Lau, Kai-Ming Li
State-of-the-art technologies to predict and optimally design outdoor noise barriers, e.g., T-shape, etc., façade, and other noise control devices
Rocket noise environments
(Joint with Physical Acoustics and INCE)
Organized by: Kent L. Gee, Jeremy Kenny
Characterization and prediction of rocket noise sources and propagation environments
Soundscape concert
(Joint with Architectural Acoustics)
Organized by: Brigitte Schulte-Fortkamp, Alexander U. Case
Experiencing the soundscape intensity in specific spaces and evaluations
Standards in psychoacoustics
(Joint with ASA Committee on Standards, Psychological and Physiological Acoustics, Animal Bioacoustics and INCE)
Organized by: Klaus Genuit
Outdoor noise and the need for the use of the new standards in psychoacoustics
Ventilation, fan, and duct noise control
(Joint with Architectural Acoustics and INCE)
Organized by: Lixie Huang
Latest developments in fan/ventilation noise source identification, aeroacoustic characterization and noise abatement techniques, with or without a duct component. Duct acoustics (theoretical or experimental), new duct noise control and modeling techniques. Industrial problems and solutions
Wind turbine noise
(Joint with ASA Committee on Standards and INCE)
Organized by: Robert D. Hellweg, Ken Kaliski, Mark Storm
Latest issues in wind turbine noise: Low-frequency noise, impact assessment and measurement methods

NOISE-CON 2010

Acoustics of technology enabled spaces
(Joint with Architectural Acoustics)
Organized by: Jeff Babich, Greg Coudriet
Acoustics of technology-enabled spaces including video-conference rooms, collaborative work-spaces, presentation spaces, 3-D visualization spaces, student learning commons, and digital libraries
Active noise control
(Joint with Noise and Signal Processing in Acoustics)
Organized by: Marty Johnson
Active control of noise and vibration generated by sources of noise and received in open and enclosed spaces
Aircraft exterior noise
(Joint with Noise and Physical Acoustics)
Organized by: Pierre Lempereur, Eric Boeker
Measurements and modeling of noise generation by and radiated outside aircraft, including aircraft engines
Aircraft interior noise
(Joint with Noise and Structural Acoustics and Vibration)
Organized by: Vincent Cotoni
Concerned with prediction and control of aircraft interior noise with emphasis on (i) case studies of noise control applications, (ii) development and application of new methods for prediction of interior noise, and (iii) measurement methods for interior noise or exterior excitation
Acoustics of energy efficient building systems
(Joint with Noise and Architectural Acoustics)
Organized by: Jeff Fullerton
Discussion about energy efficient building systems, which are dramatically changing the acoustics of the built environment
Automotive and powertrain noise and vibration
(Joint with Noise)
Organized by: Gordon Ebbitt, Terrence Connelly
Modeling and measurement of noise and vibration generated in automobiles, e.g. noise, vibration and harshness (NVH), including engineering controls
Building design and construction for effective acoustic performance
(Joint with Noise and Architectural Acoustics)
Organized by: Kenric Van Wyk
Building noise control design that appears satisfactory on the surface can be undermined by lack of attention to construction details resulting in flanking, leaks and other detrimental transfer of airborne and structureborne noise
Community noise
(Joint with Noise)
Organized by: Larry Finegold
Effects of noise, noise policies, and cost studies
Consumer and industrial product noise
(Joint with Noise)
Organized by: Chuck Hayden
Noise generation and impact of noise generated by production found in the home and industry, and modeling and applications of engineering noise controls
Construction noise
(Joint with Noise)
Organized by: Erich Thalheimer
Noise generation at construction sites, the impact of constructions noise on surrounding communities, and methods for the control of construction noise
Experimental techniques and instrumentation in noise and vibration
(Joint with Engineering Acoustics)
Organized by: Jim Thompson
New and novel measurement techniques and instrumentation to quantify noise and vibration, covering a broad range of environmental, transportation, and industrial measurement situations
Flow noise
(Joint with Physical Acoustics)
Organized by: Dean Capone
Methods of describing the generation and radiation of noise generated by flow, including jets, flow over boundaries, flow in turbomachinery, flow in heating, ventilation and air conditioning systems and supersonic flow
Ground-borne noise and vibration modeling

(Joint with Noise)
Organized by: James T. Nelson
Modeling the propagation of noise and vibration through the ground from transportation sources to buildings, and methods for the control of ground-borne noise and vibration propagation
Industrial and power plant noise
(Joint with Noise)
Organized by: Frank Brittain
Modeling community and equipment noise from industrial, process, and power facilities
Information technology noise
(Joint with Noise)
Organized by: Jeff Schmidt, Terry Baird, Marco Beltman
Noise generated by information technology equipment, e.g., computers, and methods for its control
Materials for noise control—Manufacturer presentations
Organized by: Steve Roth
Presentations of the effectiveness of materials for noise control
Mufflers and silencers
(Joint with Noise)
Organized by: Mark Storm
Developments in active, passive, reactive, and dissipative sound attenuation for motorcycles, passenger cars, trucks, and buses
Noise and vibration in the mining industry
(Joint with Noise)
Organized by: David Yantek
Mechanisms of the generation by and radiation from mining equipment, impact of noise from mining equipment, and the application of engineering controls
Noise from airports
(Joint with Noise)
Organized by: Micah Downing
The assessment of the impact of noise from airports on individuals (e.g., sleep disturbance) and the surrounding community, and methods for the reduction of the impact, such as traffic control and treatments of housing
Noise from transit systems
(Joint with Noise)
Organized by: Hugh Saurenman
Noise issues from high-speed rail and strategies for accelerated noise studies
Noise source localization
(Joint with Noise and Signal Processing in Acoustics)
Organized by: Jim Thompson, Karin Haddad
Experimental techniques to localize and characterize sound sources in complex and multiple source environments
Numerical methods in noise emission and immission
(Joint with Noise and Structural Acoustics and Vibration)
Organized by: Bryce Gardner
Numerical modeling methods, e.g., finite element and boundary element methods, for describing vibration of and radiation from complex sources of noise and vibration
Recreational noise
(Joint with Noise and ASA Committee on Standards)
Organized by: Paul Burge
Noise from, or in, recreational venues or sources such as race tracks, gun clubs, sport venues, outdoor performance venues, amusement parks, and nightclubs
Sound propagation in the atmosphere
(Joint with Noise and Physical Acoustics)
Organized by: Keith Attenborough
Modeling and measurement of sound propagation in the atmosphere
Sound quality
(Joint with Noise and Psychological and Physiological Acoustics)
Organized by: Wade Bray
Analytic approaches, techniques; case studies of sound quality issues resolved/measured by sound quality tools and techniques
Statistical energy analysis and energy methods
(Joint with Noise and Structural Acoustics and Vibration)
Organized by: Vincent Cotoni
Statistical energy analyses (SEA) and other energy-based methods of modeling the propagation of vibration and noise through coupled complex structures and surrounding acoustic media
Tire/pavement noise
(Joint with Noise and Structural Acoustics and Vibration)
Organized by: Courtney B. Burroughs, Rob Rasmussen
Mechanisms of generations of noise by tire/pavement interaction, including modeling and measurements of tire noise and the affects of pavements surfaces on tire noise
Vibration damping for noise control
(Joint with Noise and Structural Acoustics and Vibration)
Organized by: Mark Downing
The following technical areas are of particular interest: (i) case studies of vibration damping and materials, (ii) the theory of mechanisms for vibration damping, (iii) measurement of damping effectiveness in structures and, (iv) the application of dynamic absorbers, constrained layer damping, and other mechanisms to structures and panels
Vibrations and structureborne noise in buildings
(Joint with Noise and Structural Acoustics and Vibration)
Organized by: Jeff Kwolkowski, Christopher Moran
Transmission of vibrations through building structures and radiation of noise from building structures
Workshop: Public outreach on community noise
Organized by: Larry Finegold
Workshop to discuss outreach to the public on the effects of community noise and methods of reducing these effects

Physical Acoustics (PA)

Shock wave and high strain rate probes of materials
Organized by: Albert Migliori
Shocks from gas guns, or Hopkinson bar high strain rate probes explore important properties of materials under extreme conditions
Sonic boom
(Joint with Structural Acoustics and Vibration and Noise)
Organized by: Victor W. Sparrow, Natalia V. Sizov, Kimberly Lefkowitz
Propagation of sonic booms through the atmosphere, structural transmission of boom noise into buildings, and human response to sonic boom. Experimental measurements, numerical modeling, and sonic boom signature, reproduction, and acceptability studies
Ultrasonic, nonlinear acoustics, acousto-optics, and engineering acoustics in honor of Mack A. Breazeale
(Joint with Engineering Acoustics)
Organized by: Lev A. Ostrovsky, Laszlo Adler
Dedicated to the memory of Professor Mack Breazeale, scientist and gentleman. Recognition of his pioneering work in ultrasonics, nonlinear acoustics, acousto-optics, and engineering acoustics, plus personal reminiscences

Psychological and Physiological Acoustics (PP)

Application of auditory models to hearing aids
(Joint with Engineering Acoustics)
Organized by: Brent W. Edwards
Exploration of the many ways in which auditory models can be applied to the various aspects of hearing aid application: Diagnostics, fitting algorithms, audio signal processing, sound quality and speech intelligibility outcome measures
Music and the brain
(Joint with Musical Acoustics)
Organized by: Xiaoqin Wang
Neural mechanisms for music processing in the brain
Noise-induced hearing loss: From physiology to prevention
(Joint with Noise and INCE)
Organized by: Sharon G. Kujawa, Beth Cooper
Physiological bases and perceptual consequences and prevention of noise-induced hearing loss

Signal Processing in Acoustics (SP)

Battlefield acoustics
(Joint with Engineering Acoustics and Physical Acoustics)
Organized by: Michael V. Scanlon, David H. Chambers
Measurement, propagation, and processing of acoustic signals in a battlefield environment

Classification methods in acoustics and non-Gaussian noise
(Joint with Underwater Acoustics and Animal Bioacoustics)
Organized by: R. Lee Culver
Seeking new or classification methods being applied to acoustic, applications, especially involving non-Gaussian statistics
Maximum entropy and Bayesian signal processing
(Joint with Underwater Acoustics and Architectural Acoustics)
Organized by: Ning Xiang, Zoi-Heleni Michalopoulou
Application of maximum entropy principles and Bayesian probabilistic inference in acoustical signal processing, inversion, uncertainty estimates
Sparse approximations in signal processing
(Joint with Underwater Acoustics)
Organized by: James C. Preisig
New methods of sampling signals to increase resolution with fewer numbers of samples

Speech Communication (SC)

Exploring the relationship between cognitive processes and speech perception: Part II
(Joint with Psychological and Physiological Acoustics)
Organized by: Amee P. Shah
Building on the introductory session from the 2009 Portland meeting, advancements made on the understanding of the relationship between cognitive processes and speech perception will be explored. The session will specifically focus on identifying methods, variables, and theories that help test this relationship
Machine learning techniques for speech recognition
(Joint with Signal Processing in Acoustics)
Organized by: Mark A. Hasegawa-Johnson
Juxtapose papers exemplifying the breadth and depth of the symbiosis between machine learning and speech recognition and discuss the grand challenges that drive current research
Speech and noise
Organized by: Carol Y. Espy-Wilson
Methods for coping with speech in noise for automatic speech recognition and human perception
Speech for tracking human health state, performance, and emotional state
Organized by: Suzanne E. Boyce
Methods for detecting, monitoring, and/or tracking changes in cognitive, emotional and health states that are reflected in speech production and perception

Structural Acoustics and Vibration (SA)

Computer modeling of structural acoustics and vibration for complex structures
Organized by: James E. Phillips
Applications of computer analysis techniques to simulate the dynamic interaction of multiple structural acoustic subsystems
Damping mechanisms
(Joint with Physical Acoustics)
Organized by: J. Gregory McDaniel
Mechanisms that affect the amplitudes of vibration and radiation of sound from complex mechanical structures
Noise control of small submersible vehicles
(Joint with Noise, Underwater Acoustics, Animal Bioacoustics, and INCE)
Organized by: Joseph M. Cuschieri
Prediction, measurements, and noise abatement of submersible vehicles
Nondestructive testing of materials
(Joint with Engineering Acoustics)
Organized by: Hasson Tavossi
Theory and application of sound wave and ultrasonic techniques used in the detection and characterization of damage, cracks, or corrosion in materials, including damage detection in aircraft components
Space vehicle vibroacoustics
(Joint with INCE)
Organized by: Dean E. Capone
Prediction, analysis, and mitigation of vibration and noise in aerospace vehicles

Underwater Acoustics (UW)

Acoustic particle velocity and vector fields: Signal processing and communication applications
(Joint with Acoustical Oceanography and Signal Processing in Acoustics)
Organized by: Moshen Badiey, A. Abdi
Theory and application of particle velocity and vector fields for signal processing (SONAR, beamforming,…), underwater channel, as well as the emerging field of communication via particle velocity and vector sensors (transceiver design, channel modeling,…)
Deep water ambient noise
(Joint with Acoustical Oceanography)
Organized by: Michael J. Buckingham, Martin Siderius
Measurements and modeling of directionality, depth dependence, and coherence of ambient noise in deep water

Other Technical Events and Information

Hot Topics

A "Hot Topics" session sponsored by the Tutorials Committee will cover the fields of Musical Acoustics, Speech Communication, and Structural Acoustics and Vibration.

Distinguished Lecture

The Technical Committee on Structural Acoustics and Vibration will sponsor a Distinguished Lecture by Robin Langley of Cambridge University titled "Recent advances in the vibration analysis of structures with uncertain properties." It has long been recognized that the vibration and acoustical performance of a complex manufactured system, such as a automobile or an aircraft, can be sensitive to relatively small imperfections in the system properties. This sensitivity can lead to performance levels that are below the targets met by a nominally "perfect" system, and this can result in costly post-production modifications. This situation can be avoided if the response variability can be predicted (and therefore controlled) at the design stage, and this lecture will review recent advances in prediction capabilities. Both parametric and non-parametric models of uncertainty will be reviewed, and methods of propagating this uncertainty through a complex system model will be described. Particular emphasis will be placed on medium to high frequency vibration prediction, and a range of industrial applications of the theory will be described.

Exposition

A major technical exposition will be jointly sponsored by ASA and INCE. The exhibits feature 45 display booths, which will include computer-based instrumentation, multi-channel analyzers, sound quality systems, software for noise and vibration control analyses, acoustical materials, passive noise control devices, active control systems, and other products. The exposition, conveniently located near the meeting rooms, will open Monday evening, 19 April, and will close Wednesday at noon. Morning and afternoon refreshments will be available in the exposition area. Details regarding the exposition can be obtained by contacting Richard Peppin, Exposition Manager, Institute of Noise Control Engineering, c/o Scantek, Inc. 6450 A Dobbin Rd., Columbia, MD 21045, (T) 410-290-7726, (F) 410-290-9167, peppinr@verizon.net.

Topical Meeting on Signal Processing of Subtle and Complex Acoustic Signals in Animal Communication

The goal of a three-meeting series of sessions organized jointly by the Animal Bioacoustics and Signal Processing Technical Committees has been to match signal processing solutions with current animal bioacoustics research, adopting the philosophy that diverse problems may have similar solutions. Signal processors were introduced to issues in the analysis of subtle or complex features of animal calls in Portland (session 5aAB). Analytical methodologies will be discussed at the San Antonio meeting. In Baltimore, a one-day colloquium and discussion on the topic of "Signal Processing of Subtle and Complex Acoustic Signals in Animal Communication" will be held, sponsored jointly by Animal Bioacoustics and Signal Processing. In Baltimore, the sessions will be organized as a Topical Meeting to bring Animal Bioacoustics and Signal Processing specialists together to discuss ways of advancing the analysis of acoustic signals in animal commu-

nication. Focus will be on problems that have been solved by listeners up to now, such as parsing streams of calls collected by autonomous recording systems or identifying salient features in animal calls. After an introductory invited talk, subtopic sessions will focus on automated parsing of streams of real-world acoustic signals, caller localization, and feature extraction. Emphasis will be on developing recommendations for improving the state of the art and making signal processing methods more accessible to animal bioacousticians.

Technical Tours

Two technical tours are planned for the Baltimore meeting. On Monday afternoon Neil Shade will lead an architectural acoustics walking tour of the Mount Vernon arts district. Stops will include churches and concert halls with docents at each location. Cost is $18 for round trip bus transportation between the meeting hotel and tour location. This tour is limited to 30 people.

On Wednesday evening Murray Korman, with the help of professors and midshipmen, will lead a tour of laboratories at the U.S. Naval Academy in Annapolis. Participants will be able dine at the Officers Club at their own expense (approximate cost of entrees is $17 to $22). This tour is also limited to 30 people and cost is $18 for round-trip bus transportation.

Register for one or both tours online at ⟨http://asa.aip.org⟩.

Open Meetings of Technical Committees

Technical Committees will hold open meetings on Tuesday, Wednesday, and Thursday evenings at 7:30 p.m. These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings including proposals for special sessions, workshops and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussions.

Gallery of Acoustics

The Technical Committee on Signal Processing in Acoustics will sponsor the eleventh Gallery of Acoustics at the Baltimore meeting. The objective of the Gallery is to enhance ASA meetings by providing a setting for researchers to display their work to all meeting attendees in a forum emphasizing the diversity, interdisciplinary, and artistic nature of acoustics. The Gallery of Acoustics provides a means by which we can all share and appreciate the natural beauty, aesthetic, and artistic appeal of acoustic phenomena. This is a forum where science meets art.

The Gallery will consist of a multimedia collection of images, videos, audio clips, and narrations, of images and/or sounds generated by acoustic processes or resulting from signal and image processing of acoustic data. Images and videos can consist of actual visualizations of acoustic processes, or of aesthetically and technically interesting images resulting from various signal and image processing techniques and data visualization. Audio clips and segments should also have aesthetic, artistic, and technical appeal.

Entries must be submitted electronically, either by e-mail attachment, CD, or DVD. The allowed electronic formats are:

Images and Photographs must be in one of the following formats:

PDF, EPS, TIFF, JPG (although lossless formats are encouraged)

Video (3-minute limit STRICTLY ENFORCED)

QuickTime, MPEG (with QuickTime compatible CODEC)

Audio Clips (3-minute limit STRICTLY ENFORCED):

AU, WAV, MP3, AIFF

Each entry will be an individual chapter on a single multimedia DVD. Written posters, descriptions, and abstracts will be posted on the Gallery of Acoustics display surrounding the video monitor.

All entries must be accompanied by all authors' names and affiliations, a brief description of the entry and importance or interest of the entry (no more than 1000 words), and statement of permission to publish the entry in complete form or in parts.

The meeting attendees will vote on the entries on the basis of aesthetic/artistic appeal, ability to convey and exchange information, and originality. A cash prize of $500 will be awarded to the winning entry.

Note that authors must give permission to ASA for publication in complete form or in part to be eligible.

The relevant deadlines are as follows:

4 January 2010: Deadline for notice of intent to submit. Include a title, an abstract, and complete author list with full contact information. Please indicate the author who will be the primary point of contact.

1 February 2010: Deadline for the receipt of all entries and materials.

Entries, requests for information and all other communications regarding the Gallery should be directed to:

Dr. Sean K. Lehman
L-290, Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550-9234 USA
(925) 423-3580
(925) 423-3144 FAX
lehman2@llnl.gov

Getting Past the Aggravation and Getting your Institutional Review Board (IRB) Applications Approved

The IRB application process can appear to be a mysterious black box into which perfectly good studies are submitted and bizarre questions and requests emerge. This workshop will demystify the process and help you improve communication with your IRB. Topics include: Does my research need IRB approval? Isn't expedited supposed to mean fast? Do I have to use the informed consent template? What is the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and how do I make it go away? How can I enroll children without drowning in paperwork? Attendees are welcome to bring questions and cases for discussion.

Student Design Competition

The 2010 Student Design Competition will be displayed and judged at the Baltimore meeting. The Student Design Competition is intended to encourage students in the disciplines of architecture, engineering, physics, and other curriculums that involve building design and/or acoustics to express their knowledge of architectural acoustics and noise control in the design of a facility in which acoustical considerations are of significant importance.

The competition will be a poster session. Entries may be submitted by individual students or by teams of a maximum of three students. Undergraduate and graduate students are encouraged to participate. Students must be enrolled in either the Fall term of 2009 or the Spring term of 2010 to be eligible for the competition.

All competition entries will respond to a design scenario that was announced in the early fall of 2009. Information about the design scenario and registration for the competition will be available on the website of the Newman Fund, ⟨www.newmanfund.org⟩.

The Student Design Competition is sponsored by the Technical Committee on Architectural Acoustics, with support from the Wenger Foundation, the Robert Bradford Newman Student Award Fund, and the National Council of Acoustical Consultants.

Urban Design with Noise in Mind Symposium

The Acoustical Society of America is sponsoring a one-day symposium during the Baltimore meeting entitled "Urban Design with Noise in Mind." The symposium will be provided by several ASA members and public speakers as a public service to land use planners and the public at large. The morning portion of the symposium will focus on traditional methods used by land use planners to address noise impacts in the urban environment (zoning laws, noise regulations, building design codes, and noise mapping). The afternoon portion of the symposium will include discussion on the use of soundscaping methods to address noise impacts in the urban environment and a soundscape walk through the first Roman Catholic Cathedral built in the USA, the Basillica of the Assumption (which has been recently renovated), a walk through a Prayer Garden recently constructed on the cathedral site at the corner of a busy intersection, and a walk through an historic Unitarian Universalist facility located across the street from the garden. Data concerning sound levels and patron usage of all three facilities will be obtained prior to the symposium to help discuss the acoustic experience found during the soundscape walk. Representatives from the Physics Departments of local colleges and universities and ASA experts on soundscaping and active sound control will be invited to participate in the walk to discuss ways traffic noise impacts can be reduced within the facilities. All meeting attendees are invited to participate in the symposium. The fee to

attend the symposium will be waived for those registering for the conference. For additional information on the symposium, contact Kerrie Standlee at kstandlee@acoustechgroup.com.

Online Meeting Papers

The ASA provides the "Meeting Papers Online" website where authors of papers to be presented at meetings will be able to post their full papers or presentation materials for others who are interested in obtaining detailed information about meeting presentations. The online site will be open for author submissions in March. Submission procedures and password information will be mailed to authors with the acceptance notices.

Those interested in obtaining copies of submitted papers for this meeting may access the service at anytime. No password is needed. The URL is ⟨http://scitation.aip.org/asameetingpapers/⟩.

Proceedings of Meetings on Acoustics (POMA)

The upcoming meeting of the Acoustical Society of America will have a published proceedings, and submission is optional. The proceedings will be a separate volume of the online journal, "Proceedings of Meetings on Acoustics" (POMA). This is an open access journal, so that its articles are available in pdf format without charge to anyone in the world for downloading. Authors who are scheduled to present papers at the meeting are encouraged to prepare a suitable version in pdf format that will appear in POMA. The format requirements for POMA are somewhat more stringent than for posting on the ASA Online Meetings Papers Site, but the two versions could be the same. The posting at the Online Meetings Papers site, however, is not archival, and posted papers will be taken down six months after the meeting. The POMA online site for submission of papers from the meeting will be opened at the same time when authors are notified that their papers have been accepted for presentation. It is not necessary to wait until after the meeting to submit one's paper to POMA. Further information regarding POMA can be found at the site http://asa.aip.org/poma.html Published papers from previous meetings can be seen at the site http://scitation.aip.org/POMA.

Meeting Program

A complete meeting program will be mailed as Part 2 of the March issue of JASA. Abstracts will be available on the ASA Home Page ⟨http://asa.aip.org⟩ in February.

Tutorial Lecture on Animal Hearing

A tutorial presentation on "Animal Hearing" will be given by Robert J. Dooling of the University of Maryland on Monday, 19 April, at 7:00 p.m.

The auditory world of animals is, in many cases, quite different from our own, due to evolutionary pressures that have created anatomical and physiological differences in auditory structures. However, there are a number of commonalities of function and mechanisms across many species that point to general concepts of auditory perception. This tutorial lecture explores the field of animal hearing, also known as comparative psychoacoustics. Because the diversity in hearing organs among animals is considerable, the hearing of many animals is quite different from our own. We will review some of these differences in hearing across animal groups, from household pets to more exotic animals. This tutorial also highlights the specific advantages of the comparative approach and illustrates the many creative methods used for behavioral testing of hearing in animals. Species comparisons contribute to our understanding of the evolution of the auditory system, and can often clarify the relationship between structure and function in the auditory system as well as the effects of damage and repair. The study of animal hearing can lead to improved understanding of human auditory capabilities, and invites speculation about how human speech is adapted for intraspecies communication.

Lecture notes will be available at the meeting in limited supply. Those who register by 26 March are guaranteed receipt of a set of notes.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15 USD for registration received by 26 March and $25 USD at the meeting. The fee for students with current ID cards is $7 USD for registration received by 26 March and $12 USD at the meeting. To register, use the registration form in the printed call for papers or register online at ⟨http://asa.aip.org⟩.

Short Course on Array Signal Processing

Many applications in acoustics entail the remote sensing and analysis of propagating waves. Sensing systems are usually formed from multiple spatially separated sensors in order to obtain gains against spatially uncorrelated noise or spatially isolated interferences. A group of sensors is called an array. Array signal processing describes the combination of signals received at the sensors into beams primarily containing signals coming from specific angular sectors. This spatial filtering process is called beamforming and is identical to the focusing accomplished by parabolic dishes. Beamforming is used in a wide variety of applications in underwater acoustics, including active and passive sonar, marine mammal detection and localization, bottom mapping, oil exploration, and acoustic communications. Additionally, there are applications in many other fields such as radar, telecommunications, biomedical imaging, earthquake detection, and even hearing aid technology.

The course is suitable for graduate students or professionals. It is recommended that participants brush up on signal processing, linear algebra, and mathematical statistics before the course. A recommended but not required text for the course is "Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing" by Harry L. Van Trees (John Wiley & Sons, 2002).

The objective of this short course is to introduce the participant to a span of topics in array signal processing starting with the ubiquitous conventional beamformer and progressing to the more arcane adaptive beamformers and their statistical analysis. This will include exposure to the derivation of fundamental results.

Douglas Abraham obtained B.S., M.S., and Ph.D. degrees in Electrical Engineering and an M.S. degree in Statistics from the University of Connecticut. He has over twenty years of experience in the sonar field, having held positions at U.S. Navy, NATO, and University laboratories. His research has primarily been in applying detection and estimation theory to active and passive sonar signal processing problems. He has both undergraduate and graduate teaching experience at the Electrical and Computer Engineering Department of the University of Connecticut. He has managed basic and applied research programs at the Office of Naval Research and has been active in professional service through technical-committee membership, conference and workshop organization, and as an associate editor of the IEEE *Journal of Oceanic Engineering*.

The full registration fee is $300 USD ($125 USD for students) and covers attendance, instructional materials and coffee breaks. The number of attendees will be limited so please register early to avoid disappointment. Only those who have registered by 26 March will be guaranteed receipt of instruction materials. There will be a $50 USD discount off the full registration fee (discount does not apply to student fee) for registration made prior to 26 March. Full refunds will be made for cancellations prior to 26 March. Any cancellations after 26 March will be charged a $25 USD processing fee. Register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers.

NOISE-CON 2010

The 26th annual conference of the Institute of Noise Control Engineering, NOISE-CON 2010, will run concurrently with the 159th Meeting of the Acoustical Society of America on Monday through Wednesday (19–21 April 2010), culminating with the Closing Ceremony which will take place with the ASA Plenary Session and Awards Ceremony on Wednesday afternoon (21 April 2010). The Institute of Noise Control Engineering (INCE) has planned 30 sessions jointly organized with ASA Technical Committees to form an exciting and coherent program that reflects the overlap in membership interests between the two organizations, and the spirit of cooperation that led to the decision to hold this joint meeting. An additional 14 session are being organized by ASA Technical Committees and co-sponsored by INCE.

Note that there will be one registration fee for both conferences, so NOISE-CON 2010 participants are encouraged to take the opportunity to learn about some of the work being done in other areas of acoustics, not usually part of regular NOISE-CON technical programs, by attending the sessions taking place on Thursday and Friday.

All presentations in NOISE-CON 2010, including those in sessions

jointly organized with the ASA Technical Committees, must be accompanied by a written paper that will be published by INCE in the NOISE-CON 2010 Proceedings.

Conference Proceedings

The Proceedings of NOISE-CON 2010 will be published on a CD-ROM and will contain additional NOISE-CON proceedings for the past five years. The CD-ROM will be available at the conference to conference registrants for $30 USD, and will be available after the conference through the INCE/USA page at the Atlas Bookstore for $70 USD. To obtain a list of proceedings and other publications available now, go to ⟨www.bookmasters/marktplc/00726.htm⟩.

INCE Fundamentals Exam Preparation and Optional Exam

Instructors: James Barnes, Acentech.jbarnes@acentech.com
Eric Wood, Acentech.ewood@acentech.com
Day: Sunday, 18 April 2010
Time: 9:00 a.m.—1:00 p.m. (course)
2:00 p.m.—4:00 p.m. (fundamentals exam)
Location: Baltimore Marriott Waterfront Hotel
Cost: $100 INCE members; $150 nonmembers (early registration before 19 March 2010)
$150 INCE members; $200 nonmembers (after 19 March 2010)
Registration: Course registration details can be found on the following website: www.inceusa.org/NC10

Description: If you are considering taking the INCE fundamentals exam as one way to become a full member of INCE, this course is aimed at helping you understand and prepare for the exam. The exam is a two-hour, closed-book, multiple-choice examination with 75 questions drawn from the fundamentals of acoustics and noise control engineering. Qualitative questions with descriptive responses and quantitative questions requiring only simple calculations are included in the examination. The purpose of the examination is to evaluate an individual's background in the field of noise control engineering. In this preparation course you will learn about the nature of the exam questions and responses and the wide range of concepts and topics covered. A minimum of mathematics will be included.

If you feel comfortable with the material presented in the course, you may stay and take the actual fundamentals exam from 2:00 p.m. to 4:00 p.m.

Jim Barnes is a senior engineer and directs the environmental, transportation, and industrial acoustics consulting services at Acentech. He has more than 30 years of consulting experience in industrial noise and vibration control. His consulting projects have included interior and community noise control studies for existing power and industrial plants, prediction of construction and operation noise levels, and ambient sound studies for proposed industrial sites and transportation corridors for environmental impact statements. In addition, many of his projects have included evaluation and resolution of potential vibration problems at proposed sites for microelectronics and optics facilities.

Eric Wood is a principal at Acentech where he directs and provides technical contributions to engineering and environmental projects related primarily to the measurement, evaluation, and control of noise and vibration during the design, construction, and operation of major energy systems and transportation and industrial facilities. During thirty-five years of consulting practice he has contributed to hundreds of technical reports and publications. He is vice president of membership for INCE-USA and a former member of their Board of Directors. He is treasurer of the INCE Foundation and a Fellow of the Acoustical Society of America.

INCE Short Course on Aircraft Noise Modeling

Day: Sunday, 18 April 2010
Time: 9:00 a.m.—5:00 p.m.
Location: Baltimore Marriott Waterfront Hotel
Cost: $200 INCE members; $250 nonmembers (early registration before 19 March 2010)
$250 INCE members; $300 nonmembers (after 19 March 2010)
Registration: Course registration details can be found on the following website: www.inceusa.org/NC10

Description: Aviation noise planning relies heavily on semi-empirical noise models, which begin with measured source levels and use varying degrees of analytic relations to propagate noise into the community. This course examines the structure and algorithms of traditional integrated noise models (such as the FAA's INM, AEDT and NIRS, and the Department of Defense's NOISEMAP model), and those of modern time simulation models (such as Wyle Laboratories' NMSim and the Department of Defense's Advanced Acoustic Model). Assumptions, evolution and practical considerations of noise models will be reviewed. The course will address the following topics:
The nature of aircraft noise, and how it is measured and packaged for noise modeling
Practical representation of aircraft operations and trajectories
Propagation of sound from the aircraft to the ground
Algorithms of integrated noise models
Algorithms of simulation models
Unique models for military airspace
Noise metrics

Kenneth Plotkin is Chief Scientist, Wyle Acoustics and Research Group. He has over 35 years of experience in measuring, modeling and analyzing aircraft noise. He is the author of NMSim, a noise simulation model, and SVERIM, an integrated model used in Sweden, and has participated in the development of a number of other transportation noise models. He is one of the originators of the L_{dnmr} noise metric used for analysis of noise from low altitude military aircraft, and has served his time recording aircraft noise and deciphering the results into useful source models. He has applied noise models to environmental assessments of numerous projects.

Other instructors will include members of Wyle Laboratories Research Staff who have written and maintained noise models, who have collected aircraft noise source data, and who have used the models in real life.

Special Meeting Features

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups using economical ground transportation will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student depends on the number of requests received. To apply for a subsidy, submit a proposal by e-mail to be received by 15 March to: Jolene Ehl, jehl@aip.org. The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto; whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

Young Investigator Travel Grant

The Committee on Women in Acoustics (WIA) is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the Baltimore meeting. Applicants must have completed their doctorate in the past five years and plan to present a paper at the meeting. Additionally, eligible applicants cannot currently be students, or have previously received the award. Each award will be of the order of $300 with two awards anticipated. Awards will be presented by check at the WIA luncheon. Both men and women may apply. Applicants should submit a request for support, a copy of the abstract presentation, and a current resume/vita which includes information on their involvement in the field of acoustics and in the ASA. Submission by e-mail is preferred to Jennell Vick, jennell@utdallas.edu. Deadline for receipt of applications is 8 March 2010.

Students Meet Members for Lunch

The ASA Education Committee provides a way for a student to meet one-on-one with a member of the Acoustical Society over lunch. The purpose is to make it easier for students to meet and interact with members at ASA meetings. Each lunch pairing is arranged separately. Students who wish to participate should contact David Blackstock, University of Texas at Austin, by e-mail: ⟨dtb@mail.utexas.edu⟩. Please provide your name, university, department, degree you are seeking (BS, MS, or PhD), research field, acoustical interests, and days you are free for lunch. The sign-up

deadline is ten days before the start of the meeting, but an earlier sign-up is strongly encouraged. Each participant pays for his/her own meal.

Plenary Sessions, Awards Ceremony, Fellows' Luncheon, and Social Events

Buffet socials with cash bar will be held on Tuesday and Thursday at the Baltimore Marriott Waterfront Hotel.

The ASA Plenary session will be held Wednesday afternoon, 21 April, where Society awards will be presented and recognition of newly-elected Fellows will be announced.

NOISE-CON 2010 plenary sessions will be held Monday, Tuesday, and Wednesday mornings, 19–21 April. The plenary speakers are Kenneth J. Plotkin, "Sonic Boom: From Bang to Puff" and Lily M. Wang, "Effects of Building Mechanical System Noise on Worker Performance and Perception."

A Fellows Luncheon will be held Thursday, 22 April, at 12:00 noon. Dr. Alan Leshner, CEO of the American Association for the Advancement of Science, will give a presentation on science policy. This luncheon is open to all attendees and their guests. Register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Wednesday, 21 April. Those who wish to attend this luncheon must register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers. The fee is $20 (students $10) for pre-registration by 26 March and $25 (students $10) at the meeting.

Transportation and Travel Information

Air Transportation

Baltimore is served by all major airlines through three airports: Thurgood Marshall Baltimore-Washington International (BWI—12 miles SW), Ronald Reagan Washington National (DCA—45 miles S), and Washington Dulles International (IAD—52 miles SW). Information for these airports is available on the web at ⟨www.bwiairport.com⟩, ⟨www.metwashairports.com/national⟩, and ⟨www.metwashairports.com/dulles⟩.

Ground Transportation

There are a variety of ground transportation options between BWI, DCA, or IAD and the Baltimore Marriott Waterfront Hotel. In addition Baltimore has rail connections with several cities via Amtrak and the Maryland Area Regional Commuter (MARC) Train Service.

Shared Ride Shuttle: The Baltimore Marriott Waterfront has airport shuttle service to BWI on request for $13 USD one-way. Super Shuttle service is available to/from all three airports and cost depends on the number of passengers. For additional information visit ⟨www.supershuttle.com⟩; reservations can also be made by calling 800 BLUE-VAN (258-3826) or 410-859-3427.

Taxicabs: Regular taxicabs are available outside of baggage claim at all three airports. The cost from the airport to the Baltimore Marriott Waterfront is approximately $30 USD from BWI, $110 USD from DCA, and $150 USD from IAD. Enviroride ⟨http://www.enviroride.net/⟩ airport transfers are available by reservation (phone 301-549-4111 or toll free 1-866-929-4202). Approximate cost is $30 USD from BWI and $100 USD from DCA or IAD.

Light Rail: Light Rail service ⟨www.mtamaryland.com⟩ is available from BWI with free bus transfer from the terminal. It stops at the Baltimore Convention Center, corner of Howard and Pratt Streets, about 14 blocks from the hotel (8 blocks east on Pratt Street, 5 blocks south on President Street and 1 block west on Aliceanna Street).

Rail: Amtrak train service is available from many east coast cities to Penn Station located in downtown Baltimore about 1 mile north of the Hotel. Additional information and reservations are available at ⟨www.amtrak.com⟩. MARC trains travel between Penn Station (Penn Line) and Union Station in Washington, DC, or between Camden Yards (Camden Line) and Union Station, Washington, DC (see www.mtamaryland.com for more information). Union Station in Washington, DC, is accessible via light rail from Ronald Reagan National Airport (DCA). Map and schedules are available at ⟨www.wmata.com⟩.

Parking at the Baltimore Marriott Waterfront Hotel

The on-site parking fee at the Baltimore Marriott Waterfront is $7 USD hourly or $25 USD daily. Valet parking is available for $38 USD daily. The parking garage does not accommodate oversized vehicles.

Room Sharing

ASA will compile a list of those who wish to share a hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker preference, not later than 15 March to the Acoustical Society of America, by e-mail: asa@aip.org. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

Baltimore has four distinct seasons with warm weather prevailing from April through October. In April the daily average temperature ranges from 43 to 65 °F, so come prepared for cool weather as well as some warm sunny days. Local weather is available by phone at 410-936-1212.

Hotel Reservation Information

A block of guest rooms at discounted rates has been reserved for meeting participants at the Baltimore Marriott Waterfront Hotel. Early reservations are strongly recommended. Note that the special ASA meeting rates are not guaranteed after 26 March 2010. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates.

The Baltimore Marriott Waterfront is located on the edge of Baltimore's historic Inner Harbor and steps away from downtown.

Please make your reservation directly with the Baltimore Marriott Waterfront Hotel. When making your reservation, you must mention the Acoustical Society of America to obtain the special ASA rates. Alternatively, reservations can be made directly online at the website listed below, which has been set up specifically for the Acoustical Society of America, and has the conference rates and all applicable information incorporated into it.

Baltimore Marriott Waterfront Hotel
700 Aliceanna Street
Baltimore, MD 21202 USA
Tel: 1-410-385-3000
Toll Free: 1-800-228-9290
Fax: 1-410-895-1900
http://www.marriott.com/hotels/travel/bwiwf-baltimore-marriott-waterfront/
Online: http://www.marriott.com/hotels/travel/bwiwf?groupCode=aasaasa&app=resvlink&fromDate=&toDate=

Note: Your reservation must include dates between 16 April and 25 April in order to reserve at the special ASA rates.

Room rate (excluding taxes, currently 13.5%)
Single/Double/Triple/Quadruple Occupancy: $199 USD

Assistive Listening Devices

Anyone planning to attend the meeting who will require the use of an assistive listening device, is requested to advise the Society in advance of the meeting: Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, asa@aip.org.

Child Care

Information about child care services at the Baltimore meeting will be provided on the ASA website.

Child Care Grants

The Committee on Women in Acoustics (WIA) is sponsoring a Child Care Grant to help with child care costs associated with bringing a child to the ASA Baltimore meeting. Applicants must plan to present a paper at the Baltimore meeting. Each award will be of the order of $300 USD with three awards anticipated. Awards will be presented by check at the WIA luncheon. Both men and women may apply. Applicants should submit a request for support, clearly outlining the need for child care support and estimating the cost of such care, a copy of the abstract for their presentation at the meeting, and a current resume/vita which includes information on their involvement

in the field of acoustics and in the ASA. Submission by e-mail is preferred to Andone Lavery at alavery@whoi.edu. Deadline for receipt of applications is 15 March 2010.

Accompanying Persons Program

Spouses and other visitors are welcome at the Baltimore meeting. The registration fee is $60 USD for preregistration by 26 March and $95 USD at the meeting.

A hospitality room for accompanying persons will be open at the Baltimore Marriott Waterfront from 8:00 a.m. to 10:30 a.m., Monday through Friday. A speaker from the Baltimore Visitors Association will present local attractions, events and a brief history of Baltimore on Monday morning. Most all attractions in the Inner Harbor are within walking distance of the hotel or accessible by water taxi. A Harbor Pass for discounted admission to five top attractions is available for advance purchase by phone at 1-877-Baltimore. Visit ⟨http://baltimore.org/harborpass⟩ for additional information. Please check the ASA website at ⟨http://asa.aip.org/meetings.html⟩ for updates about the accompanying persons program.

Registration Information

The registration desk at the meeting will open on Monday, 19 April, at the Baltimore Marriott Waterfront. Register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers. **If your registration is not received at the ASA headquarters by 26 March you must register onsite.**

Registration fees are as follows:

Category	Preregistration by 26 March	Onsite Registration
Acoustical Society or INCE Members	$395	$455
Acoustical Society or INCE Members One-Day Attendance*	$200	$260
Nonmember	$445	$505
Nonmembers One-Day Attendance*	$225	$285
Nonmember Invited Speakers One-Day Attendance*	Fee waived	Fee waived
Nonmember Invited Speakers (Includes one-year ASA membership upon completion of an application)	$110	$110
ASA Early Career Associate or Full Members (For ASA members who transferred from ASA student member status in 2008, 2009, or 2010)	$200	$260
ASA or INCE Student Members (with current ID cards)	Fee waived	$25
Nonmember Students (with current ID cards)	$50	$60
Nonmember Undergraduate Students (with current ID cards indicating undergraduate status)	Fee waived	$60
Emeritus members of ASA or INCE (Emeritus status pre-approved by ASA or INCE)	$60	$95
Accompanying Persons (Registrants who will not participate in the technical sessions)	$60	$95

Note: A $25 FEE WILL BE CHARGED FOR CANCELLATIONS AFTER 26 MARCH.

*One-day registration is for participants who will attend the meeting for only one day. If you will be at the meeting for more than one day either presenting a paper and/or attending sessions, you must register and pay the full registration fee.

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2010) of membership. Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but **nonmember invited speakers** may register for one-day only without charge. A nonmember invited speaker who pays the full-week registration fee, will be given one free year of membership upon completion of an ASA application form.

159th Meeting Local Committee

Mardi C. Hastings, General Chair; Juan I. Arvelo, Jr., Technical Program Chair; Arthur N. Popper, Special Events; Catherine H. Frazier, Technical Tours.

NOISE-CON 2010 Local Committee

Michael L. Lucas, General Chair, Courtney B. Burroughs, Technical Program Chair; Richard J. Peppin, Exhibiton Manager

Calendar of Meetings and Congresses

2010

6–9 January	Sanya, China. 2nd International Conference on Vibro-Impact Systems. Web: www.neu.edu.cn
8–11 March	Berlin, Germany. Meeting of the German Association for Acoustics DAGA 2010. Web: www.daga-tagung.de/2010
15–19 March	Dallas, TX, USA. International Conference on Acoustics, Speech, and Signal Processing. Web: http://icassp2010.org
15–19 March	Dallas, TX, USA. International Conference on Acoustics, Speech, and Signal Processing. Web: http://icassp2010.org
7–9 April	Cambridge, UK. The David Weston Sonar Performance Assessment Symposium
19–23 April	Baltimore, MD, USA. Joint meeting: 159th Meeting of the Acoustical Society of America and Noise Con 2010. Web: http://asa.aip.org/meetings.html
27–30 April	Ghent, Belgium. Institute of Acoustics/Belgian Acoustical Association Joint Meeting. Web: www.ioa.org.uk/viewupcoming.asp
6–7 May	Paris, France. 2nd International Symposium on Ambisonics and Spherical Acoustics. Web: http://ambisonics10.ircam.fr
9–11 June	Aalborg, Denmark. 14th Conference on Low Frequency Noise and Vibration. Web: http://lowfrequency2010.org
13–16 June	Lisbon, Portugal. INTERNOISE2010. Web: www.internoise2010.org
5–9 July	Istanbul, Turkey. 10th European Conference on Underwater Acoustics. Web: http://ecua-2010-istanbul.org
23–27 August	Seattle, USA. 11th International Conference on Music Perception and Cognition. Web: http://www.musicperception.org/resources/ICMPC11_Flyer.pdf
23–27 August	Sydney, Australia. International Congress on Acoustics 2010. Web: www.ica2010sydney.org
29–31 August	Melbourne, Australia. International Symposium on Room Acoustics (ISRA2010). Web: http://web.arch.usyd.edu.au/~densil/ISRA/
14–18 September	Kyoto, Japan. 5th Animal Sonar Symposium. Web: http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html
15–18 September	Ljubljania, Slovenia. Alp-Adria-Acoustics Meeting joint with EAA. E-mail: mirko.cudina@fs.uni-lj.si
26–30 September	Makuhari, Japan. Interspeech 2010-ICSLP. Web: www.interspeech2010.org

14–16 October	Niagara-on-the Lake, Ont., Canada. Acoustics Week in Canada. Web: http://caa-aca.ca/E/index.html
11–14 October	San Diego, California, USA. IEEE 2010 Ultrasonics Symposium. E-mail: bpotter@vectron.com
15–19 November	Cancun, Mexico. Second Pan-American/Iberian Meeting on Acoustics (Joint meeting of the Acoustical Society of America, Iberoamerican Congress of Acoustics, Mexican Congress on Acoustics. Web: http://asa.aip.org/meetings.html
19–20 November	Brighton, UK. Reproduced Sound 25. Web: www.ica.org.uk/viewupcoming.asp

2011

22–27 May	Prague, Czech Republic. International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2011). Web: http://www.icassp2011.com/
27 June–1 July	Aalborg, Denmark. Forum Acusticum 2011. Web: www.fa2011.org
16–21 July	Williamstown, Massachusetts, USA. 11th International Mechanics of Hearing Workshop. Web: www.mechanicsofhearing.org/
24–28 July	Tokyo, Japan. 19th International Symposium on Nonlinear Acoustics (ISNA 19). Web: TBA
27–31 August	Florence, Italy. Interspeech 2011. Web: www.interspeech2011.org
4–7 September	Osaka, Japan. Internoise 2011. Web: www.internoise2011.com
5–8 September	Gdansk, Poland. International Congress on Ultrasonics. Web: http://icu2011.ug.edu.pl/index.html

2013

2–7 June	Montréal, Canada. 21st International Congress on Acoustics (ICA 2013) (Joint meeting: International Congress on Acoustics, Acoustical Society of America, Canadian Acoustical Association). Web: www.ica2013montreal.org

ADVANCED-DEGREE DISSERTATIONS IN ACOUSTICS

***Editor's Note**: Abstracts of Doctoral and Masters theses will be welcomed at all times. Please note that they must be limited to 200 words, must include the appropriate PACS classification numbers, and formatted as shown below. If sent by postal mail, note that they must be double spaced. The address for obtaining a copy of the thesis is helpful. Submit abstracts to: Acoustical Society of America, Thesis Abstracts, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 117474502, e-mail: asa@aip.org*

Design and evaluation of digital signal processing algorithms for acoustic feedback and echo cancellation [43.38.Tj, 43.60.Dh] —Toon van Waterschoot, *Faculty of Engineering, Katholieke Universiteit Leuven Leuven, Belgium,, March 2009 (Ph.D.).* This thesis deals with several open problems in acoustic echo cancellation and acoustic feedback control. Our main goal has been to develop solutions that provide a high performance and sound quality, and behave in a robust way in realistic conditions. This can be achieved by departing from the traditional ad-hoc methods, and instead deriving theoretically well-founded solutions, based on results from parameter estimation and system identification. In the development of these solutions, the computational efficiency has permanently been taken into account as a design constraint, in that the complexity increase compared to the state-of-the-art solutions should not exceed 50% of the original complexity.

Full text available at http://hdl.handle.net/1979/2599

Advisor: Marc Moonen

Time domain normal mode analysis of underwater acoustic wave propagation for a single layered acoustic channel in two dimensional Cartesian coordinates [43.20.Bi, 43.20.Mv, 43.30.Bp] —Hüseyin Özkan Sertlek, *Electronics Engineering Department, Gebze Institute of Technology Kocaeli, Turkey, October 2008, (M.S.).* A new analytical time domain Normal Mode solution in a single layer acoustic waveguide in Cartesian coordinates is presented in the thesis. The method is based on the separation of variables technique for the two dimensional time domain wave equation. After separation, Sturm-Liouville type ordinary differential equation for depth dependency and inhomogeneous Klein-Gordon type partial differential equation for time and range dependency are obtained. Whenever the ordinary differential equation is solved in a classical manner with the application of Dirichlet boundary condition on the upper and lower boundaries of the waveguide, Inhomogeneous Klein-Gordon type partial differential equation is solved by using the Green function technique in time domain. Then the exact solution satisfying the causality principle is constructed for the acoustic pressure. The main advantage of the proposed causal solution is not necessary to use the Fourier Transformation to obtain time domain response of arbitrary acoustical source signals having the monochromatic and Gaussian pulse type time dependency in the waveguide. The excellent agreement for the comparisons of the range dependent transmission losses is observed between the proposed solution and the KRAKEN program.

Advisor: Serkan Aksoy

Modelling the human cochlea [43.64.Kc] —Emery Mayon Ku, *Institute of Sound and Vibration Research, University of Southampton, UK, 2008 (Ph.D).* One of the salient features of the human cochlea is the incredible dynamic range it possesses—the loudest bearable sound is 10,000,000 times greater than the softest detectable sound; this is in part due to an active process. More than twelve thousand hairlike cells known as outer hair cells are believed to expand and contract in time to amplify cochlear motions. However, the cochlea's response is more than just the sum of its parts: the local properties of outer hair cells can have unexpected consequences for the global behavior of the system. One such consequence is the existence of otoacoustic emissions (OAEs), sounds that (sometimes spontaneously!) propagate out of the cochlea to be detected in the ear canal. In this doctoral thesis, a classical, lumped-element model is used to study the cochlea and to simulate click-evoked and spontaneous OAEs. The original parameter values describing the microscopic structures of the cochlea are re-tuned to match several key features of the cochlear response in humans. The frequency domain model is also recast in a formulation known as state space; this permits the calculation of linear instabilities given random perturbations in the cochlea which are predicted to produce spontaneous OAEs.

Full text available at http://eprints.soton.ac.uk/64535/

Advisors: Prof. S. J. Elliott and Dr. B. Lineton

The influence of structural-acoustic coupling on the dynamic behaviour of a one- dimensional vibro-acoustic system [43.40.Rj] —Gihwan Kim, *Institute of Sound and Vibration Research, University of Southampton, UK, January 2009 (Ph.D).* The aim of this thesis is to investigate the structural-acoustic coupling effects on the dynamic behavior of a vibro-acoustic system under passive/active control. The simplest model of a vibro-acoustic system one can consider is a one-dimensional acoustic cavity driven by a single-degree-of-freedom (SDOF) structure. This simple model is used to demonstrate the physical characteristics of the coupling phenomenon. This simple analytical model can provide various degrees of structural-acoustic coupling, which are dependent upon (i) the structural-acoustic stiffness ratio, (ii) structural-acoustic natural frequency ratio, (iii) structural damping, and (iv) acoustic damping. In this case, although the geometric coupling factor is not included because the SDOF structure has a single mode, 80 percent of the factors that determine the degree of coupling can be accounted for by the simple analytical model. The coupling mechanism, in the simple vibro-acoustic system, is investigated using the mobility-impedance approach. In order to provide the threshold of the degree of coupling, a coupling factor is calculated in terms of non-dimensional structural-acoustic parameters. Vibro-acoustic responses are represented by the acoustic potential energy in the cavity and the kinetic energy of the structure coupled to the acoustic cavity.

Full text available at http://eprints.soton.ac.uk/65676/

Advisor: Dr. T. Waters

A methodology for developing high damping materials with application to noise reduction of railway track [43.50.Lj, 43.40.Tm] —Nazirah Ahmad, *Institute of Sound and Vibration Research, University of Southampton, UK, February 2009 (Ph.D).* A methodology is developed that allows a material to be formulated for a particular damping application where temperature-dependence has to be taken into account. This is applied to a tuned absorber system used for damping the vibration of a railway track. This is required to be effective over a temperature range -20 to 40 deg C. The time-temperature superposition principle is used to convert frequency-dependence to temperature-dependence for a notional material with constant loss factor. This is used in the prediction of decay rates and thereby noise reduction. In addition, a weighted noise reduction is studied by using measured rail temperature distributions. Two types of viscoelastic material, butyl and EPDM rubbers with various amount of fillers and plasticisers, are investigated. The properties of both rubbers have been measured over the range of temperatures for frequencies 300-3000 Hz. For butyl, the best combination of filler and plasticiser gives temperature weighted noise reductions up to 5.9 dB(A). The best EPDM compound gives a temperature-weighted noise reduction up to 6.2 dB(A). Comparing these two rubbers, EPDM is more suitable for low temperatures below 10 deg and butyl is more suitable for higher temperatures above 10 deg.

Advisor: Prof. D. J. Thompson

BOOK REVIEWS

P. L. Marston
Physics Department, Washington State University, Pullman, Washington 99164

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

Viscoelastic Waves in Layered Media

Roger D. Borcherdt

Cambridge University Press, New York, 2009. 305 pp. Price: $140 (hardcover). ISBN: 978-0-521-89853-9

This is a fascinating book that deals with the theory and application of wave propagation in lossy materials and it is highly recommended reading for anyone who deals with seismic or acoustic waves in real media. Although the theory is developed for viscoelastic waves (compressional and shear waves in lossy solids), the analysis and results easily carry over into viscoacoustic waves (compressional waves only in lossy fluids). For example, in both lossy solids and lossy fluids the usually linear particle motions for compressional body waves are in general elliptical.

If you question the relevance of this theory just look at Fig. 12 of Brekhovskikh (1960) that compares the modulus and phase of the elastic reflection coefficient at a water-aluminum interface with laboratory measurements. Brekhovskikh noted that elastic wave theory (lossless media) was inadequate to predict the modulus of the reflection coefficient at the compressional and shear critical angles. For lossless media it is common knowledge that the magnitude of the reflection coefficient at super-critical angles of incidence (sub-critical grazing angles) is unity, corresponding to total internal reflection. For lossy media, however, the magnitude of the reflection coefficient drops dramatically, even to zero for some values of shear attenuation, within a few degrees beyond the shear critical angle (below the shear grazing angle). Figure 6.2.8 of Borcherdt shows similar but more detailed results from laboratory measurements made by Becker and Richardson (1969) for a water-stainless steel boundary near the shear critical angle. The agreement of the observations with the theory for lossy media is striking. Unfortunately this is the only comparison of the lossy theory with actual observations in the book.

Most of us begin to learn wave propagation in solids and fluids by considering lossless media. Much of the intuition instilled at these early stages simply does not apply for the lossy media that constitute the real earth and actual materials. Borcherdt makes an excellent case for studying the lossy theory first from which the results for lossless media are a simple subset. Just a few minutes perusing Chap. 8 on the Rayleigh wave at the free surface of a viscoelastic half-space should be enough to convince even the most cynical reader. For example, the amplitude of the vertical particle motion does not in general decay monotonically away from the free surface, and the axes of the particle motion ellipse are inclined with respect to the free surface.

One must be careful not to confuse homogeneous and inhomogeneous waves, homogeneous and inhomogeneous media, and homogeneous and inhomogeneous differential equations. One does not usually think about inhomogeneous waves in uniform materials but they are common at the interfaces between materials. In elastic materials, homogeneous waves have the surfaces of constant amplitude parallel to surfaces of constant phase (for example, body waves in a uniform solid); inhomogeneous waves have the surfaces of constant amplitude normal to the surfaces of constant phase (for example, free surface Rayleigh waves). "Wave inhomogeneity" is expressed as an angle, gamma, which is the angle between the propagation vector, normal to surfaces of constant phase, and the attenuation vector, normal to surfaces of constant amplitude. The single most important conceptual hurdle in this theory is the notion that inhomogeneous waves propagate in homogeneous media, that is, they are a solution to the wave equation for homogeneous media. The direction of maximum attenuation is not in general collinear with the direction of propagation.

Readers familiar with wave propagation in elastic layered media will be comfortable with the concept of vertically and horizontally polarized shear waves (SV and SH respectively), which, in anisotropic materials, generally propagate at different velocities, resulting in "shear wave splitting." As this book clearly describes, even in a homogeneous, isotropic, but viscoelastic material there are also two shear waves, SI and SII, which have elliptical and linear particle motions, respectively, and which propagate independently. For homogeneous, viscoelastic materials and equivalent degrees of inhomogeneity (gamma), the wave speeds for SI and SII waves are the same but the reciprocal quality factors are different!

Of particular interest to the marine seismology and bottom-interacting ocean acoustic communities there is a section (6.2.3) specifically on the consequences of viscoelasticity for reflection from the seafloor. Although brief, there is enough material here to justify the importance of the viscoelastic theory.

The book has limitations. On multiple occasions I found the index to be inadequate. Try, for example, to find the definitions of ''absorption coefficient'' and ''intrinsic absorption.'' The index is not much help although it does point you to the definition of absorption coefficient for one dimensional media on page 17. On page 34 one finds that for three dimensional (3D) media the "maximum attenuation" is the magnitude of the attenuation vector but this is not what you need. What you need is the "attenuation coefficient in the direction of phase propagation" for 3D media which is defined on page 63, where it is given as a function of the magnitude of the maximum attenuation and the degree of wave inhomogeneity (gamma, not material inhomogeneity). Unfortunately the index points you to page 64 where you must infer the definition of "absorption coefficient" from the figure. One has to guess that attenuation coefficient and absorption coefficient are the same thing. "Attenuation coefficient" is not even in the index.

It is usual practice in seismology to call the reciprocal quality factor, $1/Q$, the "internal friction" or "intrinsic attenuation" (with different values for compressional and shear waves) but you will find neither term in the index. "Intrinsic absorption," also not in the index, is used without definition on page 57. Strictly speaking all three terms refer to the reciprocal quality factors for "homogeneous" (compressional and shear) waves as indirectly referred to on pages 62 and 63. Surprisingly, for "inhomogeneous" waves, the reciprocal quality factors for Type I and Type II S waves are not equal even for homogeneous materials. It is important to understand these concepts in order to understand the theory and applications in the whole book, so it is a shame that these key pages are not indicated in the index.

This is not the place to look for a number of common topics in viscoelasticity: for example, (1) definitions of attenuation coefficients as used commonly in ocean acoustics which have various units such as Np/m, dB/m, dB/wavelength, or dB/(km Hz); (2) the viscoelasticity associated with porous solids (Biot–Stoll theory); (3) the constant Q absorption band model; or (4) the spectral ratio method. This book focuses on plane, harmonic (continuous wave) solutions to the wave equation, which are appropriate for layered media, but interesting theories for pulse propagation in lossy media are not discussed. For example, there is no mention of the Kramers–Krönig relation, ray theory, or the attenuation of head waves.

Students should be warned about the confusion in nomenclature for relaxation times in the general literature. The relaxation time of strain at constant stress is τ_σ in Nowick and Berry (1972), τ_ε in Aki and Richards (1980) and Carcione (2001), and τ_e in Borcherdt. The relaxation time of stress at constant strain is τ_ε in Nowick and Berry (1972), τ_σ in Aki and Richards (1980) and Carcione (2001), and τ_p in Borcherdt.

The book seems to use "anelastic" and "viscoelastic" interchangeably. Although both have a time dependent stress-strain relation, anelastic materials are a subset of viscoelastic materials that return to their initial state (no permanent deformation). A general viscoelastic material does not have complete recoverability.

 0001-4966/2009/126(6)/3374/2/$25.00

Throughout the book "Voight" is spelled "Voigt." Not even Voigt spelled his name with an "h."

The book has an extremely short reference list for such a rich topic.

It is fair to say that the community has not rigorously pursued the consequences of viscoelastic theory, even though the fundamentals have been known for over 40 years. Whenever particle motions are directly measured in the field they are ubiquitously elliptical and often for shear waves this ellipticity is attributed to the SV and SH waves arriving nearly simultaneously as occurs for many anisotropic materials. Fully considering the effects of viscoelasticity on polarizations is rarely carried out.

Every scientist dealing with the propagation of elastic and acoustic waves in lossy materials should be familiar with the concepts and results discussed in this book. The book would be a valuable contribution to a professional library. It is too narrow, however, to be the sole textbook for a course on wave propagation in lossy media.

RALPH A. STEPHEN
Department of Geology and Geophysics,
Woods Hole Oceanographic Institution,
Woods Hole, Massachusetts 02543-1542

Aki, K., and Richards, P. G. (**1980**). *Quantitative Seismology: Theory and Methods* (Freeman, San Francisco, CA).
Becker, F. L., and Richardson, R. L. (**1969**). "Critical-angle reflectivity," J. Acoust. Soc. Am. **45**, 793–794.
Brekhovskikh, L. M. (**1960**). *Waves in Layered Media* (Academic, New York).
Carcione, J. M. (**2001**). *Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic and Porous Media* (Pergamon, Amsterdam).
Nowick, A. S., and Berry, B. S. (**1972**). *Anelastic Relaxation in Crystalline Solids* (Academic, New York).

Seismic Interferometry

Gerard T. Schuster

Cambridge University Press, Cambridge, 2009. 260 pp. Price: $140.00 (hardcover). ISBN-10: 0521871247

This book is a timely contribution to the emerging field of seismic interferometry by presenting this line of research and its applications in an organized fashion. In recent years the extraction of the Green's function from field fluctuations has received considerable attention in a variety of fields that include ocean acoustics, ultrasound, seismology, structural engineering, and medical imaging. Because of the rapid growth of this area of research in different fields, the technique in known under a variety of names that include "Green's function extraction," "the virtual source method," and "seismic interferometry." This variety of names hides the underlying principle that one takes signals recorded at receivers and processes those signals in such a way that one obtains the signal as if one of the receivers acts as a source. This is a powerful principle because in many applications it is much easier to place a receiver within the medium than a source.

Dr. Schuster has been pioneering this principle in exploration seismology, and by writing his textbook he makes the fundamentals of this method, as well as numerous application, available in a unified way. After a brief overview of the method and its mathematical foundations, Dr. Schuster spends the largest part of the book on showing how seismic interferometry can be used to map data acquired in one type of survey into other types of surveys. How does this work in practice? Suppose, for example, that one records the waves excited by shots at the surface on geophones placed in a borehole. The waves propagating from the surface may traverse inhomogeneous regions, which may make it difficult to use those waves for imaging. Using seismic interferometry one can obtain the waves that propagate between the sensors in the borehole. This has the advantage for imaging that the detrimental imprint of a complex overburden on the images is removed. In practical applications this approach is much easier and economical than installing a source in the borehole, because such installation is expensive and wrought with practical problems. This is but one of the many data transformations that Dr. Schuster treats in a unified approach to seismic interferometry. The transformations between data of different acquisition geometries form the main part of the book. In addition he shows how interferometry can also be applied to traveltime data, an approach that has not found many applications yet. Most of the theory and examples shown are applicable to controlled source data, but Dr. Schuster devotes a chapter to extracting the Green's function from noise, with drill-bit noise and crustal seismology as examples. He finally shows how interferometry can also be used for source location.

Although Dr. Schuster does present the underlying mathematics of seismic interferometry, the style of presentation of the book relies on intuition and physical insight rather than on rigorous mathematics. For example, he treats the subsurface as an acoustic medium instead of an elastic medium. Such a simplification prevents the concepts of seismic interferometry to be cluttered by mathematical detail. In order to get the ideas across, the author shows numerous numerical examples and data applications that are very convincing. There are two good reasons for this approach. First, although the theory of seismic interferometry is well-developed, it relies on assumptions regarding the sources of field fluctuations that are, in general, not satisfied. This means that one needs to adapt the mathematical formulation of seismic interferometry to the practicalities of a given data acquisition. Second, Dr. Schuster is a master in "making things work." He is one of the pioneers in seismic interferometry and arguably is the researcher that has created most applications of this technique to real data. He has an extremely well-developed physical intuition, which has allowed him to make this technique practical for application to real data. It is fitting that his book reflects this pragmatic approach based on physical intuition, because Dr. Schuster has shown that this approach has been instrumental in the numerous applications he has developed.

This book is a useful resource for researchers in seismic interferometry, regardless of whether the applications are in seismology, acoustics, structural engineering, or other fields. The book could well be used for a graduate-level course on this topic. It provides MATLAB scripts that make it possible to play with many of the examples that are given in the book; such hands-on experience would be especially valuable in a course based on this book. The book is well-written and offers valuable insights. I very much recommend it for researchers and graduate students working in this rapidly growing field.

ROEL SNIEDER
Colorado School of Mines
rsnieder@mines.edu

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:

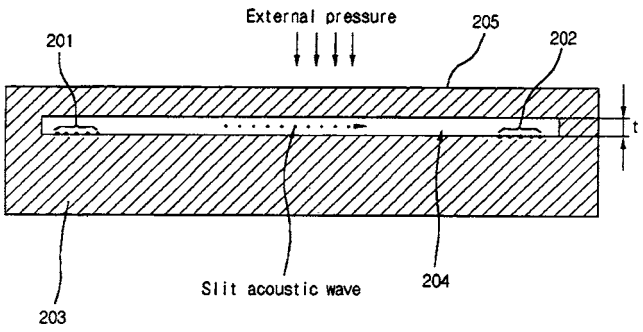
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,459,828

43.35.Zc SAW SENSOR DEVICE USING SLIT ACOUSTIC WAVE AND METHOD THEREOF

Valentin Cherednick *et al.*, assignors to LG Innotek Company, Limited
2 December 2008 (Class 310/313 R); filed in Republic of Korea 8 June 2002

A miniature pressure and viscosity sensor is claimed where the viscosity of liquid 204 passing through slot *t* is determined by the phase shift of an acoustic wave, launched, and received by interdigital transducers 203 and 202 on piezoelectric substrate 203. Alternatively, a pressure applied onto surface 205 narrows the slit to cause a phase shift of the same transmitted acoustic wave. Frequency "calculations" are made, but it is not clear whether such frequencies are those imposed, or optimized, or are regenerative. Claim is also made for measuring the dielectric permittivity of the material in the slit. The apparent signal processing in all cases is comparison of measured wave velocity and phase shift results to a library of previously recorded empirical results from known liquid samples.—AJC

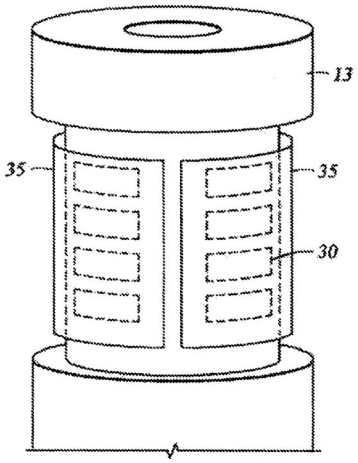


7,460,435

43.38.Hz ACOUSTIC TRANSDUCERS FOR TUBULARS

Fernando Garcia-Osuna and Jean Pierre Masson, assignors to Schlumberger Technology Corporation
2 December 2008 (Class 367/25); filed 30 November 2004

Forms of transducers suited to excite borehole acoustic modes and suited to placement within well logging head 13 are claimed. Arc-shaped elements 30 are installed on a cylindrical surface 35. Several configurations are claimed that include clearance for well fluids.—AJC

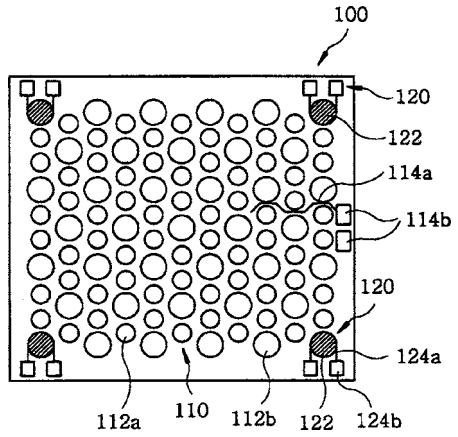


7,460,439

43.38.Hz ULTRASONIC TRANSDUCER FOR RANGING MEASUREMENT WITH HIGH DIRECTIONALITY USING PARAMETRIC TRANSMITTING ARRAY IN AIR AND A METHOD FOR MANUFACTURING SAME

Wonkyu Moon and Haksue Lee, assignors to Postech Foundation
2 December 2008 (Class 367/92); filed in Republic of Korea 13 July 2006

Directional range-finding parametric transducer array 100 and its manufacturing technique are claimed. Multiple first frequency transducers 112a and second frequency transducers 112b are interspersed as a "mixing array" on substrate 110. Radiated high intensity ultrasound beams mix in a



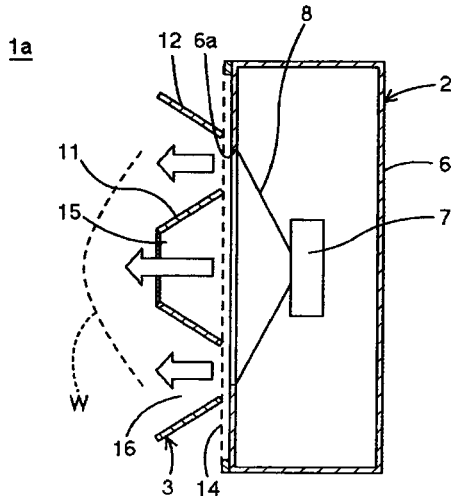
nonlinear air medium to produce sound at the difference frequency. Such difference frequency sound that encounters a distant reflecting object will produce a reflection that can be detected. Manufacture and assembly methods for array 100 are claimed.—AJC

7,466,837

43.38.Ja DIFFUSER AND SPEAKER USING THE SAME

Takeshi Nakamura, assignor to Murata Manufacturing Company, Limited
16 December 2008 (Class 381/337); filed in Japan 12 August 2003

The patent appears to describe an alternate version of how cone- and dome-type loudspeakers work ("planar" waves for the former and "breathing actions" for the latter), and thus various and sundry embodiments of



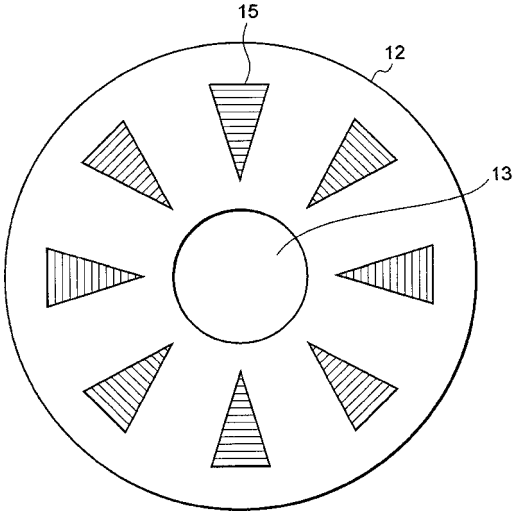
diffusers and speakers are described to provide "pseudo-spherical wave sound waves" for use with cone-type speakers to, at least to this reviewer, approximate dome-type sound radiation. Fundamentals of Acoustics, 4th ed., Chap. 8 (or similar texts) provides a version of things that many may find preferable.—NAS

7,546,897

43.38.Ja SPEAKER DIAPHRAGM AND METHOD OF MANUFACTURING SPEAKER DIAPHRAGM

Toru Takebe *et al.*, assignors to Sony Corporation
16 June 2009 (Class 181/167); filed in Japan 21 July 2005

One of the oldest tricks in loudspeaker design involves adding ribs, bumps, patches, or coating to selected areas of the diaphragm to smooth out high frequency breakup modes. This patent describes a more sophisticated technique that can be used in the fabrication of injection molded polymeric



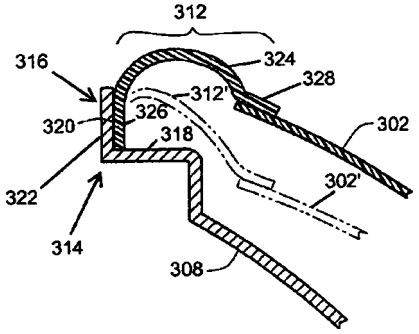
diaphragms. After a diaphragm is molded, selected areas are heated or partially melted using a computer controlled laser beam. The physical result is a "…changed portion including a plurality of peaks and valleys in and on the cone surface."—GLA

7,548,631

43.38.Ja SPEAKER SURROUND STRUCTURE FOR MAXIMIZING CONE DIAMETER

Ronnie S. Espiritu, assignor to Harman International Industries, Incorporated
16 June 2009 (Class 381/398); filed 30 June 2003

Many loudspeaker cones are supported by a flexible half-roll surround. The surround is attached to the frame by a flat outer portion that adds about 10% to the overall diameter of the speaker. In this interesting variant th



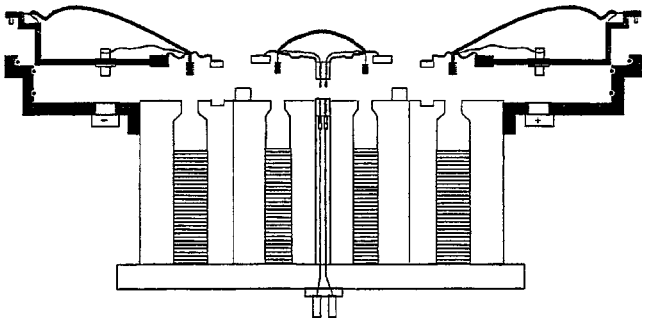
e surround is more like a 3/4 roll and attaches to vertical rim 316. Most of the 16 patent claims are pure vaporware but the basic concept has merit. A half-dozen variants are described.—GLA

7,551,746

43.38.Ja DRIVE UNIT FOR ELECTRO-ACOUSTIC CONVERTER WITH RADIALLY MAGNETIZED PERMANENT MAGNETIC BARS

Anders Sagren, Uppsala, Sweden
23 June 2009 (Class 381/182); filed 1 June 2005

The inner (high frequency) and outer (low frequency) voice coil gaps of this coaxial loudspeaker can be made to lie in the same plane, as shown. The idea is not new. The novel feature of this design is that pole pieces and magnets are all cylindrical, and the magnets (the darker areas) are magnetized radially. Although not shown in the diagram, portions of the pole pieces can be cut or drilled to provide air passages for cooling.—GLA

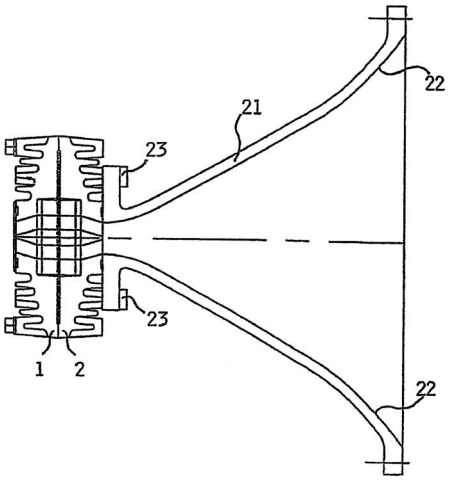


7,558,395

43.38.Ja LOUDSPEAKER

Philip Derek Eduard De Haan, assignor to Alcons Audio B.V.
7 July 2009 (Class 381/338); filed in Netherlands 3 March 2003

Some audiophiles may remember the Beveridge electrostatic loudspeaker. To counteract the directional properties of a large vibrating diaphragm, multiple channels condensed the sound energy into a narrow vertical slot. This patent shows how the same principle can be applied to a planar moving coil loudspeaker. Non-obvious perhaps but hardly a revolutionary insight.—GLA



7,453,770

43.38.Kb METHOD FOR DIRECTIONAL LOCATION AND LOCATING SYSTEM

Hans-Ueli Roeck, assignor to Phonak AG
18 November 2008 (Class 367/118); filed 15 August 2007

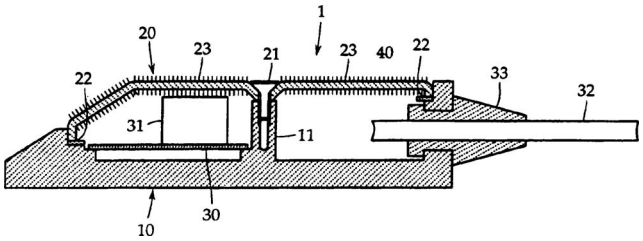
A directional sensor (localization) computation algorithm is claimed using the output of at least two hearing aid microphones.—AJC

7,471,802

43.38.Kb BOUNDARY MICROPHONE

Hiroshi Akino and Satoshi Uchimura, assignors to Kabushiki Kaisha Audio-Technica
30 December 2008 (Class 381/355); filed in Japan 30 September 2004

Pile flocked fabric 23 surrounds microphone 30 when mounted to case 1. Fabric 23 is comprised of a nylon pile with a thickness of 1–3 denier (1 denier=1 g of fiber/9000 m), a length of 0.5–1 mm, and is electrostatically charged so that the piles are essentially perpendicular to the surface of the microphone cover. By virtue of the pile thickness and size, as well as the electrostatic charge, the fabric is essentially hydrophobic. As such, the cover is water resistant and also reduces wind noise.—NAS

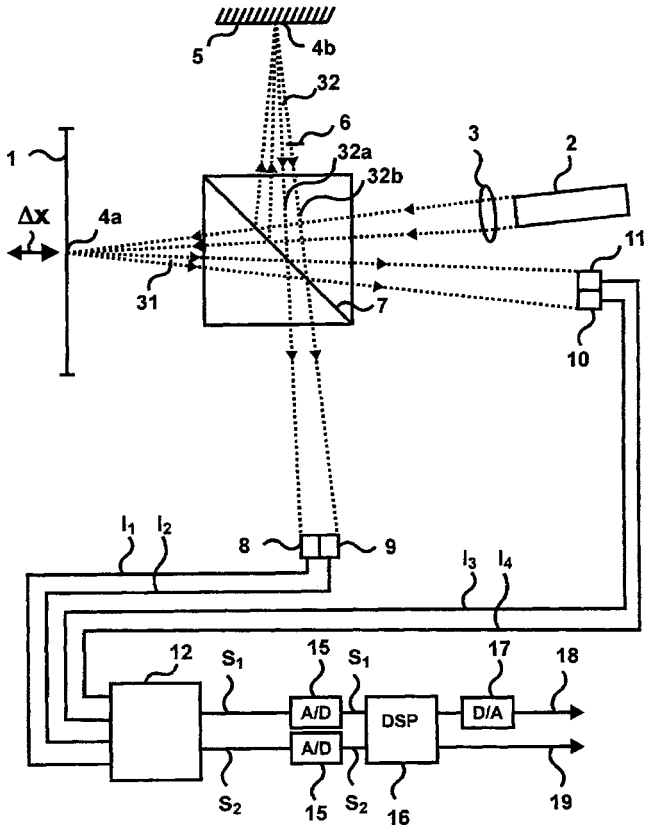


7,521,668

43.38.Kb OPTICAL AUDIO MICROPHONE ARRANGEMENT INCLUDING A MICHELSON TYPE INTERFEROMETER FOR PROVIDING A PHASE DIFFERENCE BETWEEN DIFFERENT PARTS OF LIGHT BEAMS

Ismo Kauppinen, assignor to Noveltech Solutions Limited
21 April 2009 (Class 250/231.1); filed in Finland 17 March 2006

A microphone constructed per the title is said to have better dynamic range, higher sensitivity, as well as better frequency response and greater



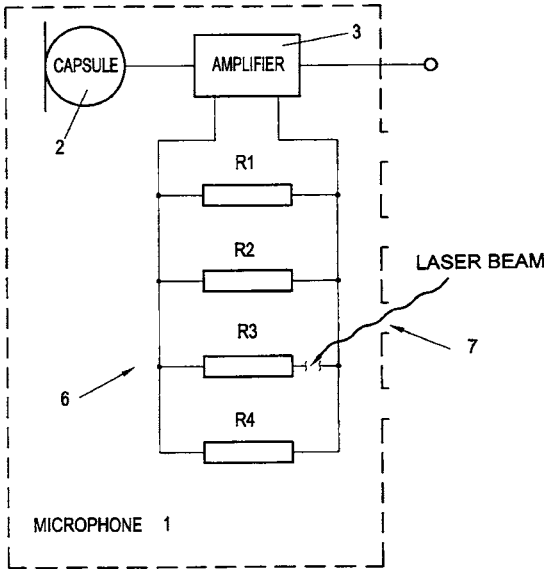
accuracy than prior art optical microphones. Another benefit is less error due to the "tilt" of the mirrors.—NAS

7,522,737

43.38.Kb MICROPHONES WITH EQUAL SENSITIVITY

Thomas Solderits, assignor to AKG Acoustics GmbH
21 April 2009 (Class 381/111); filed in Austria 1 October 2002

Transducers can have variations in their sensitivity using currently available manufacturing techniques. Arrays of such transducers will have inconsistencies in their performance due to the difference in sensitivity for each element in the array. By using an amplifier 3, the gain of which can be



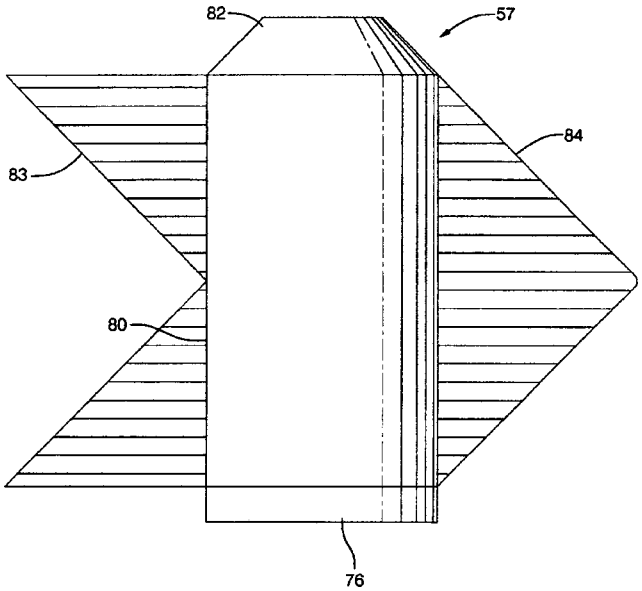
programmed by selectively disconnecting resistors or other passive components in array 6, via laser beam 7 or other means, the sensitivity of microphone 1 consisting of capsule 2, amplifier 3, and component array 6 can be controlled within specified limits.—NAS

7,526,094

43.38.Kb NORMALIZATION AND CALIBRATION OF MICROPHONES IN SOUND-INTENSITY PROBES

Robert Hickling, Huntington Woods, Michigan
28 April 2009 (Class 381/91); filed 2 May 2006

Normalizer-calibrator apparatus 57, comprised of microphone tube 80, fixture 76, loudspeaker 82, and quarter-wavelength attenuator tubes 83 and



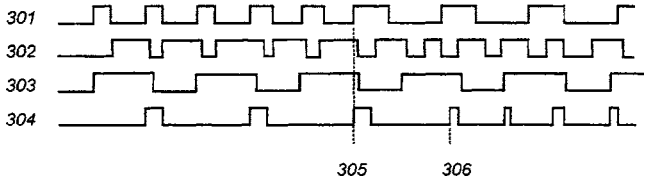
84 to suppress the inevitable standing waves that develop in tube 80, can be used to calibrate the microphones used in sound-intensity probes. The longest tube is about one-half the length of tube 80. Fixture 76 holds the microphones to be calibrated as well as the "stable" reference comparison microphone.—NAS

7,554,398

43.38.Lc INTEGRATION IMPROVEMENT TECHNIQUE FOR SWITCHING AMPLIFIERS

Larry Kirn *et al.*, Austin, Texas
30 June 2009 (Class 330/251); filed 7 July 2006

Consider a powered loudspeaker that uses Class D amplification for high, mid, and low frequency channels. The integrated three-channel amplifier has a single power supply and uses a single clock to synchronize waveform sampling in all three channels. This patent teaches that such a design is



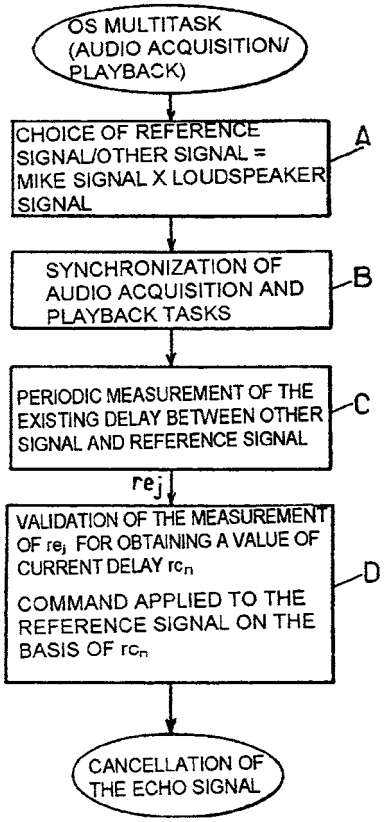
less than optimal and that better performance can be realized by using a different sampling rate for each channel and deliberately un-synchronizing pulse leading edges. The various sampling rates are all integer multiples and, in one variant, can be selected dynamically based on the spectral content of the signal.—GLA

7,555,116

43.38.Lc REAL TIME PROCESSING AND MANAGEMENT METHOD FOR CANCELLING OUT THE ECHO BETWEEN A LOUDSPEAKER AND A MICROPHONE OF A COMPUTER TERMINAL

Grégolre Le Tourneur and David Deleam, assignors to France Telecom
30 June 2009 (Class 379/406.03); filed in France 14 December 1999

Real time audio processing can be tricky with multi-tasking computer



operating systems. This patent points out that a similar problem crops up-

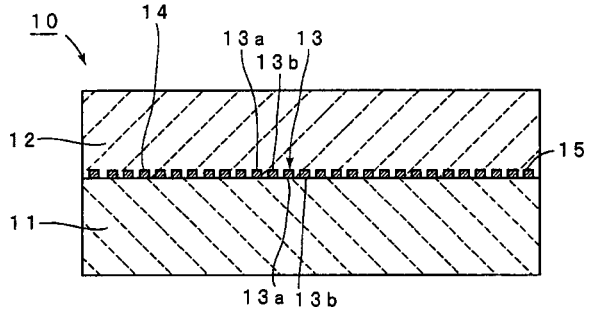
when the computer is used for teleconferencing. Audio teleconferencing requires echo cancellation, and echo cancellation involves precise control of signal delay. It would be convenient to include cancellation in the computer program rather than requiring external hardware. As stated in the patent, "The object of the present invention is to remedy the drawbacks and problems of the prior art systems, with a view to allowing effective installation of echo cancellation functions in software form so as to cater for real time processing on computer terminals of any type." Computer programmers will find the patent easy to follow; others may prefer to leave this field to the experts.—GLA

7,456,544

43.38.Rh BOUNDARY ACOUSTIC WAVE DEVICE

Hajime Kando, assignor to Murata Manufacturing Company, Limited
25 November 2008 (Class 310/313 R); filed in Japan 26 May 2005

A boundary acoustics wave device of low loss and capable of operating over a wide frequency range is claimed. Interdigital transducer fingers 13 are disposed on piezoelectric substrate 11, such as lithium niobate, and



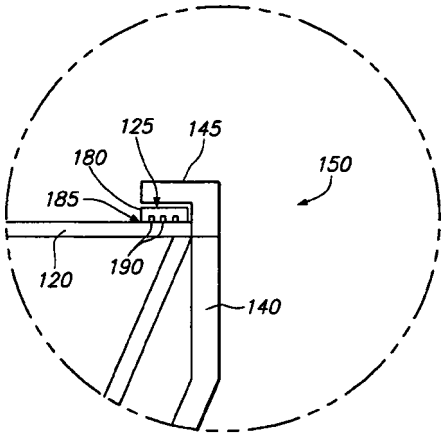
covered with low density medium 12, such as silicon dioxide. The interdigital transducers are composed of a thin high density (e.g., gold) base layer 16 on substrate 11 and a thick low density (e.g., aluminum) top layer 17.—AJC

7,456,825

43.38.Rh ACOUSTIC TOUCH SENSOR WITH LOW-PROFILE DIFFRACTIVE GRATING TRANSDUCER ASSEMBLY

Joel C. Kent *et al.*, assignors to Tyco Electronics Corporation
25 November 2008 (Class 345/177); filed 6 October 2006

Touch screen 120 surface acoustic wave transducer coupling device 180-190 and its manufacturing method is claimed. Diffraction grating 190



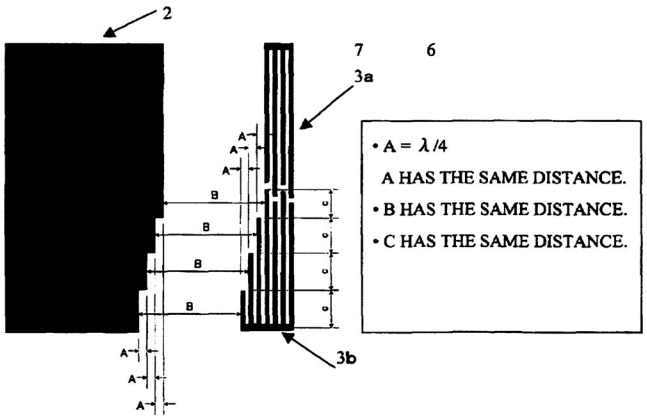
optimally couples acoustic energy from transducer 180 into touch screen 120.—AJC

7,462,973

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Hirotada Wachi *et al.*, assignors to Fujitsu Media Devices Limited
9 December 2008 (Class 310/313 B); filed in Japan 1 October 2003

An improved surface acoustic wave miniature rf band pass filter is claimed where spurious out of band transmittance is reduced by rearranging



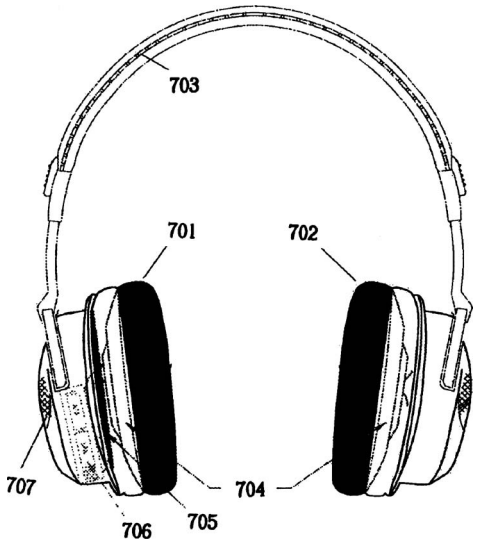
the lengths C and spacings B of interdigital transducer fingers 3a with respect to shield 2.—AJC

7,548,617

43.38.Si BLUETOOTH EARPHONE

Shun Ming Yuen, assignor to Innovation Sound Technology Company, Limited
16 June 2009 (Class 379/430); filed in China 19 August 2005

The basic premise of this patent seems clear enough. It describes a bluetooth wireless receiver whose output can be switched between conventional earphones and a pair of larger loudspeakers. The patent attempts at length to demonstrate that the method described is not obvious, but the actual embodiment is unclear. Muffs 701 and 702 contain conventional



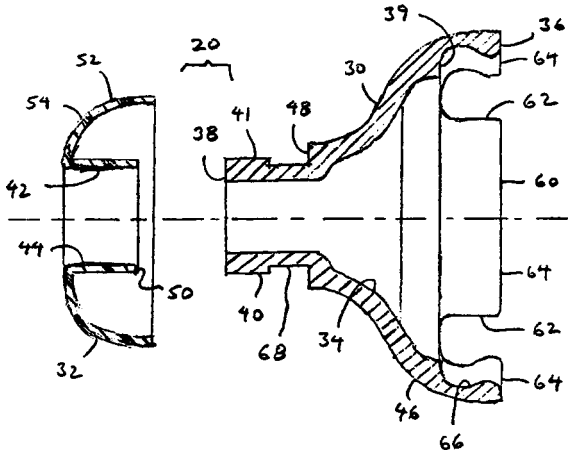
headphone receivers 704. It appears that they also contain a set of larger loudspeakers 706. At this point we are left with more questions than answers. How are the larger loudspeakers heard? Is the headset simply draped around the user's neck? Has any thought been given to the added weight and battery drain?—GLA

7,548,629

43.38.Si EARPIECE ADAPTER FOR AN EARPHONE OR A HEADPHONE

Paul Griffin, assignor to Griffin Technology Incorporated
16 June 2009 (Class 381/382); filed 8 March 2005

Suppose you own a set of ordinary headphones but would prefer to use insertable earbuds. Not to worry. The device shown (which looks like a miniature plumber's helper) transforms one into the other. The most noteworthy feature of the design is that 17 patent claims were granted.—GLA

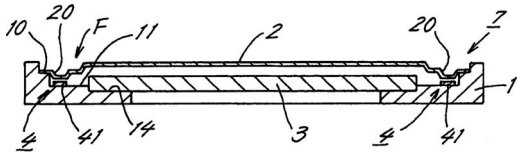


7,548,766

43.38.Si FLAT TYPE SPEAKER UNIT, AND ELECTRONIC APPLIANCE HAVING THIS UNIT

Yuki Takahata *et al.*, assignors to Sanyo Electric Company, Limited
16 June 2009 (Class 455/569.1); filed in Japan 25 April 2003

Earlier patents have suggested various ways in which the display panel of a cellular telephone can be made to function as a loudspeaker. In one arrangement, the display is mounted behind a transparent diaphragm that can be vibrated to produce sound. The patent at hand describes an improved variant of that design. The edge portions 20 of diaphragm 2 are thinner than the central portion, allowing it to vibrate easily yet remain fairly rigid. Each driving motor 41 consists of a shallow coil (possibly rectangular) attached to the diaphragm and energized by stray flux from an equally shallow magnet. Although not very efficient, this kind of gapless moving coil transducer is inexpensive and easy to assemble, making it well-suited to this application.—GLA

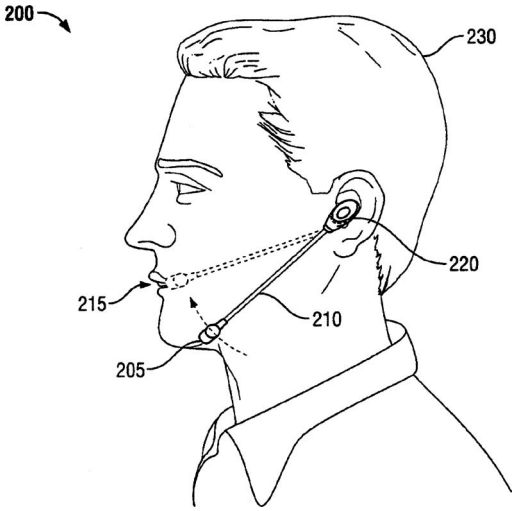


7,551,940

43.38.Si TWO-WAY VOICE COMMUNICATION DEVICE HAVING EXTERNAL ACOUSTIC NOISE REDUCTION

Mary Paulson *et al.*, assignors to Etymotic Research, Incorporated
23 June 2009 (Class 455/550.1); filed 8 January 2004

Insertable earphone 220 seals the user's ear canal, thus attenuating external sounds by at least 7 dB. The friction fit is tight enough to support lightweight microphone boom 210, which can be positioned by simply rotating the earphone in the ear. The overall design includes a number of other optional features intended to minimize airborn and mechanical noise.—GLA

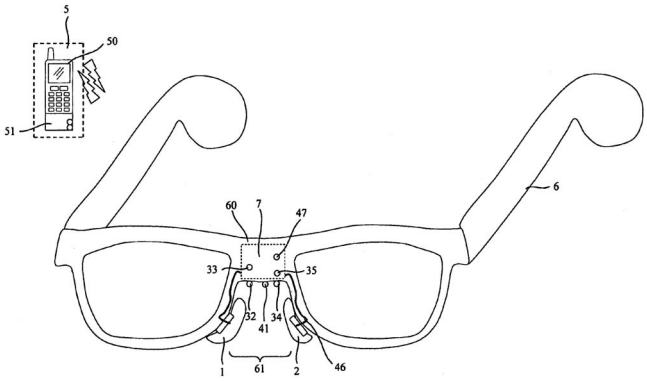


7,555,136

43.38.Si NASAL BONE CONDUCTION WIRELESS COMMUNICATION TRANSMITTING DEVICE

Dave Wang, assignor to Victorion Technology Company, Limited
30 June 2009 (Class 381/380); filed in Taiwan 25 June 2004

Some cellular telephone designs make use of bone conduction transducers to minimize noise pickup. One drawback is that audio quality depends on correct, repeatable transducer placement. This invention avoids the problem by incorporating vibration transducers in eyeglass nose pads. The transducers can function as transmitters or receivers, allowing the device to be used for two-way communication.—GLA



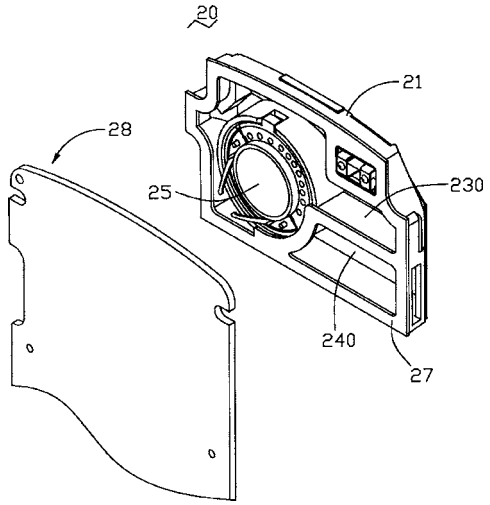
7,556,121

43.38.Si SPEAKER SET AND MOBILE PHONE INCORPORATING THE SAME

Tsung-Lung Yang and Hsien-Sheng Pei, assignors to Foxconn Technology Company, Limited
7 July 2009 (Class 181/145); filed 15 December 2006

Like some other patented designs, this cellular telephone uses the air space between the antenna and the main circuit board to house the loudspeaker. In this case the front of the loudspeaker drives a Helmholtz resonator and the rear drives two additional resonant chambers, all nested into a very small space not much larger than the speaker itself. According to the patent, the first chamber resonates at 2941 Hz and the second chamber resonates at 3851 Hz. (The third chamber apparently does not generate an identifiable third resonance.) This kind of performance is somehow intended to

"…provide as full a harmonic content as is contained in the original sound."—GLA

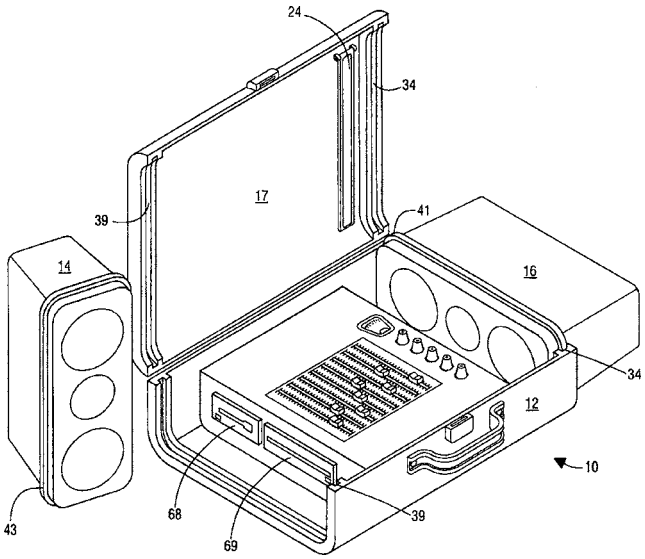


7,469,052

43.38.Tj PORTABLE SOUND SYSTEM, APPARATUS, AND METHOD

Hartley D. Peavey, assignor to Peavey Electronics Corporation
23 December 2008 (Class 381/334); filed 18 June 2004

Portable sound systems that are relatively small and self-contained are ubiquitous these days for "preachers and politicians," especially for those that travel from area to area. One problem with many current models is that the latches used to affix the speakers 14 to central body 12 can rattle. The inventor solves this problem, and the consequent problem of what happens when a latch breaks by fitting case 12 with slots 29 into which flanges 43 of speaker 14 fit when the system is assembled for travel. In addition, the central body can also be used as a lectern. The patent is clearly written and straightforward. The inventor is the founder and principal for a large musical instrument and sound equipment company.—NAS

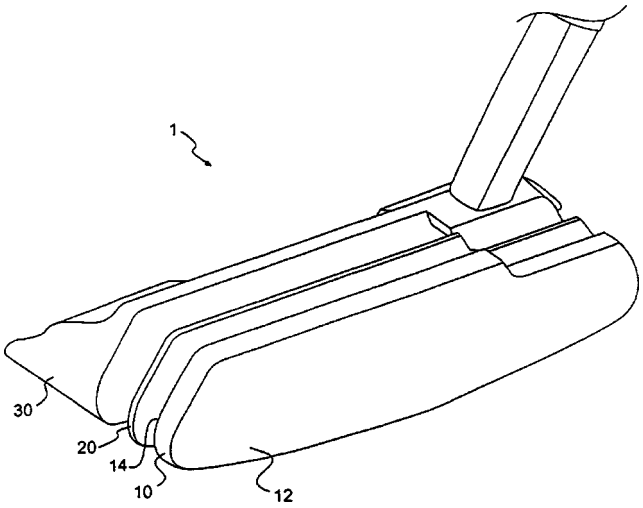


7,473,186

43.40.Kd PUTTER WITH VIBRATION ISOLATION

Christopher B. Best *et al.*, assignors to Acushnet Company
6 January 2009 (Class 473/329); filed 24 October 2006

Vibration dampening element 20 is placed between face member 10 and body member 30 to attenuate vibrations when striking a golf ball. The dampening member extends across a large percentage, up to and including all of the face member, to ensure that vibrations are attenuated regardless of what part of the club face strikes the ball. The color of vibration member 20 can be such that it contrasts with the colors of the face member and the body member to aid in lining up the shot.—NAS

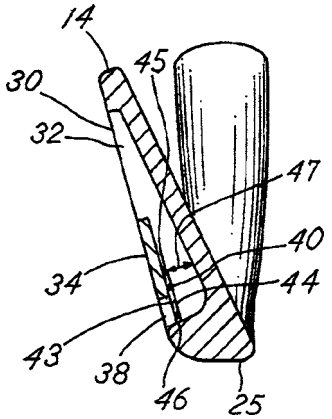


7,476,162

43.40.Kd GOLF CLUB HEAD HAVING A BRIDGE MEMBER AND A DAMPING ELEMENT

John T. Stites *et al.*, assignors to Nike, Incorporated
13 January 2009 (Class 473/332); filed 4 August 2004

The design and placement of bridge 34, first cavity 32, wall 40, and second cavity 30, among other parts, can lower the center of gravity of a golf club head 14, which is just one factor that determines whether a golf ball is propelled in the direction intended by the player using an iron type club. Parts 34, 40, as well as a viscoelastic component that can be placed between bridge 34 and rear face 40 (about where item 47 is shown) can be chosen to provide additional damping to the assembly and thereby change the feel and sound of the club when striking the ball.—NAS



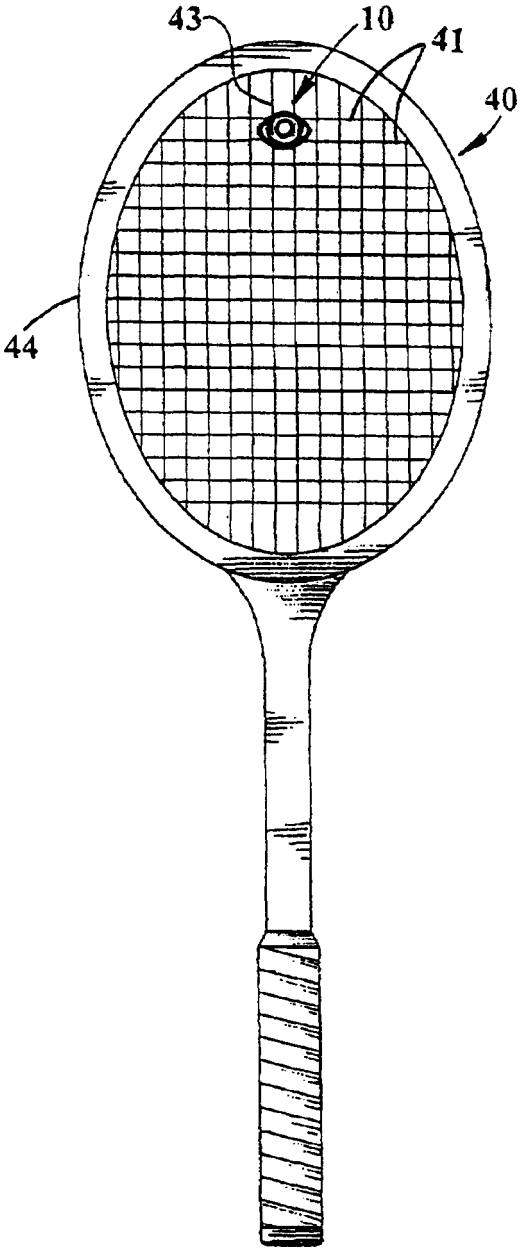
7,530,910

43.40.Kd VIBRATION DAMPENING DEVICE FOR A STRUNG SPORTS RACQUET

Gene Niksich, Alpharetta and Jim Harp, Cumming, both of Georgia
12 May 2009 (Class 473/522); filed 14 June 2007

Vibration dampening element 10 provides a "beneficial sensory effect"

by emitting either light or sound, in two embodiments, when the racket hits a tennis ball.—NAS



7,554,237

43.40.Tm CENTRIFUGAL-EFFECT VIBRATION GENERATOR HAVING COAXIAL CONTRAROTATING ROTORS

Sylvain Clary, assignor to Eurocopter
30 June 2009 (Class 310/114); filed in France 25 May 2005

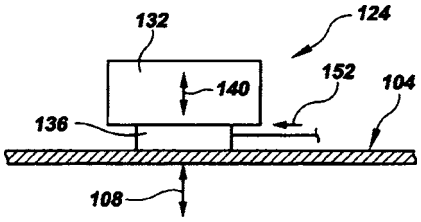
The vibration generator described in this patent is intended to counteract the vibrations generated by helicopter rotors and the like. Two coaxial unbalanced rotors are driven to counter-rotate in a housing by means of an electromagnetic arrangement that makes one rotor act much like the stator of a motor while making the other rotor act like the rotor of the motor. Electromagnetic interaction of one of the rotors with the housing permits adjustment of the phase and direction of the net unbalance force. Actuation of both rotors by a single driving system enables compact construction of this vibration generator.—EEU

7,461,728

43.40.Vn ACTIVE VIBRATION DAMPING SYSTEM

Dryver R. Huston *et al.*, assignors to University of Vermont and State Agricultural College
9 December 2008 (Class 188/378); filed 1 March 2005

An "integrated" active mass damper (AMD) system 100 is claimed where structure 104 undergoes vibration 108 that is sensed by a transducer on 104 (not shown). This sensed signal is fed to a control circuit that in turn



sends current 152 to actuator 136 that drives mass 140 to oppose vibration 108. Various other vibration modes 108 are considered, and other transducer forms 136 are proposed for appropriate AMD service.—AJC

7,548,854

43.50.Fe ARCHITECTURAL SOUND ENHANCEMENT WITH PRE-FILTERED MASKING SOUND

Kenneth P. Roy *et al.*, assignors to AWI Licensing Company
16 June 2009 (Class 704/226); filed 28 March 2002

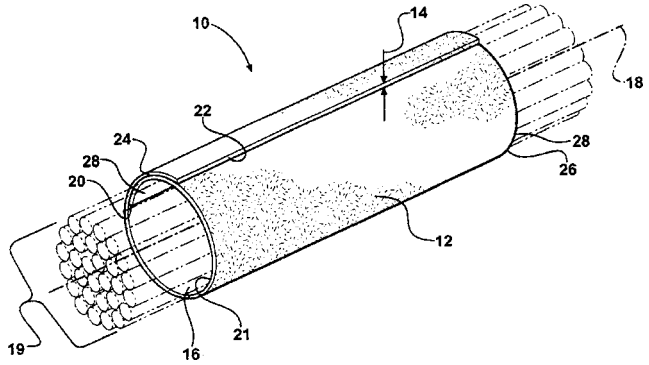
The patent outlines a method for providing various sound signals (masking sound, background music, paging, etc.) via flat panel radiators (ceiling loudspeakers) through a programmable processor control system (accessible by remote input devices). Overall, the system can support multiple zones, and the signal is equalized in third-octaves.—CJR

7,523,532

43.50.Gf NON-WOVEN SELF-WRAPPING ACOUSTIC SLEEVE AND METHOD OF CONSTRUCTION THEREOF

Hiroki Yamaguchi *et al.*, assignors to Federal Mogul World Wide, Incorporated
28 April 2009 (Class 28/110); filed 9 April 2007

A way and means of constructing a self-curling resilient acoustic sleeve 10 using a non-woven material is described, preferably polyethylene terephthalate (PET), wherein the PET fibers are intertwined (felted) and used for wall 12. A mix of 90% standard melt PET and 10% low melt PET



is said to provide "enhanced performance." A needle apparatus for felting sheet PET is described. Data for both noise and vibration reduction offered by the device are presented which are said to show superior performance for

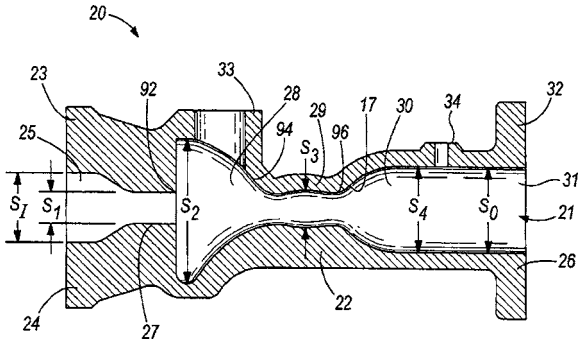
the sleeve when the preferred surface density felted PET mix is used, compared to other materials.—NAS

7,549,509

43.50.Gf DOUBLE THROAT PULSATION DAMPENER FOR A COMPRESSOR

Michael John Lucas *et al.*, assignors to Ingersoll-Rand Company
23 June 2009 (Class 181/249); filed 21 April 2005

This pulsation attenuator, intended for use at the outlet of oil-free screw compressors, in essence is a low-pass acoustic muffler consisting of a



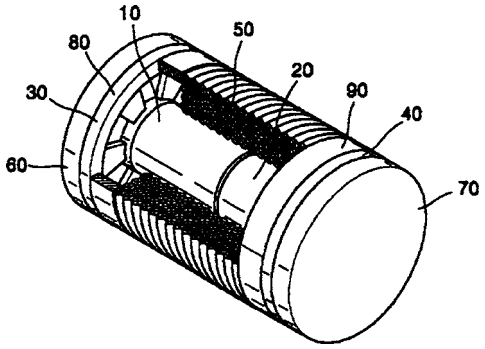
series of contraction and expansion chambers. The system is analyzed and the dimension (areas S_n and lengths) are chosen on the basis of well-known relations applicable to reactive mufflers.—EEU

7,547,275

43.66.Ts MIDDLE EAR IMPLANT TRANSDUCER

Jin Ho Cho *et al.*, assignors to Kyungpook National University Industrial Collaboration Foundation
16 June 2009 (Class 600/25); filed in Republic of Korea 25 October 2003

Distortion generated by a middle ear implant is said to be caused by transducer vibratory characteristics being different from those of the middle ear. This distortion is said to be reduced by using two vibrating members, each attached to the edges of two permanent magnets. The magnets are



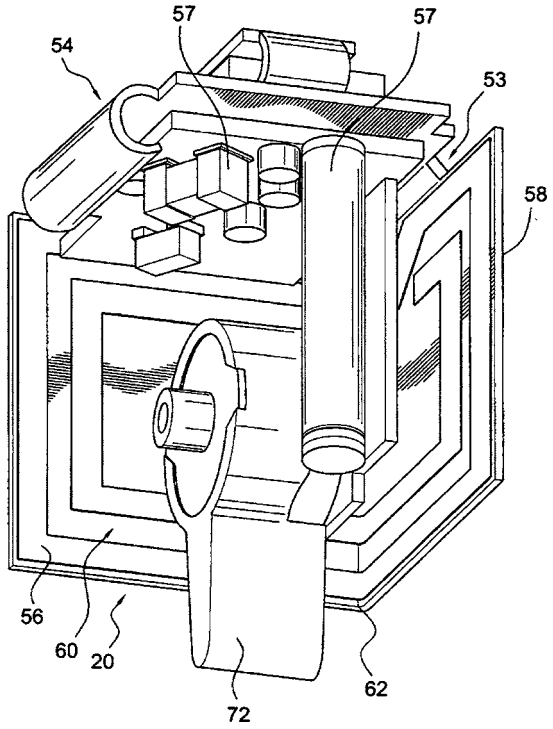
oriented such that the same poles face each other and are enclosed by a coil that is separated by a predetermined distance from the outer surfaces of the magnets. Resonant frequency and vibration displacement are adjusted by varying the thickness of support films.—DAP

7,548,211

43.66.Ts WIRELESS AUDIO SIGNAL RECEIVER DEVICE FOR A HEARING INSTRUMENT

Rainer Platz, assignor to Phonak AG
16 June 2009 (Class 343/718); filed 30 March 2006

A magnetic antenna, separately packaged with signal processing electronics and an interface to a hearing aid input (for example, via a direct audio input connection), is formed as a loop on a printed circuit board



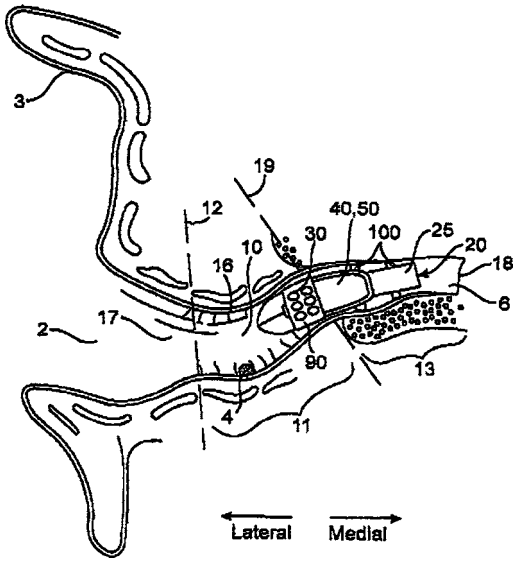
having a partially flexible substrate. The antenna has first and second planes oriented 60–120° with respect to each other to prevent nulls in the energy pickup.—DAP

7,551,747

43.66.Ts PERFORATED CAP FOR A HEARING AID

Timothy Cuongdung Huynh *et al.*, assignors to InSound Medical, Incorporated
23 June 2009 (Class 381/325); filed 14 February 2005

A perforated protective cap with a protective coating resides in the ear canal on the end of a deeply placed, completely in-the-canal hearing aid.



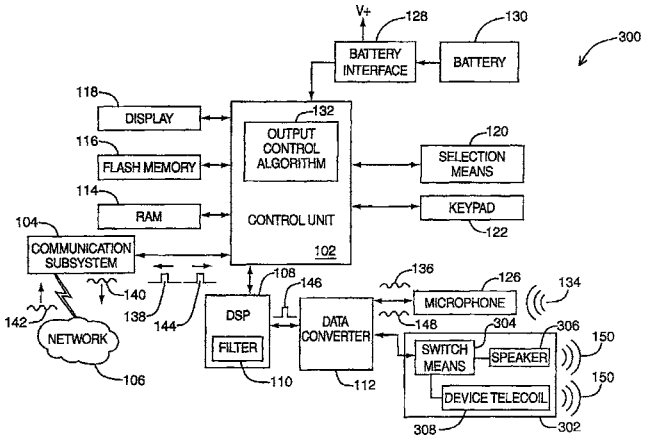
The perforations in the cap provide drainage and aeration to reduce the relative humidity surrounding the hearing aid when it is in the ear canal. Further, the cap perforations are sized and placed to help prevent moisture and debris from affecting the device's acoustical performance, while allowing sound to enter the microphone without significant attenuation.—DAP

7,551,942

43.66.Ts HEARING AID COMPATIBILITY IN A WIRELESS COMMUNICATIONS DEVICE

Larry Edward Hawker and Scott MacIntosh, assignors to Research In Motion Limited
23 June 2009 (Class 455/556.1); filed 30 July 2004

To improve coupling from wireless communication devices to hearing aids, a control unit configures the output stage and produces appropriate equalization parameters for a processed signal in the wireless device to



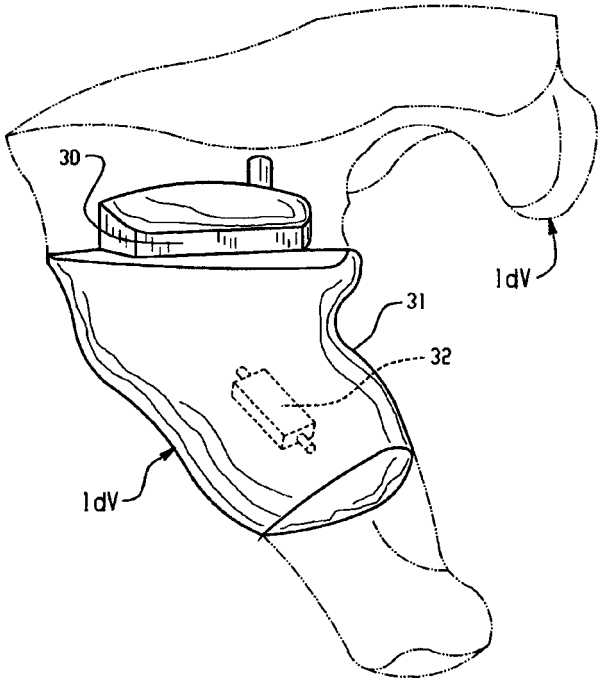
generate acoustic and/or magnetic signals for hearing aid microphone mode, telecoil mode, and a microphone/telecoil mode.—DAP

7,555,356

43.66.Ts METHOD FOR MANUFACTURING A BODY-WORN ELECTRONIC DEVICE ADAPTED TO THE SHAPE OF AN INDIVIDUAL'S BODY AREA

Samuel Hans Martin Roth *et al.*, assignors to Phonak AG
30 June 2009 (Class 700/118); filed 3 April 2003

During earshell manufacturing for custom hearing aids, one anatomical feature is automatically selected, and the location of its digitized three-



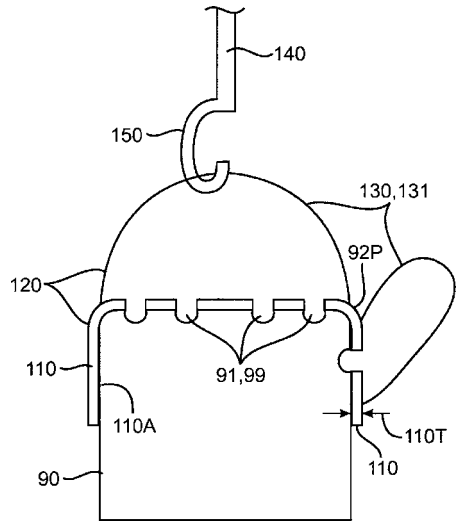
dimensional (3D) model that best fits in the digitized 3D model of the hearing aid wearer's body is automatically determined using predetermined geometric rules. Included are anatomical features such as ear canal entrance, first or second ear canal bends, tragus, and anti-tragus. The best-fitting location is then used to detail a cast and/or the digitized 3D model of the wearer's body using an expert system of recorded knowledge.—DAP

7,558,394

43.66.Ts SYSTEMS AND METHODS FOR IN SITU CERUMEN REMOVAL FROM HEARING DEVICES

Alex Tilson *et al.*, assignors to InSound Medical, Incorporated
7 July 2009 (Class 381/325); filed 25 April 2006

A method is described for removing cerumen from a deep fitting, extended-wear hearing device without disturbing its position or removing it from the ear canal. A removal tool has a hook that engages a loop on a



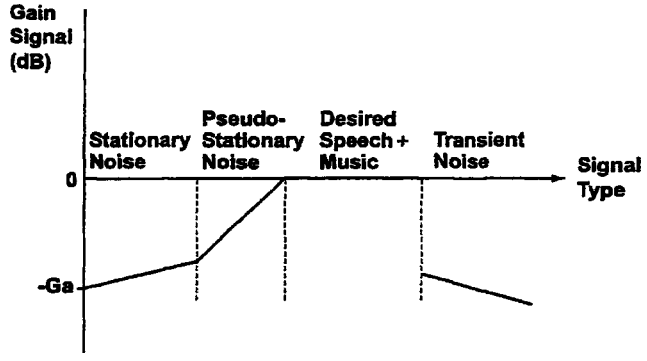
removable perforated cap that covers the hearing device. The bottom portion of the cap consists of flexible arms that slide over the side of a hearing device. Cerumen comes out with the cap as it is pulled off the hearing device and out of the ear canal.—DAP

7,558,636

43.66.Ts APPARATUS AND METHOD FOR ADAPTIVE SIGNAL CHARACTERIZATION AND NOISE REDUCTION IN HEARING AIDS AND OTHER AUDIO DEVICES

Horst Arndt and Henry Luo, assignors to Unitron Hearing Limited
7 July 2009 (Class 700/94); filed 21 March 2002

After conversion to the frequency domain, a decision is made in multiple frequency bands as to whether the input signal is desirable audio or undesirable noise based on combining an index reflecting the changes in



intensity over a first time period with an index indicating the signal temporal modulation frequency over a second time period. The result of this decision

determines whether the incoming audio signal is amplified or attenuated.—DAP

7,478,043

43.72.Ar ESTIMATION OF SPEECH SPECTRAL PARAMETERS IN THE PRESENCE OF NOISE

Robert David Preuss, assignor to Verizon Corporate Services Group, Incorporated
13 January 2009 (Class 704/233); filed 5 June 2003

This patent teaches a method for comparing speech spectra with spectral models stored in a codebook, using discriminant analysis to find the optimal model for a speech segment. It is shown how to implement an efficient computation of the discriminant function by applying the "preconditioned conjugate gradient" process. This latter method is heavily promoted in the patent as though it were the chief innovation, and the term is used in the claims. It is perhaps surprising, then, that this entire patent describes an application of an algorithm that has its own Wikipedia page. It can only be concluded that the patent was granted because it was judged to show a non-obvious way of employing a public domain method.—SAF

7,475,016

43.72.Ja SPEECH SEGMENT CLUSTERING AND RANKING

Maria E. Smith and Jie Z. Zeng, assignors to International Business Machines Corporation
6 January 2009 (Class 704/258); filed 15 December 2004

The patented system is designed to identify badly mislabeled speech sounds in a concatenative text-to-speech synthesis speech corpus, which are a common byproduct of automatic extraction algorithms that are used to segment the speech corpus. The mislabelings most commonly result from misalignment. A key guiding principle is that "the more severely misaligned a speech segment is, the more likely it is that the speech segment is part of a cluster of misaligned speech segments." This "finding" by the inventors appears to be the basis for granting the patent. Several embodiments are patented, which can include identifying clusters of potentially misaligned speech segments within an ordered sequence. How it would do so is never made precise.—SAF

7,478,039

43.72.Ja STOCHASTIC MODELING OF SPECTRAL ADJUSTMENT FOR HIGH QUALITY PITCH MODIFICATION

Ioannis G Stylianou and Alexander Kain, assignors to AT&T Corporation
13 January 2009 (Class 704/207); filed 9 May 2005

A particular method for modifying recorded speech segments for concatenative synthesis is described. The patent exploits findings which show that it is better to modify the spectral envelope along with the pitch of a segment. The system performs this modification after learning the general correlation between the pitch and the "line spectral frequencies" for the particular target speaker in the synthesis database.—SAF

7,467,086

43.72.Ne METHODOLOGY FOR GENERATING ENHANCED DEMIPHONE ACOUSTIC MODELS FOR SPEECH RECOGNITION

Xavier Menendez-Pidal *et al.*, assignors to Sony Corporation
16 December 2008 (Class 704/255); filed 16 December 2004

A method is described by which simple acoustic hidden Markov models for speech recognition can be optimized through training. The models are trained to be either "preceding dominant" or "succeeding dominant" demiphones, using rudimentary decision tree training. This idea seems simple and rather neat; it is not shown whether it helps, however.—SAF

7,478,041

43.72.Ne SPEECH RECOGNITION APPARATUS, SPEECH RECOGNITION APPARATUS AND PROGRAM THEREOF

Osamu Ichikawa, assignors to International Business Machines Corporation
13 January 2009 (Class 704/233); filed in Japan 14 March 2002

The strangely redundant title foreshadows the contents of this patent, which does not appear to contain a single innovative step. Broadly, a process is patented which involves speech recognition using a signal input through a microphone array, which uses apparently standard delay-and-sum methods to perform localization and noise suppression. Is the combination of microphone array techniques and speech recognition techniques supposed to be innovative?—SAF

7,475,013

43.72.Pf SPEAKER RECOGNITION USING LOCAL MODELS

Ryan Rifkin, assignor to Honda Motor Company, Limited
6 January 2009 (Class 704/238); filed 26 March 2004

A speaker recognition method is patented which seems very general and also relies greatly on public domain technology. It is not at all clear what the patentable innovation is. The system gathers spectrum feature data from the speech, but the nature of the features is not fully stipulated (typical cepstral vectors are mentioned as one option). It organizes the data into a structured space, but it is not specified precisely how. It uses the data structure in some unspecified way to extract "approximate nearest neighbors." It uses a Parzen window method (which has its own Wikipedia page) to estimate the ordering of approximate nearest neighbors. On the basis of this information, a speaker can be identified.—SAF

7,549,961

43.80.Vj SYSTEM AND METHOD SUPPORTING IMAGING AND MONITORING APPLICATIONS

Juinjet Hwang, assignor to SonoSite, Incorporated
23 June 2009 (Class 600/440); filed 30 July 2004

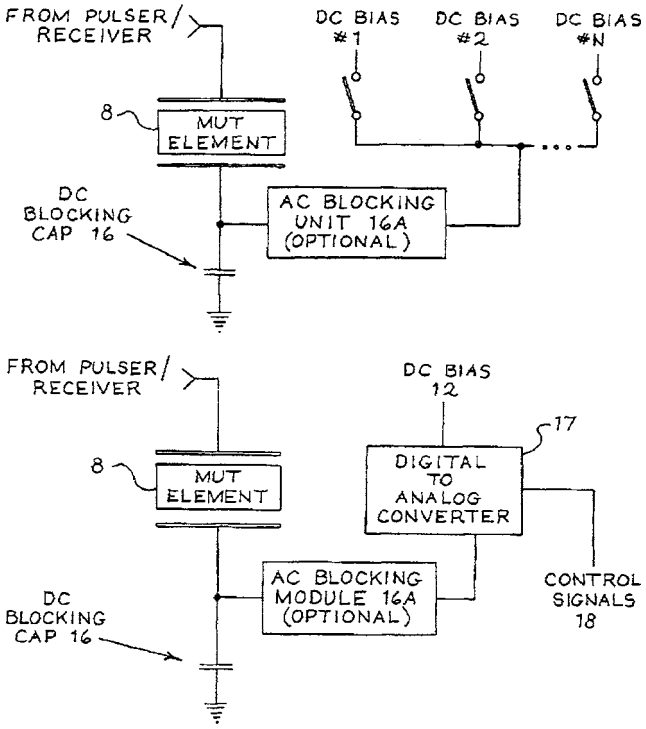
Ultrasound images are acquired by a portable system that supports simultaneous imaging and monitoring applications. The system includes an external controller for a communication device such as a medical data assistant (MDA) or a personal digital assistant (PDA) and a bus for mating with the communication device. Images can be displayed on the PDA or MDA and vital-sign monitoring can be performed. The images or monitoring information can be transferred for review at a remote location.—RCW

7,549,962

43.80.Vj DIAGNOSTIC MEDICAL ULTRASOUND SYSTEMS AND TRANSDUCERS UTILIZING MICRO-MECHANICAL COMPONENTS

William R. Dreschel *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
23 June 2009 (Class 600/443); filed 18 February 2003

Micro-mechanical components such as switches, relays, and inductors are used in an ultrasound transducer probe head, the transducer connector, or elsewhere in the imaging system to reduce size and cost while enhancing signal integrity and operational flexibility.—RCW



7,556,602

43.80.Vj BREAST CANCER SCREENING WITH ADJUNCTIVE ULTRASOUND MAMMOGRAPHY

Shih-Ping Wang *et al.*, assignors to U-Systems, Incorporated
7 July 2009 (Class 600/437); filed 31 May 2002

A sequence of images, each with a beam that is thin in the dimension perpendicular to the image plane, is acquired by mechanically scanning a transducer. From the sequence of images, an image with a thick dimension perpendicular to the image plane is assembled. Any of the images can be used for computer-aided detection or diagnosis. Vibration images can also be obtained and similarly processed.—RCW

7,556,603

43.80.Vj ULTRASONIC CELLULAR TISSUE SCREENING SYSTEM

Kevin M. Kelly and Roger Royce, assignors to SonoCine, Incorporated
7 July 2009 (Class 600/437); filed 29 August 2005

An ultrasound probe is progressively moved over a breast with the nipple covered by a pad while the breast and pad are held in place by a fabric covering.—RCW