CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 120, No. 1 July 2006

(Continued)

CONTENTS—*Continued from preceding page*

UNDERWATER SOUND [30]

TRANSDUCTION [38]

NOISE: ITS EFFECTS AND CONTROL [50]

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

JASA EXPRESS LETTERS

CUMULATIVE AUTHOR INDEX

Dual-electrode pitch discrimination with sequential interleaved stimulation by cochlear implant users

Bom Jun Kwon and Chris van den Honert
Cochlear Americas, 400 Inverness Parkway, Suite 400, Englewood, Colorado 80112
bjkwon@gmail.com

Abstract: Cochlear implant users may perceive intermediate place-pitches between those elicited by the individual electrodes when two electrodes are stimulated simultaneously or sequentially. This study examined pitch discrimination between adjacent electrodes using sequential dual-electrode stimulation in terms of the sensitivity index, d', which was obtained by adding d's from intermediate dual-electrode stimuli. Loudness was balanced for each tested pair and the intensities were roved. Twelve ears with the Nucleus® 24 or Freedom implants demonstrated a wide range of d', from 0.7 to 9.6. "Virtual channels" can be implemented through nonsimultaneous stimulation, with comparable pitch discrimination to that observed with simultaneous stimulation.

PACS numbers: 43.66.Ts, 43.66.Hg, 43.66.Fe [QF]
Date Received: March 15, 2006 **Date Accepted:** June 13, 2006

1. Introduction

Cochlear implants (CIs) provide a hearing sensation for individuals with a severe or profound degree of hearing loss. In an attempt to increase place-pitch resolution beyond the discrete set of pitches associated with the physical electrodes, researchers have proposed simultaneous activation of multiple electrodes (Townshend *et al.*, 1987), so as to create a current field by superposition of the two fields of the individual electrodes. Wilson *et al.* (1993) coined the term "virtual channels" to describe use of the resulting intermediate pitch percepts to provide increased spatial resolution. Subsequently, McDermott and McKay (1994) demonstrated that intermediate pitches could also be perceived when two electrodes were stimulated sequentially with a short temporal gap between the pulses (less than 1 ms). All five subjects in that study were able to discriminate pitch of a dual-electrode stimulus from those of single-electrode stimuli in tonotopic order on at least one pair of electrodes. However, detailed data of pitch discrimination (e.g., how many pitches the subject could distinguish) from multiple subjects were unavailable, as the study provided the data of only one subject showing discrimination of six pitches between electrodes. Recently, Donaldson *et al.* (2005) measured dual-electrode place-pitch discrimination thresholds with six subjects with the Clarion CII device, where the electrode spacing was approximately 1 mm. They estimated two to nine discriminable pitches between adjacent electrodes by extrapolating the measured thresholds. The study confirmed that a substantially greater number of informational channels than the number of physical electrodes could be implemented in a CI system. It also implies that the number of usable virtual channels is, in fact, determined by CI user's ability rather than the capacity of the system, as a large degree of variability in individual performance has been observed in the studies mentioned earlier.

In sequential stimulation, the percept of intermediate pitch may be based on interaction of neural signals at the brainstem or higher, as there is no vector summation of current fields. A peripheral mechanism may also be involved. For instance, with a sufficiently short gap ($\ll$1 ms), some neurons located between the electrodes would be in a refractory state, while others may be sensitized by the first pulse, and more susceptible to stimulation by a subsequent pulse on an adjacent electrode. Perhaps the proportion of fibers in a refractory state at each

location would vary with regard to the current proportion on electrodes and somewhat influence the pitch sensation. At any rate, it was not clear whether sequential stimulation is as effective in delivering intermediate pitches as simultaneous stimulation. Thus, in the present study, it was of primary interest to measure pitch discrimination in detail with sequential dual-electrode stimulation and to compare it with published results using simultaneous stimulation.

With the same ultimate goal of estimating the number of discriminable pitch steps between adjacent electrodes, the present study differed in methodology from the study by Donaldson *et al.* (2005). They measured the just-noticeable-difference (jnd) in a discrimination task, where a single-electrode stimulus was always the reference (fixed) signal and a dual-electrode stimulus was systematically varied. The jnd, measured in terms of α, the relative current proportion of basal electrode to the total dual-electrode currents, ranged from 0.11 to 0.64. Extrapolation of this jnd measure across the entire range of α led to the conclusion that dual-electrode stimuli could produce two to nine discriminable pitches between the adjacent single-electrode pair. This conclusion was based on the assumption that the performance function (d' in discrimination) was linear with the variable, α. The assumption seems reasonable, as several d' curves in their Fig. 1 showed the linearity. It is uncertain, however, whether this reasoning is valid for subjects demonstrating good performance (such as D01 and D08), without the linearity seen in the whole range of α. In the present study, the values of d' were directly obtained from pitch-judgment tests between necessary pairs of the dual-electrode stimuli and were summed to obtain the index indicating a perceptual distance between the adjacent electrodes. This approach would be particularly useful to estimate the number of discriminable pitches for subjects with good performance, as all of the pairs of dual-electrode stimuli in the range were to be actually tested.

2. Methods and procedure

2.1 Revisiting d′ analysis

The sensitivity index, d', indicates perceptual distance between two stimuli. Under the assumption that the pitch percept is unidimensional and the distribution of pitch judgment is Gaussian, d' is additive (Nelson *et al.*, 1995). In other words, when perceptual distance between stimulus X and Y is measured in terms of d', it should be consistent with the summation of d's measured between the stimulus X and A and between the stimulus A and Y (assuming that A exists between X and Y on the unidimensional space). In the study by Nelson *et al.*, the additivity was not fully utilized for the subjects demonstrating excellent place-pitch sensitivity due to the ceiling effect, which necessitated substitution of d' of 3.29 when a perfect score (100%) was obtained in a pitch ranking test. The value of d' for a perfect score is mathematically infinite, making it unattainable in reality under the assumption of Gaussian distribution. Therefore, when the measured score was 100% after a certain number of trials (40 in the study), Nelson *et al.* approximated it at 99% instead of 100%, as the comparison of individual subjects' performance across electrode locations was of interest. In the present study, a more elaborate scheme was used to estimate d' between "perfectly" discriminable single electrodes, as the goal was to examine the discrimination of place-pitch in detail with dual-electrode stimuli between adjacent electrodes.

In this scheme, if the subject's judgment of pitch with a pair of stimuli ($\{X, Y\}$) was perfect (a 100% score) after a given number of trials, a new stimulus (A) was constructed between X and Y, and subsequently pitch comparison was tested for the new subpairs ($\{X, A\}$) and $\{A, Y\}$. We refer to this process as *subdivision* of the X-Y interval. If the score from a tested pair was less than 100%, d' could be determined for the pair (Hacker and Ratcliff, 1979); if not, a further subdivision was done to create more difficult pairs, until the tested pair led to a score less than 100%. With this approach, the summation of all d's measured represents a good estimate of perceptual distance for place-pitch between the two single-electrode stimuli. It should be noted that d' is independent of the point of subdivision. For instance, the interval could be subdivided at either at A or B, as seen in the following equation:

$$d'_{X,Y} = d'_{X,A} + d'_{A,Y} = d'_{X,B} + d'_{B,Y}, \quad (1)$$

provided that both A and B are on the unidimensional space and all scores are less than 100%. In reality, the subdivision point (A or B) should be chosen with reasonable care. If the dividing point is located too close to either end, the estimate of performance, and hence d', might not be reliable. For instance, if A is too close to X, then $d'_{X,A}$ approaches zero, but the estimate obtained from a small number of trials (such as 20) is not reliably zero. Therefore, an excessive subdivision should be avoided; i.e., each d' obtained should not be near zero.

Collins and Throckmorton (2000) indicated, with a multi-dimensional scaling analysis, that perceptual features of loudness-matched single-electrode stimuli on different electrodes could be better represented by two, rather than one, dimensions, challenging the presumed unidimensionality. However, in a similar investigation of perceptual dimensionality, McKay *et al.* (1996) demonstrated that dual-electrode stimuli lay along the line connecting corresponding single-electrode stimuli, implying that dual-electrode stimuli could be represented by a single dimension in the range between corresponding two single-electrode stimuli, if the interelectrode delay is short (less than 1 ms). Considering that Collins and Throckmorton's indication was based on the observation of the wide range of the electrode array, it appears to be reasonable to accept the assumption of the *local* unidimensionality between adjacent electrodes, as implied by McKay *et al.* In addition, during a pilot study with two subjects, we cross-checked the results obtained through the procedure of subdivision (proposed in this study) with those obtained with a cascade of adaptive procedures, where each successive adaptive procedure (Levitt, 1971) was done with the reference stimulus at the last-measured discrimination threshold (a cumulative d' could be obtained from each adaptive run with a two alternative-forced choice and two-down, one-up paradigm producing the percent correct of 71%, i.e., d' of 0.78). There was generally close agreement between the two results. In another preliminary study, we directly compared the summation of d's obtained with different choices of subdivision points (in six different electrode pairs with four subjects) and found good consistency. Thus, the additivity seen in Eq. (1), the fundamental basis of the present study, was justified by theoretical considerations elaborated earlier and our own validation.

2.2 Procedure

The following three electrode pairs were tested to obtain d': E19/18, E11/10, and E4/3. The current was specified in units of current level (CL), which is an integer, logarithmic unit between 0 (10 μA) and 255 (1.75 mA). One CL step corresponds to a 0.176-dB change in current. The stimulation was always charge-balanced, biphasic, and monopolar. The pulse width was 25 μs and an interphase gap was 8 μs. Each stimulus comprised two 500-ms bursts on the two adjacent electrodes, interleaved with a gap of 19 μs. Eleven subjects participated in the study; one of them, R4, was a bilateral user; therefore, a total of 12 ears were tested. Among those, there were two ears with a Nucleus CI24 implant with a straight, banded electrode array, and ten ears with a Contour perimodiolar electrode array (8 Nucleus Contour™ and 2 Nucleus Freedom™ implants[1]). The electrode spacing was constant at 0.75 mm for the straight array and varied for each electrode pair for the Contour array (0.73, 0.57, and 0.47 mm, for E4/3, E11/10, and E19/18, respectively). Each subject had at least 6 months of device experience. The stimulation rate was chosen from the subject's clinical MAP, ranging from 500 to 1800 Hz, to use the rate most familiar to individual subjects, with the expectation that place-pitch would not be greatly influenced by the stimulation rates used in the study. Custom software written in C++ based on the NIC™ 1.0 library (distributed by Cochlear Americas, Englewood, CO) generated and delivered the stimuli from a PC, administered the experiment, and collected the subject's response.

The task was pairwise pitch comparison, where the subject was presented with a pair of stimuli in a random order and was asked to indicate which interval was higher in pitch. Subjects were allowed to repeat the presentation as many times as desired before they responded. A percent correct score was obtained from 20 trials without response feedback for each pair of stimuli. The "pitch steering" of a dual-electrode stimulus was characterized by Δ,

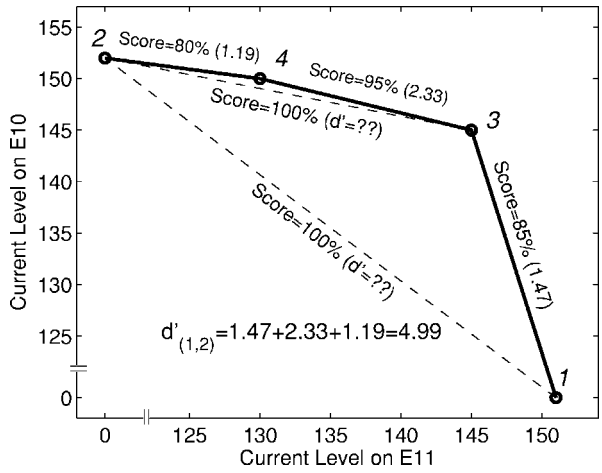


Fig. 1. Stimulus pairs tested plotted on the current level space for subject R29. The abscissa and ordinate indicate current level on electrodes 11 and 10, respectively. The subdivision was made whenever a 100% score was obtained (see text for detail). The two stimuli at each line were loudness balanced prior to testing.

the difference in CL, between the two electrodes. Namely, a Δ of 0 indicated that the stimulus had the same currents on both electrodes, a negative value of Δ indicated that the current on the apical side electrode was greater than the basal side (with pitch "steered" lower), and vice versa. Therefore, the subject's response was considered as correct (i.e., tonotopically consistent) when the subject indicated the stimulus with higher Δ as "higher pitch." For every session, loudness was first balanced for the tested stimuli (see below). Furthermore, the currents of stimuli were randomly roved by as much as ±2 CLs in each presentation, to reduce a systematic bias due to any residual imbalance of loudness. Prior to data collection, subjects were given a brief—no more than 15 min[2]—practice session with feedback using not more than two dual-electrode stimuli. The practice sessions were meant to provide inexperienced subjects with a necessary guideline as to how the judgment could be made, ensuring that the concepts of pitch and loudness were adequately differentiated.

To maintain the continuum of stimulus variation, dual-electrode stimuli were always used in the present study, i.e., single-electrode stimuli were actually created by setting the intensity for the other unused electrode at 0 CL, i.e., 10 μA. The initial stimulus level was chosen to be at 2 CLs below the C-level on the electrode from the subject's MAP. Subdivision was done whenever the score was 100%, as described earlier. For instance (as illustrated in Fig. 1), a pitch testing session was initially done with a loudness-balanced pair of CL 153 on E11 and 152 on E10. The score was 100%, and the pair was subdivided at a Δ of 0. Loudness of the stimulus with a Δ of 0 was balanced against stimulus 1 by adjusting current levels on both electrodes together up or down, while maintaining the Δ value fixed at 0. As a result, stimulus 3 was constructed with a CL of 146 on both electrodes and was subsequently tested with stimulus 1. The score was 85%, producing d' of 1.47. The score of the complementary pair (stimulus 3 and stimulus 2) was 100%; therefore, another subdivision was made at a Δ of 20, i.e., stimulus 4 (again with loudness balancing), thereby new pairs were produced (stimuli 3 and 4, stimuli 4 and 2). As the scores were nonperfect with these pairs, no further subdivision was necessary. The summation of d' values, 4.99, was then taken as the final estimate of d' for this electrode pair, E11/10. As expected, the better the subject performed in pitch discrimination, the longer it took to obtain d', because more pairs needed to be tested.

3. Results and discussion

Figure 2 displays the d' values for all the conditions of the electrode locations and subjects. First of all, performance varied widely across not only subjects but also the electrode locations. The

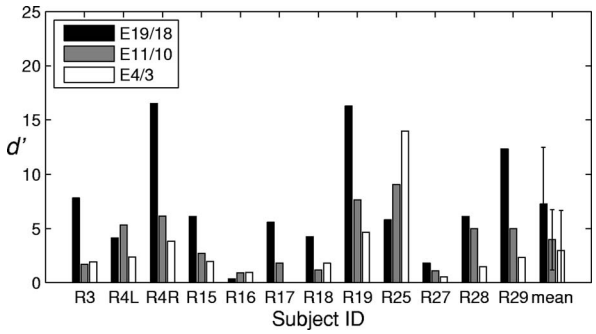


Fig. 2. Bar graph of d' for 12 ears and three different electrode pairs. The data is displayed by each subject. The data of the bilateral recipient, R4, are displayed as R4L and R4R, for the left and right sides, respectively. The bar of the E4/3 condition for R17 is not visible in the figure because the d' value was near zero.

average individual d' values across the electrode locations ranged from 0.7 (R16) to 9.6 (R25). Pitch discrimination was generally best at the apical location and degraded as the pair moved to basal locations. Overall averages of d' values across subjects for each electrode pair were 7.26, 3.97, and 2.97 for E19/18, E11/10, and E4/3, respectively. The general pattern of relatively poorer performance at the basal location was consistent with previous studies (e.g., McDermott and McKay, 1994). First of all, poor neural survival at the base might have attributed to this trend. Second, the choice of stimulus levels in this experiment might also have been attributed to the trend, as known from the previous studies (Pfingst *et al.*, 1999; Donaldson *et al.*, 2005). Because the choice of stimulus levels in the present study was based on the individual's clinical MAP, it was unclear how well the loudness was balanced across the electrode array. Subjects often reported that the basal stimuli were "not as loud" as the apical ones—but "equally comfortable" (note that CI users often prefer to set the loudness on basal electrodes relatively lower in their MAPs). However, without knowing precisely how the loudness varied across the array, it is difficult to assess accurately the effects of stimulation levels on the poor performance on the basal locations. Notably, relatively poor pitch discrimination on the basal location does not necessarily suggest that CI users practically have more trouble with high frequencies than low frequencies when listening to a speech sound through the device. The frequency range covered by E4 and E3 is approximately between 4000 and 6000 Hz, whereas the range of electrodes E19 and E18 is between 400 and 900 Hz. For speech understanding, pitch discrimination between E19 and E18 is likely to be more critical than that between E4 and E3. Therefore, this observed difference in performance of pitch discrimination across the electrode array, even if it still existed under the conditions of equal loudness, would not necessarily suggest a shortcoming in speech representation with current CI processing systems.

A value of d' simply indicates perceptual distance or an index for distinctiveness. In order to interpret this to the number of discriminable steps (jnd's), a criterion needs to be chosen (the number of jnd's would vary depending on the criterion). With the criterion used by Donaldson *et al.* (2005; $d' = 1.16$), the numbers of jnd's become 6.3, 3.4, and 2.6 for the apical, middle, and basal pair, respectively. The number of jnd's for each individual varied from 0.6 (R16) to 8.3 (R25), which is comparable to the results of Donaldson *et al.* (note that the Nucleus electrode array has narrower electrode spacing). The results in the present study affirm that intermediate pitches, upon which virtual channels are predicated, can be elicited between electrodes for most subjects through nonsimultaneous interleaved stimulation and the achievable pitch resolution is comparable to that achieved through simultaneous stimulation. The concept of simultaneous dual-electrode stimulation was intriguing, because the profile of neural activation is not limited by physical electrodes and could be created for specific needs. However, as sequential dual-electrode stimulation provides the same degree of pitch discrimination for CI users, the relative advantage of simultaneous implementation of dual-electrode stimulation is yet to be seen.

It should be kept in mind that encoding of frequency distinctions beyond the resolution

of existing physical electrodes has been *implicitly* implemented in the form of sequential dual-electrode stimulation in pulsatile processing strategies that are clinically used, because of the overlapping filter skirts of bandpass filters involved in the spectral analysis. As the input frequency of a tone moves from one band into the next, the output from the lower band decreases while that of the upper band increases. The present results provide information as to how much this implicit coding of frequency could be (or is actually being) utilized by CI users. It is apparent that it varies greatly across individuals. If the concept of pitch steering were to be implemented *explicitly* (i.e., the changes in input frequencies are encoded with dual-electrode stimuli and perceived as intended), a good scheme to reconcile such a wide individual variability would be desirable.

Acknowledgments

The authors are grateful to Mario Svirsky, Peter Busby, Qian-Jie Fu, and three anonymous reviewers for their feedback and constructive criticisms during the preparation of the manuscript.

References

[1]R3 and R4 (left side) were implanted with CI24/straight and R28 and R29 were with Freedom/Contour (the rest were the CI24/Contour).

[2]R16 and R27 needed more time for practice due to their poor performance, yet the practice did not improve the scores.

Collins, L. M., and Throckmorton, C. S. (**2000**). "Investigating perceptual features of electrode stimulation via a multidimensional scaling paradigm," J. Acoust. Soc. Am. **108**, 2353–2365.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **101**, 440–455.

Donaldson, G. S., Kreft, H. A., and Litvak, L. (**2005**). "Place-pitch discrimination of single—versus dual-electrode stimuli by cochlear implant users (L)," J. Acoust. Soc. Am. **118**, 623–626.

Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d′ for M-alternative forced choice," Percept. Psychophys. **26**, 168–170.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

McDermott, H. J., and McKay, C. M. (**1994**). "Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea," J. Acoust. Soc. Am. **96**, 155–162.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1996**). "The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees," J. Acoust. Soc. Am. **99**, 1079–1090.

Nelson, D. A., van Tasell, D., Schroder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.

Pfingst, B. E., Holloway, L. A., Zwolan, T. A., and Collins, L. M. (**1999**). "Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants," Hear. Res. **134**, 105–115.

Townshend, B., Cotter, N., van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

Wilson, B. S., Zerbi, M., and Lawson, D. (1993). "Speech processors for auditory prostheses," NIH Contract N01-DC-2-2401, 3rd Quarterly Progress Report, 1 February–30 April 1993.

Nonreciprocal Green's function retrieval by cross correlation

Kees Wapenaar
Department of Geotechnology, Delft University of Technology, P.O. Box 5028, 2600 GA Delft, The Netherlands
c.p.a.wapenaar@tudelft.nl

Abstract: The cross correlation of two recordings of a diffuse acoustic wave field at different receivers yields the Green's function between these receivers. In nearly all cases considered so far the wave equation obeys time-reversal invariance and the Green's function obeys source-receiver reciprocity. Here the theory is extended for nonreciprocal Green's function retrieval in a moving medium. It appears that the cross correlation result is asymmetric in time. The causal part represents the Green's function from one receiver to the other whereas the acausal part represents the time-reversed version of the Green's function along the reverse path.

PACS numbers: 43.20.Bi, 43.60.Ac, 43.60.Tj [ANN]
Date Received: March 12, 2006 **Date Accepted:** June 13, 2006

1. Introduction

It has been shown by many researchers in geophysics, ultrasonics, and underwater acoustics that the cross correlation of acoustic wave fields recorded by two different receivers yields the response at one of the receiver positions as if there was a source at the other.[1–7] Various theories have been developed to explain this phenomenon, ranging from diffusion theory for enclosures,[8,9] multiple scattering theory and stationary phase theory for random media,[10–12] and reciprocity theory for deterministic and random media.[13–15] In nearly all cases it is assumed that the medium is lossless and nonmoving, which is equivalent with assuming that the underlying wave equation is invariant for time reversal. Moreover, in all cases the Green's functions obey source-receiver reciprocity. The time-reversal invariance together with the source-receiver reciprocity property has been elegantly exploited in an intuitive derivation,[16] building on earlier work on time-reversed acoustic focusing.[17] In a medium with losses the wave equation is no longer invariant for time reversal, but, as long as the medium is not moving, source-receiver reciprocity still holds. When the losses are not too high, the cross-correlation method yields a Green's function with correct travel times and approximate amplitudes.[18,19] On the other hand, in a moving medium, both the time-reversal invariance and source-receiver reciprocity break down. It has previously been shown that, with some modifications, time-reversed acoustic focusing can still work in a moving medium.[20,21] In this paper we derive a theory for nonreciprocal Green's function retrieval by cross correlation in a moving medium.

2. Nonreciprocal Green's function representation

The basis for our derivation is a reciprocity theorem, where "reciprocity" should be interpreted in a broader sense than source-receiver reciprocity. In general a reciprocity theorem relates two independent acoustic states in one and the same domain.[22,23] One can distinguish between reciprocity theorems of the convolution type and of the correlation type.[24] In the following we derive a correlation-type reciprocity theorem for a moving, arbitrary inhomogeneous, lossless acoustic medium, and show step-by-step how this leads to a simple expression for nonreciprocal Green's function retrieval by cross correlation.

Consider an acoustic wave field, characterized by the acoustic pressure $p(\mathbf{x},t)$ and particle velocity $v_i(\mathbf{x},t)$, propagating in a lossless inhomogeneous flowing medium with mass density $\rho(\mathbf{x})$, compressibility $\kappa(\mathbf{x})$, and stationary inhomogeneous flow velocity $v_k^0(\mathbf{x})$. Throughout this paper we assume that the spatial variations of the flow velocity are small in

comparison with those of the wave field. For this situation the equation of motion and the stress-strain relation read $\rho D_t v_i + \partial_i p = 0$ and $\kappa D_t p + \partial_i v_i = q$, respectively, where $q(\mathbf{x},t)$ is a source distribution in terms of volume injection rate density, ∂_i is the partial derivative in the x_i direction, and D_t is the material time derivative,[25] defined as $D_t = \partial_t + v_k^0 \partial_k$. We define the temporal Fourier transform of a space- and time-dependent quantity $p(\mathbf{x},t)$ as $\hat{p}(\mathbf{x},\omega) = \int \exp(-j\omega t) p(\mathbf{x},t)\, \mathrm{d}t$. In the space-frequency domain the equation of motion and the stress-strain relation thus become $\rho(j\omega + v_k^0 \partial_k)\hat{v}_i + \partial_i \hat{p} = 0$ and $\kappa(j\omega + v_k^0 \partial_k)\hat{p} + \partial_i \hat{v}_i = \hat{q}$, respectively.

We introduce two independent acoustic states, which will be distinguished by subscripts A and B, and consider the following combination of wave fields in both states: $\hat{p}_A^* \hat{v}_{i,B} + \hat{v}_{i,A}^* \hat{p}_B$, where the asterisk denotes complex conjugation. In the following we assume that the medium parameters and flow velocities in both states are identical; only the sources and wave fields are different (but a more general derivation is possible[26]). The correlation-type reciprocity theorem is obtained by applying the differential operator ∂_i, according to $\partial_i\{\hat{p}_A^* \hat{v}_{i,B} + \hat{v}_{i,A}^* \hat{p}_B\}$, substituting the equation of motion and the stress-strain relation for states A and B, integrating the result over a spatial domain V with boundary S and outward pointing normal vector $\mathbf{n} = (n_1, n_2, n_3)$ and applying the theorem of Gauss. This gives

$$\int_V \{\hat{q}_A^* \hat{p}_B + \hat{p}_A^* \hat{q}_B\}\mathrm{d}^3\mathbf{x} = \oint_S \{\hat{p}_A^* \hat{v}_{i,B} + \hat{v}_{i,A}^* \hat{p}_B\} n_i \, \mathrm{d}^2\mathbf{x} + \int_V v_k^0 \{\kappa \partial_k (\hat{p}_A^* \hat{p}_B) + \rho \partial_k (\hat{v}_{i,A}^* \hat{v}_{i,B})\}\mathrm{d}^3\mathbf{x}. \tag{1}$$

This relation is independent of the choice of S; moreover, the medium and flow velocity can be inhomogeneous inside as well as outside S. In comparison with the convolution-type reciprocity theorem, Eq. (1) is remarkably simple. The convolution-type theorem can only be simplified to a form similar to Eq. (1) by choosing opposite flow velocities in the two states.[26–29] In the correlation-type theorem of Eq. (1) the flow velocities in both states are identical.

Next we choose impulsive point sources in both states, according to $\hat{q}_A(\mathbf{x},\omega) = \delta(\mathbf{x} - \mathbf{x}_A)$ and $\hat{q}_B(\mathbf{x},\omega) = \delta(\mathbf{x} - \mathbf{x}_B)$, with $\mathbf{x}_A$ and $\mathbf{x}_B$ both in V. The wave field in state A can thus be expressed in terms of a Green's function, according to

$$\hat{p}_A(\mathbf{x},\omega) = \hat{G}^{p,q}(\mathbf{x},\mathbf{x}_A,\omega), \tag{2}$$

$$\hat{v}_{i,A}(\mathbf{x},\omega) = \hat{G}_i^{v,q}(\mathbf{x},\mathbf{x}_A,\omega). \tag{3}$$

The superscripts refer to the observed wave field quantity at $\mathbf{x}$ and the source type at $\mathbf{x}_A$, respectively. Similar expressions hold for the wave field in state B. Substitution into Eq. (1) gives

$$\hat{G}^{p,q}(\mathbf{x}_A,\mathbf{x}_B,\omega) + \{\hat{G}^{p,q}(\mathbf{x}_B,\mathbf{x}_A,\omega)\}^* = \sum_{n=1}^{4} I_n, \tag{4}$$

where

$$I_1 = \oint_S \{\hat{G}^{p,q}(\mathbf{x},\mathbf{x}_A,\omega)\}^* \hat{G}_i^{v,q}(\mathbf{x},\mathbf{x}_B,\omega) n_i \, \mathrm{d}^2\mathbf{x}, \tag{5}$$

$$I_2 = \oint_S \{\hat{G}_i^{v,q}(\mathbf{x},\mathbf{x}_A,\omega)\}^* \hat{G}^{p,q}(\mathbf{x},\mathbf{x}_B,\omega) n_i \, \mathrm{d}^2\mathbf{x}, \tag{6}$$

$$I_3 = \int_V v_k^0 \kappa \partial_k [\{\hat{G}^{p,q}(\mathbf{x},\mathbf{x}_A,\omega)\}^* \hat{G}^{p,q}(\mathbf{x},\mathbf{x}_B,\omega)] \mathrm{d}^3\mathbf{x}, \tag{7}$$

$$I_4 = \int_V v_k^0 \rho \partial_k [\{\hat{G}_i^{v,q}(\mathbf{x},\mathbf{x}_A,\omega)\}^* \hat{G}_i^{v,q}(\mathbf{x},\mathbf{x}_B,\omega)] \mathrm{d}^3\mathbf{x}. \quad (8)$$

Equations (4)–(8) show how the Green's function in a medium with flow can, in principle, be obtained from cross correlations of Green's functions in the same medium. However, application of these equations requires the measurements of different types of Green's function. In the following we make a number of approximations which make these expressions suited for practical applications.

First we assume that the medium at and outside S is homogeneous, so that the Green's functions in I_1 and I_2 represent outgoing waves at S. Moreover, we assume that the flow velocity at S is small in comparison with the propagation velocity c, i.e., $|v_k^0 n_k|/c \ll 1$. We express the Green's function $\hat{G}_i^{v,q}$ in terms of $\hat{G}^{p,q}$ using the approximation $\hat{G}_i^{v,q} n_i \approx (1/\rho c)\hat{G}^{p,q}$. This is the high-frequency approximation for a normally outward propagating ray in a nonflowing medium. It involves an amplitude error for non-normal outward propagating rays in a flowing medium, but it handles the phase correctly (in the high-frequency regime). By using this approximation we avoid the need of determining the inhomogeneous propagation and flow models, tracing the rays and computing the propagation angles at S. With this approximation we find[30]

$$I_1 \approx I_2 \approx \frac{1}{\rho c} \oint_S \hat{G}^*(\mathbf{x},\mathbf{x}_A,\omega)\hat{G}(\mathbf{x},\mathbf{x}_B,\omega)\mathrm{d}^2\mathbf{x}. \quad (9)$$

Here and in the following $\hat{G}$ stands for $\hat{G}^{p,q}$. To show that I_3 and I_4 are small, we assume that the spatial variations of the medium parameters (as well as those of the flow velocity) are small in comparison with those of the wave field. Using the theorem of Gauss and $\kappa = 1/\rho c^2$ we may thus rewrite I_3 as

$$I_3 \approx \frac{1}{\rho c} \oint_S \hat{G}^*(\mathbf{x},\mathbf{x}_A,\omega)\hat{G}(\mathbf{x},\mathbf{x}_B,\omega)\frac{v_k^0 n_k}{c}\mathrm{d}^2\mathbf{x}. \quad (10)$$

Using the aforementioned assumption $|v_k^0 n_k|/c \ll 1$ we thus find $I_3 \ll I_1$. In a similar way we find $I_4 \ll I_1$. In the following we replace the right-hand side of Eq. (4) by $2I_1$, with I_1 approximated by Eq. (9).

Next we interchange the source and receiver coordinates in the Green's functions. According to the flow reversal theorem[26–29] this is allowed if we simultaneously revert the flow direction, i.e., if we replace $v_k^0(\mathbf{x})$ by $-v_k^0(\mathbf{x})$. We apply this to all Green's functions in Eq. (4), with the right-hand side approximated by $2I_1$, hence

$$\hat{G}(\mathbf{x}_B,\mathbf{x}_A,\omega) + \hat{G}^*(\mathbf{x}_A,\mathbf{x}_B,\omega) \approx \frac{2}{\rho c} \oint_S \hat{G}^*(\mathbf{x}_A,\mathbf{x},\omega)\hat{G}(\mathbf{x}_B,\mathbf{x},\omega)\mathrm{d}^2\mathbf{x}, \quad (11)$$

where all Green's functions are now defined in a medium with flow velocity $-v_k^0(\mathbf{x})$. The minus sign is not important; what matters is that the flow velocity is the same for all Green's functions in this equation. From here onward we define $w_k^0(\mathbf{x}) = -v_k^0(\mathbf{x})$ as the actual flow velocity. Hence, Eq. (11) applies to the actual situation and, with hindsight, Eqs. (4)–(10) apply to the situation with the reversed flow velocity $-w_k^0(\mathbf{x})$.

Applying an inverse Fourier transform to Eq. (11) yields

$$G(\mathbf{x}_B,\mathbf{x}_A,t) + G(\mathbf{x}_A,\mathbf{x}_B,-t) \approx \frac{2}{\rho c} \oint_S G(\mathbf{x}_A,\mathbf{x},-t) * G(\mathbf{x}_B,\mathbf{x},t)\mathrm{d}^2\mathbf{x}, \quad (12)$$

where the asterisk denotes temporal convolution. The right-hand side represents an integral of cross correlations of observations of the acoustic pressure in a moving medium at $\mathbf{x}_A$ and $\mathbf{x}_B$,

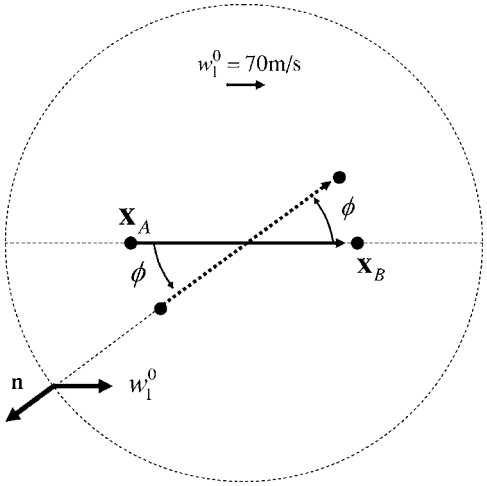


Fig. 1. Two receivers at $\mathbf{x}_A$ and $\mathbf{x}_B$ in a moving medium with constant flow velocity. The receivers are surrounded by 360 noise sources on a circle.

respectively, due to impulsive sources of volume injection rate at $\mathbf{x}$ on S; the integration takes place along the source coordinate $\mathbf{x}$. The left-hand side is the superposition of the response from $\mathbf{x}_A$ to $\mathbf{x}_B$ and the time-reversed response from $\mathbf{x}_B$ to $\mathbf{x}_A$. Note the similarity with the expressions for the situation of a nonmoving medium.[13–16,30,31] However, unlike for the situation of a nonmoving medium, our result is asymmetric in time. $G(\mathbf{x}_B,\mathbf{x}_A,t)$ is obtained by taking the causal part of the left-hand side of Eq. (12), $G(\mathbf{x}_A,\mathbf{x}_B,t)$ by time-reverting the acausal part.

Until now we assumed that the sources on S are impulsive point sources, of which the responses are measured independently. Let us now consider noise sources $N(\mathbf{x},t)$ that act simultaneously for all $\mathbf{x}$ on S. For the observed wave field at $\mathbf{x}_A$ we write $p^{\text{obs}}(\mathbf{x}_A,t) = \oint_S G(\mathbf{x}_A,\mathbf{x},t) * N(\mathbf{x},t)\mathrm{d}^2\mathbf{x}$; a similar expression holds for the observed wave field at $\mathbf{x}_B$. We assume that any two noise sources $N(\mathbf{x},t)$ and $N(\mathbf{x}',t)$ with $\mathbf{x} \neq \mathbf{x}'$ are uncorrelated and that their autocorrelation $C(t)$ is independent of $\mathbf{x}$. Hence, we assume that the source distribution on S obeys the relation $\langle N(\mathbf{x},-t) * N(\mathbf{x}',t)\rangle = \delta(\mathbf{x}-\mathbf{x}')C(t)$, where $\langle\cdot\rangle$ denotes a spatial ensemble average.[6,12–16] Equation (12) can thus be rewritten as

$$\{G(\mathbf{x}_B,\mathbf{x}_A,t) + G(\mathbf{x}_A,\mathbf{x}_B,-t)\} * C(t) \approx \frac{2}{\rho c}\langle p^{\text{obs}}(\mathbf{x}_A,-t) * p^{\text{obs}}(\mathbf{x}_B,t)\rangle. \tag{13}$$

According to this equation the cross correlation of the observed noise fields at $\mathbf{x}_A$ and $\mathbf{x}_B$ in a moving medium yields the Green's function from $\mathbf{x}_A$ to $\mathbf{x}_B$ plus the time-reversed Green's function from $\mathbf{x}_B$ to $\mathbf{x}_A$, convolved with the autocorrelation of the noise sources. Note the resemblance with the retrieval of the Green's function in a diffuse wave field in a nonmoving medium.[4–16] Again the main difference is the temporal asymmetry of the correlation result in a moving medium versus the symmetry of that in a nonmoving medium.

3. Numerical example

We illustrate Eq. (13) with a 2-D numerical example. Consider a homogeneous medium with propagation velocity $c=350$ m/s and a constant flow in the x_1 direction, with flow velocity $w_1^0=70$ m/s (see Fig. 1). Hence, the Mach number, defined as $\mathrm{M}=w_1^0/c$, equals 0.2. Following a similar derivation as for the 3-D situation[28] we obtain for the 2-D Green's function $\hat{G}(\mathbf{x},\mathbf{x}_A,\omega)=\rho(j\omega+w_1^0\partial_1)\hat{\mathcal{G}}(\mathbf{x},\mathbf{x}_A,\omega)$, with

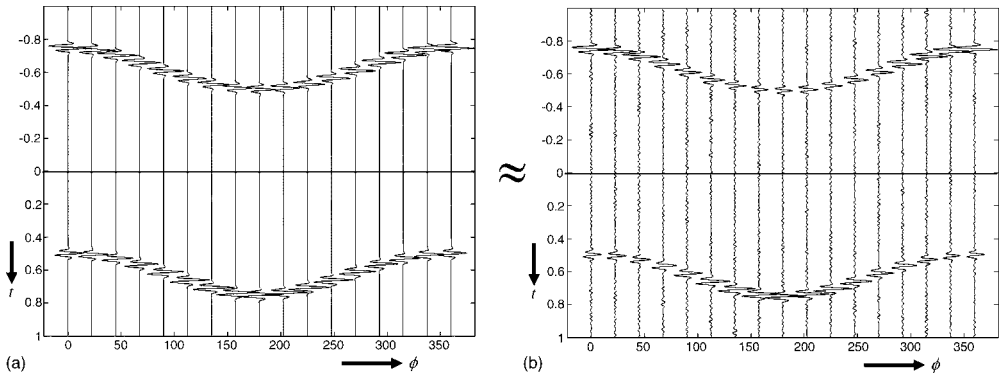


Fig. 2. (a) Left- and (b) right-hand side of Eq. (13) for different values of ϕ (the angle between the flow velocity and the line through $\mathbf{x}_A$ and $\mathbf{x}_B$, see Fig. 1).

$$\hat{\mathcal{G}} = -\frac{j\exp(j\omega \mathrm{M}(x_1 - x_{1,A})/c(1-\mathrm{M}^2))}{4\sqrt{1-\mathrm{M}^2}}\mathrm{H}_0^{(2)}\left(\frac{\omega R}{c(1-\mathrm{M}^2)}\right), \tag{14}$$

where $R=\sqrt{(x_1-x_{1,A})^2+(1-\mathrm{M}^2)(x_2-x_{2,A})^2}$ and $\mathrm{H}_0^{(2)}$ is the zeroth-order Hankel function of the second kind. Using this expression we model the response of 360 uncorrelated noise sources on a circle with a radius of 470 m (the noise is filtered around a central frequency of 30 Hz). We consider two receivers at $\mathbf{x}_A$ and $\mathbf{x}_B$, separated by a distance $d=210$ m, each registering 9600 s of noise. In the first instance the line through $\mathbf{x}_A$ and $\mathbf{x}_B$ is aligned with the flow velocity, hence ϕ in Fig. 1 equals zero. The cross correlation of the noise registrations, i.e., the right-hand side of Eq. (13), is represented by the first trace in Fig. 2(b) (at $\phi=0$). The numerical experiment is repeated for different angles ϕ between the flow velocity and the line through $\mathbf{x}_A$ and $\mathbf{x}_B$; the cross-correlation results are represented by the other traces in Fig. 2(b). The Green's functions convolved with $C(t)$ in the left-hand side of Eq. (13) are shown for the same range of angles ϕ in Fig. 2(a). For $\phi=0$ the traveltime of the causal and acausal Green's functions are given by $d/c(1+\mathrm{M})=210/420=0.5$ s and $-d/c(1-\mathrm{M})=-210/280=-0.75$ s, respectively. For $\phi=90°$ the travel times are $\pm d/c\sqrt{1-\mathrm{M}^2}=\pm 0.612$ s. Note that the travel times of the cross-correlation results accurately match those of the Green's functions for all angles ϕ. The amplitudes of the Green's functions are less accurately recovered by the cross-correlation procedure (see Fig. 3). The amplitude errors are explained as follows. The main contributions to the integrals come from those sources on the circle where the line through $\mathbf{x}_A$ and $\mathbf{x}_B$ intersects the circle[12,30] (see Fig. 1). Ignoring I_3 and I_4 with respect to I_1 and I_2 introduces a relative amplitude error in the order of $-v_k^0 n_k/c = w_1^0 n_1/c$, with $n_1=-\cos\phi$ (see Fig. 1). Evaluated as a function of ϕ we thus find for the relative amplitude error $w_1^0 n_1/c=-0.2\cos\phi$, which is approximately what we observe in Fig. 3.

4. Conclusion

We have shown that the nonreciprocal Green's function in a moving medium can be recovered from cross correlations of impulse responses [Eq. (12)] or noise measurements [Eq. (13)] at two receivers. The sources are assumed to be distributed along an arbitrary surface enclosing the two receivers. Unlike in the situation of a nonmoving medium, the cross-correlation result is asymmetric in time. The theory holds for a lossless arbitrary inhomogeneous medium with stationary inhomogeneous flow. The main underlying assumptions (in addition to those for a nonmoving medium) are that the spatial variations of the flow velocity are small in comparison with those of the wave field and that the flow velocity is small in comparison with the propagation velocity (small Mach number). The cross-correlation method accurately recovers the travel

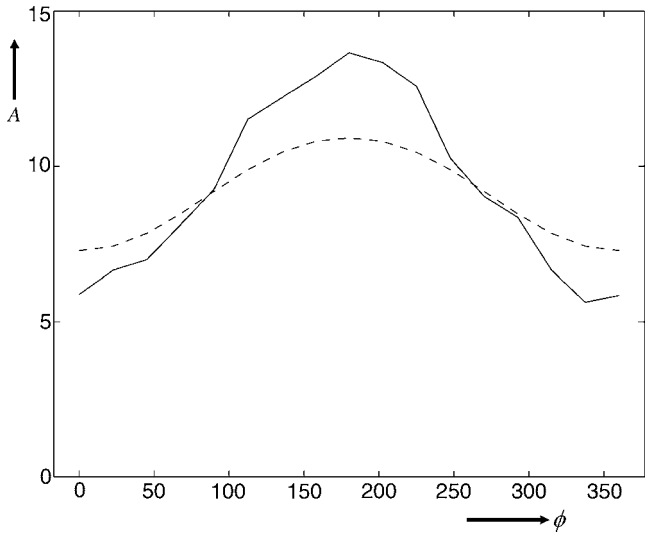


Fig. 3. Maximum amplitudes of the causal part of Figs. 2(a) (dashed) and 2(b) (solid).

times of the nonreciprocal Green's function. When the autocorrelation of the sources is known, the amplitudes are recovered with relative errors that are in the order of the Mach number. This error is negligible in comparison with the amplitude error that occurs when the sources are unknown. Hence, for practical situations (unknown source amplitudes, irregular source distribution, etc.), the accuracy of the retrieved nonreciprocal Green's function in a moving medium is of the same order as that of the retrieved reciprocal Green's function in a nonmoving medium.

References

[1]J. F. Claerbout, "Synthesis of a layered medium from its acoustic transmission response," Geophysics **33**, 264–269 (1968).
[2]T. L. Duvall, S. M. Jefferies, J. W. Harvey, and M. A. Pomerantz, "Time-distance helioseismology," Nature (London) **362**, 430–432 (1993).
[3]J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge **18**(8), 957–960 (1999).
[4]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301-1–134301-4 (2001).
[5]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[6]P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[7]N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[8]O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[9]R. L. Weaver and O. I. Lobkis, "On the emergence of the Green's function in the correlations of a diffuse field: pulse-echo using thermal phonons," Ultrasonics **40**, 435–439 (2002).
[10]B. A. van Tiggelen, "Green function retrieval and time reversal in a disordered world," Phys. Rev. Lett. **91**, 243904-1–243904-4 (2003).
[11]A. E. Malcolm, J. A. Scales, and B. A. van Tiggelen, "Extracting the Green function from diffuse, equipartitioned waves," Phys. Rev. E **70**, 015601(R)-1–015601(R)-4 (2004).
[12]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610-1–046610-8 (2004).
[13]K. Wapenaar, D. Draganov, J. Thorbecke, and J. Fokkema, "Theory of acoustic daylight imaging revisited," in 72nd Annual Meeting of the Society of Exploration Geophysicists (SEG), Salt Lake City, 2002, edited by Mike H. Powers (SEG, Tulsa, 2002), pp. 2269–2272.
[14]K. Wapenaar, "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301-1–254301-4 (2004).
[15]R. L. Weaver and O. I. Lobkis, "Diffuse fields in open systems and the emergence of the Green's function (L),"

J. Acoust. Soc. Am. **116**, 2731–2734 (2004).
[16]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 (L)," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[17]M. Fink, "Time reversed acoustics," Phys. Today **50**, 34–40 (1997).
[18]P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).
[19]E. Slob, D. Draganov, and K. Wapenaar, "GPR without a source," in Eleventh International Conference on Ground Penetrating Radar, Columbus, Ohio, 2006, edited by C.-C. Chen and J. Daniels (2006).
[20]D. R. Dowling, "Phase-conjugate array focusing in a moving medium," J. Acoust. Soc. Am. **94**, 1716–1718 (1993).
[21]P. Roux, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, and M. Stevenson, "A nonreciprocal implementation of time reversal in the ocean," J. Acoust. Soc. Am. **116**, 1009–1015 (2004).
[22]A. T. de Hoop, "Time-domain reciprocity theorems for acoustic wave fields in fluids with relaxation," J. Acoust. Soc. Am. **84**, 1877–1882 (1988).
[23]J. T. Fokkema and P. M. van den Berg, *Seismic Applications of Acoustic Reciprocity* (Elsevier, Amsterdam, 1993).
[24]N. N. Bojarski, "Generalized reaction principles and reciprocity theorems for the wave equations, and the relationship between the time-advanced and time-retarded fields," J. Acoust. Soc. Am. **74**, 281–285 (1983).
[25]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[26]K. Wapenaar and J. Fokkema, "Reciprocity theorems for diffusion, flow and waves," J. Appl. Mech. **71**, 145–150 (2004).
[27]L. M. Lyamshev, "On some integral relationships in acoustics of moving medium," Dokl. Akad. Nauk SSSR **138**, 575–578 (1961).
[28]L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II. Point Sources and Bounded Beams* (Springer, Berlin, 1992).
[29]B. P. Belinskiy, *On Some General Mathematical Properties of the System: Elastic Plate-Acoustic Medium* (World Scientific, Singapore 2001), pp. 193–218.
[30]K. Wapenaar, J. Fokkema, and R. Snieder, "Retrieving the Green's function in an open system by cross-correlation: a comparison of approaches (L)," J. Acoust. Soc. Am. **118**, 2783–2786 (2005).
[31]D.-J. van Manen, J. O. A. Robertsson, and A. Curtis, "Modeling of wave propagation in inhomogeneous media," Phys. Rev. Lett. **94**, 164301-1–164301-4 (2005).

ACOUSTICAL NEWS—USA

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

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

New Fellow of the Acoustical Society of America

Dajun Tang—For contributions to seabed scattering.

Regional Chapter News

North Texas Chapter

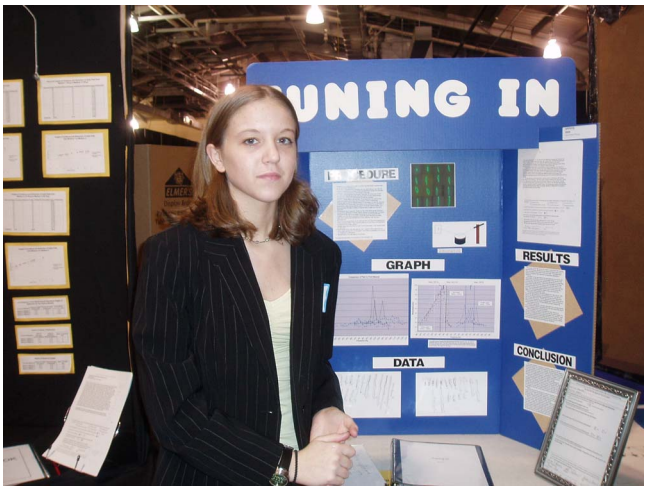
More than 800 students from schools in eight North Texas counties competed in the 49th Beal Bank Regional Science and Engineering Fair. This year two Senior and two Junior Division contestants won Outstanding Acoustics Project Awards.

Senior Division Winner Ms. Kelsey Smith, with the guidance of Ms. Cathy Bambanek, Clark High School, Plano ISD, investigated effects of drum head thickness and mass on resonant frequency. She built one-, two-, and three-layered heads (to approximate an ear drum). She recorded and compared movements at and near resonant frequencies by reflecting the light of a laser pointer from each drum head onto photographic film.

Senior Division Winner Mr. David Liu, with the guidance of Ms. Julie John, Williams High School, Plano ISD, meticulously evaluated the audio codecs of iPods and Windows-based players along four dimensions.

Junior Division Team Winners Mr. Vineet Gopal and Mr. Chase Hoskins, with the support of Ms. Rachel Robb, Haggard Middle School, Plano, put their fingers on the rims of restaurant stemware, added measures of water to construct a pitch scale, converted pitch to frequency, then brute-force-constructed an equation to correlate frequency and amount of water.

Peter Assmann, Laurie Bornstein, Greg Hughes, Ben Seep, and Michael Daly represented the Society.



Kelsey Smith, Senior Division winner.

Madras Regional Chapter

The Madras-India Regional Chapter of the Acoustical Society of America (MIRC-ASA) held four meetings in 2005, on 30 January, 22 June, 6 August, and 15–16 December, and a Science and Engineering Fair on Acoustics was held 17–18 February.

Three students, V. H. Priyadarsani (first place), M. Banurekha, and K. S. Mercyemilet, received student awards at the 30 January meeting.

Professor S. Swarnamani and Professor R. I. Sujith of IIT Madras delivered invited lectures at the 22 June meeting held at Tamil Nadu Science & Technology Center in Chennai.

At the 6 August meeting, six students received awards including Sanjith Gopalkrishnan (first place), Sowmya Murali (second place), M. Choundappan, Guruprasad S. Srivastava, M. Sridurga, and M. Kartick (see Fig. 1). Sanjith Gopalkrishnan was presented with the all-time best project and presentation award in the 11-year history of the Chapter (see Fig. 2). Awards were presented by H. S. Paul, MIRC Chapter Representative and Treasurer.

Dr. Ranjan Moodithaya of the National Aerospace Lab., Bangalore, and Professor U. S. P. Seth of IIT Madras delivered an invited lecture at the joint meeting of the Acoustical Society of India (ASI) and the MIRC-ASA at the 16 December meeting. The International Research Institute for the Deaf (IRID) presented a Silver Medal to Dr. V. Bhunjanga Rao, Vice President of ASI and Associate Director, NSTL, Visakhapatnam. Five research students were presented best paper awards during the joint meeting at Bangalore including Suryanarayana A. N. Prasad, T. Anathi, Nithya Subramanian, Abhijit Sarkar, and M. Sesagiri Rao.

The Science and Engineering Fair on Acoustics was held 17–18 February at the International Research Institute for the Deaf. The award ceremony was organized by Dr. M. Kumaresan, President of MIRC-ASA and Director of IRID, and held at the Quaid E. Millat Government College for Women in Chennai on 19 February. Dr. Kumaresan succeeded in attracting 44 projects for the fair, which is the largest number of participants at fairs conducted over the past 12 years. The award ceremony began with the delivery of a distinguished lecture by Professor B. V. A. Rao, VIT, Vellore, titled "Acoustics in India." Professor A. Harold Marshall, Marshall Day Acoustics, delivered the Mira Paul memorial distinguished lecture titled "Acoustical Design Process." The MIRC-ASA Gold Medal award was presented to Professor Marshall following his lecture (see Fig. 3). Professor

 0001-4966/2006/120(1)/1/3/$22.50

FIG. 1. Recipients of student awards at 6 August Chapter meeting: (Standing) M. Karthick, Sanjith Gopalkrishnan, G. S. Srivastava, M. Choundappan, M. Sridurga, and M. Sowmya (Seated) C. P. Vendhan (MIRC Secretary), H. S. Paul (MIRC Chapter Representative and Treasurer), Dhilsha Rajappan (MIRC member-at-Large), and A. Ramachandraiah (MIRC Vice President).

FIG. 2. Sanjith Gopalkrishnan (l) receives best project and presentation award from H. S. Paul (r).

FIG. 3. A. Harold Marshall receives Mira Paul memorial Gold Medal of IRID from H. S. Paul (r).

FIG. 4. (l-r) H. S. Paul (Treasurer & Chapter Representative), A. Ramachandraiah (Vice President), C. P. Vendhan (Secretary), A. H. Marshall, B. V. A. Rao (Member-at-Large), C. Anandam (Member-at-Large), and M. Kumaresan (President).

FIG. 5. Dipti Agrawal (m) receives first-rank (senior) award from H. S. Paul (l) and M. Kumaresan (r).

FIG. 6. H. Mercy (m) receives second-rank (senior) award from H. S. Paul (l) and M. Kumaresan (r).

FIG. 7. D. Sai Kiran (m) receives first-rank (junior) award from H. S. Paul (l) and M. Kumaresan (r).

Marshall, who donated his lecture honorarium to the IRID fund, met with the trustees of IRID after he received the Gold Medal (see Fig. 4).

Nine students were selected to receive MIRC-ASA awards including Dipti Agrawal (first place) (see Fig. 5), H. Mercy (second place) (see Fig. 6), M. V. Indumathi, A. Devarani, R. Mohana, K. Salma, D. S. Kiran (junior first place) (see Fig. 7), M. Karthick, and K. Anitha Rani. Three students received IRID awards, which are awards for the best students in each discipline of acoustics: V. Revathi, S. Nandakumari, and B. Sarina.

Hari S. Paul

USA Meetings Calendar

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

2006

17–21 Sept. INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA [www.interspeech2006.org ⟨http://www.interspeech2006.org/⟩]

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

2007

4–8 June 153rd Meeting of the Acoustical Society of America, Salt Lake City, UT [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

27 Nov.–2 Dec. 154th Meeting of the Acoustical Society of America, New Orleans, LA (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

2008

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

Cumulative Indexes to the Journal of the Acoustical Society of America

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

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

ACOUSTICAL STANDARDS NEWS

Susan B. Blaeser, Standards Manager
ASA Standards Secretariat

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

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes Catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please, contact Susan B. Blaeser.
Comments are welcomed on all material in Acoustical Standards News.
This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: http://asa.aip.org.

Standards Meetings Calendar—National

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

•***Tuesday, 28 November 2006***
ASACOS Steering Committee
•***Wednesday, 29 November 2006***
ASA Committee on Standards (ASACOS). Meeting of the Committee that directs the Standards Program of the Acoustical Society.

Accredited Standards Committee on Acoustics, S1

(J. P. Seiler, Chair; G. S. K. Wong, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance, and comfort.

S1 Working Groups

S1/Advisory—Advisory Planning Committee to S1 (G. S. K. Wong)
S1/WG1—Standard Microphones and their Calibration (V. Nedzelnitsky)
S1/WG4—Measurement of Sound Pressure Levels in Air (M. Nobile)
S1/WG5—Band Filter Sets (A. H. Marsh)
S1/WG17—Sound Level Meters and Integrating Sound Level Meters (B. M. Brooks)
S1/WG19—Insertion Loss of Windscreens (A. J. Campanella)
S1/WG20—Ground Impedance (K. Attenborough, Chair; J. Sabatier, Vice Chair)
S1/WG22—Bubble Detection and Cavitation Monitoring (Vacant)
S1/WG25—Specification for Acoustical Calibrators (P. Battenberg)
S1/WG26—High Frequency Calibration of the Pressure Sensitivity of Microphones (A. Zuckerwar)
S1/WG27—Acoustical Terminology (J. Vipperman)
S1/WG28—Passive Acoustic Monitoring for Marine Mammal Mitigation for Seismic Surveys (A. Thode)

S1 Inactive Working Groups

S1/WG16—FFT Acoustical Analyzers (R. J. Peppin, Chair)

S1 STANDARDS ON ACOUSTICS

ANSI S1.1-1994 (R2004) American National Standard Acoustical Terminology
ANSI S1.4-1983 (R2006) American National Standard Specification for Sound Level Meters
ANSI S1.4A-1985 (R2006) Amendment to ANSI S1.4-1983
ANSI S1.6-1984 (R2006) American National Standard Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements
ANSI S1.8-1989 (R2006) American National Standard Reference Quantities for Acoustical Levels
ANSI S1.9-1996 (R2006) American National Standard Instruments for the Measurement of Sound Intensity
ANSI S1.11-2004 American National Standard Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters
ANSI S1.13-2005 American National Standard Measurement of Sound Pressure Levels in Air
ANSI S1.14-1998 (R2003) American National Standard Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields, 25 MHz to 1 GHz
ANSI S1.15-1997/Part 1 (R2006) American National Standard Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones
ANSI S1.15-2005/Part 2 American National Standard Measurement Microphones, Part 2: Primary Method for Pressure Calibration of Laboratory Standard Microphones by the Reciprocity Technique
ANSI S1.16-2000 (R2005) American National Standard Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones
ANSI S1.17-2004/Part 1 American National Standard Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air
ANSI S1.18-1999 (R2004) American National Standard Template Method for Ground Impedance
ANSI S1.20-1988 (R2003) American National Standard Procedures for Calibration of Underwater Electroacoustic Transducers
ANSI S1.22-1992 (R2002) American National Standard Scales and Sizes for Frequency Characteristics and Polar Diagrams in Acoustics
ANSI S1.24 TR-2002 ANSI Technical Report Bubble Detection and Cavitation Monitoring
ANSI S1.25-1991 (R2002) American National Standard Specification for Personal Noise Dosimeters
ANSI S1.26-1995 (R2004) American National Standard Method for Calculation of the Absorption of Sound by the Atmosphere
ANSI S1.40-1984 (R2001) American National Standard Specification for Acoustical Calibrators
ANSI S1.42-2001 (R2006) American National Standard Design Response of Weighting Networks for Acoustical Measurements
ANSI S1.43-1997 (R2002) American National Standard Specifications for Integrating-Averaging Sound Level Meters

 0001-4966/2006/120(1)/5/9/$22.50

Accredited Standards Committee on Mechanical Vibration and Shock, S2

(R. J. Peppin, Chair; D. J. Evans, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

S2 Working Groups

S2/WG1—S2 Advisory Planning Committee (D. J. Evans)
S2/WG2—Terminology and Nomenclature in the Field of Mechanical Vibration and Shock and Condition Monitoring and Diagnostics of Machines (D. J. Evans)
S2/WG3—Signal Processing Methods (T. S. Edwards)
S2/WG4—Characterization of the Dynamic Mechanical Properties of Viscoelastic Polymers (W. M. Madigosky, Chair; J. Niemic, Vice Chair)
S2/WG5—Use and Calibration of Vibration and Shock Measuring Instruments (D. J. Evans, Chair; B. E. Douglas, Vice Chair)
S2/WG6—Vibration and Shock Actuators (G. Booth)
S2/WG7—Acquisition of Mechanical Vibration and Shock Measurement Data (B. E. Douglas)
S2/WG8—Analysis Methods of Structural Dynamics (B. E. Douglas)
S2/WG9—Training and Accreditation (R. Eshleman, Chair)
S2/WG10—Measurement and Evaluation of Machinery for Acceptance and Condition (R. Eshleman, Chair; H. Pusey, Vice Chair)
S2/WG10/Panel 1—Balancing (R. Eshleman)
S2/WG10/Panel 2—Operational Monitoring and Condition Evaluation (R. Bankert)
S2/WG10/Panel 3—Machinery Testing (R. Eshleman)
S2/WG10/Panel 4—Prognosis (R. Eshleman)
S2/WG10/Panel 5—Data Processing, Communication, and Presentation (K. Bever)
S2/WG11—Measurement and Evaluation of Mechanical Vibration of Vehicles (A. F. Kilcullen)
S2/WG12—Measurement and Evaluation of Structures and Structural Systems for Assessment and Condition Monitoring (B. E. Douglas, Chair)
S2/WG13—Shock Test Requirements for Commercial Electronic Systems (P. D. Loeffler)
S2/WG39 (S3)—Human Exposure to Mechanical Vibration and Shock—Parallel to ISO/TC 108/SC 4 (D. D. Reynolds, Chair; R. Dong, Vice Chair)

S2 Inactive Working Group

S2/WG54—Atmospheric Blast Effects (J. W. Reed)

S2 STANDARDS ON MECHANICAL VIBRATION AND SHOCK

ANSI S2.1-2000 ISO 2041:1990 Nationally Adopted International Standard Vibration and Shock—Vocabulary
ANSI S2.2-1959 (R2006) American National Standard Methods for the Calibration of Shock and Vibration Pickups
ANSI S2.4-1976 (R2004) American National Standard Method for Specifying the Characteristics of Auxiliary Analog Equipment for Shock and Vibration Measurements
ANSI S2.7-1982 (R2004) American National Standard Balancing Terminology
ANSI S2.8-1972 (R2006) American National Standard Guide for Describing the Characteristics of Resilient Mountings
ANSI S2.9-1976 (R2006) American National Standard Nomenclature for Specifying Damping Properties of Materials
ANSI S2.16-1997 (R2006) American National Standard Vibratory Noise Measurements and Acceptance Criteria of Shipboard Equipment
ANSI S2.17-1980 (R2004) American National Standard Techniques of Machinery Vibration Measurement
ANSI S2.19-1999 (R2004) American National Standard Mechanical Vibration—Balance Quality Requirements of Rigid Rotors, Part 1: Determination of Permissible Residual Unbalance, Including Marine Applications
ANSI S2.20-1983 (R2006) American National Standard Estimating Airblast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects
ANSI S2.21-1998 (R2002) American National Standard Method for Preparation of a Standard Material for Dynamic Mechanical Measurements
ANSI S2.22-1998 (R2002) American National Standard Resonance Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials
ANSI S2.23-1998 (R2002) American National Standard Single Cantilever Beam Method for Measuring the Dynamic Mechanical Properties of Viscoelastic Materials
ANSI S2.24-2001 (R2006) American National Standard Graphical Presentation of the Complex Modulus of Viscoelastic Materials
ANSI S2.25-2004 American National Standard Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships
ANSI S2.26-2001 (R2006) American National Standard Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment
ANSI S2.27-2002 American National Standard Guidelines for the Measurement and Evaluation of Vibration of Ship Propulsion Machinery
ANSI S2.28-2003 American National Standard Guidelines for the Measurement and Evaluation of Vibration of Shipboard Machinery
ANSI S2.29-2003 American National Standard Guidelines for the Measurement and Evaluation of Vibration of Marine Shafts on Shipboard Machinery
ANSI S2.31-1979 (R2004) American National Standard Method for the Experimental Determination of Mechanical Mobility, Part 1: Basic Definitions and Transducers
ANSI S2.32-1982 (R2004) American National Standard Methods for the Experimental Determination of Mechanical Mobility, Part 2: Measurements Using Single-Point Translational Excitation
ANSI S2.34-1984 (R2005) American National Standard Guide to the Experimental Determination of Rotational Mobility Properties and the Complete Mobility Matrix
ANSI S2.42-1982 (R2004) American National Standard Procedures for Balancing Flexible Rotors
ANSI S2.43-1984 (R2005) American National Standard Criteria for Evaluating Flexible Rotor Balance
ANSI S2.46-1989 (R2005) American National Standard Characteristics to be Specified for Seismic Transducers
ANSI S2.48-1993 (R2006) American National Standard Servo-Hydraulic Test Equipment for Generating Vibration—Methods of Describing Characteristics
ANSI S2.60-1987 (R2005) American National Standard Balancing Machines—Enclosures and Other Safety Measures
ANSI S2.61-1989 (R2005) American National Standard Guide to the Mechanical Mounting of Accelerometers

Standards on Human Exposure to Vibration

ANSI S3.18-2002/ISO 2631-1:1997 Nationally Adopted International Standard Mechanical vibration and shock—Evaluation of human exposure to whole-body vibration—Part 1: General requirements
ANSI S3.18-2003/ISO 2631-4: 2001 Nationally Adopted International Standard Mechanical vibration and shock—Evaluation of human exposure to whole body vibration—Part 4: Guidelines for the evaluation of the effects of vibration and rotational motion on passenger and crew comfort in fixed-guideway transport systems
ANSI S2.70-2006 American National Standard Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand
ANSI S2.71-1983(R2006) American National Standard Guide to the Evaluation of Human Exposure to Vibration in Buildings
ANSI S3.40-2002 ISO 10819:1996 Nationally Adopted International Standard Mechanical vibration and shock—Hand-arm vibration—Method for the measurement and evaluation of the vibration transmissibility of gloves at the palm of the hand

Accredited Standards Committee on Bioacoustics, S3

(R. F. Burkard, Chair; C. A. Champlin, Vice Chair)

Scope: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance, and comfort.

S3 Working Groups

S3/Advisory—Advisory Planning Committee to S3 (R. F. Burkard)
S3/WG35—Audiometers (R. L. Grason)
S3/WG36—Speech Intelligibility (R. S. Schlauch)
S3/WG37—Coupler Calibration of Earphones (B. Kruger)
S3/WG39—Human Exposure to Mechanical Vibration and Shock (D. D. Reynolds)
S3/WG43—Method for Calibration of Bone Conduction Vibrators (J. Durrant)
S3/WG48—Hearing Aids (D. A. Preves)
S3/WG51—Auditory Magnitudes (R. P. Hellman)
S3/WG56—Criteria for Background Noise for Audiometric Testing (J. Franks)
S3/WG59—Measurement of Speech Levels (M. C. Killion and L. A. Wilber, Co-Chairs)
S3/WG60—Measurement of Acoustic Impedance and Admittance of the Ear (Vacant)
S3/WG62—Impulse Noise with Respect to Hearing Hazard (J. H. Patterson)
S3/WG67—Manikins (M. D. Burkhard)
S3/WG72—Measurement of Auditory Evoked Potentials (R. F. Burkard)
S3/WG76—Computerized Audiometry (A. J. Miltich)
S3/WG78—Thresholds (W. A. Yost)
S3/WG79—Methods for Calculation of the Speech Intelligibility Index (C. V. Pavlovic)
S3/WG81—Hearing Assistance Technologies (L. Thibodeau and L. A. Wilber, Co-Chairs)
S3/WG82—Basic Vestibular Function Test Battery (C. Wall, III)
S3/WG83—Sound Field Audiometry (T. R. Letowski)
S3/WG84—Otoacoustic Emission (G. R. Long)
S3/WG86—Audiometric Data Structures (W. A. Cole and B. Kruger, Co-Chairs)
S3/WG87—Human Response to Repetitive Mechanical Shock (N. Alem)
S3/WG88—Standard Audible Emergency Evacuation and Other Signals (I. Mande)
S3/WG89—Spatial Audiometry in Real and Virtual Environments (J. Besing)
S3/WG90—Animal Bioacoustics (A. E. Bowles)
S3/WG91—Text-to-Speech Synthesis Systems (A. K. Syrdal and C. Bickley, Co-Chairs)
S3/WG92—Effects of Sound on Fish and Turtles (R. R. Fay and A. N. Popper, Co-Chairs)
S2/WG39 (S3)—Human Exposure to Mechanical Vibration and Shock—Parallel to ISO/TC 108/SC 4 (D. D. Reynolds)

S3 Liaison Group

S3/L-1 S3 U. S. TAG Liaison to IEC/TC 87 Ultrasonics (W. L. Nyborg)

S3 Inactive Working Groups

S3/WG71—Artificial Mouths (R. L. McKinley)
S3/WG80—Probe-Tube Measurements of Hearing Aid Performance (W. A. Cole)
S3/WG58—Hearing Conservation Criteria

S3 STANDARDS ON BIOACOUSTICS

ANSI S3.1-1999 (R2003) American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms
ANSI S3.2-1989 (R1999) American National Standard Method for Measuring the Intelligibility of Speech over Communication Systems
ANSI S3.4-2005 American National Standard Procedure for the Computation of Loudness of Steady Sound
ANSI S3.5-1997 (R2002) American National Standard Methods for Calculation of the Speech Intelligibility Index
ANSI S3.6 2004 American National Standard Specification for Audiometers
ANSI S3.7-1995 (R2003) American National Standard Method for Coupler Calibration of Earphones
ANSI S3.13-1987 (R2002) American National Standard Mechanical Coupler for Measurement of Bone Vibrators
ANSI S3.20-1995 (R2003) American National Standard Bioacoustical Terminology
ANSI S3.21-2004 American National Standard Methods for Manual Pure-Tone Threshold Audiometry
ANSI S3.22-2003 American National Standards Specification of Hearing Aid Characteristics
ANSI S3.25-1989 (R2003) American National Standard for an Occluded Ear Simulator
ANSI S3.35-2004 American National Standard Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-Ear Working Conditions
ANSI S3.36-1985 (R2006) American National Standard Specification for a Manikin for Simulated *in situ* Airborne Acoustic Measurements
ANSI S3.37-1987 (R2002) American National Standard Preferred Earhook Nozzle Thread for Postauricular Hearing Aids
ANSI S3.39-1987 (R2002) American National Standard Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance)
ANSI S3.41-1990 (R2001) American National Standard Audible Emergency Evacuation Signal
ANSI S3.42-1992 (R2002) American National Standard Testing Hearing Aids with a Broad-Band Noise Signal
ANSI S3.44-1996 (R2006) American National Standard Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment
ANSI S3.45-1999 American National Standard Procedure for Testing Basic Vestibular Function
ANSI S3.46-1997 (R2002) American National Standard Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids

Accredited Standards Committee on Noise, S12

(R. D. Hellweg, Chair; W. J. Murphy, Vice Chair)

Scope: Standards, specifications, and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation, and control; including biological safety, tolerance and comfort, and physical acoustics as related to environmental and occupational noise.

S12 Working Groups

S12/Advisory—Advisory Planning Committee to S12 (R. D. Hellweg)
S12/WG3—Measurement of Noise from Information Technology and Telecommunications Equipment (K. X. C. Man)
S12/WG11—Hearing Protector Attenuation and Performance (E. H. Berger)
S12/WG12—Evaluation of Hearing Conservation Programs (J. D. Royster, Chair; E. H. Berger, Vice Chair)
S12/WG13—Method for the Selection of Hearing Protectors that Optimize the Ability to Communicate (D. Byrne)
12/WG14—Measurement of the Noise Attenuation of Active and/or Passive Level Dependent Hearing Protective Devices (J. Kalb, Chair; W. J. Murphy, Vice Chair)
S12/WG15—Measurement and Evaluation of Outdoor Community Noise (P. D. Schomer)
S12/WG18—Criteria for Room Noise (R. J. Peppin)
S12/WG23—Determination of Sound Power (R. J. Peppin and B. M. Brooks, Co-Chairs)
S12/WG31—Predicting Sound Pressure Levels Outdoors (R. J. Peppin)
S12/WG32—Revision of ANSI S12.7-1986 Methods for Measurement of Impulse Noise (A. H. Marsh)
S12/WG33—Revision of ANSI S5.1-1971 Test Code for the Measurement of Sound from Pneumatic Equipment (B. M. Brooks)

S12/WG36—Development of Methods for Using Sound Quality (G. L. Ebbit and P. Davies, Co-Chairs)

S12/WG38—Noise Labelling in Products (R. D. Hellweg and J. Pope, Co-Chairs)

S12/WG40—Measurement of the Noise Aboard Ships (S. Antonides, Chair; S. Fisher, Vice Chair)

S12/WG41—Model Community Noise Ordinances (L. Finegold, Chair; B. M. Brooks, Vice Chair)

S12/WG43—Rating Noise with Respect to Speech Interference (M. Alexander)

S12/WG44—Speech Privacy (G. C. Tocci, Chair; D. Sykes, Vice Chair)

S12 Liaison Groups

S12/L-1—IEEE 85 Committee for TAG Liaison—Noise Emitted by Rotating Electrical Machines (Parallel to ISO/TC 43/SC 1/WG 13) (R. G. Bartheld)

S12/L-2—Measurement of Noise from Pneumatic Compressors Tools and Machines (Parallel to ISO/TC 43/SC 1/WG 9) (Vacant)

S12/L-3—SAE Committee for TAG Liaison on Measurement and Evaluation of Motor Vehicle Noise (Parallel to ISO/TC 43/SC 1/WG 8) (R. F. Schumacher and J. Johnson)

S12/L-4—SAE Committee A-21 for TAG Liaison on Measurement and Evaluation of Aircraft Noise (J. Brooks)

S12/L-5—ASTM E-33 on Environmental Acoustics (to include activities of ASTM E33.06 on Building Acoustics, parallel to ISO/TC 43/SC 2 and ASTM E33.09 on Community Noise) (K. P. Roy)

S12/L-6—SAE Construction-Agricultural Sound Level Committee (I. Douell)

S12/L-7—SAE Specialized Vehicle and Equipment Sound Level Committee (T. Disch)

S12/L-8—ASTM PTC 36 Measurement of Industrial Sound (R. A. Putnam, Chair; B. M. Brooks, Vice Chair)

S12 Inactive Working Groups

S12/WG27—Outdoor Measurement of Sound Pressure Level (G. Daigle)

S12/WG8—Determination of Interference of Noise with Speech Intelligibility (L. Marshall)

S12/WG9—Annoyance Response to Impulsive Noise (L. C. Sutherland)

S12/WG19—Measurement of Occupational Noise Exposure (J. P. Barry and R. Goodwin, Co-Chairs)

S12/WG29—Field Measurement of the Sound Output of Audible Public-Warning Devices (Sirens) (P. Graham)

S12/WG 34—Methodology for Implementing a Hearing Conservation Program (J. P. Barry)

S12/WG37—Measuring Sleep Disturbance Due to Noise (K. S. Pearsons)

S12 STANDARDS ON NOISE

ANSI S12.1-1983 (R2006) American National Standard Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources

ANSI S12.2-1995 (R1999) American National Standard Criteria for Evaluating Room Noise

ANSI S12.3-1985 (R2006) American National Standard Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment

ANSI S12.5-1990 (R1997) American National Standard Requirements for the Performance and Calibration of Reference Sound Sources

ANSI S12.6-1997 (R2002) American National Standard Methods for Measuring the Real-Ear Attenuation of Hearing Protectors

ANSI S12.7-1986 (R2006) American National Standard Methods for Measurements of Impulse Noise

ANSI S12.8-1998 (R2003) American National Standard Methods for Determining the Insertion Loss of Outdoor Noise Barriers

ANSI S12.9-1988/Part 1 (R2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 1

ANSI S12.9-1992/Part 2 (R2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 2: Measurement of Long-Term, Wide-Area Sound

ANSI S12.9-1993/Part 3 (R2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 3: Short-Term Measurements with an Observer Present

ANSI S12.9-2005/Part 4 American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 4: Noise Assessment and Prediction of Long-Term Community Response

ANSI S12.9-1998/Part 5 (R2003) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 5: Sound Level Descriptors for Determination of Compatible Land Use

ANSI S12.9-2000/Part 6 (R2005) American National Standard Quantities and Procedures for Description and Measurement of Environmental Sound, Part 6: Methods for Estimation of Awakenings Associated with Aircraft Noise Events Heard in Homes

ANSI S12.10-2002/ISO 7779:1999 Nationally Adopted International Standard Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment

ANSI S12.11-2003/Part 1/ISO 10302: 1996 (MOD) American National Standard Acoustics—Measurement of noise and vibration of small air-moving devices—Part 1: Airborne noise emission

ANSI S12.11-2003/Part 2 American National Standard Acoustics—Measurement of Noise and Vibration of Small Air-moving Devices—Part 2: Structure-Borne Vibration

ANSI S12.12-1992 (R2002) American National Standard Engineering Method for the Determination of Sound Power Levels of Noise Sources Using Sound Intensity

ANSI S12.13 TR-2002 ANSI Technical Report Evaluating the Effectiveness of Hearing Conservation Programs through Audiometric Data Base Analysis

ANSI S12.14-1992 (R2002) American National Standard Methods for the Field Measurement of the Sound Output of Audible Public Warning Devices Installed at Fixed Locations Outdoors

ANSI S12.15-1992 (R2002) American National Standard For Acoustics *B* Portable Electric Power Tools, Stationary and Fixed Electric Power Tools, and Gardening Appliances—Measurement of Sound Emitted

ANSI S12.16-1992 (R2002) American National Standard Guidelines for the Specification of Noise of New Machinery

ANSI S12.17-1996 (R2006) American National Standard Impulse Sound Propagation for Environmental Noise Assessment

ANSI S12.18-1994 (R2004) American National Standard Procedures for Outdoor Measurement of Sound Pressure Level

ANSI S12.19-1996 (R2006) American National Standard Measurement of Occupational Noise Exposure

ANSI S12.23-1989 (R2006) American National Standard Method for the Designation of Sound Power Emitted by Machinery and Equipment

ANSI S12.30-1990 (R2002) American National Standard Guidelines for the Use of Sound Power Standards and for the Preparation of Noise Test Codes

ANSI S12.42-1995 (R2004) American National Standard Microphone-in-Real-Ear and Acoustic Test Fixture Methods for the Measurement of Insertion Loss of Circumaural Hearing Protection Devices

ANSI S12.43-1997 (R2002) American National Standard Methods for Measurement of Sound Emitted by Machinery and Equipment at Workstations and other Specified Positions

ANSI S12.44-1997 (R2002) American National Standard Methods for Calculation of Sound Emitted by Machinery and Equipment at Workstations and other Specified Positions from Sound Power Level

ANSI S12.50-2002/ISO 3740:2000 Nationally Adopted International Standard Acoustics—Determination of sound power levels of noise sources—Guidelines for the use of basic standards

ANSI S12.51-2002/ISO 3741:1999 Nationally Adopted International Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision method for reverberation rooms

ANSI S12.53/1-1999 (R2004)/ISO 3743-1:1994 Nationally Adopted International Standard Acoustics—Determination of sound power levels of noise sources—Engineering methods for small, movable sources in reverberant fields—Part 1: Comparison method for hard-walled test rooms

ANSI S12.53/2-1999 (R2004)/ISO 3743-2:1994 Nationally Adopted International Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, mov-

able sources in reverberant fields—Part 2: Methods for special reverberation test rooms

ANSI S12.54-1999 (R2004)/ISO 3744:1994 Nationally Adopted International Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering method in an essentially free field over a reflecting plane

ANSI S12.56-1999 (R2004)/ISO 3746:1995 Nationally Adopted International Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Survey method using an enveloping measurement surface over a reflecting plane

ANSI S12.57-2002/ISO 3747:2000 Nationally Adopted International Standard Acoustics—Determination of sound power levels of noise sources using sound pressure—Comparison method *in situ*

ANSI S12.60-2002 American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools

ANSI S12.65-2006 American National Standard for Rating Noise with Respect to Speech Interference

ASA Committee on Standards (ASACOS)

ASACOS (P. D. Schomer, Chair and ASA Standards Director)

U. S. Technical Advisory Groups (TAGS) for International Standards Committees

ISO/TC 43 Acoustics, **ISO/TC 43 /SC 1** Noise (P. D. Schomer, U.S. TAG Chair)

ISO/TC 108 Mechanical Vibration and Shock (D. J. Evans, U.S. TAG Chair)

ISO/TC 108/SC2 Measurement and Evaluation of Mechanical Vibration and Shock as Applied to Machines, Vehicles and Structures (A. F. Kilcullen and R. F. Taddeo U.S. TAG Co-Chairs)

ISO/TC 108/SC3 Use and Calibration of Vibration and Shock Measuring Instruments (D. J. Evans, U.S. TAG Chair)

ISO/TC 108/SC4 Human Exposure to Mechanical Vibration and Shock (D. D. Reynolds, U.S. TAG Chair)

ISO/TC 108/SC5 Condition Monitoring and Diagnostic Machines (D. J. Vendittis, U.S. TAG Chair)

ISO/TC 108/SC6 Vibration and Shock Generating Systems (G. Booth, U.S. TAG Chair)

IEC/TC 29 Electroacoustics (V. Nedzelnitsky, U.S. Technical Advisor)

Standards News from the United States

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

American National Standards Call for Comment on Proposals Listed

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

ASA (ASC S2) (Acoustical Society of America)

(Comment deadline: 8 May 2006)

REAFFIRMATIONS

BSR S2.16-1997 (R200x), Vibratory Noise Measurements and Acceptance Requirements for Shipboard Equipment [Reaffirmation of ANSI S2.16-1997 (R2001)]

This standard contains guidelines for limiting the machinery and operating equipment vibration on board ships for the purposes of habitability and mechanical suitability. The mechanical guidelines result in a suitable environment for installed equipment and preclude many major vibration problems such as unbalance, misalignment, or other damage to the machinery and operating equipment.

BSR S2.26-2001 (R200x), Vibration Testing Requirements and Acceptance Criteria for Shipboard Equipment (Reaffirmation of ANSI S2.26-2001)

This standard describes procedures for vibration testing of shipboard equipment, specifying amplitude, frequency, and endurance requirements

ASA (ASC S2) (Acoustical Society of America)

(Comment deadline: 15 May 2006)

REVISIONS

BSR S2.70-200x, Guide for the Measurement and Evaluation of Human Exposure to Vibration Transmitted to the Hand [Revision of ANSI S3.34-1986 (R1997)]

Specifies recommended method for measurement, data analysis, vibration and health risk assessments, and reporting of human exposure to hand-transmitted vibration. Specifies format for measurement, data analysis, vibration and health risk assessments, and reporting of hand-transmitted vibration, periodic or random, in three orthogonal axes, in the frequency range from 5.6 to 1400 Hz. Three normative annexes address risk assessments, mitigation, training, and medical surveillance.

REAFFIRMATIONS

BSR S2.48-1993 (R200x), Servo-Hydraulic Test Equipment for Generating Vibration—Methods of Describing Characteristics [Reaffirmation of ANSI S2.48-1993 (R2001)]

Provides method for specifying the characteristics of servo-hydraulic test equipment for generating vibration and serves as a guide to the selection of such equipment. It applies to servo-hydraulic vibration generators and power amplifiers, individually and in combination. Provides means to assist a prospective user to calculate and compare the performance of equipment provided by two or more manufacturers, even if the vibration generator and the power amplifier are from different manufacturers.

WITHDRAWALS

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

This standard provides guidelines for the measurement of building vibrations and evaluation of their effects on buildings. It is intended to establish the basic principles for carrying out vibration measurements and processing data, with regard to evaluating vibration effects on buildings. The evaluation of the effects of building vibration is primarily directed at structural response, and includes appropriate analytical methods where the frequency, duration, and amplitude can be defined.

ASA (ASC S3) (Acoustical Society of America)

(Comment deadline: 24 April 2006)

REAFFIRMATIONS

BSR S3.36-1985 (R200x), Specification for a Manikin for Simulated in-situ Airborne Acoustic Measurements [Reaffirmation of ANSI S3.36-1985 (R2001)]

This standard describes a manikin for airborne acoustic measurements. It comprises a head with external ears and ear canals, and a torso that simulates a median human adult. It is intended primarily as an instrument for measuring the acoustic gain of hearing aids under simulated *in situ* conditions. Both geometric and acoustical response descriptions are given.

BSR S3.44-1996 (R200x), Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment [Reaffirmation of ANSI S3.44-1996 (R2001)]

This standard presents the statistical relationship between noise exposures and changes in hearing threshold levels for a noise exposed population, and can be applied to the calculation of the risk of incurring hearing handicap from sustained daily exposure to noise. It provides guidance to the measurement of noise exposure. Unlike its ISO counterpart, it allows assessment of noise exposure using a time/intensity trading relation other than a 3-dB increase per halving of exposure time.

ASA (ASC S12) (Acoustical Society of America)

(Comment deadline: 24 April 2006)

REAFFIRMATIONS

BSR S12.1-1983 (R200x), Guidelines for the Preparation of Standard Procedures to Determine the Noise Emission from Sources [Reaffirmation of ANSI S12.1-1983 (R2001)]

Standard contains guidelines for preparation of procedures (standards, test codes, recommended practices, etc.) for determination of noise emission from sources. Included are general questions that need to be considered during development of a measurement procedure. Guidelines on the following subjects are included: prefactory material, measurement conditions, measurement operations, data reduction, preparation of a test report, and guidelines for selection of a descriptor for noise emission.

BSR S12.3-1985 (R200x), Statistical Methods for Determining and Verifying Stated Noise Emission Values of Machinery and Equipment [Reaffirmation of ANSI S12.3-1985 (R2001)]

This standard defines the preferred methods for determining and verifying noise emission values for machinery and equipment that are stated in product literature or labeled by other means.

BSR S12.17-1996 (R200x), Impulse Sound Propagation for Environmental Noise Assessment [Reaffirmation of ANSI S12.17-1996 (R2001)]

Describes engineering methods to calculate propagation of high-energy impulsive sounds through the atmosphere for purposes of assessment of environmental noise. The methods yield estimates for the mean C-weighted sound exposure level of impulsive sound at distances between source and receiver ranging from 1 to 30 km. Equations to estimate standard deviation about the mean C-weighted sound exposure levels are provided. The methods apply for explosive masses between 50 g and 1000 kg.

BSR S12.19-1996 (R200x), Measurement of Occupational Noise Exposure [Reaffirmation of ANSI S12.19-1996 (R2001)]

The standard presents methods that can be used to measure a person's noise exposure received in a workplace. The methods have been developed to provide uniform procedures and repeatable results for the measurement of occupational noise exposure.

BSR S12.23-1989 (R200x), Sound Power Emitted by Machinery and Equipment [Reaffirmation of ANSI S12.23-1989 (R2001)]

Standard describes a method for expressing the noise emission of machinery and equipment in a convenient manner. Standard applies to all machinery and equipment that is essentially stationary in nature and for which overall A-weighted sound power is a meaningful descriptor of noise emission. Standard is intended to facilitate preparation of equipment specifications, labels, or other documentation that expresses in quantitative terms the noise emission of machinery or equipment.

CEA (Consumer Electronics Association)

(Comment deadline: 5 June 2006)

NEW STANDARDS

BSR/CEA 2010-200x, Standard Method of Measurement for Powered Subwoofers (New standard)

This standard defines a method for measuring the audio performance of powered subwoofer

BSR/CEA 2019-200x, Testing and Measurement Methods for Audio Amplifiers (New standard)

This standard defines a method for measuring and reporting the output power of home and/or professional audio amplifiers, including home theater systems, products with integrated audio amplifiers such as radio receivers, TV sets, and computers, and stand-alone audio amplifiers.

Projects Withdrawn from Consideration

An accredited standards developer may abandon the processing of a proposed new or revised American National Standard or portion thereof if it has followed its accredited procedures. The following projects have been withdrawn accordingly:

CEA (Consumer Electronics Association)

BSR/CEA 490-B-200x, Standard Test Methods of Measurement for Audio Amplifiers (New standard)

American National Standards

30 Day Notice of Withdrawal: ANS 5 to 10 years past approval date

In accordance with clause 4.7.1 Periodic Maintenance of American National Standards of the ANSI Essential Requirements, the following American National Standards have not been reaffirmed or revised within the five-year period following approval as an ANS. Thus, they shall be withdrawn at the close of this 30-day public review notice in Standards Action.

ANSI/ASTM E336-1997, Test Method for Measurement of Airborne Sound Insulation In Buildings

ANSI/ASTM E596-1997, Test Method for Laboratory Measurement of the Noise Reduction of Sound-Isolating Enclosures

ANSI/ASTM E1007-1997, Test Method for Field Measurement of Tapping Machine Impact Sound Transmission through Floor-Ceiling Assemblies and Associated Support Structures

ANSI/ASTM E1042-1997, Classification for Acoustically Absorptive Materials Applied by Trowel or Spray

ANSI/ASTM E1050-1997, Test Method for Impedance and Absorption of Acoustical Materials Using a Tube, Two Microphones and a Digital Frequency Analysis System

ANSI/ASTM E1065-1999, Guide for Evaluating Characteristics of Ultrasonic Search Units

ANSI/ASTM E1124-1997, Test Method for Field Measurement of Sound Power Level by the Two-Surface Method

ANSI/ASTM E1179-1997, Specification for Sound Sources Used for Testing Open Office Components and Systems

ANSI/ASTM E1222-1997, Test Method for Laboratory Measurement of the Insertion Loss of Pipe Lagging Systems

ANSI/ASTM E1265-1997, Test Method for Measuring Insertion Loss of Pneumatic Exhaust Silencers

ANSI/ASTM E1289-1997, Specification for Reference Specimen for Sound Transmission Loss

ANSI/ASTM E1374-1997, Guide for Open Office Acoustics and Applicable ASTM Standards

ANSI/ASTM E1433-1997, Guide for Selection of Standards on Environmental Acoustics

ANSI/ASTM E1503-1997, Test Method for Conducting Outdoor Sound Measurements Using a Digital Statistical Analysis System

ANSI/ASTM E1573-1997, Test Method for Evaluating Masking Sound in Open Offices Using A-Weighted and One-Third Octave Band Sound Pressure Levels

ANSI/ASTM E1617-1997, Practice for Reporting Particle Size Characterization Data

ANSI/ASTM E1620-1997, Terminology Relating to Liquid Particles and Atomization

ANSI/ASTM E1686-1997, Guide for Selection of Environmental Noise Measurements and Criteria

ANSI/ASTM E1704-1997, Guide for Specifying Acoustical Performance of Sound-Isolating Enclosures

ANSI/ASTM E1779-1997, Guide for Preparing a Measurement Plan for Conducting Outdoor Sound Measurements

ANSI/ASTM E1780-1997, Guide for Measuring Outdoor Sound Received from a Nearby Fixed Source

Project Initiation Notification System (PINS)

ANSI procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards.

The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic main-

tenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

IPC (IPC—Association Connecting Electronics Industries)

BSR/IPC J-STD-001DS-200x, Space Applications Electronic Hardware Addendum to IPC J-STD-001D (Supplement to ANSI/IPC J-STD-001D-2005)

This addendum provides additional requirements over those published in J-STD-001D to ensure the reliability of soldered electrical and electronic assemblies that must survive the vibration and thermal cyclic environments getting to and operating in space. Where content criteria are not supplemented, the Class 3 requirements of IPC J-STD-001D apply.

BSR/IPC WHMA-A-620AS-200x, Space Applications Electronic Hardware Addendum to IPC/WHMA-A-620A (Supplement to ANSI/IPC WHMA-A-620-2002)

This addendum provides additional requirements over those published in IPC/WHMA-A-620 to ensure the reliability of soldered electrical and electronic assemblies that must survive the vibration and thermal cyclic environments getting to and operating in space. Where content criteria are not supplemented, the Class 3 requirements of IPC/WHMA-A-620 apply.

SMACNA (Sheet Metal and Air-Conditioning Contractors' National Association)

BSR/SMACNA 007-200x, Residential Comfort System Installation Standards Manual (New standard)

The document will provide installation standards for residential heating, ventilating, and air conditioning (HVAC) systems. It will include the most current mechanical and control technology so contractors and designers can consider design aspects, construct, and install from the simplest to state-of-the-art HVAC systems. Forced-air heating, heat pumps, automatic controls and thermostats, flues, vents, sound and vibration, air cleaning, and other subjects and technologies appropriate for this new century will be included.

AWWA (American Water Works Association)

BSR/AWWA C750-200x, Transit-Time Flowmeters in Full Closed Conduits (Revision of ANSI/AWWA C750-2003)

This standard describes transit-time ultrasonic flowmeters for water supply service application. An ultrasonic flowmeter is a meter that uses acoustic energy signals to measure fluid velocity. There are currently two distinct types of ultrasonic flowmeters available: Doppler effect and transit time. The Doppler-effect meter is used extensively for fluids containing solid particles or gases, and the transit-time flowmeter is used in a wide variety of applications in the water industry. Project need: The purpose of this standard is to provide purchasers, manufacturers, and suppliers with the minimum requirements for transit-time flowmeters, including components, performance, calibration, and verification.

Final actions on American National Standards

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

ASA (ASC S1) (Acoustical Society of America)

REAFFIRMATIONS

ANSI S1.4-1983 (R2006), Specification for Sound Level Meters [Reaffirmation of ANSI S1.4-1983 (R2001)] (21 March 2006)

ANSI S1.4a-1985 (R2006), Amendment to ANSI S1.4-1983—Specification for Sound Level Meters [Reaffirmation of ANSI S1.4a-1985 (R2001)] (21 March 2006)

ANSI S1.6-1984 (R2006), Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements [Reaffirmation of ANSI S1.6-1984 (R2001)] (21 March 2006)

ANSI S1.8-1989 (R2006), Reference Quantities for Acoustical Levels [Reaffirmation of ANSI S1.8-1989 (R2001)] (21 March 2006)

ANSI S1.9-1996 (R2006), Instruments for the Measurement of Sound Intensity [Reaffirmation of ANSI S1.9-1996 (R2001)] (21 March 2006)

ANSI S1.42-2001 (R2006), Design Response of Weighting Networks for Acoustical Measurements [Reaffirmation of ANSI S1.42-2001] (21 March 2006)

ANSI S1.15, Part 1-1997 (R2006), Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones [Reaffirmation of ANSI S1.15, Part 1-1997 (R2001)] (21 March 2006)

WITHDRAWALS

ANSI S1.15, Part 1-1997 (R2006), Measurement Microphones, Part 1: Specifications for Laboratory Standard Microphones [Reaffirmation of ANSI S1.15, Part 1-1997 (R2001)] (21 March 2006)

ASA (ASC S2) (Acoustical Society of America)

REAFFIRMATIONS

ANSI S2.8-1972 (R2006), Guide for Describing the Characteristics of Resilient Mountings [Reaffirmation of ANSI S2.8-1972 (R2001)] (21 March 2006)

ANSI S2.9-1976 (R200x), Nomenclature for Specifying Damping Properties of Materials [Reaffirmation of ANSI S2.9-1976 (R2001)] (21 March 2006)

ANSI S2.20-1983 (R2006), Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects [Reaffirmation of ANSI S2.20-1983 (R2001)] (21 March 2006)

ANSI S2.24-2001 (R2006), Graphical Presentation of the Complex Modulus of Viscoelastic Materials (Reaffirmation of ANSI S2.24-2001) (21 March 2006)

WITHDRAWALS

ANSI S2.13-Part 1-1996, Mechanical Vibration of Non-Reciprocating Machines—Measurements on Rotating Shafts and Evaluation—Part 1: General Guidelines [Withdrawal of ANSI S2.13-Part 1-1996 (R2001)] (21 March 2006)

ANSI S2.41-1985, Mechanical Vibration of Large Rotating Machines with Speed Range from 10 to 200 rev/s—Measurement and Evaluation of Vibration Severity in situ [Withdrawal of ANSI S2.41-1985 (R2001)] (21 March 2006)

ASA (ASC S12) (Acoustical Society of America)

REVISIONS

ANSI S12.65-2006, Rating Noise with Respect to Speech Interference [Revision and redesignation of ANSI S3.14-1977 (R1997)] (28 February 2006)

IEEE (ASC C63) (Institute of Electrical and Electronics Engineers)

REVISIONS

ANSI C63.19-2006, Methods of Measurement of Compatibility between Wireless Communications Devices and Hearing Aids (Revision of ANSI C63.19-2001) (6 April 2006)

Standards News from Abroad

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

International Organization for Standardization (ISO)

Newly Published ISO and IEC Standards

Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization

ISO Standards

ACOUSTICS (TC 43)

ISO 10848-1:2006, Acoustics—Laboratory measurement of the flanking transmission of airborne and impact sound between adjoining rooms—Part 1: Frame document

ISO 10848-2:2006, Acoustics—Laboratory measurement of the flanking transmission of airborne and impact sound between adjoining rooms—Part 2: Application to light elements when the junction has a small influence

ISO 10848-3:2006, Acoustics—Laboratory measurement of the flanking transmission of airborne and impact sound between adjoining rooms—Part 3: Application to light elements when the junction has a substantial influence

ISO 17201-4:2006, Acoustics—Noise from shooting ranges—Part 4: Prediction of projectile sound

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 18436-1/Cor1:2006, Condition monitoring and diagnostics of machines—Requirements for training and certification of personnel—Part 1: Requirements for certifying bodies and the certification process—Corrigendum

ISO/DIS 19499, Mechanical vibration—Balancing and balancing standards—Introduction (22 July 2006)

ISO 20283-3:2006, Mechanical vibration—Measurement of vibration on ships—Part 3: Preinstallation vibration measurement of shipboard equipment

ISO Draft Standard

ACOUSTICS (TC 43)

ISO/DIS 3382-2, Acoustics—Measurement of room acoustic parameters—Part 2: Reverberation time in ordinary rooms (3 June 2006)

ISO/DIS 3743-1, Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Engineering method for small, movable sources in reverberant fields—Part 1: Comparison method for a hard-walled test room (8 July 2006)

ISO/DIS 3744, Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Engineering method for an essentially free field over a reflecting plane (1 July 2006)

ISO/DIS 5130, Acoustics—Measurements of sound pressure level emitted by stationary road vehicles (6 February 2006)

ISO 7779/DAmd2, Revision of measurement surfaces, procedures for equipment installation/operation and identification of prominent discrete tones (10 June 2006)

ISO/DIS 20906, Acoustics—Unattended monitoring of aircraft sound in the vicinity of airports (9 September 2006)

TECHNICAL SYSTEMS AND AIDS FOR DISABLED OR HANDICAPPED PERSONS (TC 173)

ISO/DIS 23600, Assistive products for persons with vision impairments and persons with vision and hearing impairments—Acoustic and tactile signals for pedestrian traffic lights (15 July 2006)

IEC Draft Standard

88/260/FDIS, Amendment 1 to IEC 61400-11 Ed.2: Wind turbine generator systems—Part 11: Acoustic noise measurement techniques (12 May 2006)

IEC Technical Specifications

ULTRASONICS (TC 87)

IEC/TS 62306 Ed. 1.0 b: 2006, Ultrasonics—Field characterisation—Test objects for determining temperature elevation in diagnostic ultrasound fields

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

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

U.S. TAG	ISO and IEC documents
S1	**Systematic Review of ISO12124:2001** "Acoustics–Procedures for the measurement of real-ear acoustical characteristics of hearing aids"
	Second IEC/CD 60318-1 (29/593/CD) "Electroacoustics-Simulators of human head and ear-Part 1: Ear simulator for the calibration of supra-aural and circumaural earphones" (Revision of IEC 60318-1:1998 and IEC 60318-2:1998)
S1/S3	Systematic Review of **ISO 8253-3:1996** "Acoustics—Audiometric test methods—Part 3: Speech audiometry"
S2	Systematic Review of **ISO9611:1996** "Acoustics—Characterization of sources of structure-borne sound with respect to sound radiation from connected structures—Measurement of velocity at the contact points of machinery when resiliently mounted"
S12	Systematic Review of **ISO532:1975** "Acoustics—Method for calculating loudness level"
	ISO/FDIS17201-2 "Acoustics—Noise from shooting ranges—Part 2: Estimation of muzzle blast and projectile sound by calculation"
	Systematic Review of **ISO 2923:1996** "Acoustics—Measurement of noise on board vessels"
	ISO/DIS3743-1 "Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Engineering method for small, movable sources in reverberant fields—Part 1: Comparison method for a hard-walled test room"
	Systematic Review of **ISO3891:1978** "Acoustics—Procedures for describing aircraft noise heard on the ground"
	Systematic Review of **ISO5129:2001** "Acoustics—Measurement of sound pressure levels in the interior of aircraft during flight"
	Systematic Review of **ISO5131:1996** "Acoustics—Tractors and machinery for agriculture and forestry—Measurement of noise at the operator's position—Survey method"
	Systematic Review of **ISO9613-2:1996** "Acoustics—Attenuation of sound during propagation outdoors—Part 2: General method of calculation"
	Systematic Review of **ISO9614-2:1996** "Acoustics—Determination of sound power levels of noise sources using sound intensity—Part 2: Measurement by scanning"
	Systematic Review of **ISO11689:1996** "Acoustics—Procedure for the comparison of noise-emission data for machinery and equipment"
	Systematic Review of **ISO11690-2:1996** "Acoustics—Recommended practice for the design of low-noise workplaces containing machinery—Part 2: Noise control measures"
	Systematic Review of **ISO11820:1996** "Acoustics—Measurements on silencers in situ"
	Systematic Review of **ISO11957:1996** "Acoustics—Determination of sound insulation performance of cabins—Laboratory and in situ measurements"

U.S. TAG	ISO and IEC documents
	Systematic Review of **ISO12001:1996** "Acoustics—Noise emitted by machinery and equipment—Rules for the drafting and presentation of a noise code"
	Systematic Review of ISO **14257:2001** "Acoustics—Measurement and parametric description of spatial sound distribution curves in workrooms for evaluation of their acoustical performance"
	Systematic Review of ISO **16:1975** "Acoustics—Standard tuning frequency (standard musical pitch)"
	ISO 7779:1999/DAmd 2—Draft Amendment—Acoustics—Measurement of airborne noise emitted by information technology and telecommunications equipment—Amendment 2: Revision of measurement surfaces, procedures for equipmentinstallation/operation and identification of prominent discrete tones

U.S. TAG	ISO and IEC documents
	ISO/DIS20906 "Acoustics—Unattended monitoring of aircraft sound in the vicinity of airports"

BOOK REVIEWS

P. L. Marston
Physics Department, Washington State University, Pullman, Washington 99164

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

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

Architectural Acoustics

Marshall Long

Elsevier Academic Press, Burlington MA, 2006. 844 pp. Price: $99.00 (hardcover). ISBN: 0124555519

Architectural Acoustics is a new book that the author suggests is appropriate as an undergraduate text. It is in fact much more than this, and while it includes the type of material one might expect in a textbook, it is also a very comprehensive compendium of information that would be a valuable resource for an acoustical consultant. It has a very elegant cover, is printed on good quality paper, and includes many clearly drawn figures and graphs. The 22 chapters include a broad range of topics that cover both basic fundamentals and practical design issues. Although titled *Architectural Acoustics*, the book includes a wide range of material related to architectural and building acoustics, as well as related subjects such as environmental noise and sound system design.

The first chapter gives a fascinating run through the history of architectural acoustics. It is a concise but broad introduction including mention of the historical development of music and theatre as well as purely acoustical issues. For those who want to read more, there are many references that can be pursued. This is followed by a chapter on acoustical fundamentals and another on human perception and reaction to sound. This latter chapter reveals one of the weaknesses of this book. There is much of historical interest, such as the work of French and Steinberg and the Articulation Index (AI), but no mention of the Speech Intelligibility Index that replaced AI nearly 10 years ago. There is much mention of material from the EPA Levels Document (from the early 1970s), but little reference to more recent work on the effects of environmental noise. There are errors too, such as the definition of a rhyme-type speech intelligibility test at the bottom of page 92. On the other hand, there is invaluable information from less accessible references such as the 1983 report by Chanaud.

Chapter 4 on acoustics measurements and noise metrics is quite comprehensive and includes both indoor and outdoor noise metrics and procedures. It is at times perhaps too comprehensive, introducing archaic measures without comment. For example, the Noise Pollution Level and the Traffic Noise Index (page 136), originally developed in the United Kingdom, have long since ceased to be used because they are not particularly good predictors of human response to environmental noises. On page 154, it is suggested that a variation of the Speech Transmission Index should be developed for evaluating sound systems. Such a measure exists and is referred to as STIpa as defined in the IEC 60268-16 standard.

The inclusion of a chapter on environment noise may at first seem out of place in a book on architectural acoustics, but most concerns about environmental noise are to protect people in buildings. I found this chapter a useful addition that presents much material that is not available in other books. The basics of noise from moving sources, outdoor propagation, and noise barriers are all presented. The emphasis is on road traffic noise and aircraft noise, and more fundamental issues about outdoor propagation are not so extensively considered.

Chapter 6 goes from the basics of wave acoustics to practical results on the radiation of sound from loudspeakers. A chapter on the basic principles of sound absorption, reflection, and transmission at solid surfaces follows. This is again quite comprehensive and the inclusion of the work of Rindel on reflector panels is another of several gems of information that the book presents. Rindel's work was published in difficult to find conference publications, yet seems to be of great practical value. This is followed by a chapter on the fundamentals of sound in enclosed spaces, which serves as an introduction to the acoustics of duct systems as well as to room acoustics. There is mention of standard reverberation chamber tests but no discussion of the limitations of these tests.

Two chapters on airborne sound transmission are included. The first introduces the fundamentals and depends heavily on Sharpe's work; the second discusses the many practical issues in real buildings. New equations for sound transmission are developed that include the direct sound component in the receiving space. This is not the usual approach and there is no discussion as to how closely a real wall approximates the assumed rigid piston model that is introduced for the direct sound. There are many tables of sound transmission loss values in this chapter and in a later chapter on sound transmission through floors. This is very useful, but I personally would find it easier to appreciate this information if it were in graphical format.

A good introduction to the basics of vibration and vibration isolation is given in Chap. 11. This extends to floor vibrations and human perception of them and also includes an interesting description of tuned mass dampers in large towers. Chapters follow this on transmission through floor systems, noise in mechanical systems, and sound attenuation in ducts. These all have an applied practical focus and include information from a number of other very useful references.

At this point the book moves on from components and fundamentals to more complete building systems. The chapter on multifamily buildings includes much qualitative advice but in many cases there are no quantitative details. A chapter on office buildings follows and includes much useful information on speech privacy, much of which is derived from the 1983 report by Chanaud. There are some inconsistencies and a lack of discussion of conflicting sources of information. For example, in Table 16.2 AI=0.1 is termed *normal privacy*, but in Figs. 16.15 and 16.17 normal privacy corresponds to AI=0.15. This chapter also includes extensive discussions of various practical issues related to mechanical systems in office buildings.

Chapters on speech in rooms, sound systems, and rooms for music follow. The chapter on speech in rooms contains an incorrect description of the "cocktail party effect." This well-known psychoacoustic phenomenon refers to our ability to focus listening attention on a single talker in a noisy mix of conversations and background noise, and not to the increase in speech levels in such an environment. The chapter on sound systems presents a huge amount of practical information but not necessarily in the form of a step-by-step design process. The chapter on rooms for music does mention various ''newer'' room acoustics measures but does not mention the ISO 3382 standard that includes the definitions of many of these measures. There are a number of specific concert halls and opera house examples that come from the books by Beranek and Barron.

The various components of multipurpose auditoria are extensively discussed in the following chapter. This information ranges from key components such as seating, orchestra shells, and orchestra pits, to obscure details such as that in some orthodox Jewish synagogues the use of condenser microphones but not dynamic microphones is acceptable, which is somehow related to ancient traditions about not lighting fires on the Sabbath. The chapter on studios and listening rooms includes a very short but interesting history of the development of audio engineering. This is another chapter where much of the information is not readily available elsewhere and so is particularly valuable to readers. The final chapter on room acoustic modeling provides an introduction to a topic that is currently changing rapidly.

There are a number of small errors in the text including missing references and incorrectly referenced figures, etc. However, most do not present significant problems for the reader. Occasionally words are not as precise as one might like (e.g., on page 68 equal level contours are referred to as equal loudness contours). In general, the book is well written and explanations are clear and easily understandable. An impressive 15 pages of useful references are included. However, of the many references only a very small percentage were published in the last 10 years. I think this points to a major weakness of the book. It includes an immense compilation of information but is not as up to date as one might like. The style and organization of the work leads to some duplication between chapters. However, this makes each chapter a more complete discussion of each topic with less need to refer to other chapters.

For me the major feature of *Architectural Acoustics* is the comprehensive range of useful information compled into one book. The book is particularly helpful when it reproduces material from several very helpful sources that are not readily available to most readers. Its weakest points are the shortage of references to newer material and the tendency to present large amounts of information from a variety of sources without much discussion of the differences and connections among the various details. In spite of my nit-picking, the book is a valuable new contribution and seems to be a bargain for the immense amount of material it includes.

JOHN BRADLEY
Institute for Research in Construction,
National Research Council,
1200 Montreal Rd.,
Ottawa, K1A 0R6 Canada

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

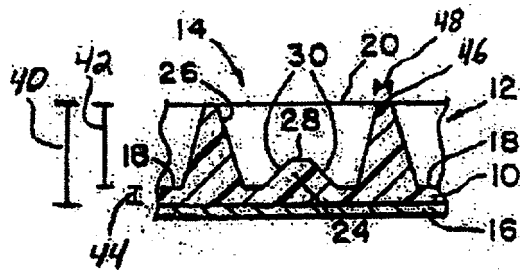
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,954,315

43.20.EI NIGHT VISION AND AUDIO SIGNAL REDUCTION SYSTEM

Richard J. Tracy, assignor to Illinois Tool Works Incorporated
11 October 2005 (Class 359/707); filed 25 June 2004

This patent describes a surface that is said to be nonreflective, or, at least, diffusive, for a range of wavelengths of both EM and acoustic radiation, including infrared and ultraviolet. For photons, the surface needs only



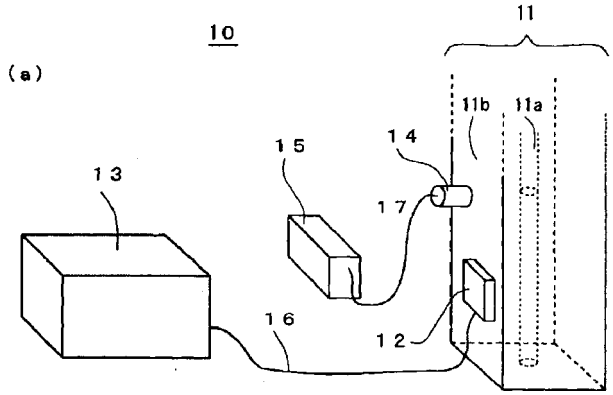
to appear black, although this may not be easy over the frequency range quoted. In the preferred embodiment, dimension **40** would be on the order of 0.008 in. It is hard to imagine this surface as having any absorbtive acoustical effect at all at audible wavelengths.—DLR

6,962,082

43.20.Tb DEVICE AND METHOD FOR ACOUSTIC DIAGNOSIS AND MEASUREMENT BY PULSE ELECTROMAGNETIC FORCE

Mitsuo Hashimoto and Masanori Takanabe, assignors to Amic Company, Limited
8 November 2005 (Class 73/579); filed in Japan 17 November 2000

This device sends a magnetic pulse into a structure of reinforced concrete and detects an acoustic signal produced by the eddy-current response



of the reinforcing materials to the magnetic pulse. The analysis is said to provide information on the state of corrosion, adhesion, cover depth, and the diameter of the reinforcing rods.—DLR

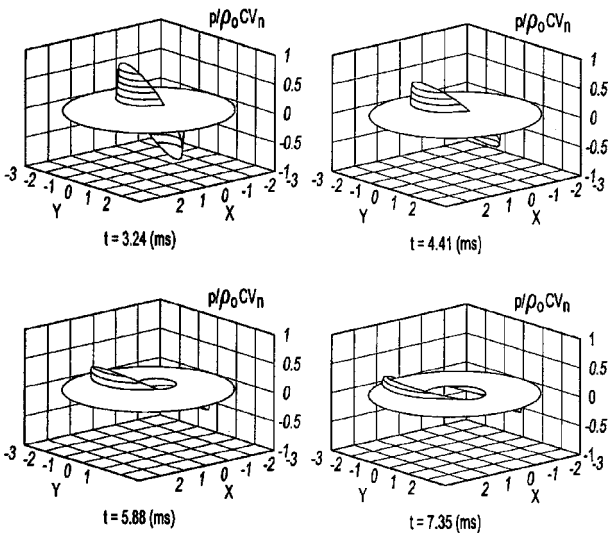
6,996,481

43.20.Ye RECONSTRUCTION OF TRANSIENT ACOUSTIC RADIATION FROM A FINITE OBJECT SUBJECT TO ARBITRARILY TIME-DEPENDENT EXCITATION

Sean F. Wu, assignor to Wayne State University
7 February 2006 (Class 702/39); filed 8 January 2004

The patent deals with complex methods of imaging three-dimensional motion of an object via a scanning microphone array that picks up and

 0001-4966/2006/120(1)/15/14/$22.50



analyzes the acoustical field generated by the object. A typical display of such motion is shown in the figure. The technique may be used for many problems in engineering structural analysis.—JME

6,995,707

43.30.Wi INTEGRATED MARITIME PORTABLE ACOUSTIC SCORING AND SIMULATOR CONTROL AND IMPROVEMENTS

Christopher R. Karabin *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
7 February 2006 (Class 342/357.09); filed 18 March 2004

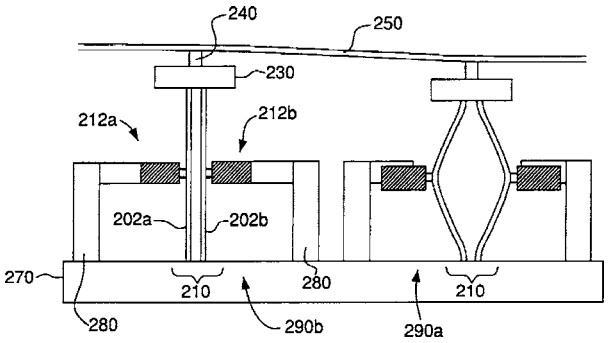
The system described represents a portable marine scoring and simulation system that can be deployed in an ocean area for marine combat training. It comprises three or more buoys, each with a GPS receiver, a rf radio system, an acoustic analysis system, and a microprocessor. The acoustic analysis system captures the acoustic signature of some explosive or nonexplosive ordnance impacting the ocean in the area bounded by the set of buoys. When an acoustic signature is captured, the rf radio system transmits the time of capture plus the GPS position of the buoy to the system controller. When three or more buoys transmit a captured acoustic signature, the system controller computes the location of the ordnance impact. The system is portable, which allows it to be deployed in any environmentally suitable area of the ocean and to be recovered for transport and subsequent use.—WT

6,995,895

43.38.Ar MEMS ACTUATOR FOR PISTON AND TILT MOTION

Dennis S. Greywall, assignor to Lucent Technologies Incorporated
7 February 2006 (Class 359/290); filed 5 February 2004

This patent discloses the use of flat plates as mechanical levers to accomplish large excursion drive using short-throw transducers such as piezoelectric devices. The figure shows one such manifestation of this concept, with two identical pushers **210** side-by-side, acting to tip and/or bend an upper flexible membrane **250** that they are attached to. The transducers



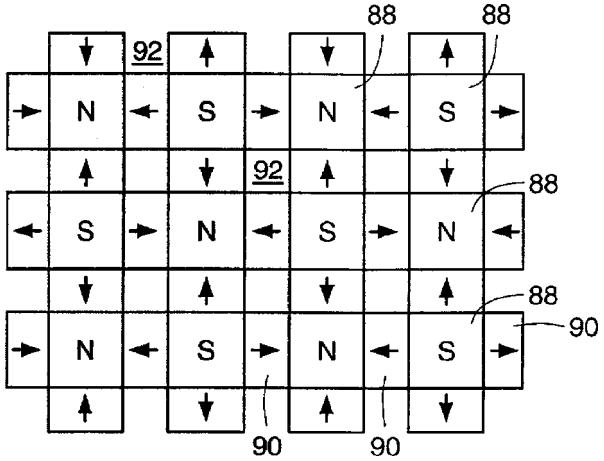
are embedded in **212a** and **212b**, and act through mechanical lever arms **210** to effect a tilt in the membrane. One cited advantage for such an arrangement is that the driving elements are in the plane of the wafer they are fabricated from, allowing a single-wafer fabrication of an out-of-plane MEMS actuator. A few other designs for such an out-of-plane actuator are shown and fabrication details are given.—JAH

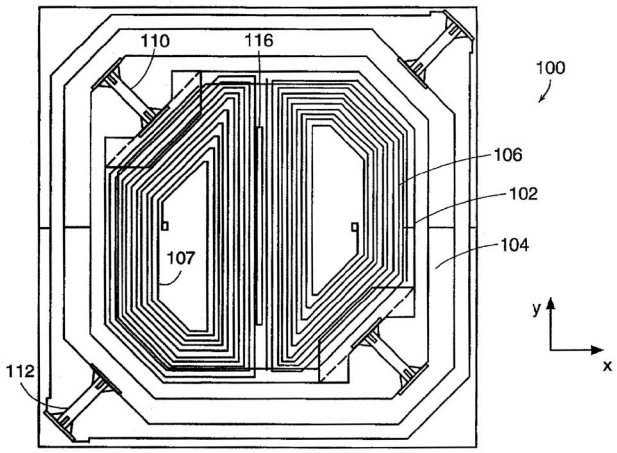
6,989,921

43.38.Dv MAGNETICALLY ACTUATED MICRO-ELECTRO-MECHANICAL APPARATUS AND METHOD OF MANUFACTURE

Jonathan Bernstein *et al.*, assignors to Corning Incorporated
24 January 2006 (Class 359/290); filed 24 August 2001

This patent describes a MEMS mirror array that is actuated electrodynamically rather than via the usual electrostatic or thermal means. Planar coils sit above an array of magnets as seen in the figures and current in the coils is used to tilt the mirrors about two orthogonal gimbals. The gimbals and coils are machined from a single silicon wafer, which seems to be a





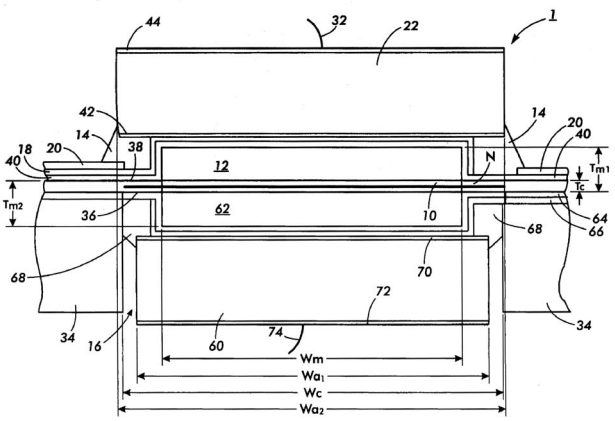
selling point of this design. The inventors state that this electrodynamic actuation arrangement is more suitable for low-voltage actuation than the electrostatic types that have swept the marketplace. Perhaps time will tell whether there is a niche application for this design. The patent is clearly written but a little vague on the physical basis for the design.—JAH

6,987,348

43.38.Fx PIEZOELECTRIC TRANSDUCERS

Steven A. Buhler *et al.*, assignors to Palo Alto Research Center Incorporated
17 January 2006 (Class 310/330); filed 16 September 2003

The inventors describe a piezoelectric laminate transducer that has the piezoelectric elements on the outside rather than the inside of the assembly, as is the case in more common types of laminate transducers, e.g., flextensional types. Apparently from the wording of the abstract, the inventors are



aiming for reversible applications, but they afford no evidence that the resulting transducers have high electromechanical coupling constants useful for reversible applications. The piezoelectric layers are **22** and **60**, surrounding air gaps **12** and **62** created by shells of insulator **40** and metal interconnect **18**. The complexity and bonding issues associated with this design would seem to be prohibitive. While the patent is clearly written, it is mostly about the materials and fabrication process.—JAH

6,987,433

43.38.Fx FILM ACOUSTICALLY-COUPLED TRANSFORMER WITH REVERSE C-AXIS PIEZOELECTRIC MATERIAL

John D. Larson III and Yury Oshmyansky, assignors to Agilent Technologies, Incorporated
17 January 2006 (Class 333/189); filed 29 April 2004

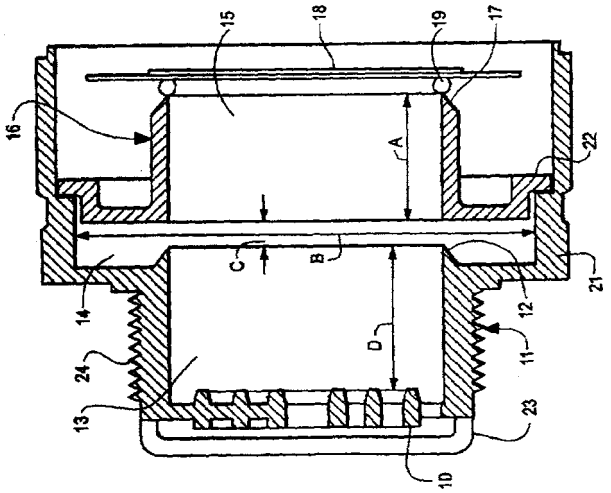
This patent discloses an interesting piezoelectric transformer based on stacked film bulk acoustic wave resonators (FBARs). The transformer is intended for cell phone use, as it is operated in the 1–2-GHz frequency range and the piezoelectric elements are only 700 nm thick. The inventors describe how they cancel parasitic capacitive coupling between the input and output circuits by inverting the electrical signal to one half of it and simultaneously inverting the poling axis to that side, in analogy to the way that ordinary transformers often have layered, interleaved windings. The patent is clearly written and informative, but lacking any performance data.—JAH

6,987,445

43.38.Fx WATER RESISTANT AUDIBLE SIGNAL

George A. Burnett *et al.*, assignors to Mallory Sonalert Products, Incorporated
17 January 2006 (Class 340/387.1); filed 22 September 2000

Mallory has been manufacturing piezoelectric "beep" generators for many years. When these are used in exposed locations, water can accumulate in the housing and corrode metal parts. Mounting the device face-down



inhibits water accumulation, but may not be the best orientation for acoustic performance. Adding a layer of porous, water-repellant material **23** to the sound exit might be a good idea and that is exactly what has been patented.—GLA

6,990,046

43.38.Fx SONAR TRANSDUCER

Jozef J. Gluszyk, Houston, Texas
24 January 2006 (Class 367/174); filed 15 August 2003

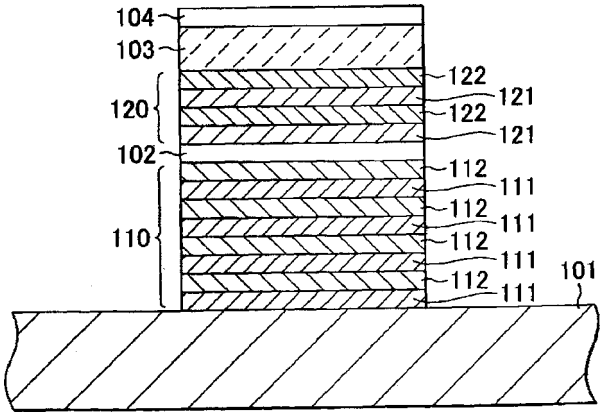
An ultrasonic probe for measuring the level of fluid, solid, or gas interfaces within some vessel or container comprises a routine piezoelectric disc mounted at the end of a shaft which is closed by a suitable diaphragm at the interface end with the vessel. A simple measurement of the travel times of the radiated acoustic wave and the waves reflected from any of these interfaces within the vessel provides values for the levels of these interfaces.—WT

6,995,497

43.38.Fx FILM BULK ACOUSTIC RESONATOR

Kenji Inoue, assignor to TDK Corporation
7 February 2006 (Class 310/320); filed in Japan 29 October 2003

A film bulk acoustic wave resonator is described that is said to have reduced interelectrode capacitance due to a quarter wave stack **111** and **112** that is interposed in the acoustic wave path. This is not novel, and has been



done many times as a matter of course, in practice on the mm length scale during transducer fabrication using ceramic or metal backing plates with insulating washers. Then again, there could be more here than meets the eye.—JAH

6,990,209

43.38.Hz HIGH DIRECTIVITY MICROPHONE ARRAY

Martin Reed Bodley *et al.*, assignors to GN Netcom, Incorporated
24 January 2006 (Class 381/172); filed 28 October 2002

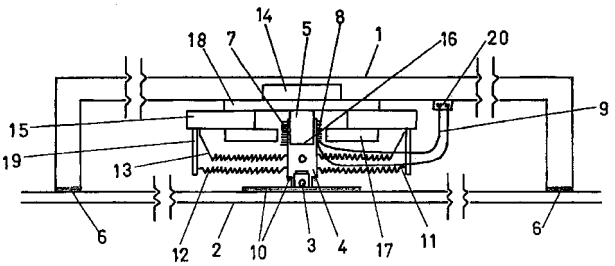
The usual handheld cellphone or other digital device normally has a single microphone, which, due to proximity of the talker, provides sufficient signal-to-noise ratio for virtually all applications. This patent describes an array of microphones, positioned in the upper and lower sections of the fold-down apparatus and thus producing a fairly complex microphone array. There are 23 drawings attached, but there is no indication of actual performance improvement to be expected relative to the traditional approach.—JME

6,965,679

43.38.Ja EQUALIZABLE ELECTRO-ACOUSTIC DEVICE USED IN COMMERCIAL PANELS AND METHOD FOR CONVERTING SAID PANELS

Alejandro José Pedro Lopez Bosio and Hernán Humberto Rojas Castillo, both of Saragossa, Spain
15 November 2005 (Class 381/152); filed 17 October 2000

A moving-coil motor is attached to wall panels, which is said to turn the commercial building panels into "flat radiators of high-fidelity sound with a response of 40–18,000 Hz ±3 dB and an efficiency of 86 dBWm." Although the patent acknowledges some prior art, specifically that of Sound Advance Systems, other prior art, such as the iceiling line from Armstrong World Industries (although an Armstrong-type mineral fiber is mentioned),



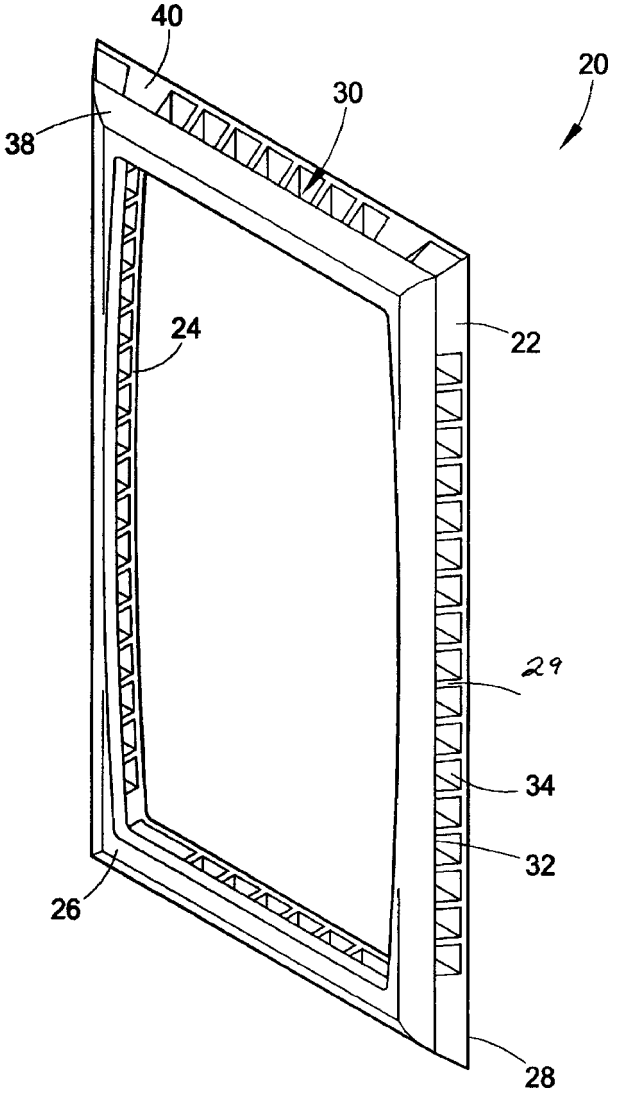
the surfeit of patents from NXT, etc., are not mentioned in any way. How the described motor can provide either the claimed frequency response or the claimed efficiency is not clearly described. If using basic commercially available building materials as an audio transducer is of interest to the reader, the reviewer suggests a search on motors that use magnetostrictive conversion methods, such as those now available using Terfenol-d.—NAS

6,993,145

43.38.Ja SPEAKER GRILLE FRAME

Christopher Combest, assignor to Multi-Service Corporation
31 January 2006 (Class 381/391); filed 26 June 2003

Speaker grill frames are typically at least 10 mm thick and it is well known that reflections from the interior edges of the frame can noticeably degrade high-frequency performance. One solution is to mount the frame on



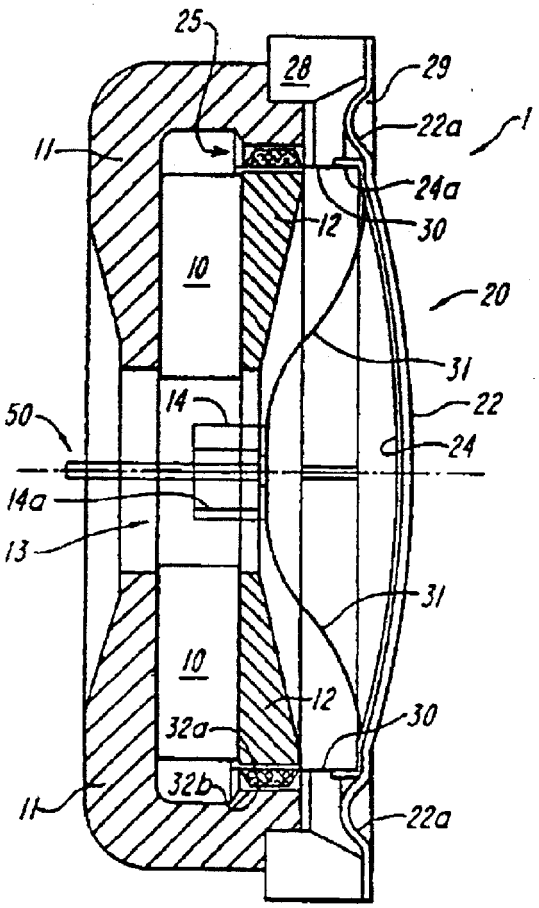
a step-down shelf so that the top of the frame forms an extension of the baffle surface. This patent describes an alternative approach in which a porous frame allows free passage of sound waves, as shown.—GLA

6,993,147

43.38.Ja LOW COST BROAD RANGE LOUDSPEAKER AND SYSTEM

Godehard A. Guenther, San Francisco, California
31 January 2006 (Class 381/407); filed 31 March 2003

This appears to be a fairly conventional, edge-driven dome that might be used as a tweeter although the inventor envisions operation down to 200 Hz or so. A rare-earth magnet **10** is mounted in a cast pot **11** and fitted with



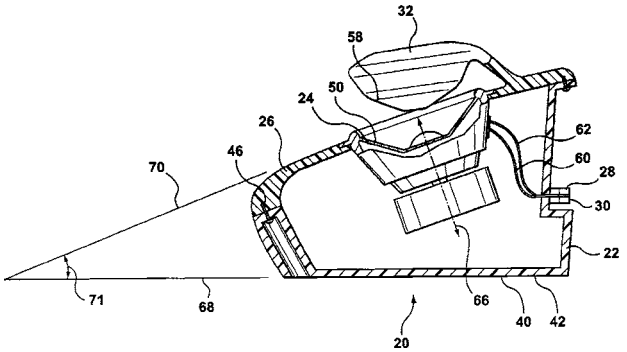
a shaped center pole-piece **12**. (There will be a brief pause while loudspeaker manufacturers chuckle at the term ''low cost.'') The sole novel feature appears to be the use of metallic strips bonded to internal flexible bands **31**. These connect one or more voice coils **32a**, **32b** to fixed terminals **14**.—GLA

6,996,243

43.38.Ja LOUDSPEAKER WITH SHAPED SOUND FIELD

Andrew C. Welker and John Tchilinguirian, assignors to Audio Products International Corporation
7 February 2006 (Class 381/160); filed 4 March 2003

Inventors have been placing conical reflectors in front of loudspeakers for more than 50 years. In the case at hand, the reflector is tilted with respect to the speaker axis, thus producing an asymmetric coverage pattern in the



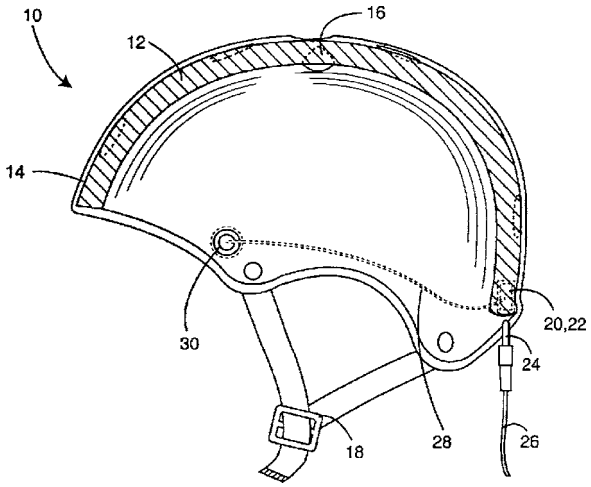
2–5-kHz band. A stacked, coaxial array can be used to extend the upper-frequency limit.—GLA

6,970,691

43.38.Si SPORTS HELMET HAVING INTEGRAL SPEAKERS

Spencer J. Thompson, Newbury, California
29 November 2005 (Class 455/344); filed 28 May 2002

Small loudspeakers **30** are mounted within the liner **12** of a sports helmet **10**. The location of the loudspeakers is said to ''provide audio to the helmet wearer without blocking surrounding sound, and without affecting



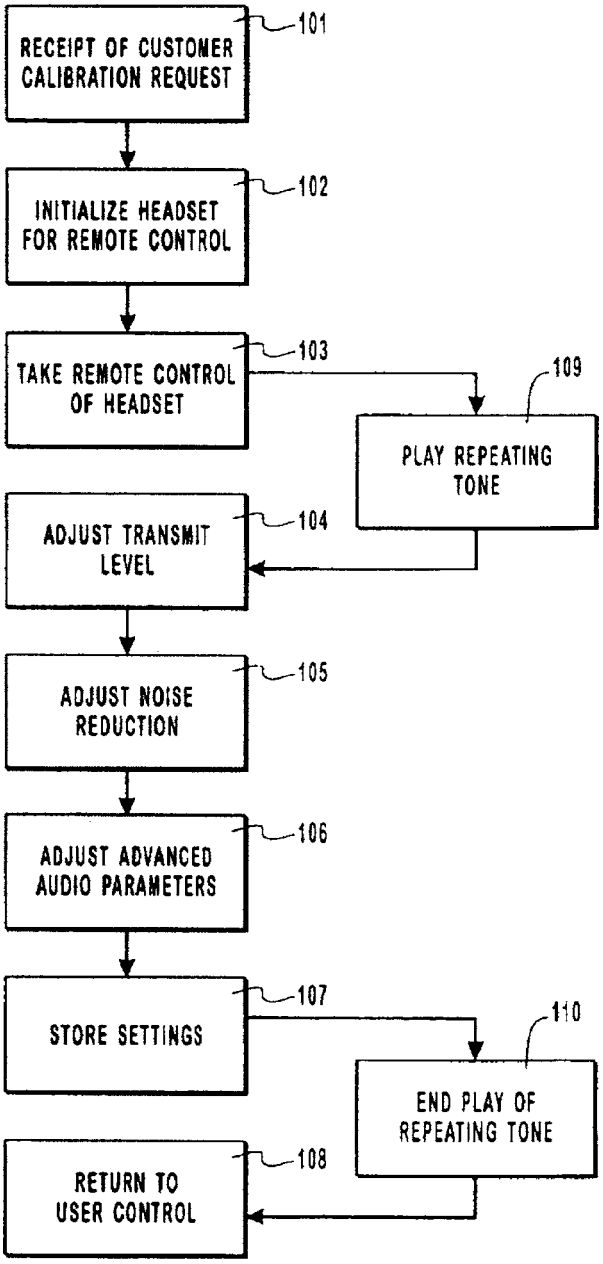
the safety aspect of the helmet,'' which may be true as long as the level of the sound from the loudspeakers is not raised to a level that would overcome the obstruction between the loudspeaker and the helmet wearer's ears, i.e., the helmet wearer's head.—NAS

6,990,190

43.38.Si METHOD AND SYSTEM FOR REMOTE TELEPHONE CALIBRATION

Daniel W. Mauney *et al.*, assignors to GN Jabra Corporation
24 January 2006 (Class 379/392.01); filed 27 March 2002

If you buy a hands-free telephone headset and plug it into your desk base-unit, a number of adjustments must be set manually to achieve satisfactory operation. These include signal level, speaker volume, and echo cancellation. This patent describes a user-friendly method for achieving optimum settings through remote control via the telephone line. Most of the



patent consists of a long computer program, followed by 33 patent claims. Believe it or not, there was a time when all telephone equipment in the United States was designed to work together and there would have been no need for such a process.—GLA

6,993,349

43.38.Si SMART RINGER

Paul Martinez and Scott Beith, assignors to Kyocera Wireless Corporation
31 January 2006 (Class 455/456.4); filed 18 July 2001

A cellular telephone may be used in almost any sonic environment. It would be desirable to automatically adjust audio and ringing loudness in response to background noise and a number of patents deal with this problem. In this case, however, computerized analyses of noise level, frequency spectrum, and rhythmic variation are performed and an optimum ring signal is then generated. This would seem to be somewhat self-defeating since the user is conditioned to respond to a familiar, identifiable signal.—GLA

6,990,205

43.38.Vk APPARATUS AND METHOD FOR PRODUCING VIRTUAL ACOUSTIC SOUND

Jiashu Chen, assignor to Agere Systems, Incorporated
24 January 2006 (Class 381/17); filed 20 May 1998

The patent deals with the processing of multiple monophonic signals, via HRTFs, for assignment to specific directions in a binaural headphone/stereo loudspeaker playback configuration. Applications include ''computer gaming, 3D audio, stereo sound enhancement, reproduction of multiple channel sound, virtual cinema sound, and other applications where spatial auditory perspective of 3D space is desired.''—JME

6,990,211

43.38.Vk AUDIO SYSTEM AND METHOD

Jeffrey C. Parker, assignor to Hewlett-Packard Development Company, L.P.
24 January 2006 (Class 381/310); filed 11 February 2003

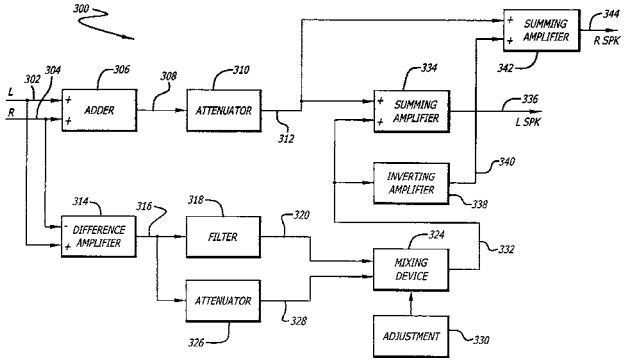
This patent is one of many that deal with adjustment of playback parameters in a surround-sound loudspeaker array to match the position and orientation of a given listener. Presumably, the operant factors of signal panning, signal level, signal delay, and signal equalization are embodied here.—JME

6,996,239

43.38.Vk SYSTEM FOR TRANSITIONING FROM STEREO TO SIMULATED SURROUND SOUND

Bradley C. Wood, assignor to Harman International Industries, Incorporated
7 February 2006 (Class 381/17); filed 2 May 2002

Given a standard home or business computer setup with stereo loudspeakers and a single subwoofer, it is proposed that multiple audio inputs can be selectively, and continuously, varied from mono, through stereo, to



simulated surround sound presentation—all via manipulation of relative levels, equalization, delay, and HRTFs. The figure shows basically how this is done. Note that variable elements **318**, **324**, and **330** are essential to the process.—JME

6,957,581

43.40.Le ACOUSTIC DETECTION OF MECHANICALLY INDUCED CIRCUIT DAMAGE

Peter Gilgunn, assignor to Infineon Technologies Richmond, LP
25 October 2005 (Class 73/587); filed 29 October 2003

The described system would detect abnormalities in the transmission of an acoustic signal in order to detect cracks or other faults in the silicon wafer during a semiconductor manufacturing process. Curiously, the patent

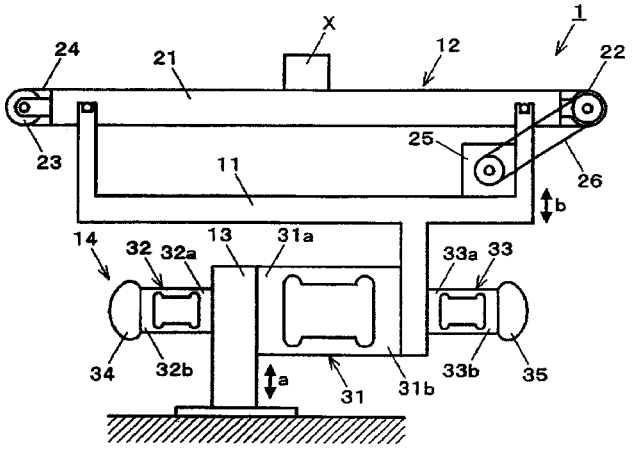
deals almost exclusively with mechanical details, such as transducer placement. Signal analysis is not mentioned.—DLR

6,987,227

43.40.Le WEIGHT DETECTING APPARATUS WITH VIBRATIONAL SENSORS ATTACHED TO BOTH THE FREE END AND THE FIXED END OF THE LOAD CELL

Yukio Wakasa, assignor to Ishida, Company, Limited
17 January 2006 (Class 177/25.13); filed in Japan 2 April 2003

A load cell **31** is to determine the weight of an object **X** moving along a conveyor belt **12**. The conveying system is supported on a frame **11** and driven by a motor **25**. Load cells **32**, capped with a known mass **34**, and attached to the support **13**, generate a signal that corresponds to the support



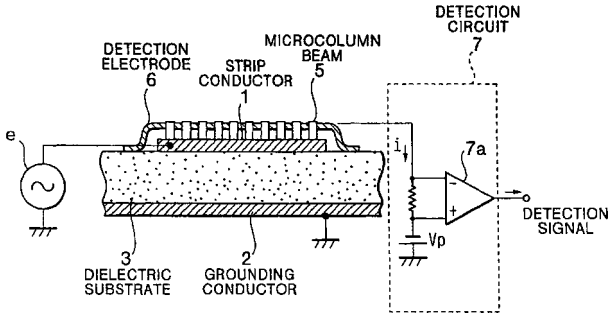
vibrations. Similarly, load cell **33**, capped with mass **35**, and attached to the conveyor frame **11**, produces a signal that corresponds to the conveyor frame vibrations, including those due to the drive motor. The signals derived from these two mass-capped load cells are combined with that from the primary load cell **31** for better determination of the weight.—EEU

6,995,633

43.40.Sk MICROMACHINE VIBRATION FILTER

Kunihiko Nakamura and Yoshito Nakanishi, assignors to Matsushita Electric Industrial Company, Limited
7 February 2006 (Class 333/186); filed in Japan 13 February 2002

The inventor discloses several types of resonant detectors for converting electrical signals into laser or capacitance modulation via the electric-field-induced motion of mechanical resonators. In one representative embodiment, the electrical signal is applied to stripline **1** which is bridged by a set of mechanical resonator filters **5**. These mechanical filters then supposedly move in response to the (current? or voltage?) signals on the stripline, causing induced currents in the amplifier **7a**. How this works is hard to see, because there are no electrostatic or magnetic biases applied or mentioned in



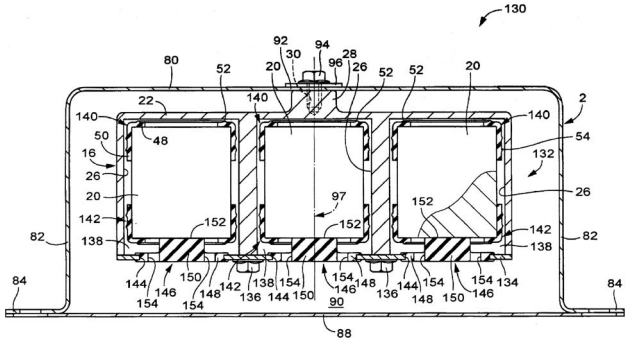
the text. There are also no piezoelectric or other electromechanical transducers mentioned, though at one point, it is mentioned that the columns are affected by the electric fields in the transmission line. The whole discussion is so vague and muddled that it is incomprehensible and the reader is left puzzled. It is best to pass this one by.—JAH

6,991,077

43.40.Tm VIBRATION DAMPING DEVICE

Hajime Maeno and Katsuhiko Katagiri, assignors to Tokai Rubber Industries, Limited
31 January 2006 (Class 188/380); filed in Japan 28 September 2001

This patent describes dynamic absorbers in which several masses are mounted resiliently in an elastically supported frame. In one embodiment, a frame **22** contains three cylindrical cavities in which are mounted cylindrical masses **20**. The frame is supported on a leaf spring **80**, attached to vibrating object **88**. The masses rest on elastomeric elements **150** and are cushioned



by elastomeric sleeves **140** and **142**, which fit around the masses, but have some clearance between them and the walls of the cavities in which they are located, including at the tops of the masses. Thus, the masses can move vertically on their supports at small amplitudes and also can impact against the cavity walls at larger amplitudes.—EEU

6,987,346

43.58.Kr ENERGY TRAP TYPE PIEZOELECTRIC RESONATOR COMPONENT

Mitsuhiro Yamada *et al.*, assignors to Murata Manufacturing Company, Limited
17 January 2006 (Class 310/320); filed in Japan 3 June 2003

This patent describes an "energy trapping" resonator design that works on the third harmonic of the thickness resonance of a piezoelectric

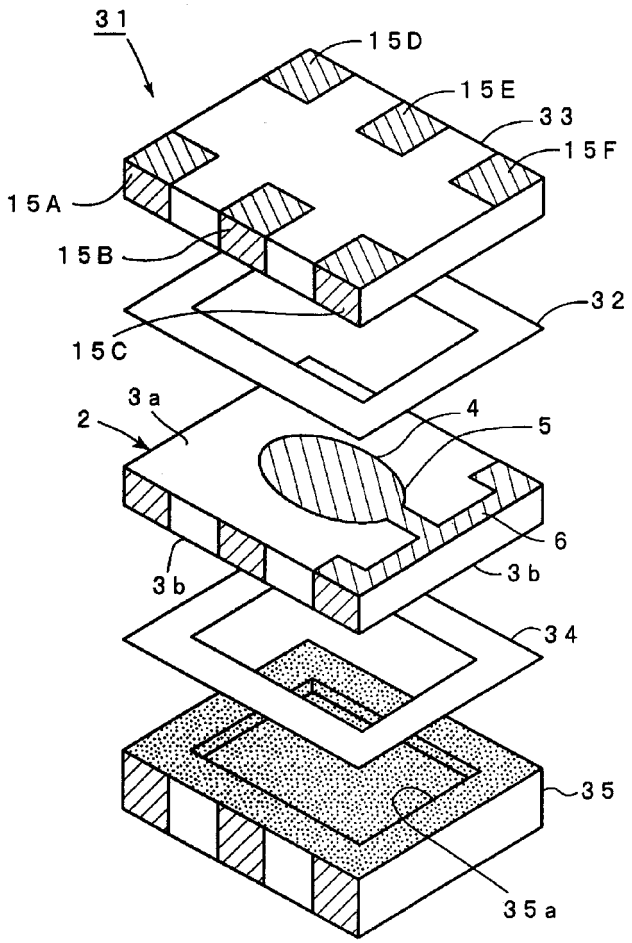


plate. The exploded view shows the essentials: piezo plate **2** is electroded in an elliptical pattern **4** on its top and bottom. The elliptical pattern aids in energy trapping to increase the Q of the resonator, while, at the same time, clamping and gluing around the edges causes suppression of the fundamental resonance mode. The patent is concise and well written, but lacking in details of energy trapping.—JAH

6,987,347

43.58.Kr PIEZOELECTRIC RESONATOR COMPONENT

Masakazu Yoshio *et al.*, assignors to Murata Manufacturing Company, Limited
17 January 2006 (Class 310/320); filed in Japan 30 May 2003

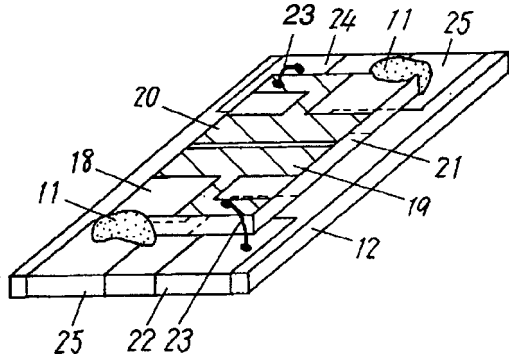
This patent is very similar to United States Patent 6,987,346, reviewed above. It differs only in that it explains the use of certain electrode geometries to suppress the fundamental-mode resonance. The inventors conclude here that circular electrodes are useful, whereas the previously referenced patent asserted that elliptical electrodes were optimal. Stay tuned for the sequel.—JAH

6,992,424

43.58.Wc PIEZOELECTRIC VIBRATOR LADDER-TYPE FILTER USING PIEZOELETRIC VIBRATOR AND DOUBLE-MODE PIEZOLECTRIC FILTER

Yukinori Sasaki *et al.*, assignors to Matsushita Electric Industrial Company, Limited
31 January 2006 (Class 310/360); filed in Japan 19 February 2001

This patent discloses the use of lithium tantalate as a useful material for high-stability piezoelectric resonators. The inventors use the anisotropic crystalline behavior to advantage by rotating the crystalline axis about the



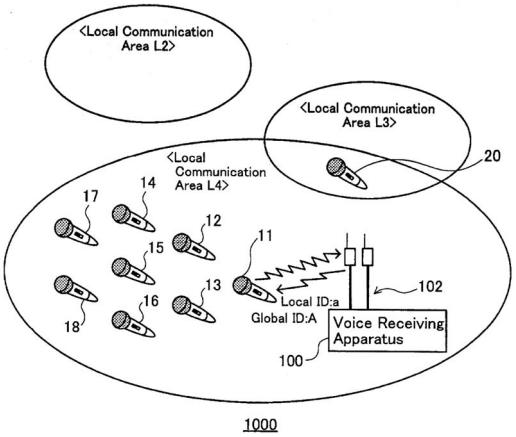
electric field direction to get a zero temperature coefficient. Along with that, they disclose a similar arrangement suited for use with ladder filter networks. The figure shows an embodiment of this design, which is really quite normal in all respects but the material used.—JAH

6,987,949

43.60.Dh WIRELESS MICROPHONE SYSTEM, VOICE RECEIVING APPARATUS, AND WIRELESS MICROPHONE

Shohei Taniguchi and Kenji Matsumoto, assignors to Matsushita Electric Industrial Company, Limited
17 January 2006 (Class 455/62); filed in Japan 31 October 2001

Despite continuing government restrictions on available bandwidth and power output, modern wireless microphone technology has attained broad acceptance in virtually all areas of communications and sound reinforcement. This patent extends the range of application to two-way linking



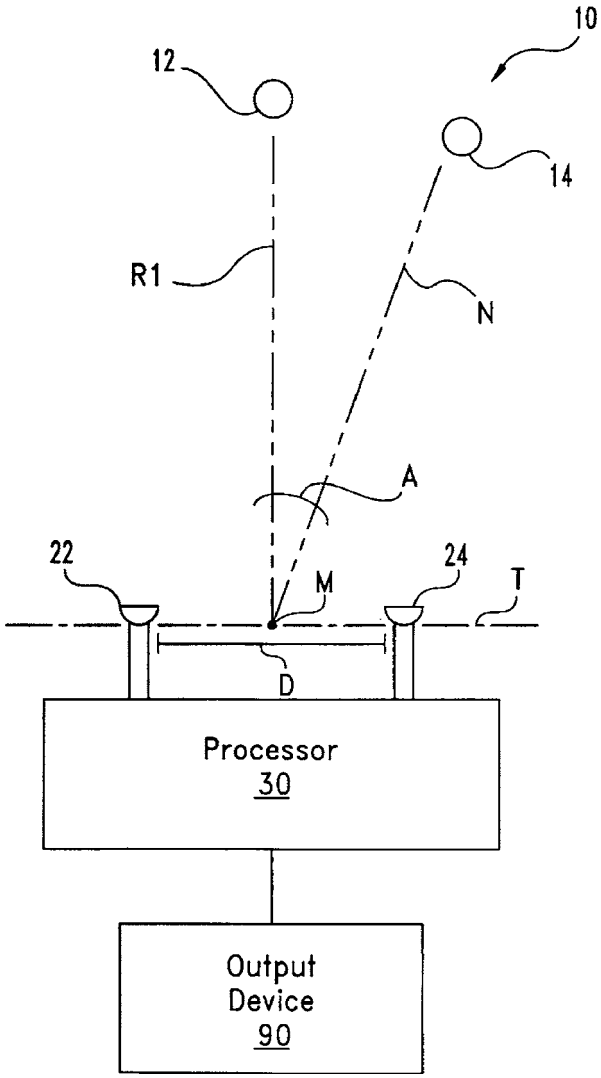
of each microphone with the target receiver, enabling useful communication and instructions from the receiving position to each microphone. Such housekeeping data as connection integrity, operating levels, and the like can be relayed to the users, as needed.—JME

6,987,856

43.60.Fg BINAURAL SIGNAL PROCESSING TECHNIQUES

Albert S. Feng *et al.*, assignors to Board of Trustees of the University of Illinois
17 January 2006 (Class 381/92); filed 16 November 1998

This complex patent describes a two-microphone array for ''zeroing in'' on a targeted on-axis sound source, simultaneously minimizing noise from off-axis sources. This is accomplished by sequentially delaying the



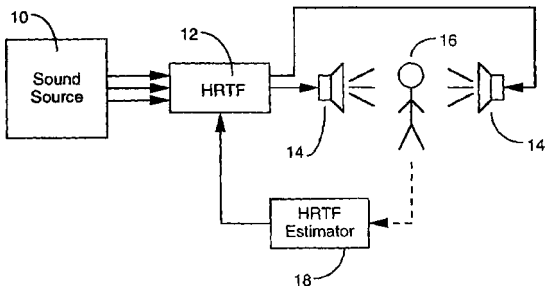
samples and analyzing them in several frequency bands, ultimately reducing the noise components to an acceptable level. The technique has obvious application for the hearing-impaired, using microphones positioned, left and right, in a set of spectacle frames.—JME

6,996,244

43.66.Pn ESTIMATION OF HEAD-RELATED TRANSFER FUNCTIONS FOR SPATIAL SOUND REPRESENTATIVE

Malcolm Slaney *et al.*, assignors to Vulcan Patents LLC
7 February 2006 (Class 381/303); filed 6 August 1999

The HRTFs for a subject can be derived from the set of HRTFs for a generic model, as compared to the complex differential HRTFs for that



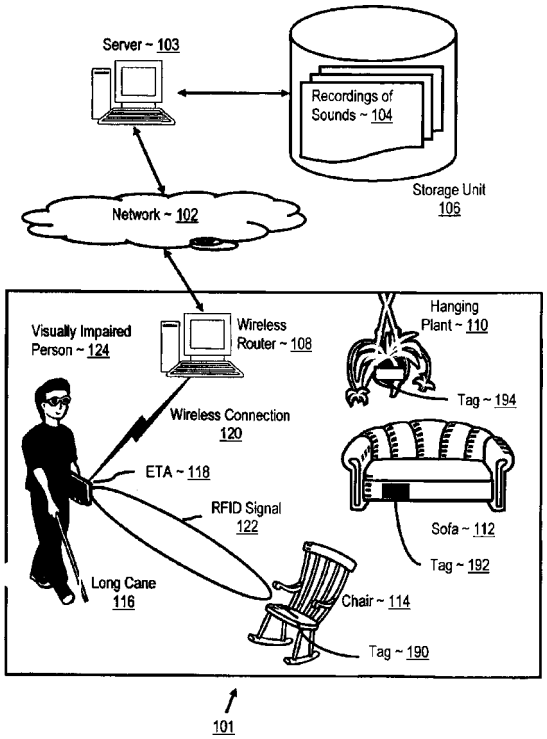
subject. This method simplifies the measurement process and has been noted earlier. The figure shows the basic analysis model.—JME

6,992,592

43.66.Ts RADIO FREQUENCY IDENTIFICATION AIDING THE VISUALLY IMPAIRED WITH SOUND SKINS

Michael Gilfix and Jerry Walter Malcolm, assignors to International Business Machines Corporation
31 January 2006 (Class 340/825.19); filed 6 November 2003

This electronic traveler aid (ETA) for a visually impaired person uses radio frequency identification (RFID) to identify objects and let the user



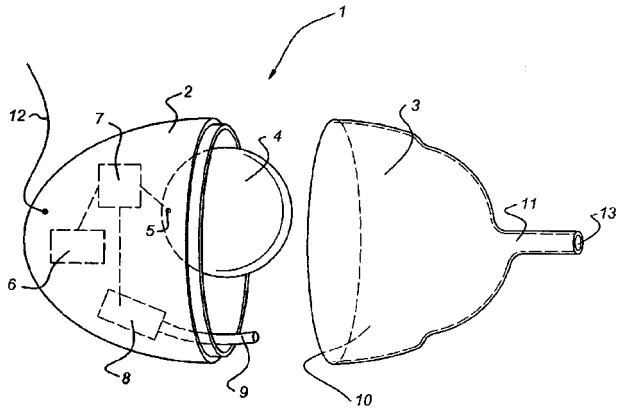
know what the objects are through audio prompts. The ETA stores a recording of a sound representing at least one attribute of an object having a RFID tag. Receipt of an appropriate RFID code retrieves the recording from storage and plays the recording through an audio interface of the ETA. The sound recordings may be stored locally or remotely and may be indexed and retrieved from storage according to a RFID code, a classification code for the object, and a type code (a sound skin identifier) for the recording.—DRR

6,993,142

43.66.Ts HEARING AID

Lourens George Bordewijk, assignor to Audilux Science B.V.
31 January 2006 (Class 381/323); filed in the Netherlands 3 December 1999

The sound output channel of a deep-fitting hearing aid is said to be more resistant to plugging due to ear-canal secretions by means of a two-piece housing. A battery chamber serves as a buffer between the secretions



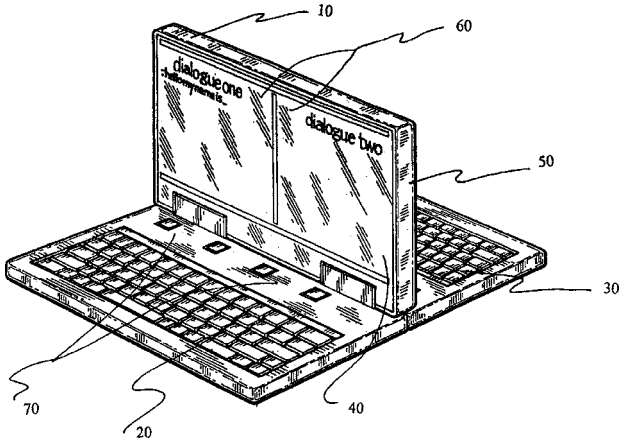
and the sound outlet. One end of a flexible tube is placed on the loudspeaker output and the other end opens into the battery chamber.—DAP

6,993,474

43.66.Ts INTERACTIVE CONVERSATIONAL SPEECH COMMUNICATOR METHOD AND SYSTEM

David G. Curry, Sedalia and Jason R. Curry, Kansas City, both of Missouri
31 January 2006 (Class 704/3); filed 17 May 2001

This device was developed with the view to changing the way hearing-impaired individuals communicate on a global scale by using an interactive Speech Communicator (sComm) system. The device may also be used by speakers of a foreign language. Using the sComm system, the hearing- or speaking-impaired user would be able to converse seamlessly without a translator and businessmen of different cultures and languages would be able to converse in conference rooms around the world without a human interpreter. The sComm system includes a custom laptop configuration having a two-sided display screen, a keyboard on each side of the screen, and a



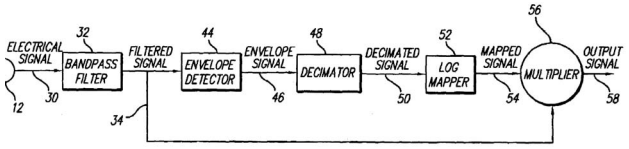
translation system for translating text from the original language into another language. The display screen will also have a split configuration, i.e., a double screen depicting chat boxes, each chat box dedicated to a user.—DRR

6,996,438

43.66.Ts ENVELOPE-BASED AMPLITUDE MAPPING FOR COCHLEAR IMPLANT STIMULUS

Andrew W. Voelkel, assignor to Advanced Bionics Corporation
7 February 2006 (Class 607/56); filed 14 October 2003

A log mapping of the current levels for stimulation of an electrode array in multiple frequency bands is obtained using an envelope-based amplitude-



mapping technique. A compressive function transforms a decimated envelope that is derived from multiple bandpass-filter outputs.—DAP

6,981,569

43.66.Vt EAR CLIP

Paul J. Stilp, Aliquippa, Pennsylvania
3 January 2006 (Class 181/129); filed 22 April 2003

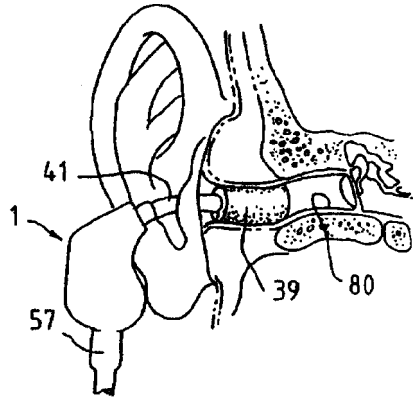
With two fingers, reach behind your earlobe, push it up, and pinch, trying to seal the external auditory canal. This is the principle behind the EAR CLIP. One wonders how to determine the NRR of this device.—JE

6,993,144

43.66.Yw INSERT EARPHONE ASSEMBLY FOR AUDIOMETRIC TESTING AND METHOD FOR MAKING SAME

Donald L. Wilson and Steven J. Iseberg, assignors to Etymotic Research, Incorporated
31 January 2006 (Class 381/380); filed 28 September 2000

A receiver (speaker) in a housing is driven by electrical signals from an audiometer and is acoustically coupled via a tube-nipple assembly to a flexible tube attached to a flexible eartip. An acoustic damper is located in



the inner portion of the tube nipple to cancel low-frequency resonance. An angled configuration of the tube nipple allows the housing to hang from the ear.—DAP

6,961,695

43.70.Jt GENERATING HOMOPHONIC NEOLOGISMS

Stephen Graham Copinger Lawrence, assignor to International Business Machines Corportion
1 November 2005 (Class 704/10); filed in the United Kingdom 26 July 2001

This mechanism (computer program) generates new words which sound the same as an existing word. For example, given ''was,'' the system is said to produce ''waz,'' ''woz,'' ''whaz,'' and ''whoz.'' An occurrence table allows the system to decide that ''woz'' would be the ''most likely'' spelling. There is no mention of why anyone would want to do this.—DLR

6,954,726

43.72.Ar METHOD AND DEVICE FOR ESTIMATING THE PITCH OF A SPEECH SIGNAL USING A BINARY SIGNAL

Cecilia Brandel and Henrik Johannisson, assignors to Telefonaktiebolaget L M Ericsson (Publ)
11 October 2005 (Class 704/217); filed in the European Patent Office 6 April 2000

This patent appears to disclose a completely standard run-of-the-mill LPC-autocorrelation pitch tracker, presented with a minimum of obscurative hand waving. The one possible novelty is that the well-known autocorrelation method is combined with an equally well known single-bit implementation of the correlation signal.—DLR

6,990,447

43.72.Ew METHOD AND APPARATUS FOR DENOISING AND DEVERBERATION USING VARIATIONAL INFERENCE AND STRONG SPEECH MODELS

Hagai Attias *et al.*, assignors to Microsoft Corportion
24 January 2006 (Class 704/240); filed 15 November 2001

The idea here is to move beyond the typical adaptive filtering scheme and somehow base denoising adaptation on what the speech is actually doing during the frame sequence. A probability distribution for speech model parameters (linear-predictive modeling is suggested) is used to identify a statistical distribution of denoised values for these parameters by adjusting the latter to improve a variational inference to better approximate the joint distribution of the speech model and the denoised values given the noisy signal. The variation of an ''improvement function'' is used to approximate this posterior joint probability because the perfect solution is intractable. It is suggested to employ this technique during the expectation step of an E-M model training run. Specific methods and equations are provided within.—SAF

6,993,480

43.72.Ew VOICE INTELLIGIBILITY ENHANCEMENT SYSTEM

Arnold I. Klayman, assignor to SRS Labs, Incorporated
31 January 2006 (Class 704/226); filed 3 November 1998

This patent provides an extensive amount of text and diagrams in the description of a system that is really rather simple. A speech enhancer is envisioned for public-address systems, etc., in which the signal is filtered through a transfer function that approximates the complement to the Fletcher-Munson curves, which quantify the frequency filtering that is performed by the average human auditory system. The signal is subject to fairly flat-response amplification when the signal volume is low, but moving more toward de-emphasizing the low and ultra-high frequencies as the signal volume gets louder in order to preserve the intelligible range while limiting the overall sound power to reduce ear strain.—SAF

6,996,524

43.72.Ew SPEECH ENHANCEMENT DEVICE

Ercan Ferit Gigi, assignor to Koninklijke Philips Electronics N.V.
7 February 2006 (Class 704/226); filed in the European Patent Office 9 April 2001

This patent suggests yet another technique (there have been a number of these in the past year) for performing frequency-domain noise reduction of speech. The technique is provided in some detail, yet the patent speaks for itself in saying ''the [background noise subtractor] is basically a frequency domain adaptive filter.'' Each successive signal frame is filtered, but since the filter characteristics may change, an overlap-add scheme is employed to prevent discontinuities at frame boundaries. Although the filter updating procedure is fully disclosed, how it is initialized (i.e., what is the background noise of the first frame) is unclear.—SAF

6,996,526

43.72.Fx METHOD AND APPARATUS FOR TRANSCRIBING SPEECH WHEN A PLURALITY OF SPEAKERS ARE PARTICIPATING

Sara H. Basson *et al.*, assignors to International Business Machines Corporation
7 February 2006 (Class 704/231); filed 2 January 2002

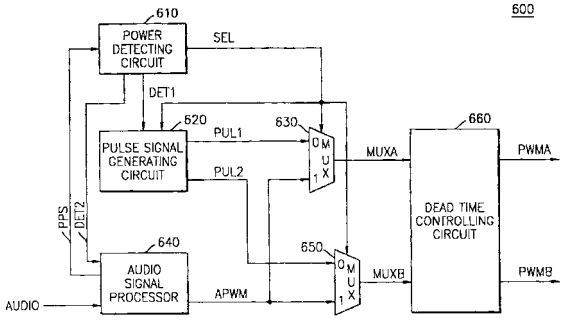
A method is described for applying standard speech recognition approaches in the service of transcribing a conference or other situation involving numerous speakers. For known speakers, a different speech-recognition system for each speaker is to be employed, each using an appropriate distinct speaker model. In one variation, speakers are identified in the signal using referenced prior-art techniques, and the correct speech-recognition system is then invoked. In another variation, all systems simultaneously decode the speech, and the one with the highest confidence score can be presumed to be that corresponding to whoever actually spoke.—SAF

6,987,418

43.72.Gy SOUND SIGNAL GENERATING APPARATUS AND METHOD FOR REDUCING POP NOISE

Il Joong Kim and Goog Chun Cho, assignors to Samsung Electronics Company, Limited
17 January 2006 (Class 330/10); filed in the Republic of Korea 2 May 2003

To prevent overshoots during power on and power off, transitions of a first pulse signal are delayed so as to not overlap with transitions of a second



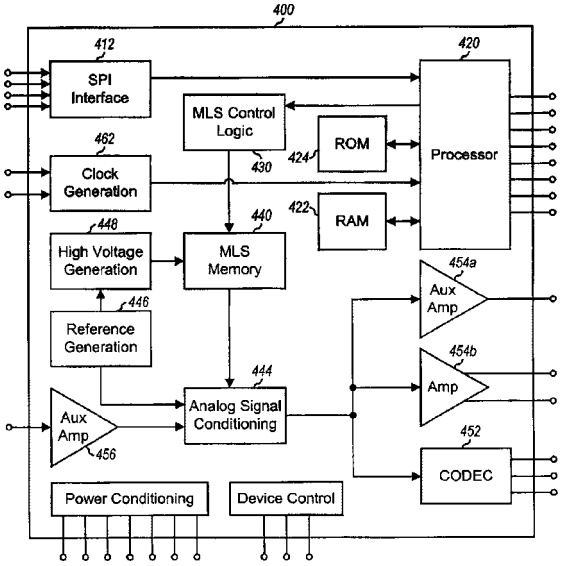
pulse signal. Reduced-width pulses are used in conjunction with a synchronized mute circuit and are the last pulse signals received by the pulse-width-modulating output amplifier prior to power turnoff.—DAP

6,959,279

43.72.Ja TEXT-TO-SPEECH CONVERSION SYSTEM ON AN INTEGRATED CIRCUIT

Geoffrey Bruce Jackson *et al.*, assignors to Winbond Electronics Corporation
25 October 2005 (Class 704/258); filed 26 March 2002

This patent describes a single-chip speech storage system originally patented by Information Storage Devices, later known simply as ISD. That company was later purchased by the present assignee, which, in turn, has developed a number of new applications for the storage technology. The



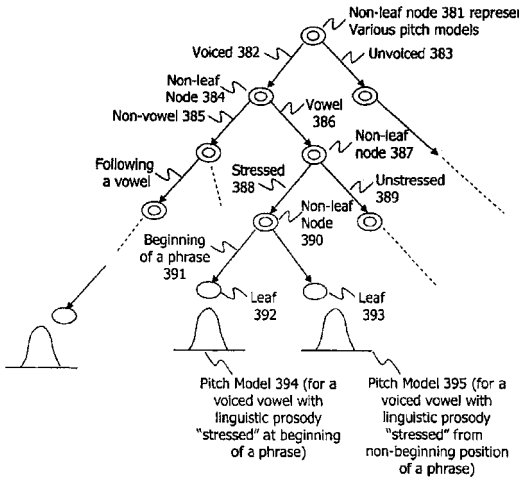
multilevel storage technology **MLS** consists of storing an analog speech sample in each memory cell of a flash or other memory system. Samples can be stored with an analog accuracy corresponding to roughly eight bits of resolution.—DLR

6,961,704

43.72.Ja LINGUISTIC PROSODIC MODEL-BASED TEXT TO SPEECH

Michael S. Phillips *et al.*, assignors to Speechworks International, Incorporated
1 November 2005 (Class 704/268); filed 31 January 2003

This is a prosody generator model intended for use with a speech synthesizer. In preparation, a marked training-data set is used to construct and catalog pitch, energy, and duration prototypes for various stress-marked



syntactic structures. During a symbolic markup phase of input text processing, syntactic structure is used to assign a single stress level to selected syllables. The parameter prototypes are then searched for the most nearly applicable fit and synthesis-parameter sequences are generated.—DLR

6,989,740

43.72.Ja ADVANCED AUDIO SAFETY APPARATUS

Joseph A. Tabe, Silver Spring, Maryland
24 January 2006 (Class 340/463); filed 13 February 2002

This patent suggests broadcasting a spoken warning from vehicles during dangerous situations, most typically truck reversing and school bus unloading. Going beyond the ubiquitous ''beep-beep'' of today, the recorded human voice will convey the message ''Attention! Please stand clear, this refuse truck is backing.'' Similarly, school buses will preferably intone: ''Please stop at 25 feet; this vehicle is coming to a complete stop.''—SAF

6,961,705

43.72.Ne INFORMATION PROCESSING APPARATUS, INFORMATION PROCESSING METHOD, AND STORAGE MEDIUM

Seiichi Aoyagi *et al.*, assignors to Sony Corporation
1 November 2005 (Class 704/275); filed in Japan 25 January 2000

Intended for Internet browser operation, and, more specifically, as a means to collect user information to produce more relevant search results, this system would recognize the user's speech, perform a linguistic analysis, update a user-information database, and consult a dialog-management unit to construct replies. The single independent claim requires that the system track and maintain records of the length of time during which the input contains references to particular topics. These records form the basis of the user-information database.—DLR

6,990,179

43.72.Ne SPEECH RECOGNITION METHOD OF AND SYSTEM FOR DETERMINING THE STATUS OF AN ANSWERED TELEPHONE DURING THE COURSE OF AN OUTBOUND TELEPHONE CALL

Lucas Merrow *et al.*, assignors to Eliza Corporation
24 January 2006 (Class 379/69); filed 31 August 2001

In order for service providers to be more efficient in contacting customers via automatic telephone calls, a speech recognition system determines if a live person answers the telephone, and, if so, if the target person has answered. If an answering machine or another person is detected, a prerecorded message is left for the target person.—DAP

6,996,519

43.72.Ne METHOD AND APPARATUS FOR PERFORMING RELATIONAL SPEECH RECOGNITION

Horacio E. Franco *et al.*, assignors to SRI International
7 February 2006 (Class 704/9); filed 28 September 2001

The proposal here has a hallmark of ingenuity, namely that it seems like too good an idea for everyone to have overlooked until now. The typical multipass speech-recognition scheme is altered by performing a database search with the initial pass results, with the goal of obtaining a variety of information about the indicated vocabulary domain. This information can then be used to select a more appropriate language model or acoustic models to be applied during subsequent recognition passes, thereby reducing the

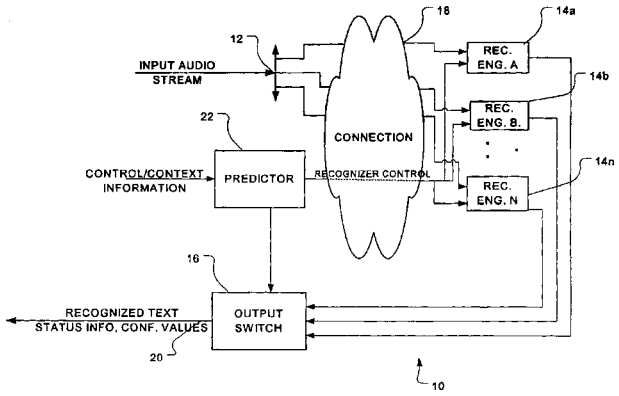
search space and, it is hoped, improving recognition accuracy (though no validation is reported). The examples include recognition of street addresses, where, if a city name is recognized on the first pass, subsequent search passes can then be restricted to models of the street names in that area.—SAF

6,996,525

43.72.Ne SELECTING ONE OF MULTIPLE SPEECH RECOGNIZERS IN A SYSTEM BASED ON PERFORMANCE PREDECTIONS RESULTING FROM EXPERIENCE

Steven M. Bennett and Andrew V. Anderson, assignors to Intel Corporation
7 February 2006 (Class 704/231); filed 15 June 2001

A method is proposed for selecting the most appropriate speech recognizer for an application (e.g., dictation) from multiple speech recognizers.



A predictor controls the routing of the input stream to the enabled recognizer. The predictor may use a recognizer confidence measure and computational requirements to select a recognizer and track performance over time.—DAP

6,990,446

43.72.Pf METHOD AND APPARATUS USING SPECTRAL ADDITION FOR SPEAKER RECOGNITION

Xuedong Huang and Michael D. Plumpe, assignors to Microsoft Corporation
24 January 2006 (Class 704/240); filed 10 October 2000

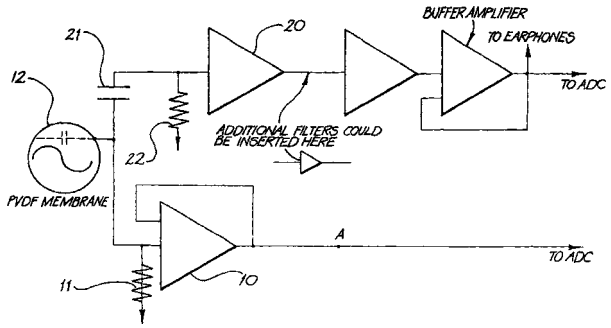
A problem for speaker recognition algorithms is often created by simple mismatches between training and test signal environments, in particular, different levels of background noise. This ingenious little idea proposes to avoid known problems with prior ''spectral subtraction'' attempts to match a test signal to a training signal, by instead matching the signals by adding to the mean and variance of multiple frequency components in both training and test signals. Since no spectral subtraction is ever performed, no important information can be lost.—SAF

6,988,993

43.80.Qf BIOPHYSICAL SENSOR

Colin Edward Sullivan and Ricardo Bianchi, assignors to Australian Centre for Advanced Medical Technology Limited
24 January 2006 (Class 600/528); filed in Australia 22 June 2000

This is a biophysical sound-detecting sensor designed to yield a sensed output and provide two or more separate signal processing paths for processing the sensed outputs into output signals. The output signals can, for



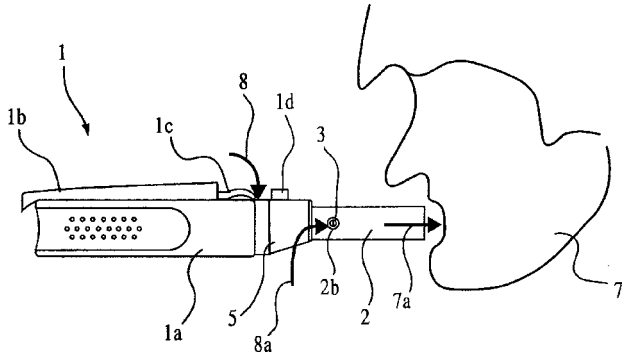
example, relate to specific frequency bands within the raw biological data. One embodiment of this device is an electronic stethoscope in which the sensing element is a PVDF (a polymer with piezoelectric properties) membrane. The signal processing entails an operational amplifier connected to the PVDF membrane, followed by a unity-gain buffer amplifier to allow a direct connection to headphones.—DRR

6,990,976

43.80.Qf ASTHMA DRUG INHALER WITH WHISTLE

Akihiko Miyamoto, Ibaraki, Japan
31 January 2006 (Class 128/200.23); filed in Japan 22 October 2002

A whistle is incorporated into an asthma drug inhaler to indicate that inhalation is properly done. The whistle is attached to a small opening for the air intake which may be part of a mouthpiece located on an inhalation



passage of a finely powdered drug. The whistle makes a sound when the inhalation is properly executed, but this may constitute an annoying factor when the inhaler is used in quiet surroundings.—DRR

6,986,740

43.80.Vj ULTRASOUND CONTRAST USING HALOGENATED XANTHENES

H. Craig Dees *et al.*, assignors to Xantech Pharmaceuticals, Incorporated
17 January 2006 (Class 600/458); filed 9 December 2002

The primary active component in this contrast agent is a halogenated xanthene such as Rose Bengal or a halogenated xanthene derivative such as a functional derivative of Rose Bengal.—RCW

6,988,990

43.80.Vj AUTOMATIC ANNOTATION FILLER SYSTEM AND METHOD FOR USE IN ULTRASOUND IMAGING

Lihong Pan *et al.*, assignors to General Electric Company
24 January 2006 (Class 600/437); filed 29 May 2003

An ultrasound image is annotated using a keyboard or a method of speech recognition to select from a vocabulary words that describe the anatomy in the image.—RCW

6,988,991

43.80.Vj THREE-DIMENSIONAL ULTRASOUND IMAGING METHOD AND APPARATUS USING LATERAL DISTANCE CORRELATION FUNCTION

Nam Chul Kim *et al.*, assignors to Medison Company, Limited
24 January 2006 (Class 600/443); filed in the Republic of Korea 11 May 2002

The distance is estimated between consecutive b-scan images that are obtained by manual scanning. Using the estimated distance, the sequence of images is converted into a three-dimensional format. The resulting three-dimensional image is displayed.—RCW

LETTERS TO THE EDITOR

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

Polymer acoustic matching layer for broadband ultrasonic applications (L)

Hironori Tohmyoh[a)]
Department of Nanomechanics, Tohoku University, Aoba 6-6-01, Aramaki, Aoba-ku, Sendai 980-8579, Japan

(Received 6 January 2006; revised 19 April 2006; accepted 21 April 2006)

A polymer acoustic matching layer has been designed that can act as a frequency filter between water and a test sample, and a method for controlling the high-frequency components of broadband ultrasound by using the designed layer is described. Acoustic imaging of a silicon-bonding sample using a very-low-scale matching layer fabricated from poly(vinylidene chloride) is demonstrated, in which the air gaps between the layer and the sample are evacuated. The experimental results show that the layer, which was designed for signal amplification, works as a broadband filter, as predicted, as well as offering protection against water for dry acoustic imaging. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2205127]

PACS number(s): 43.35.Sx, 43.35.Yb, 43.35.Zc [TDM] Pages: 31–34

I. INTRODUCTION

Acoustic microscopy/imaging has been a standard tool for the nondestructive detection and evaluation of defects in samples and for characterizing material.[1,2] The recent trend toward micro- and nano-technologies demands higher resolution and also cross-sectional images with higher signal-to-noise ratio in advanced samples, e.g., microelectronic packages.[3–5] Pushing the operating frequency of acoustic devices up to the gigahertz range, and also the realization of broadband ultrasound,[6] have been shown to be effective in improving spatial resolution, and sometimes the resolution that can be achieved is superior to that of an optical microscope.[7,8] However, the penetration depth of the high-frequency components of ultrasound is shallow, and the area that can be evaluated is limited to only the surface or the subsurface. Moreover, conventional acoustic microscopy/imaging has suffered from an inherent problem, i.e., immersion of the sample in water, which imposes a limitation on the samples that can be analyzed.

Several techniques, e.g., the capacitance transducer,[9,10] the electromagnetic acoustic transducer,[11] and laser ultrasound,[12,13] are able to generate and receive ultrasound without using a liquid medium, but these techniques are restricted in their ability to transmit high-frequency, broadband ultrasound. The development of dry techniques has recently been reported, in which the transduction of high-frequency ultrasound is accomplished via a solid layer such as a polymer or an elastomeric layer, and the acoustic imaging of electronic packages has been successfully achieved by these dry technologies.[14] Furthermore, it has been discovered experimentally that the signal intensity of ultrasound transmitted via a solid layer sometimes exceeds that achieved by water immersion due to the ultrasonic resonance that occurs between water, the layer, and the sample.[4,15] Desilets *et al.* have reported a design method for producing acoustic thin-disk transducers, which is based on the use of quarter-wave matching layers between a piezoelectric material and the acoustic load, and the validity of their method has been verified by the experiments using ultrasound of up to several MHz in frequency.[16] Today, this method is widely used to design piezoelectric transducers for medical imaging.

In this letter, the frequency characteristic of a very-low-scale polymer layer inserted between water and a test sample is derived theoretically around a resonance frequency for broadband high-frequency ultrasonic applications. An acoustic microscopy concept utilizing the ultrasonic resonance between water, the layer, and the sample is also shown, in which the designed polymer layer realizes signal amplification as an example of frequency control for broadband ultrasound, as well as offering protection against water. As expected, it was confirmed experimentally that the designed layer works well as a frequency filter over a wide frequency range between 20 and 100 MHz.

II. DESIGN OF THE ACOUSTIC MATCHING LAYER

Let us consider a transmission system comprised of water, the polymer layer, and the sample, where their acoustic

a)Electronic mail: tohmyoh@ism.mech.tohoku.ac.jp

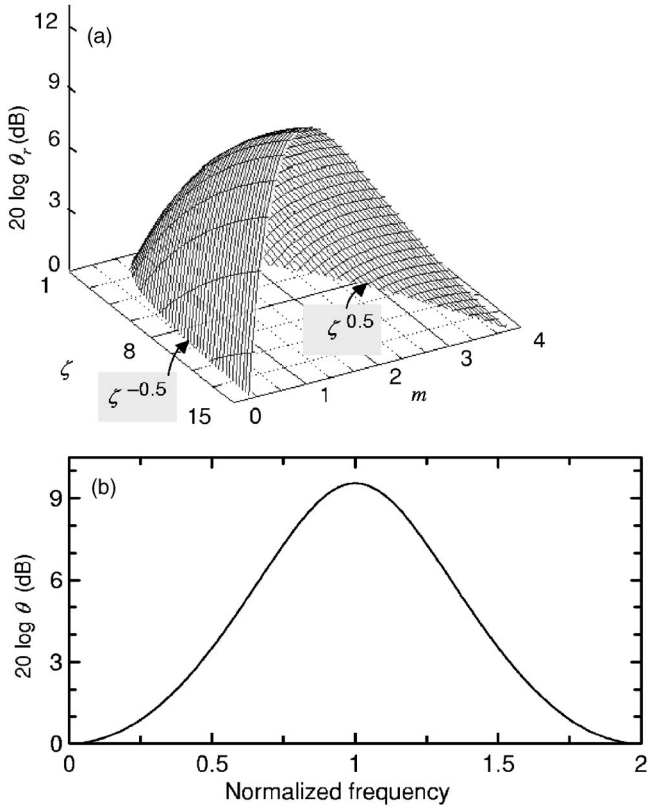


FIG. 1. Concept of the polymer acoustic matching layer. (a) Distribution of θ_r as a function of the dimensionless parameters ζ and m. (b) Behavior of θ in the case where $\theta_r=3.0$ and $h=1$.

impedances are denoted by Z_W(=1.51 MNm^{-3} s), Z_L, and Z_S, respectively. For plane acoustic waves of frequency ν that are normally incident on a planar layer, the echo transmittance of the three media, T_1, is given in Ref. 17, and it takes its maximum value at $Z_L=(Z_W Z_S)^{0.5}$, and at a layer thickness given by

$$d=(c_L/4\nu_r)h, \tag{1}$$

where c_L is the longitudinal wave velocity in the layer, ν_r is the resonance frequency, and h is an odd number. Here, d in the case where $h=1$ is well known as the condition for a quarter-wave matching layer.[16] By introducing a pair of dimensionless parameters $\zeta(=Z_S/Z_W)$ and $m[=Z_L/(Z_W Z_S)^{0.5}]$, the echo transmittance T_1 is found to be simply expressed by

$$T_1=T_2\times[\cos^2(2\pi\nu/c_L)d+A\sin^2(2\pi\nu/c_L)d]^{-1}, \tag{2}$$

where

$$A=[\zeta(1+m^2)^2]/[m(1+\zeta)]^2 \tag{3}$$

and where $T_2[=4\zeta/(1+\zeta)^2]$ is the echo transmittance in the case without the layer and is independent of ν. Therefore, at ν_r, the ratio $\theta(=T_1/T_2)$ has its maximum value $\theta_r(=A^{-1})$. The value of ν_r is governed by d and c_L, and the value of θ_r is determined by Z_W, Z_L, and Z_S. The distribution of θ_r as a function of ζ and m is shown in Fig. 1(a). The value of θ_r is greater than unity for $\zeta>1$ and $\zeta^{-0.5}<m<\zeta^{0.5}$, while at any value of m, the larger the value of ζ (which signifies a greater mismatch between Z_W and Z_S) the larger the value

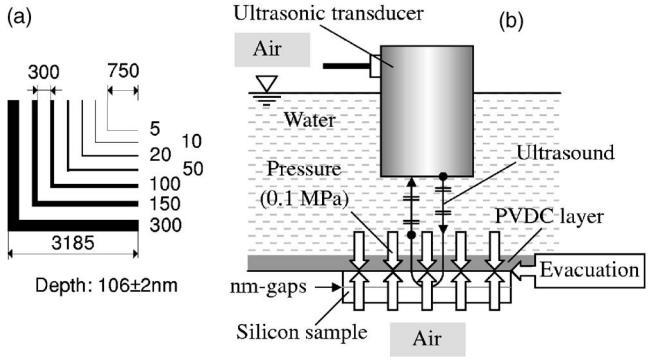


FIG. 2. Experimental details. (a) Nanometer gaps fabricated on a silicon [100] chip (units are in μm). (b) Schematic of acoustic imaging via the PVDC layer, where the air between the layer and the silicon sample is evacuated by a diaphragm-type vacuum pump.

of θ_r. For example, if we consider Plexiglass, with its low Z_S value of 3.28 MNm^{-3} s(ζ=2.2) and Si [100], with its high Z_S value of 20.74 MNm^{-3} s(ζ=13.7) as the test samples, the values of θ_r at $m=1$ for both samples become 1.2 and 3.9, respectively.

It can also be found that the behavior of θ can be expressed as

$$\theta=[\cos^2(0.5\pi h\nu/\nu_r)+\theta_r^{-1}\sin^2(0.5\pi h\nu/\nu_r)]^{-1}. \tag{4}$$

An example of the relationship between θ and the normalized frequency ν/ν_r in the case where $\theta_r=3.0$ and $h=1$ is shown in Fig. 1(b). The ratio θ is greater than unity for ν/ν_r between 0 and 2, and Eq. (4) is found to be a window function for the effective frequency range of $1-h^{-1}<\nu/\nu_r<1+h^{-1}$. The effective frequency ranges for $h=3$ and $h=5$ become $0.67<\nu/\nu_r<1.33$, and $0.8<\nu/\nu_r<1.2$, respectively. Therefore, for transmitting a broadband signal via a polymer layer, a value of d where $h=1$ is suitable.

III. EXPERIMENTAL ARRANGEMENT

The performance of the acoustic matching layer was examined by using a silicon sample with nanometer gaps. Two chips measuring $20\times20\times0.5$ mm^3 were cut from silicon [100] wafers. Gaps were patterned on one of the silicon chips by using photolithography and a fast atom beam etching technique, as shown in Fig. 2(a), where the depth of the gaps measured by a displacement meter was 106±2 nm. After all of the necessary etching on the silicon chip was completed, a test sample was fabricated by directly bonding the patterned and nonpatterned chips. The interface was physically bonded, and no reflective sources other than the nanometer gaps were introduced. This sample can provide us with valuable information, not only about spatial resolution, but also regarding the acoustic coupling at the dry interface between the polymer layer and the sample surface. A schematic of the acoustic imaging is illustrated in Fig. 2(b). A layer of poly(vinylidene chloride) (PVDC) for which $d=9$ μm was inserted between the water and the sample. The acoustical parameters of the PVDC layer were $Z_L=3.23$ MNm^{-3} s ($m=0.58$) and $c_L=1964$ m/s. Thus, the layer was expected to behave as a frequency filter between 0 and 109.1 MHz (θ_r

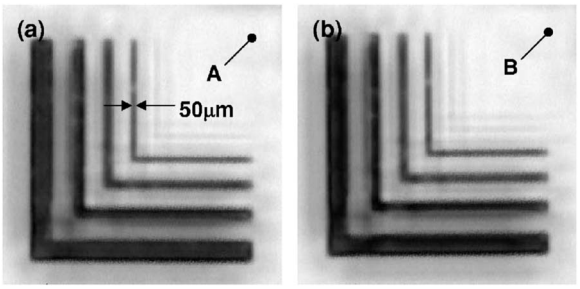


FIG. 3. Acoustic images obtained with (a) the PVDC-contact and (b) water immersion (4×4 mm).

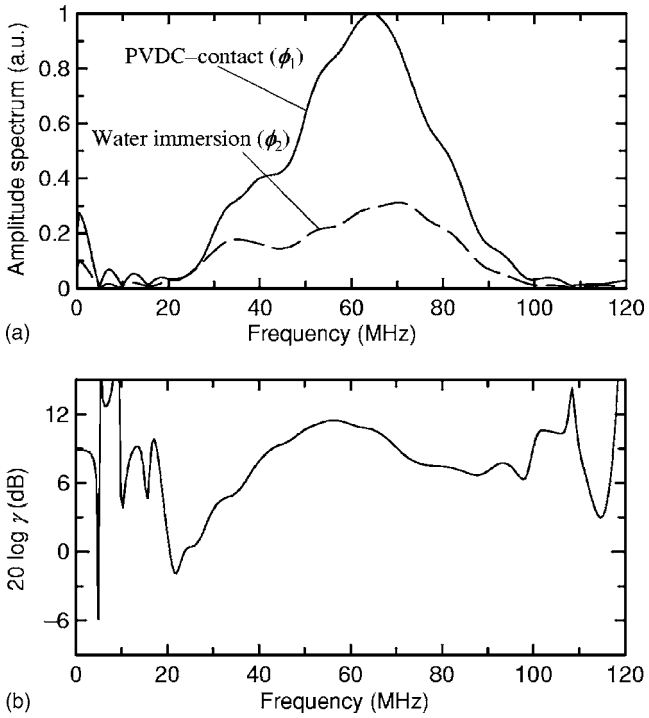


FIG. 4. (a) Amplitude spectra of the back-wall echoes of the sample with (ϕ_1) and without (ϕ_2) the PVDC layer, which were recorded at points A and B in Fig. 3. (b) Efficiency of the dry ultrasonic transmission $\gamma(=\phi_1/\phi_2)$.

=3.0, ν_r=54.6 MHz). A broadband ultrasonic transducer with a nominal frequency of 100 MHz, a piezoelectric element of 6.35-mm diameter, and a focal length of 12.7 mm was used, and the back-wall echoes of the sample with (ϕ_1) and without (ϕ_2) the PVDC layer were recorded. The surface roughness of the sample and the elastic deformation of the solid layer and the sample surface that accompany the ultrasonic transmission are liable to form air gaps at the contact interface.[14,18] To improve the acoustic coupling at the layer/sample interface, a pressure of about 0.1 MPa was applied to the interface between the layer and the sample by evacuating the air between them.[14]

IV. RESULTS AND DISCUSSIONS

Acoustic images of the sample recorded with the PVDC contact and with water immersion are shown in Figs. 3(a) and 3(b). Both images clearly show gaps that are more than 50 μm wide. In Fig. 3(a), no detail other than the nanometer gaps that were introduced can be observed, and there are no conspicuous differences between either of the acoustic images. These results verify that an acoustically coupled interface between the PVDC layer and the silicon sample can be successfully accomplished as long as there are no air gaps greater than 100 nm high and 50 μm wide at the interface.

The amplitude spectra ϕ_1 and ϕ_2, which were recorded at points A and B in Figs. 3(a) and 3(b), are shown in Fig. 4(a). Figure 4(a) clearly shows that $\phi_1 > \phi_2$ in the frequency range between 20 and 100 MHz. The peak frequencies of ϕ_1 and ϕ_2 are 64.5 and 70.3 MHz, and the −6-dB bandwidths of ϕ_1 and ϕ_2 are 33.2 and 37.1 MHz, respectively. These results suggest that the PVDC layer works as a frequency filter that effectively transmits high-frequency broadband ultrasound. The efficiency of ultrasonic transmission via such a layer in comparison with the case of immersion in water is expressed as $\gamma=\phi_1/\phi_2=\theta\xi\psi$, where ξ and ψ are related to the ultrasonic attenuation in the layer and the acoustic coupling at the layer/sample interface. The relationship between γ and ν is shown in Fig. 4(b). The maximum value of γ is 3.7 at 56.2 MHz, and these values are close to θ_r(=3.0) at ν_r(=54.6 MHz). The behavior of γ is also similar to that of θ [Fig. 1(b)] derived theoretically in the frequency range between 20 and 100 MHz, i.e., the performance of the layer as a frequency filter is predictable from Eq. (4). This indicates that low signal loss relating to the ultrasonic attenuation in the layer and good acoustic coupling at the layer/sample interface could be realized by utilizing a very thin PVDC layer and by the application of pressure to the layer/sample interface.

In this letter, an example of signal amplification of broadband ultrasound using a polymer acoustic matching layer, whose ν_r was close to the peak frequency of ϕ_2 and whose effective frequency range was close to that of ϕ_2, was demonstrated. The concept of acoustic microscopy via an acoustic matching layer allows us to achieve control over various frequency characteristics, and not only the signal amplification but also the signal modulation toward higher frequency components and widening the bandwidth are expected. For example, if the effective frequency range of θ is broader than that of ϕ_2, and the value of ν_r is higher than the peak frequency of ϕ_2, the higher frequency components of ϕ_2 will be amplified preferentially, and therefore the bandwidth of ϕ_1 will become wider than that of ϕ_2 without the layer.

V. CONCLUSIONS

The performance of a polymer layer as a frequency filter between a test sample and water was predicted, and the concept of acoustic microscopy via a polymer acoustic matching layer was shown. A very thin, poly(vinylidene chloride) layer worked well as a frequency filter in a wide frequency range of 20–100 MHz, where an acoustically coupled interface was formed between the layer and silicon by evacuating the air between them. The use of a polymer acoustic matching layer is therefore verified to be very effective for broadband ultrasonic applications under a dry environment.

ACKNOWLEDGMENTS

The author acknowledges Professor Masumi Saka for his constructive suggestions with respect to the drafting of this manuscript. This work was partly supported by the Ministry of Education, Culture, Sports, Science and Technology, Japan under Grant-in-Aid for Young Scientists (B) 17760072.

[1]Z. Yu and S. Boseck, "Scanning acoustic microscopy and its applications to material characterization," Rev. Mod. Phys. **67**, 863–891 (1995).
[2]L. Wang, "The contrast mechanism of bond defects with the scanning acoustic microscopy," J. Acoust. Soc. Am. **104**, 2750–2755 (1998).
[3]C. Miyasaka and B. R. Tittmann, "Recent advances in acoustic microscopy for nondestructive evaluation," J. Pressure Vessel Technol. **122**, 374–378 (2000).
[4]H. Tohmyoh and M. Saka, "Design and performance of a thin, solid layer for high-resolution, dry-contact acoustic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 423–438 (2004).
[5]G.-M. Zhang, D. M. Harvey, and D. R. Braden, "An improved acoustic microimaging technique with learning overcomplete representation," J. Acoust. Soc. Am. **118**, 3706–3720 (2005).
[6]J. Zhang, P. Guy, J. C. Baboux, and Y. Jayet, "Theoretical and experimental responses for a large-aperture broadband spherical transducer probing a liquid-solid boundary," J. Appl. Phys. **86**, 2825–2835 (1999).
[7]V. Jipson and C. F. Quate, "Acoustic microscopy at optical wavelength," Appl. Phys. Lett. **32**, 789–791 (1978).
[8]J. S. Foster and D. Rugar, "High resolution acoustic microscopy in superfluid helium," Appl. Phys. Lett. **42**, 869–871 (1983).
[9]W. Manthey, N. Kroemer, and V. Mágori, "Ultrasonic transducers and transducer arrays for applications in air," Meas. Sci. Technol. **3**, 249–261 (1992).
[10]D. W. Schindel, D. A. Hutchins, L. Zou, and M. Sayer, "The design and characterization of micromachined air-coupled capacitance transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 42–50 (1995).
[11]H. Ogi, M. Hirao, and T. Honda, "Ultrasonic attenuation and grain-size evaluation using electromagnetic acoustic resonance," J. Acoust. Soc. Am. **98**, 458–464 (1995).
[12]Y.-H. Liu, T.-T. Wu, and C.-K. Lee, "Application of narrow band laser ultrasonics to the nondestructive evaluation of thin bonding layers," J. Acoust. Soc. Am. **111**, 2638–2643 (2002).
[13]R. E. Green, Jr., "Non-contact ultrasonic techniques," Ultrasonics **42**, 9–16 (2004).
[14]H. Tohmyoh and M. Saka, "Dry-contact technique for high-resolution ultrasonic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 661–667 (2003).
[15]H. Tohmyoh and M. Saka, "Effective transmission of high frequency ultrasound into a silicon chip through a polymer layer," JSME Int. J., Ser. A **47**, 287–293 (2004).
[16]C. S. Desilets, J. D. Fraser, and G. S. Kino, "The design of efficient broad-band piezoelectric transducers," IEEE Trans. Sonics Ultrason. **SU-25**, 115–125 (1978).
[17]L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[18]B. Drinkwater, R. Dwyer-Joyce, and P. Cawley, "A study of the transmission of ultrasound across solid-rubber interfaces," J. Acoust. Soc. Am. **101**, 970–981 (1997).

Relationship between time reversal and linear equalization in digital communications (L)

W. J. Higley,[a] Philippe Roux, and W. A. Kuperman
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

(Received 16 August 2005; revised 5 May 2006; accepted 8 May 2006)

Iterative time reversal has been suggested as both an efficient method of creating a spatio-temporal focus and for use in telecommunications as a form of equalization. In this paper, the equivalence of a passive, i.e., via computation, iterative time reversal to the Moore-Penrose pseudo-inverse of the propagation matrix is shown. In the context of communications, however, any received signal is corrupted by noise. Therefore, a regularization term is introduced to the iterative equations, causing convergence to the canonical minimum mean-squared error linear equalizer. Hence, a relationship between time reversal and equalization is demonstrated. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2208458]

PACS number(s): 43.60.Dh, 43.60.Tj [DRD] Pages: 35–37

I. INTRODUCTION

In digital communications, the performance of a standard time reversal (TR) process has limitations because the reduction in intersymbol interference is related to a matched-filter process.[1–3] On the other hand, the equalization used in coherent communications is more closely related to inverse filtering. Though computationally more intensive than TR, the reduction in intersymbol interference is potentially much greater, particularly when the ratio of receiving elements to transmission elements is small, and equalization can therefore provide a better approach than standard TR. A recent paper by Montaldo *et al.*[4] has described an active iterative form of time reversal to achieve spatio-temporal focusing through a complex medium with greater intersymbol interference reduction than standard TR. Additionally, they have suggested that this could be used as an efficient method of equalization.[5] In certain environments, such as underwater communications, repeated propagation through the media as described in the papers is impractical. By performing the iteration passively, via computation, this problem is relieved, at the expense of the additional computation. In this paper, a relationship between TR and equalization is derived by showing that the inclusion of a regularization term in the passive iterative TR process is identical to the minimum mean-square error linear equalizer (MMSE-LE). This allows for a physical insight into the MMSE equalizer.

II. DESCRIPTION OF ITERATIVE TIME REVERSAL

One goal of communications is to have multiple sources transmit to an array of receivers. Such is the case in underwater acoustic communications with a network of autonomous underwater vehicles (AUVs), where each source corresponds to a different user, or in some forms of array-to-array communications. The goal of passive iterative time reversal is to create a set of filter banks that equalize the received signals, such that the combined impulse response of the channel and each filter bank is a spatio-temporal Kronecker delta function corresponding to each source.

The propagation between each transmitter and receiver element is described by the set of impulse responses $h_{ij}(t)$, $i=1,2,\ldots,N_R$, and $j=1,2,\ldots,N_T$. For example, if the signal sent from each transmitter is the time-dependant signal $x_j(t)$, the received signals on the array, in the absence of noise, are

$$y_i(t) = \sum_j h_{ij}(t) \otimes x_j(t), \tag{1}$$

where $\otimes$ indicates convolution. Equivalently, in the frequency domain, one may write

$$Y_i(\omega) = \sum_j H_{ij}(\omega) X_j(\omega), \tag{2}$$

where capitalization indicates the Fourier transform and ω is frequency. Writing this in matrix notation yields

$$\mathbf{Y}(\omega) = \mathbf{H}(\omega)\mathbf{X}(\omega). \tag{3}$$

When the transmitters wish to send information, they first send a known function followed by the communications sequences, $\mathbf{X}(\omega)$. The receiver array is able to extract a noisy estimate of the unitless channel transfer functions, $\mathbf{H}(\omega)$, from the known part of the signals. The second part of the received signals are the communications sequences convolved with the transfer functions, $\mathbf{H}(\omega)\mathbf{X}(\omega)$. The transfer functions extracted from the first part of the signal are time reversed to initialize the filter for the first iteration, designated $\mathbf{F}_n$, where the subscript indicates iteration number,

$$\mathbf{F}_1(\omega) = \mathbf{H}^H(\omega). \tag{4}$$

The iterative process begins by passively propagating the filter impulse responses back to the transmitters. The term "passively propagating," in this case, is taken to mean replicating, via computation on a computer, the result of physically transmitting the filter impulses and measuring the field back at the original transmitters. This is analytically

[a]Author to whom correspondence should be addressed. Electronic mail: bhigley@mpl.ucsd.edu

equivalent to an active time-reversal process when noise is not considered and reciprocity is valid.[6] This results in a combined channel/filter impulse expressed by the following equation, where superscript H indicates conjugate-transpose:

$$\mathbf{R}_1(\omega) = \mathbf{H}(\omega)\mathbf{H}^H(\omega). \tag{5}$$

Often, particularly with small arrays, this results in temporal sidelobes that act as intersymbol interference and degrade the performance of communications systems. The next step in the iterative process is to subtract this result from the objective delta functions, which are constant in frequency, yielding a difference term expressible as

$$\mathbf{D}_1(\omega) = \mathbf{I} - \mathbf{H}(\omega)\mathbf{H}^H(\omega). \tag{6}$$

The filter impulse responses are then updated by adding to them the difference term convolved with the time-reversed transfer functions previously obtained. As written in Ref. 4, the iterative procedure is the set of equations below, where the frequency dependence has been suppressed for clarity:

$$\begin{aligned} &\mathbf{F}_0 = \mathbf{0}, \\ &\mathbf{R}_n = \mathbf{H}\mathbf{F}_n, \\ &\mathbf{D}_n = \mathbf{I} - \mathbf{R}_n, \\ &\mathbf{F}_{n+1} = \mathbf{F}_n + \mathbf{H}^H\mathbf{D}_n. \end{aligned} \tag{7}$$

The difference term of the nth iteration can be shown to be

$$\mathbf{D}_n = (\mathbf{I} - \mathbf{H}\mathbf{H}^H)^n, \tag{8}$$

which causes the filter responses of the following iteration to be equal to

$$\mathbf{F}_{n+1} = \mathbf{H}^H \sum_{k=0}^{n} (\mathbf{I} - \mathbf{H}\mathbf{H}^H)^k. \tag{9}$$

The summation term in the above equation can be recognized as the Neumann expansion[7] of the matrix inverse, which states

$$\mathbf{A}^{-1} = \sum_{k=0}^{\infty} (\mathbf{I} - \mathbf{A})^k, \tag{10}$$

given that the norm of $(\mathbf{I}-\mathbf{A})$ is less than one. After many iterations, the filter responses converge to

$$\mathbf{F} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}. \tag{11}$$

This is recognized as the Moore-Penrose pseudo-inverse of the propagation matrix, $\mathbf{H}$. Finally, this filter set is applied to the received communications sequences, $\mathbf{Y}$, and the signals, $\mathbf{X}$, decoded. The problem of the estimate of the transfer function matrix, $\mathbf{H}$, being noisy is lessened in the case of active iterative time reversal compared to the passive case, as each iteration introduces a different realization of the noise process. However, active iteration is impractical in certain circumstances, such as during an at-sea experiment.[2] Additionally, in the context of communications, additive noise dominates channel estimation error, therefore the transfer functions estimates are usually assumed to be the true transfer functions. A more appropriate goal of passive time reversal would be to create a set of filter impulse responses that minimize the mean-squared error of the received communications sequences. The impulse responses that achieve this goal under a white-noise assumption are governed by the well-know minimum mean-squared error linear equalizer (MMSE-LE) expression

$$\mathbf{F} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I})^{-1}, \tag{12}$$

where σ^2 is the inverse of the signal-to-noise ratio (SNR), calculated as the ratio of the power transmitted from each transmitter, P, and the noise power, N_0, received at a single receiver.

It is possible, through the addition of a regularization term, to alter the iterative procedure of Eq. (7) so that it converges to the MMSE-LE equation stated above in Eq. (12). After this modification, the iterative procedure is written as

$$\begin{aligned} &\mathbf{F}_0 = \mathbf{0}, \\ &\mathbf{R}_n = \mathbf{H}\mathbf{F}_n, \\ &\mathbf{D}_n = \mathbf{I} - \mathbf{R}_n - \sigma^2 \sum_{k=0}^{n-1} \mathbf{D}_k, \\ &\mathbf{F}_{n+1} = \mathbf{F}_n + \mathbf{H}^H\mathbf{D}_n. \end{aligned} \tag{13}$$

The only difference is the addition of a regularization term in the third iteration equation. This can be recognized as a form of gradient descent solution, similar to the conjugate gradient method,[7] to finding the MMSE-LE filter impulse responses. Again, once a number of iterations have been performed, the filter set is applied to the received communications sequences and the signals decoded.

The MMSE-LE has advantages over the inverse filter relating to its performance in a noisy environment. By minimizing the mean-squared error of the communication sequence, it also maximizes the signal-to-interference-plus-noise ratio (SINR). Additionally, convergence is monotonic, as visible in Fig. 1, meaning there is no "best number" of iterations, such as was the case in Ref. 5. Recent results also seem to indicate that the MMSE-LE is more robust to channel estimation error.[8]

Shown in Fig. (1) is a demonstration of the convergence of iterative time-reversal output signal-to-interference-plus-noise to the optimum output SINR of the MMSE-LE. Figure 1(a) shows the output SINR as a function of number of iterations for a sample set of five measured impulse responses,[9] which are shown (each shifted in both time and amplitude for visualization purposes) in Fig. 1(b). The four pairs of curves represent four input SNRs each separated by 5 dB. After each iteration, the combined channel and filter response is converted to the time domain where SINR calculations are done. The solid lines show the output SINR of modified iterative time-reversal, whereas nonmodified iterative time reversal is shown with dashed curves. Also, the MMSE-LE filter is calculated explicitly, using the formula of Eq. (12), and the combined channel and filter response con-

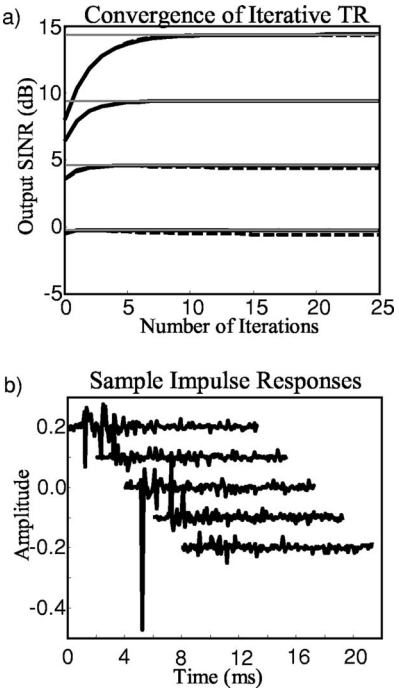


FIG. 1. Convergence of iterative time reversal. (a) The output signal-to-interference-plus-noise ratio (SINR) is shown for four different noise powers (input SNRs), each separated by 5 dB, as a function of number of iterations. The dashed lines correspond to iterative time reversal and the solid lines correspond to the modified iterative time reversal. The optimum SINR calculated explicitly with the MMSE-LE formula is shown in gray. (b) The channel used for calculation is a measured at-sea transfer function set with one input, five outputs, and 100 taps for each of the five transfer functions.

verted to the time domain where SINR calculations are done. The results are shown as a thin gray line for each noise power.

Through analysis of the iterative process, one can see the MMSE-LE as a filter set that attempts to cancel its own sidelobes in the time domain, but is regularized that the gain in the frequency domain is not too large, taking into account the fact that the communications sequence is noisy.

III. CONVERGENCE

The modified iterative time reversal procedure converges so long as the Neumann expansion of the matrix inverse is valid, that the norm of $(\mathbf{I}-\mathbf{H}\mathbf{H}^H-\sigma^2\mathbf{I})$ is less than one. This is not a restrictive constraint, as the received signal, $\mathbf{Y}$, can be multiplied by a constant to ensure this condition is met without loss of optimality, in the sense of maximizing SINR. Scaling the received signal, $\mathbf{Y}$, is equivalent to scaling both the transfer function matrix, $\mathbf{H}$, and the noise, thus leaving the SINR unchanged.

The speed of convergence is determined, as in many gradient methods, by the eigenvalue spread, in frequency, of the matrix $\mathbf{H}\mathbf{H}^H+\sigma^2\mathbf{I}$. The larger the spread, the longer the iterative algorithm takes to converge. As can be seen in Fig. 1, convergence occurs with less iteration at lower signal-to-noise ratios because the eigenvalue spread is smaller, as they are dominated by the constant noise components.

IV. CONCLUSION

It has been shown that iterative time reversal can be performed passively, resulting in a procedure that converges to the Moore-Penrose pseudo-inverse of the propagation matrix. More importantly, it has been shown that a minor modification, the inclusion of a regularization term, alters the procedure so that it converges to the MMSE linear equalizer. Thus, one can view the MMSE equalizer equivalently as a regularized iterative time reversal process. The iteration systematically reduces temporal sidelobes and the regularization limits the amplification in the frequency domain preventing noisy channels being included in the signal estimate.

ACKNOWLEDGMENTS

The authors wish to acknowledge helpful and insightful discussions with Bruce Cornuelle. This research was supported by the Office of Naval Research.

[1]T. C. Yang, "Temporal resolution of time-reversal and passive-phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**, 229–245 (2003).
[2]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **98**, 25–40 (1998).
[3]G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kuperman, and H. C. Song, "An initial demonstration of underwater acoustic communication using time reversal," IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[4]G. Montaldo, M. Tanter, and M. Fink, "Real time inverse filter focusing through iterative time reversal," J. Acoust. Soc. Am. **115**, 768–775 (2004).
[5]G. Montaldo, G. Lerosey, A. Derode, A. Tourin, J. de Rosny, and M. Fink, "Telecommunication in a disordered environment with iterative time reversal," Waves Random Media **14**, 287–302 (2004).
[6]P. Roux, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, and M. Stevenson, "A nonreciprocal implementation of time reversal in the ocean," J. Acoust. Soc. Am. **116**, 1009–1015 (2004).
[7]T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 2000).
[8]M. Ding, B. L. Evans, and I. C. Wong, "Effect of channel estimation error on bit rate performance of time domain equalizers," *Proc. IEEE Asilomar Conference on Signals, Systems and Computers, Vol. 2*, Pacific Grove, CA, Nov. 7–10, 2004, pp. 2056–2060.
[9]H. C. Song, W. S. Hodgkiss, W. A. Kuperman, M. Stevenson, and T. Akal, "Improvement of time reversal communications using adaptive channel equalizers," IEEE J. Ocean. Eng. in press (2006).

Vector intensity field scattered by a rigid prolate spheroid

Brian R. Rapids[a] and Gerald C. Lauchle
Graduate Program in Acoustics, The Pennsylvania State University, 217 Applied Science Building, University Park, Pennsylvania 16802

(Received 25 February 2005; revised 4 April 2006; accepted 27 April 2006)

The short-wavelength, steady-state vector intensity field scattered by a rigid prolate spheroid (10:1 fineness ratio) is investigated analytically. The intensity field is the product of the scalar acoustic pressure and the acoustic particle velocity fields which are computed from the Helmholtz equation in prolate spheroidal coordinates. Particular emphasis is placed on results in the forward-scattered direction. It is found that the equivalent plane wave intensity varies by less than ±0.5 dB (relative to the incident acoustic intensity) in the far-forward-scattered direction with the selected parameters. This illustrates that the forward-scattered pressure is masked by and interferes with the incident wave. The reactive intensity in the forward-scattered direction is found to be −25 dB relative to the incident active intensity, and the computed phase difference between pressure and particle velocity is 4° at ranges approaching 10 spheroid lengths at a reduced frequency of $h=41.7$; in the absence of the spheroid, the phase difference due to the incident plane wave alone is exactly 0°. The study demonstrates that unique information, extending beyond direction-of-arrival estimation, can be derived from the total acoustic intensity field and may allow inferences to be made regarding the presence or absence of the spheroid. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2206514]

PACS number(s): 43.20.Fn, 43.30.Vh, 43.30.Zk [WMC] Pages: 38–48

I. INTRODUCTION

The scattering of a scalar wave by an obstacle (having ideal boundary conditions) is a classic problem in theoretical physics. The significance of forward scattering at high frequencies has been discussed by many investigators[1–7] due to its application in imaging, radar, and sonar systems. For example, when light is obstructed by a circular disk, it is in the forward-scatter region where the Poisson spot can be observed and the classic Airy diffraction pattern is created. In atmospheric optics, forward scattering from small water droplets causes the colored rings around the moon (lunar corona). It is also the forward scatter from a sphere in the high-frequency limit that gives rise to the extinction paradox, in which the total scattering cross section of a sphere is twice the geometrical cross section of the sphere.[8] Previous investigations into the forward-scattering of waves by objects emphasize calculations—or measurements—of the scalar field magnitude. In acoustics, this amounts to computing or measuring the acoustic pressure field in the far-field region of the scattering body.

The objective of the research reported in this paper is to determine if the acoustic vector intensity field provides new or additional information regarding the presence of an insonified scatterer in the forward-scattered direction over that information obtained with scalar pressure sensors alone. The use of vector sensors in underwater acoustics has been limited primarily to applications that increase array gain by leveraging upon the directionality of the sensors,[9–11] or their use as passive sensors in surveillance and measurement of ambient acoustic noise.[12–14] Researchers in the field of noise control engineering have been making use of metrics derived from acoustic intensity measurements in air to identify abnormalities of a sound field[15] for more than a decade. The research presented here expands upon these areas by investigating the perturbation in the physical parameters of the *total acoustic vector intensity field* due to a scattering body. The theoretical analysis involves analytic computations of the scalar acoustic pressure and the acoustic particle velocity vector in order to construct the total acoustic intensity vector field. These theoretical results serve as a basis for an underwater scattering experiment which has been discussed elsewhere;[16,17] the results of that experiment will be published separately.

II. CONCEPTS IN ACOUSTIC INTENSITY

Fundamental relationships in acoustic intensity will be presented in this section, while the reader is referred to other references[18–22] for derivations which need not be reproduced here. These relationships emphasize concepts that will be employed in Sec. III to discuss the perturbation of the acoustic intensity due to the presence of a scattering body. An $e^{-i\omega t}$ time convention is used in complex notation where ω is the radian frequency and t represents time. The expressions for the components of acoustic intensity are derived using an arbitrary pressure and particle velocity to define a power factor angle as the phase difference between pressure and particle velocity. Second, the same intensity expressions are developed using an arbitrary pressure field and the linearized Euler equation to show how the active and reactive intensities are related to the mean-squared pressure field and the gradient of the mean-square pressure field, respectively.

[a]Current address: Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Rd., Laurel, MD 20723. Electronic mail: Brian.Rapids@jhuapl.edu

Lastly, the concept of complex acoustic intensity is developed and used to relate the power factor angle to the active and reactive intensities.

The derivation begins with the complex field variables for acoustic pressure and acoustic particle velocity, respectively,

$$p(\mathbf{r},t) = Pe^{i(\varphi_p(\mathbf{r})-\omega t)}, \tag{1}$$

$$\mathbf{u}(\mathbf{r},t) = \begin{bmatrix} U_i e^{i\varphi_{u_i}(\mathbf{r})}\hat{i} \\ U_j e^{i\varphi_{u_j}(\mathbf{r})}\hat{j} \\ U_k e^{i\varphi_{u_k}(\mathbf{r})}\hat{k} \end{bmatrix} e^{-i\omega t}. \tag{2}$$

The instantaneous intensity is the product of the real parts of the complex pressure and velocity,

$$\mathbf{I}_i(\mathbf{r},t) = \mathrm{Re}\{p(\mathbf{r},t)\}\mathrm{Re}\{\mathbf{u}(\mathbf{r},t)\}. \tag{3}$$

The rate of change of total energy in a volume is equal to the acoustic power flowing across the surface of the volume. Acoustic power per unit area is the instantaneous intensity. Substituting Eq.(1) and (2) into Eq.(3) results in a form of the instantaneous intensity as the sum of an in-phase and a quadrature component,

$$\mathbf{I}_i = \frac{P}{2}\begin{bmatrix} U_i \cos\varphi_{pu_i}(1+\cos(2\varphi_p - 2\omega t))\hat{i} \\ U_j \cos\varphi_{pu_j}(1+\cos(2\varphi_p - 2\omega t))\hat{j} \\ U_k \cos\varphi_{pu_k}(1+\cos(2\varphi_p - 2\omega t))\hat{k} \end{bmatrix} + \frac{P}{2}\begin{bmatrix} U_i \sin\varphi_{pu_i}\sin(2\varphi_p - 2\omega t)\hat{i} \\ U_j \sin\varphi_{pu_j}\sin(2\varphi_p - 2\omega t)\hat{j} \\ U_k \sin\varphi_{pu_k}\sin(2\varphi_p - 2\omega t)\hat{k} \end{bmatrix}. \tag{4}$$

For an arbitrary unit vector $\hat{n}$, two new variables are defined: $\varphi_{pu_n} = \varphi_p - \varphi_{u_n}$ and $\gamma = 2\varphi_p - 2\omega t$, which allow Eq. (4) to be rewritten more compactly as

$$\mathbf{I}_i = \mathbf{I}(1+\cos\gamma) + \mathbf{Q}\sin\gamma. \tag{5}$$

Equation (5) presents the instantaneous intensity as the sum of an instantaneous active intensity, $I\hat{n} = \frac{1}{4}PU_n \cos\varphi_{pu_n}\hat{n}$, and instantaneous reactive intensity, $Q\hat{n} = \frac{1}{4}PU_n \sin\varphi_{pu_n}\hat{n}$.

The phase difference between pressure and acoustic particle velocity will be referred to as the power factor angle. If the pressure and velocity oscillations are in-phase, then the power factor angle is zero and the time-averaged intensity is maximized (e.g., the case of a plane wave). Conversely, if the two oscillations are in quadrature, then the power factor angle is 90° and the time-averaged intensity is zero (e.g., the case of a standing wave). The reference of power factor angle has been adopted from the variable of the same name used for power calculations associated with steady-state voltages and currents of circuits driven by sinusoidal sources.[19,20]

An alternative form of the instantaneous intensity $\mathbf{I}_i$ can be derived by first substituting Eq. (1) into the linearized Euler equation,

$$\frac{\partial \mathbf{u}(\mathbf{r},t)}{\partial t} = -\frac{1}{\rho}\nabla p, \tag{6}$$

to obtain the particle velocity in terms of the pressure,

$$\mathbf{u}(\mathbf{r},t) = \frac{-i}{\omega\rho}[P\nabla\varphi_p - i\nabla P]e^{i(\varphi_p-\omega t)}. \tag{7}$$

The direction of the instantaneous velocity vector is governed by two gradient operators and is affected by the local spatial variations of the phase and amplitude of the pressure wave. The second form of the instantaneous intensity is thus obtained by substituting Eqs. (1) and (7) into (3),

$$\mathbf{I}_i = \frac{P^2\nabla\varphi_p}{2\omega\rho}(1+\cos\gamma) + \frac{\nabla P^2}{4\omega\rho}\sin\gamma. \tag{8}$$

Comparison of Eq. (8) with Eq. (5) reveals that the variables $\mathbf{I}$ and $\mathbf{Q}$, which contain the spatial dependencies of the instantaneous active and reactive intensities, can be alternatively described by

$$\mathbf{I}(\mathbf{r}) = \frac{P^2\nabla\varphi_p}{2\omega\rho} \tag{9}$$

$$\mathbf{Q}(\mathbf{r}) = \frac{\nabla P^2}{4\omega\rho}. \tag{10}$$

It can be seen that the direction (and to some degree the magnitude) of the active intensity is governed by the gradient of the pressure phase which is parallel to the wave vector, $\mathbf{k}$, of the harmonic wave component. Conversely, the reactive intensity vector is governed by the gradient of the squared pressure amplitude.

The separation of time and spatial dependencies allows for a third form of the instantaneous acoustic intensity to be in the form of a complex intensity, $\mathbf{I}_c = \mathbf{I} + i\mathbf{Q}$. We find

$$\mathbf{I}_i = \mathrm{Re}\{\mathbf{I}_c(1+e^{-i\gamma})\}. \tag{11}$$

The complex intensity is easily shown to be the product of the acoustic pressure and the complex conjugate of particle velocity,

$$\mathbf{I}_c = \frac{p\mathbf{u}^*}{2}. \tag{12}$$

Because the complex intensity is a complex-valued function, any of its components in the $\hat{n}$ direction may be represented as a phasor having a magnitude equal to the envelope of $I_{c,n}\hat{n}$ and a phase angle determined by I_n and Q_n according to

$$I_{c,n} = \sqrt{|I_n|^2 + |Q_n|^2}e^{i\varphi_{pu_n}}, \quad \text{where } \varphi_{pu_n} = \tan^{-1}\left(\frac{Q_n}{I_n}\right). \tag{13}$$

The phase angle associated with $I_{c,n}\hat{n}$ is the phase difference between pressure and particle velocity (i.e., the power factor angle).

The average flow of acoustic energy per unit area (active acoustic intensity) is equal to the product of the magnitudes of the pressure and particle velocities weighted by $\cos\varphi_{pu}$.[21,22] The power factor angle of the acoustic field can-

not be measured with a single scalar pressure sensor, except in the very special case of a plane wave, where $\varphi_{pu}=0$. Because of this, a plane-wave field contains no reactive intensity. However, if a scattering body perturbs the plane-wave field, then neither the reactive intensity nor the power factor angle should be identically zero in the vicinity of those perturbations. A reliable estimation of φ_{pu} (or the reactive intensity) should therefore allow an inference to be made regarding the presence of a scattering body. Section III develops these ideas further by investigating the spatial gradients in the acoustic pressure field established by the scattering of a plane wave by a rigid prolate spheroid and computing various acoustic intensity quantities.

III. SCATTERING FROM A RIGID PROLATE SPHEROID

A. Overview

The prolate spheroid is an ellipsoid of revolution whose axis of symmetry coincides with the major axis of the underlying ellipse. The prolate spheroid degenerates to a sphere as the eccentricity approaches zero ($e\rightarrow 0$), or it degenerates to an infinitely thin line segment of length equal to the interfocal distance as $e\rightarrow 1$. The importance of the prolate spheroid comes from the variety of scattering shapes attainable by the spheroid, coupled with the fact that the prolate spheroidal coordinate system is one of 11 that permits a separable solution to the scalar Helmholtz equation. The analytical method for generating an exact solution to the scattering problem is based on the partial wave expansion of the incident wave in terms of basis functions and the adoption of classical boundary conditions (Neumann, Dirichlet, or Robin). The basis functions for the azimuthal angle and radial portions of the separable solution both employ spheroidal harmonics. Computational difficulty arises when attempting to evaluate these basis functions.[23–26] As a result, analytical investigations have been limited to expanding the spheroidal harmonics in terms of the more familiar spherical harmonics and constraining the analyses to a high or low frequency limit and observation points in the far field of the scattering body.[4,27–30] Recently, new techniques have been developed[31,32] to evaluate the prolate spheroidal radial functions (and their derivatives) over a wide range of parameters with sufficient precision and accuracy for rigorous studies at high frequencies. The derivatives of the radial functions are necessary for satisfying the rigid boundary condition and for computation of the acoustic particle velocity in the field. These techniques have enabled the straightforward evaluation of the three-dimensional field scattered from a prolate spheroid in both the near and far field of the scatterer over a wide range of incident frequencies and plane-wave incidence angles.

By superposition, the total acoustic field observed when a scattering body is placed in an incident wave field is the sum of the incident and the scattered fields. Computation of the total pressure and particle velocity fields enables evaluation of both the active (convective) and reactive (nonconvective) components of the instantaneous intensity at all points around the scattering body. The scattering problem in the high-frequency (or short wavelength) limit is of particular interest because an acoustic shadow is formed on the opposite side of the body from insonification.[2,8] This shadow formation is a result of constructive and destructive interference. In fact, the strong forward scattering by the prolate spheroid in the high-frequency limit of geometrical acoustics is what gives rise to the classic extinction paradox for the degenerate case of a sphere, i.e., the total scattering cross section is twice the geometrical cross section of the sphere.[8] It is hypothesized that in these regions of interference, spatial gradients exist in the pressure field which will give rise to a nonzero power factor angle and reactive intensity according to the theory presented in Sec. II.

B. Prolate spheroidal coordinate system

Rotation of the two-dimensional elliptical coordinate system about the major axis yields the prolate spheroidal coordinate system with natural coordinates of (ξ,η,φ). The ξ coordinate $(1\leqslant\xi\leqslant\infty)$ is the radial coordinate that specifies concentric ellipsoidal surfaces. The η coordinate $(-1\leqslant\eta\leqslant 1)$ is often called the angular coordinate due to the angle, $\cos^{-1}\eta$, with which it intersects the z axis. The φ coordinate $(0\leqslant\varphi<2\pi)$ is the rotational coordinate that specifies a unique plane stemming from the z axis. This study is specifically interested in the intersection of this coordinate system with the x-z plane shown in Fig. 1. The φ coordinate is constrained to zero while $\hat{\varphi}$ is perpendicular to this plane.

Any prolate spheroid is constructed by specifying a constant $\xi=\xi_0$, which equivalently specifies the eccentricity $e=\xi_0^{-1}$ as well as the fineness (length-to-diameter) ratio $\varepsilon=L/2R=\xi_0/(\xi_0^2-1)^{1/2}$. The interfocal distance of the spheroid is given by a. In this paper, ξ_0 and ε are used to reference the surface of the rigid spheroid having a major axis of length L_0, while ξ and L/L_0 are used to denote the radial coordinate at which the various quantities are evaluated.

C. Scattered pressure field

Consider a monochromatic plane wave in a lossless, unbounded medium described by $p(\mathbf{r},t)=P_0e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}=Pe^{-i\omega t}$ that is incident upon prolate spheroid ξ_0 from an arbitrary angle defined by $\cos^{-1}\eta_{\text{inc}},\varphi_{\text{inc}}$. The medium has bulk sound speed c and bulk density ρ. The wave vector of the incident plane wave, $\mathbf{k}$, is related to the sound speed and radian frequency by $k=|\mathbf{k}|=\omega/c$. Substitution of the time-harmonic pressure field into the wave equation results in the Helmholtz equation. Its separable solution can be written in terms of the eigenfunction

$$P_{mn}=R_{mn}^{(j)}(h,\xi)S_{mn}(h,\eta)\begin{bmatrix}\cos m\varphi\\ \sin m\varphi\end{bmatrix},$$

where $h=ka/2$. The function $S_{mn}(h,\eta)$ is the prolate spheroidal angle function of the first kind of order m and degree n. The function $R_{mn}^{(j)}(h,\xi)$ is the prolate spheroidal radial function of the jth kind of order m and degree n. The radial functions may be tailored to represent standing or traveling waves. For the problem under consideration, outgoing traveling waves are most appropriate, so the radial function of the third kind is employed.[25,26]

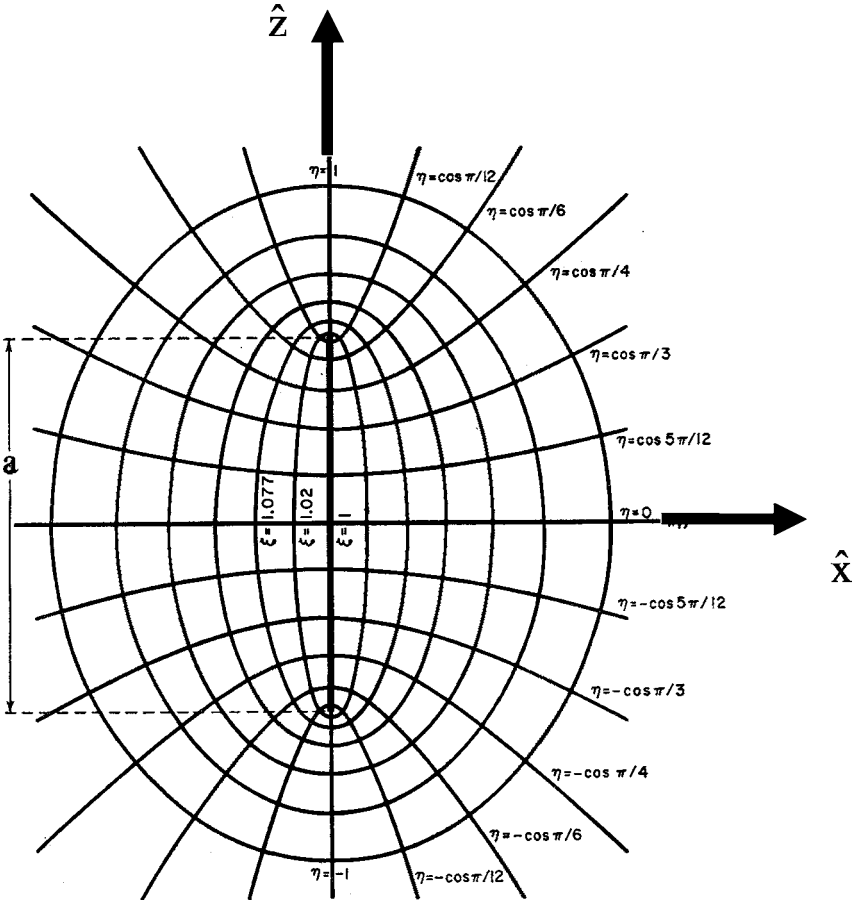


FIG. 1. The prolate spheroidal coordinate system.

The incident plane-wave pressure field, of magnitude P_0, is given by[25]

$$P_{\text{inc}}(\xi,\eta,\phi) = 2P_0 \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} i^n \frac{\varepsilon_m}{N_{mn}} S_{mn}(h,\eta_{\text{inc}}) S_{mn}(h,\eta) R_{mn}^{(1)} \times (h,\xi)\cos m(\phi-\phi_{\text{inc}}). \quad (14)$$

A Neumann boundary condition is imposed at the surface of the rigid prolate spheroid, $\xi=\xi_0$, requiring that $\{\hat{n}\cdot\nabla P_{\text{sct}} + \hat{n}\cdot\nabla P_{\text{inc}}\}|_{\xi=\xi_0}=0$. Substitution of Eq. (14) into this boundary condition allows the scattered field to be written[4,30]

$$P_{\text{sct}}(\xi,\eta,\phi) = -2P_0 \sum_{m=0}^{\infty} \sum_{n=m}^{\infty} i^n \frac{\varepsilon_m}{N_{mn}} \frac{R_{mn}^{(1)\prime}(h,\xi_0)}{R_{mn}^{(3)\prime}(h,\xi_0)} S_{mn}(h,\eta_{\text{inc}}) \times S_{mn}(h,\eta) R_{mn}^{(3)}(h,\xi)\cos m(\phi-\phi_{\text{inc}}). \quad (15)$$

In this analytical solution, the eigenfunction in the rotational coordinate has been reduced to $\cos m\varphi$ with $m \geqslant 0$ in order to satisfy a required periodicity over 2π. The Neumann factor is

$$\varepsilon_m = \begin{cases} 1 & \text{for } m=0 \\ 2 & \text{for } m \geqslant 1 \end{cases},$$

and the normalization factor N_{mn} is specified so that the prolate spheroidal angular functions have the same normalization[24,25] as the associated Legendre function $P_n^m(\eta)$.

Calculations using Eq. (15) for the far-field scattered pressure patterns in the x-z plane of a prolate spheroid defined by $\xi_0=1.005$ (i.e., $L/2R=10$) are shown in Fig. 2. Here, the plane wave originates at $\theta_{\text{inc}}=60°$, $\varphi_{\text{inc}}=0°$ (in spherical coordinates), and the incident amplitude is unity ($P_0=1$). At the lowest value of h, where the wavelength is much larger than $a/2$, the scattering is predominantly in the backscattered direction. The scattering is analogous to Rayleigh scattering from a sphere of radius $r_0=a/2$. For $h=4.2$ the dominant backscatter lobe narrows and rotates from a backscatter direction towards the direction of specular reflection. This lobe falls short of the true direction of specular reflection because the local radius of curvature of the prolate spheroid at the point of plane-wave impingement is comparable to the incident wavelength. The lobe approaches the angle of true specular reflection as $h \to 10$.

For $h>10$ the wavelength becomes smaller than the characteristic dimensions of the prolate spheroid; the asymptotic condition of geometrical acoustics begins to apply. The pressure scattered in the forward direction acquires a more significant magnitude. The forward-scattered lobe grows in magnitude (relative to the specular magnitude) and narrows in angular extent as h increases. This happens because a shadow is generated in the near field immediately behind the spheroid. The total pressure anywhere in the field exterior to the spheroid, $\xi>\xi_0$, is the sum of the incident and scattered pressures. In order for the shadow to exist in the high-frequency limit, the scattering body must scatter an acoustic wave in the forward direction that is equal in magnitude but opposite in sign from the incident wave. Destructive interference between the incident and scattered waves

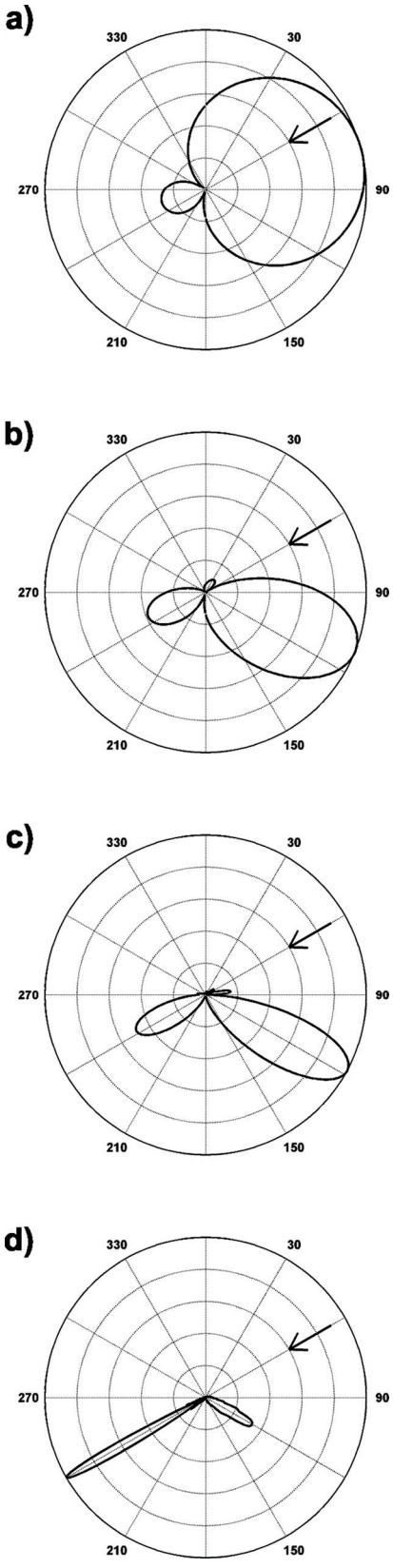


FIG. 2. (Color online) Far-field scattering patterns for a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$) insonified by a monochromatic plane wave from the direction $\theta_{inc}=60°$, $\varphi_{inc}=0°$. These patterns represent the magnitude of the scattered pressure, normalized by their respective maximum values, evaluated over $\theta=\cos^{-1}\eta$ at $\xi=50.25$. Each division represents 0.2 and the maximum radial value shown is 1. The arrow represents the wave vector $\mathbf{k}$ of the incident plane wave. The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d).

causes the acoustic shadow immediately behind the scatterer. In the degenerate case of a sphere in the far-forward-scattered direction, the scattered and incident waves would constructively interfere to form the Poisson cone.

D. Total pressure field

By the principle of superposition, the computed scattered pressure is combined with the incident wave pressure to obtain the total pressure field as presented in Fig. 3. Interference between the incident and scattered fields is apparent only when $h \gg 1$. Although the scattered pressure field interferes with the incident field at high values of h, the constructive and destructive interference generating the far-field diffraction pattern does not cause the overall pressure levels to deviate significantly from the incident pressure level. Therefore, the presence of the scattering body does not appreciably perturb the total pressure field for the selected spheroid size and frequencies.

E. Total intensity field

The scattered particle velocity $\mathbf{u}_{sct}=u_{sct}^{(\xi)}\hat{\xi}+u_{sct}^{(\eta)}\hat{\eta}+u_{sct}^{(\varphi)}\hat{\varphi}$ can be derived using the linearized Euler equation. Substitution of Eq. (15) into the Euler equation given by Eq. (6) yields the following expression for the scattered particle velocity vector:

$$\mathbf{u}_{sct}=\frac{-\mathrm{i}}{\omega\rho}\left(\frac{2}{\mathrm{a}}\sqrt{\frac{\xi^2-1}{\xi^2-\eta^2}}\frac{\partial P_{sct}}{\partial\xi}\hat{\xi}+\frac{2}{\mathrm{a}}\sqrt{\frac{1-\eta^2}{\xi^2-\eta^2}}\frac{\partial P_{sct}}{\partial\eta}\hat{\eta}+\frac{2}{\mathrm{a}\sqrt{(\xi^2-1)(1-\eta^2)}}\frac{\partial P_{sct}}{\partial\varphi}\hat{\varphi}\right). \tag{16}$$

The scattered particle velocity can be combined with the particle velocity of the incident plane wave to obtain the total particle velocity, $\mathbf{u}_{tot}(\mathbf{r},t)$. This expression, along with $p_{tot}(\xi,\eta,\varphi)$, is used to construct the complex acoustic intensity according to Eq. (12). The vector components of the magnitude and phase (between the total pressure and the total particle velocity) are computed from the complex vector intensity. The computed phase in the $\hat{\xi}$- and $\hat{\eta}$ directions are presented in Figs. 4 and 5, respectively. The power factor angle for the complex intensity in the $\hat{\varphi}$ directions is presented in Fig. 6 for completeness. The phase angle of the $\hat{\varphi}$ component of complex intensity can still be evaluated despite a vanishing small magnitude due to $\mathbf{k}\cdot\hat{\varphi}=0$.

The spatial distribution of power factor angle indicates the development of significant and localized disturbances in the phase of the complex intensity that are strongly dependent upon h. The phase of the $\hat{\xi}$ component of the complex intensity field exceeds 5° in the specularly scattered direction, even at distances $>20L_0$. Conversely, the phase of the $\hat{\eta}$ component exhibits a significant anomaly that is localized to the forward-scattered region with little perturbation in the specular direction. The observed phase differences in the two intensity components can be explained by examining the real

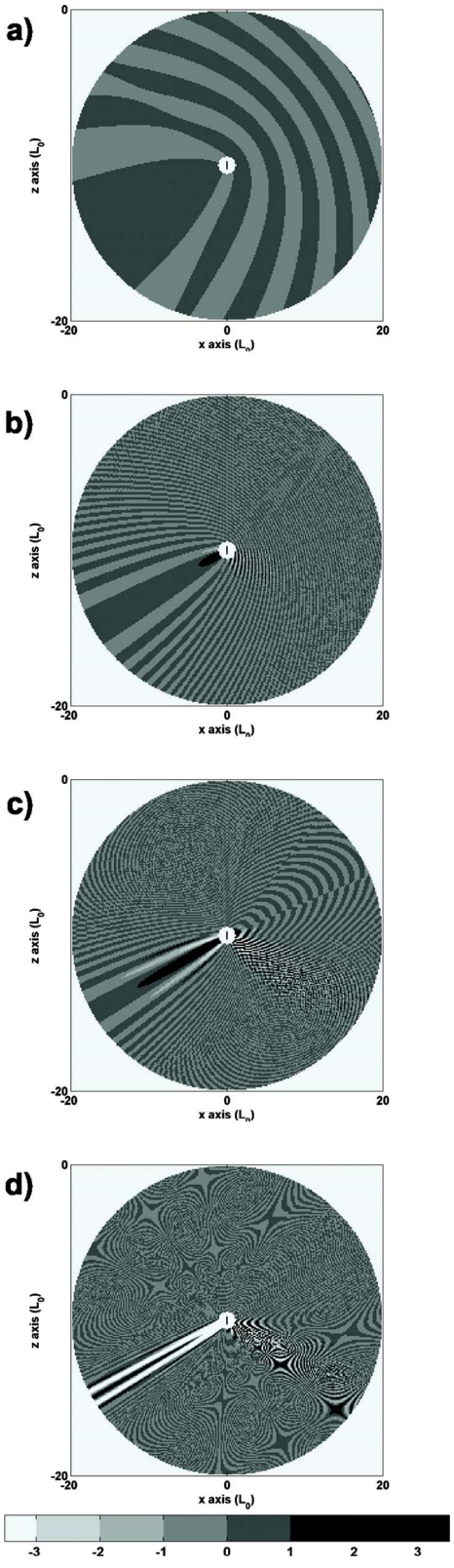


FIG. 3. (Color online) Total pressure field, $10\log_{10}(|P_{\text{tot}}/P_0|^2)$, due to the scattering of a monochromatic plane wave by a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$) insonified by a monochromatic plane wave from the direction $\theta_{\text{inc}}=60°$, $\varphi_{\text{inc}}=0°$. The field in the vicinity of the spheroid at the origin has been omitted for clarity. The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d). The x and z axes of the plots extend between $\pm 20L_0$ ($1.17<\xi<40.0$). The gray scale represents the total acoustic pressure in decibels. The development of the interference pattern in the forward scatter direction at high reduced frequencies gives rise to spatial gradients in the mean-square pressure field, causing nonzero phase angles in certain complex intensity vectors.

and imaginary parts of the complex intensity, which correspond to the active and reactive intensity, respectively.

The magnitude of the active intensity in the $\hat{\xi}$ direction is presented in Fig. 7. This is a vector quantity and therefore contains a spatial response corresponding to the dot product of the local acoustic wave vector and the unit vectors of the coordinate system. This results in a cosine shading (a dipole spatial response) which is exaggerated in Fig. 7 by the small dynamic range of the gray scale. The radial component[33] of active intensity, I_{ξ}, quickly drops to zero at those locations where the radial unit vector is orthogonal to the incident wave vector, i.e., $\mathbf{k}\cdot\hat{\xi}=0$. Since $\hat{\xi}$ and $\hat{\eta}$ are orthogonal to one another, the tangential component of active intensity is at a maximum when the radial component of active intensity is at a minimum. The $\hat{\varphi}$ component of active intensity is at a minimum for all the field points evaluated on the x-z plane because the $\hat{\varphi}$ unit vector is orthogonal to this plane which contains $\mathbf{k}$. In Sec. II, it was shown that the active intensity was proportional to the square of the pressure field. Therefore, the features that were observed in Fig. 3 are present in the active intensity as displayed in Fig. 7.

The reactive intensity is the imaginary part of the complex intensity which is proportional to the gradient of the mean-squared pressure field. The reactive intensity for the $\hat{\xi}$- and $\hat{\eta}$ directions is presented in Figs. 8 and 9, respectively. The reactive intensity in the $\hat{\varphi}$ direction is not presented because $\mathbf{k}\cdot\hat{\varphi}=0$. The $\hat{\xi}$ component of reactive intensity (Fig. 8) at $h=0.4$ indicates the development of very small levels of reactive intensity in the backscattered direction. Only when $h\gg 1$ does the $\hat{\xi}$ component of the reactive intensity become a feature that is localized in the direction of specular reflection. The justification for this is that Eq. (16) implies that the $\hat{\xi}$ component of particle velocity will exhibit significant magnitude when the radial portion of the gradient of the pressure is significant. In the steady-state scattering problem, the incident plane wave is always interfering with the specularly scattered wave in a fashion that causes the radial pressure gradient to take on significant magnitude. This development can also be observed in Fig. 4, which shows that the power factor angle for the $\hat{\xi}$ direction is localized to the same region at the larger values of h.

A similar explanation is used to explain the development of the $\hat{\eta}$ component of reactive intensity as presented in Fig. 9. This component of reactive intensity exists in the specular direction at high reduced frequencies because of the interference between the incident and scattered waves. However, reactive intensity has also developed in the forward direction at the larger values of h. The behavior mimics the development of appreciable power factor angles in the forward-scattered direction in Fig. 5. This can be explained by referring to Fig. 3 and observing that the gradient of the mean-square pressure in the $\hat{\eta}$ direction is significant due to the presence of fringes of the patterns.

The sensitivity of the phase angles corresponding to the $\hat{\eta}$ and $\hat{\varphi}$ components of total intensity to the field angle off of forward scatter and to reduced frequency for the case of $\theta_{\text{inc}}=90°$ is given in Fig. 10. The sensitivity of these phase angles to plane-wave incidence angle at constant reduced

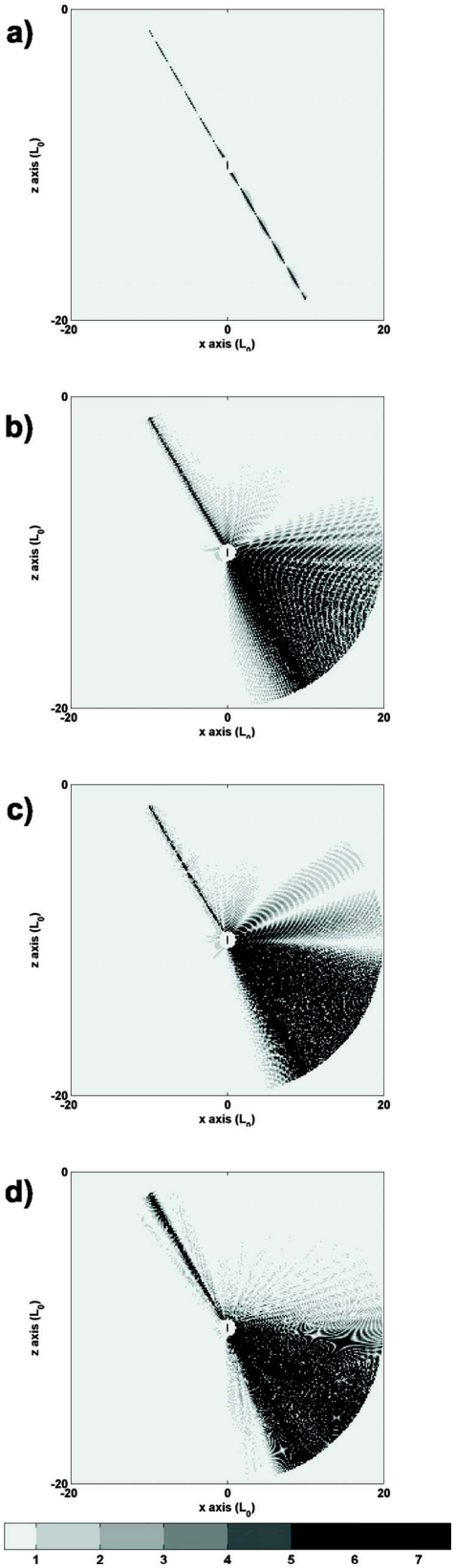


FIG. 4. (Color online) Phase angle of the $\hat{\xi}$ component of the complex intensity field when a monochromatic plane wave originating from $\theta_{\text{inc}}=60°$ is perturbed by the presence of a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$). The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d). The x and z axes of the plots extend between $\pm 20L_0(1.17<\xi<40.0)$. The gray scale represents relative phase angle in degrees. The noteworthy feature of the plots is the absence of a phase angle anomaly in the forward-scatter direction.

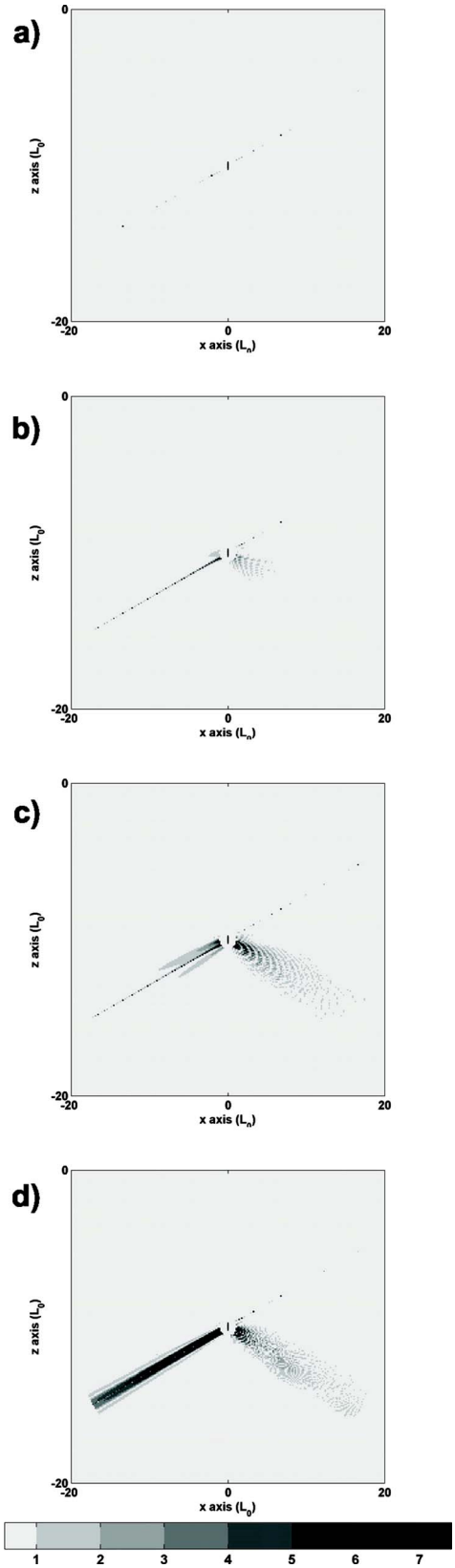


FIG. 5. (Color online) Phase of the $\hat{\eta}$ component of the complex intensity field when a monochromatic plane wave originating from $\theta_{\text{inc}}=60°$ is perturbed by the presence of a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$). The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d). The x and z axes of the plots extend between $\pm 20L_0$ $(1.17<\xi<40.0)$. The gray scale represents relative phase angle in degrees. A phase angle anomaly develops in the forward-scatter direction at high reduced frequencies.

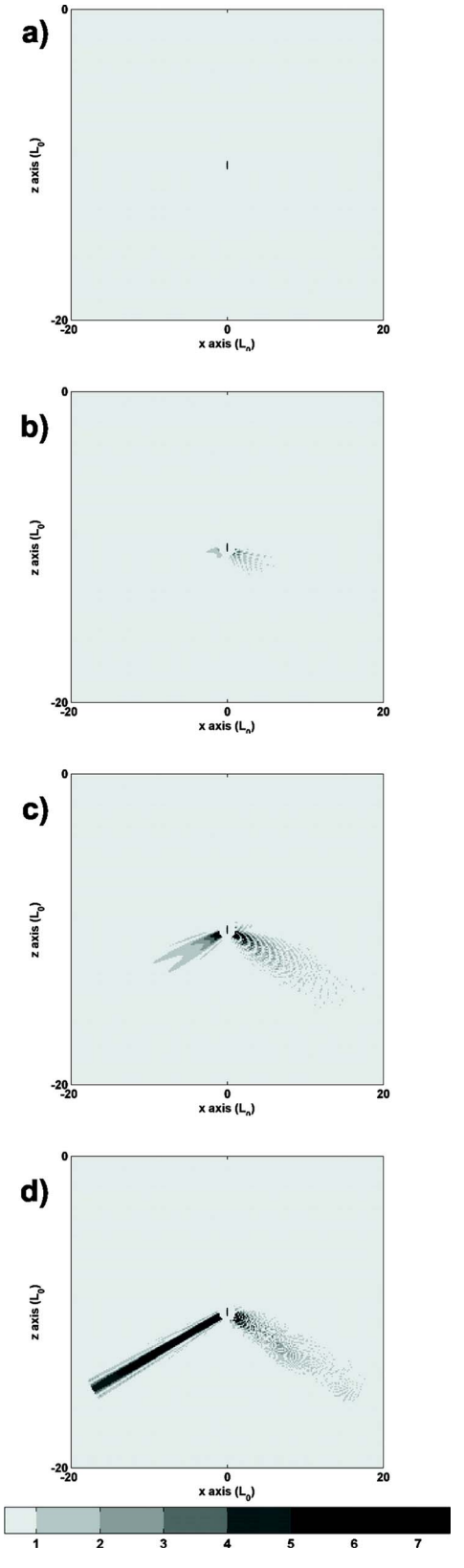


FIG. 6. (Color online) Phase of the $\hat{\varphi}$ component of the complex intensity field when a monochromatic plane wave originating from $\theta_{\text{inc}}=60°$ is perturbed by the presence of a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$). The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d). The x and z axes of the plots extend between $\pm 20L_0$ ($1.17<\xi<40.0$). The gray scale represents relative phase angle in degrees. A phase angle anomaly develops in the forward-scatter direction at high reduced frequencies. This phase angle exists numerically even though the magnitude of the intensity in this direction is vanishingly small due to the orthogonality of $\hat{\varphi}$ and the incident wave vector, $\mathbf{k}$.

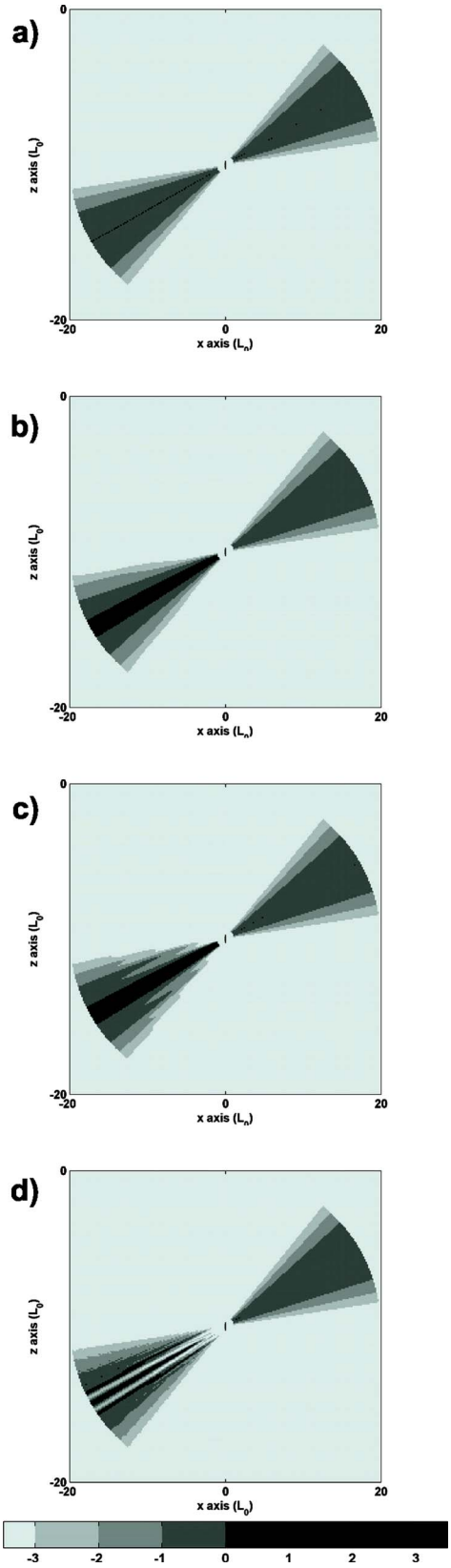


FIG. 7. (Color online) Total active intensity field, $10\log_{10}(|I_{\hat{\xi}}/I_{\text{ref}}|)$, in the $\hat{\xi}$ direction resulting from the scattering of a monochromatic plane wave originating from $\theta_{\text{inc}}=60°$ by a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$). The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d). The x and z axes of the plots extend between $\pm 20L_0$ ($1.17<\xi<40.0$). The gray scale represents the active acoustic intensity in decibels relative to incident active intensity, $I_{\text{ref}}=P_0^2/2\rho c$.

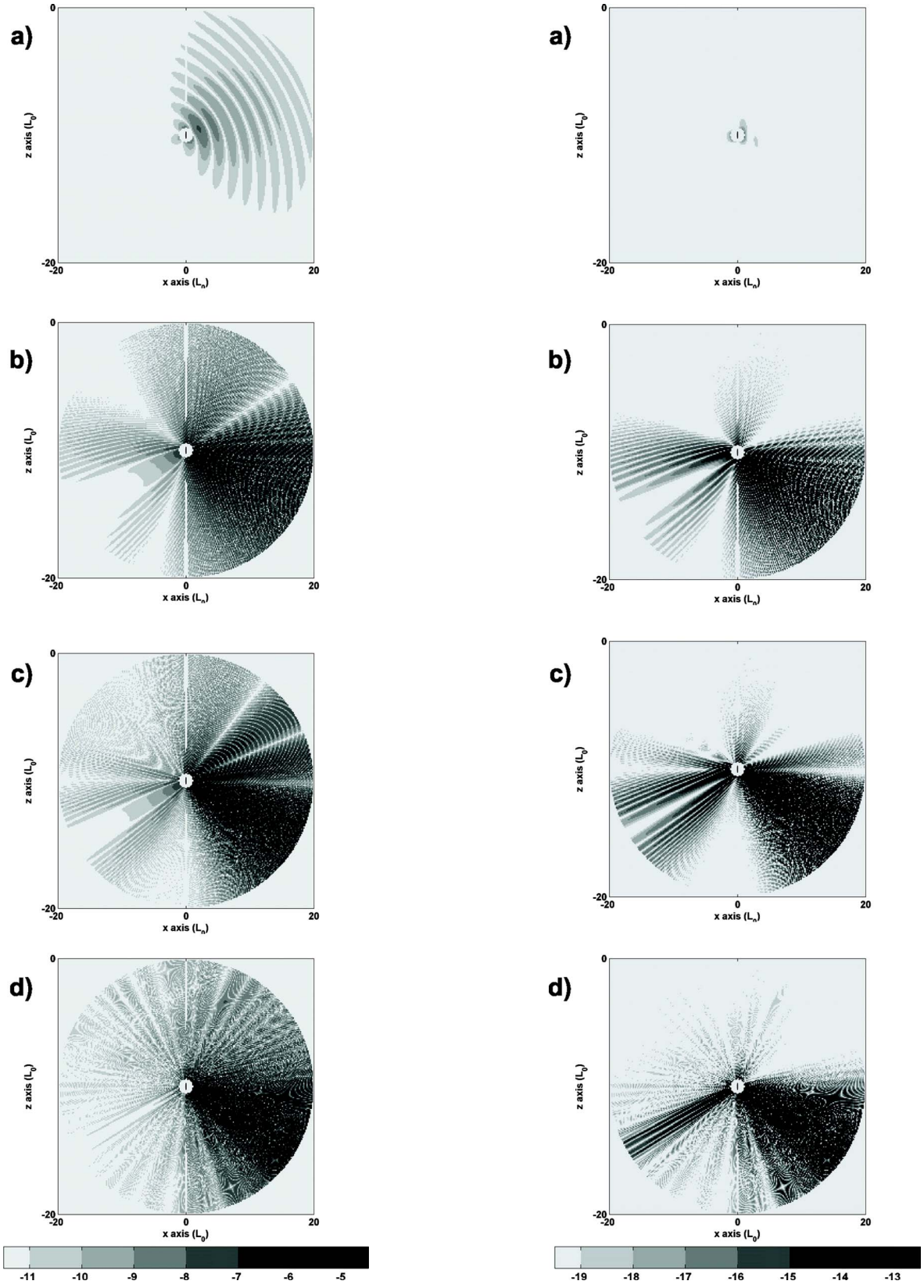


FIG. 8. (Color online) Total reactive intensity field, $10\log_{10}(|Q_{\hat{\xi}}/I_{\text{ref}}|)$, in the $\hat{\xi}$ direction resulting from the scattering of a monochromatic plane wave originating from $\theta_{\text{inc}}=60°$ by a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$). The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d). The x and z axes of the plots extend between $\pm 20L_0$ ($1.17<\xi<40.0$). The gray scale represents the reactive acoustic intensity in decibels relative to incident active intensity, $I_{\text{ref}}=P_0^2/2\rho c$.

FIG. 9. (Color online) Total reactive intensity field, $10\log_{10}(|Q_{\hat{\eta}}/I_{\text{ref}}|)$, in the $\hat{\eta}$ direction resulting from the scattering of a monochromatic plane wave originating from $\theta_{\text{inc}}=60°$ by a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$). The reduced frequency h is 0.4 (a), 4.2 (b), 8.3 (c), and 41.7 (d). The x and z axes of the plots extend between $\pm 20L_0$ ($1.17<\xi<40.0$). The gray scale represents the reactive acoustic intensity in decibels relative to incident active intensity, $I_{\text{ref}}=P_0^2/2\rho c$.

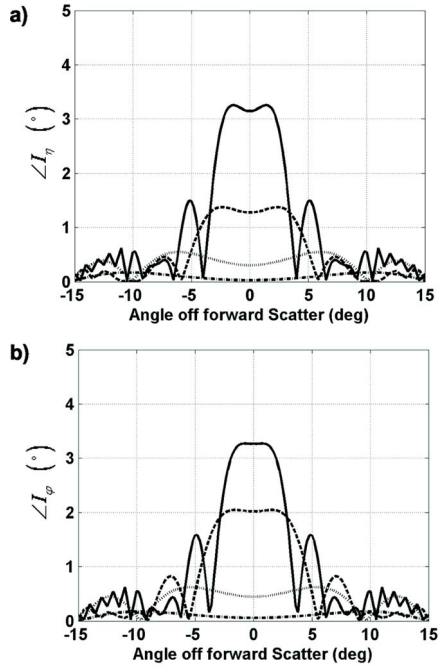


FIG. 10. (Color online) Phase angle of the $\hat{\eta}$ component (left) and the $\hat{\varphi}$ component (right) of the complex intensity field when a monochromatic plane wave originating from $\theta_{\text{inc}}=90°$ is perturbed by the presence of a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$). For both plots, the field has been evaluated at $\xi=20.1$, which corresponds to a range of $10L_0$. The reduced frequency h is 41.7 (solid), 20.8 (dashed), 8.3 (dotted), and 4.2 (dash-dot).

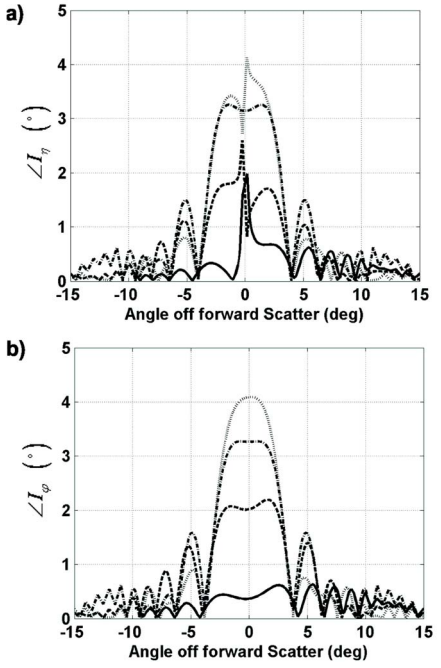


FIG. 11. (Color online) Phase angle of the $\hat{\eta}$ component (left) and the $\hat{\varphi}$ component (right) of the complex intensity field when a monochromatic plane wave is perturbed by the presence of a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$) and $h=41.7$. For both plots, the field has been evaluated at $\xi=20.1$, which corresponds to a range of $10L_0$. The incidence angle θ_{inc} is 10° (solid), 30° (dashed), 60° (dotted), and 90° (dash-dot).

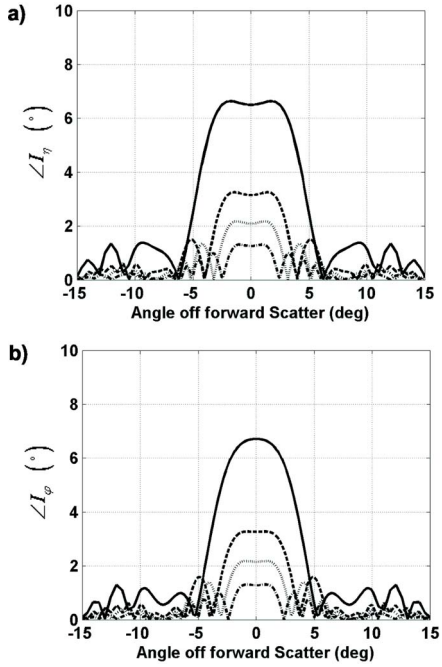


FIG. 12. (Color online) Phase angle of the $\hat{\eta}$ component (left) and the $\hat{\varphi}$ component (right) of the complex intensity field when a monochromatic plane wave originating from $\theta_{\text{inc}}=90°$ is perturbed by the presence of a rigid prolate spheroid having a fineness ratio of 10:1 ($\xi_0=1.005$) and $h=41.7$. For both plots, the ξ coordinate is 10.0 (solid), 20.1 (dashed), 30.1 (dotted), and 50.2 (dash-dot). These coordinates correspond to ranges of $5L_0$, $10L_0$, $15L_0$, and $25L_0$, respectively.

frequency is given in Fig. 11. Figure 12 shows the sensitivity of these phase angles to radial coordinate ξ with $\theta_{\text{inc}}=90°$ and $h=41.7$. In Fig. 10, the onset of the phase angle anomaly and the narrowing of its angular extent at high reduced frequencies can be observed. For $h=41.7$ the phase angle is seen to be $>3°$ at the computed range of $10L_0$ from the center of the spheroid. Figure 11 shows that, as the incident angle changes from nearly end-on at $\theta_{\text{inc}}=10°$ to broadside at $\theta_{\text{inc}}=90°$, the phase angles in the forward-scattering direction increase, and the lateral extent of this effect also increases. It is interesting to note the asymmetry of the phase angle anomaly at angles other than broadside. The decay of the phase angle anomaly with range can be observed in Fig. 12. At the range of $25L_0$, the magnitude of the central portion ($\sim 0°$) of the anomaly is only 1°, which is comparable to the phase angle of the first-order sidelobe. Given the physical interpretations presented earlier, it can be expected that the magnitude of these anomalies will increase as the reduced frequency increases.

IV. SUMMARY AND CONCLUSION

The complex-valued acoustic vector intensity field associated with the scattering of a steady-state harmonic plane wave by a rigid prolate spheroid in an unbounded, lossless medium has been investigated analytically. Calculations indicate that there are localized regions in the total acoustic field where the presence of the scattering body perturbs the active and reactive intensity, as well as the power factor angle. The regions where these perturbations are most apparent are the forward-scatter and specular-scatter regions due

to the interference of the incident and scattered waves. Perturbations in the backscatter direction would occur for the special case of broadside incidence angles, where the specular-and backscatter directions coincide.

These theoretical results suggest that for the 3D, high-frequency scattering problem considered, an acoustic field rich in phase and pressure gradients exists. The detection of these features can be accomplished using the basic principles of complex acoustic vector intensity measurements. The reactive intensity and the power factor angle can be determined directly from simultaneous measurement of the acoustic pressure and acoustic particle velocity. The observable manifestation of the phenomenon studied in this investigation is the phase difference between acoustic pressure and particle velocity, i.e., power factor angle. In the short-wavelength limit, the ability to reliably measure power factor angles of 1° to 5°, and to differentiate them from situations where this angle is theoretically 0°, would allow an investigator to infer that a scattering body is present in the line-of-sight between a source and receiver. From a practical point of view, the influence of finite statistics upon phase estimation in low signal-to-noise ratio situations will limit the applicability of this concept. The presence of secondary sources will bias the intensity vectors and even dictate the gradient of the mean-square pressure field if they are of significantly higher level than the source illuminating the spheroid. The use of arrays of intensity sensors may enable spatial filtering to reduce the effects of secondary sources but would require further investigation.

ACKNOWLEDGMENTS

The authors would like to thank A. L. Van Buren at the Naval Undersea Warfare Center for his useful discussions regarding spheroidal wave functions as well as for making his FORTRAN code available to the authors for computations of the prolate spheroidal wave functions. This work was supported by the Office of Naval Research, Code 321 MS under Grant Number N00014-01-1-0108, Dr. James F. McEachern, project monitor. This support is gratefully appreciated.

[1]B. Gillespie, K. Rolt, G. Edelson, R. Shaffer, and P. Hursky, "Littoral target forward scattering," in *Acoustical Imaging* (Plenum, New York, 1997), Vol. **23**, pp. 501–506.
[2]H. M. Nussenzveig, *Diffraction Effects in Semiclassical Scattering* (Cambridge University Press, Cambridge England, 1992).
[3]P. Ratilal and N. C. Makris, "Extinction theorem for object scattering in a stratified medium," J. Acoust. Soc. Am. **110**, 2924–2945 (2001).
[4]G. S. Sammelmann, D. H. Trivett, and R. H. Hackman, "High-frequency scattering from rigid prolate spheroids," J. Acoust. Soc. Am. **83**, 46–54 (1988).
[5]A. Sarkissian, C. F. Guammond, and L. R. Dragonette, "T-matrix implementation of forward scattering from rigid structures," J. Acoust. Soc. Am. **94**, 3448–3453 (1993).
[6]H. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and P. Guerrini, "Demonstration of a high-frequency acoustic barrier with a time-reversal mirror," IEEE J. Ocean. Eng. **28**, 246–249 (2003).
[7]N. Willis, *Bistatic Radar* (Artech House, Boston, MA, 1991).
[8]A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1989), p. 430.
[9]A. Baggeroer and H. Cox, "Comparison of the performance of vector sensors using optimum processing," J. Acoust. Soc. Am. **114**, 2427 (2003).
[10]H. Cox and A. Baggeroer, "Performance of vector sensors in noise," J. Acoust. Soc. Am. **114**, 2426 (2003).
[11]B. A. Cray and A. H. Nuttall, "Directivity factors for linear arrays of velocity sensors," J. Acoust. Soc. Am. **110**, 324–331 (2001).
[12]G. L. D'Spain, W. S. Hodgkiss, G. L. Edmonds, J. C. Nickles, F. H. Fisher, and R. A. Harriss, "Initial analysis of the data from the vertical DIFAR array," in "Proceedings of Mastering the Oceans Through Technology (OCEANS 92)," Newport, RI, 26–29 October 1992, pp. 346–351.
[13]M. A. Hawkes and A. Nehorai, "Acoustic vector-sensor correlations in ambient noise," IEEE J. Ocean. Eng. **26**, 337–347 (2001).
[14]V. A. Shchurov, "Coherent and diffusive field of underwater acoustic ambient noise," J. Acoust. Soc. Am. **90**, 991–1001 (1991).
[15]F. Jacobsen, "Sound field indicators: Useful tools," Noise Control Eng. J. **35**, 37–46 (1990).
[16]B. R. Rapids and G. C. Lauchle, "Acoustic intensity measurements involving forward scatter from prolate spheroids," J. Acoust. Soc. Am. **116**, 2528 (2004).
[17]B. R. Rapids, "Acoustic Intensity Methods in Classical Scattering," Ph.D. thesis in Acoustics, The Pennsylvania State University (2004).
[18]F. H. Fahy, *Sound Intensity*, 2nd ed. (E&FN Spon, London, UK, 1995).
[19]P. J. Westervelt, "Acoustical impedance in terms of energy functions," J. Acoust. Soc. Am. **23**, 347–348 (1951).
[20]T. K. Stanton and R. T. Beyer, "Complex wattmeter measurements in a reactive acoustic field," J. Acoust. Soc. Am. **65**, 249–252 (1979).
[21]J. A. Mann III, J. Tichy, and A. J. Romano, "Instantaneous and time-averaged energy transfer in acoustic fields," J. Acoust. Soc. Am. **82**, 17–30 (1987).
[22]J. A. Mann III and J. Tichy, "Acoustic intensity analysis: distinguishing energy propagation and wave-front propagation," J. Acoust. Soc. Am. **90**, 20–25 (1991).
[23]B. J. King, R. V. Baier, and S. Hanish, "A FORTRAN computer program for calculating the prolate spheroidal radial functions of the first and second kind and their first derivatives," NRL Rep. No. 7012 (1970).
[24]B. J. King, R. V. Baier, and S. Hanish, "A FORTRAN computer program for calculating the prolate and oblate angle functions of the first kind and their first and second derivatives," NRL Rep. No. 7161 (1970).
[25]C. Flammer, *Spheroidal Wave Functions* (Stanford University Press, Stanford, CA, 1957).
[26]A. L. Van Buren, R. V. Baier, S. Hanish, and B. J. King, "Calculation of spheroidal wave functions," J. Acoust. Soc. Am. **51**, 414–416 (1972).
[27]J. P. Barton, N. L. Wolff, H. Zhang, and C. Tarawneh, "Near-field calculations for a rigid spheroid with an arbitrary incident acoustic field," J. Acoust. Soc. Am. **113**, 1216–1222 (2003).
[28]A. Germon and G. C. Lauchle, "Axisymmetric scattering of spherical waves by a prolate spheroid," J. Acoust. Soc. Am. **65**, 1322–1327 (1979).
[29]G. C. Lauchle, "Short-wavelength acoustic backscattering by a prolate spheroid," J. Acoust. Soc. Am. **58**, 576–580 (1975).
[30]G. C. Lauchle, "Short-wavelength acoustic diffraction by prolate spheroids," J. Acoust. Soc. Am. **58**, 568–575 (1975).
[31]A. L. Van Buren and J. E. Boisvert, "Accurate calculation of prolate spheroidal radial functions of the first kind and their first derivatives," Q. Appl. Math. **60**, 589–599 (2002).
[32]A. L. Van Buren and J. E. Boisvert, "Improved calculation of prolate spheroidal radial functions of the second kind and their first derivates," Q. Appl. Math. **62**, 493–508 (2004).
[33]The $\hat{\xi}$-component is "radial" only for $\xi \gg 1$. The term is used here to offer a more familiar term borrowed from scattering in spherical coordinates.

Evaluation of layered multiple-scattering method for antiplane shear wave scattering from gratings

Liang-Wu Cai
Department of Mechanical and Nuclear Engineering, 3031 Rathbone Hall, Kansas State University, Manhattan, Kansas 66506

(Received 18 October 2005; revised 28 February 2006; accepted 28 April 2006)

The layered multiple-scattering method is based on an approximate solution for infinite gratings. In this method, an array of regularly arranged scatterers is viewed as comprising of layers of infinite grating and treated as a multiple transmission-reflection process in a multilayer panel. In this paper, this method is evaluated by comparing with exact solutions obtained by other means. One is a multiple-scattering solution. Another is the exact solution for an infinite grating, which is obtained by combining the *T*-matrix formulation of the multiple-scattering theory and an alternative representation of the Schlömilch series. The infinity nature enables the waves due to a planar incident wave to be expressed as planar waves and divided into propagating and evanescent modes. The layered multiple-scattering method accounts only for the propagating modes. Details of these modes are analyzed for a single grating, and it is concluded that only the first evanescent modes would have significant presence in a limited frequency range. The layered multiple-scattering method suggests that the only important geometric parameters for wave transmission and reflection are the grating distance and the interlayer distance. Numerical examples indicate that error due to evanescent modes might be significant due to interlayer interactions, such as critical frequencies of a stopband. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2206517]

PACS number(s): 43.20.Gp, 43.20.Fn, 43.40.Fz [DF] Pages: 49–61

I. INTRODUCTION

The infinite grating problem refers to the scattering of planar waves by an infinite number of identical scatterers periodically arranged along a straight line. Such structures, or variants with finite numbers of scatterers, appear in applications as wave filters or reflectors for many types of waves. Approximate theories have been studied by early pioneers of sound theory such as Lord Rayleigh (1907) nearly a century ago.

Many modern theories for infinite gratings can be traced back to a seminal work by Twersky (1956). In this work, following his earlier work on finite gratings (Twersky, 1952b), the Green's function method, also known as the integral equation method, is used to obtain a formalism for the scattering of electromagnetic waves by infinite gratings whose members are symmetric about the grating axis. The most important result is that the resulting waves are separated into propagating modes that propagate in a set of "diffraction angles," and nonpropagating modes of "surface waves" that localize near the grating. It also explains the so-called *Wood's anomaly* in which the scattered wave appears to be propagating along the grating when the incident wave is directed at a certain angle. Closely following Twersky's work, Millar studied the infinite and finite grating problems in greater detail in two series of articles (1961a, b, 1963a, b, 1964a, b, 1966).

Using the integral equation method, complicated scatterer geometries have been studied. Ivanov (1971) formulated the scattering by multiple gratings. Leiko and Mayatskii studied the scattering of acoustic waves by gratings of elliptical perfectly compliant (1974) and rigid (1975) cylinders. Kalhor and Ilyas (1982) studied gratings comprised of scatterers of an arbitrary cross-section embedded in a dielectric slab. Lakhtakia *et al.* have also studied a similar problem (1986a). They further extended their analysis for a circular cylindrical grating in a slab subjected to elastic antiplane shear waves (1986b) and longitudinal and shear waves (1988). Achenbach and colleagues (Achenbach and Li, 1986; Angel and Achenbach, 1987) studied gratings of cracks. Scarpetta and Sumbatyan (1995, 1997, 2002, 2003) studied the scattering of elastic wave by gratings of a variety of shapes of defects.

Twersky later published two more papers on infinite grating problem, which represent more significant analytical simplifications, although they received less references until rather recently. In the first paper (Twersky, 1961), an alternative representation for Schlömilch series using elementary functions is obtained. Schlömilch series are omnipresent in infinite grating problems where cylindrical wave functions are used. In the second paper (Twersky, 1962), a more accessible form of formulation, via the separation of variables method, is obtained for gratings of circular cylinders. This formulation resulted in a set of algebraic equations in terms of the solution to the corresponding single scattering problem. A single scattering problem refers to the one in which a single scatterer embedded in an infinite host medium is subjected to an incident wave. Extension to gratings of elliptical cross sections was obtained by Burke and Twersky (1966).

Following Twersky's formulation, Klyukin and Chabanov (1975) studied acoustic scattering by gratings of rigid cylinders. Miles (1982) combined single scattering results in the low-frequency limit obtained by Rayleigh and Lamb with Twersky's formulation to study acoustic scattering by a grat-

ing of flat plates. Heckl and colleagues (Heckl, 1992, 1994; Huang and Heckl, 1993; Mulholland and Heckl, 1994; Heckl and Mulholland, 1995) studied various configurations of sound wave scattering by gratings of circular tubes, taking into account losses due to dissipation and heat generations at interfaces. Recently, Kavakloğlu (2000, 2001, 2002) studied the scattering of out-of-plane oblique electromagnetic waves by gratings of circular dielectric cylinders.

The separation of variable approach is sometime called the waveguide approach when employed in a Cartesian coordinate system. Kristiansen and Fahy (1972) studied the acoustic scattering by multiple gratings of square columns. Vovk and colleagues (Vovk *et al.*, 1976; Vovk and Grinchenko, 1978) analyzed the scattering of acoustic waves by gratings of hollow elastic boxes. Radlinski and colleagues (Brigham *et al.*, 1977; Radlinski and Simon, 1982; Radlinski and Janus, 1986; Radlinski, 1989) studied the acoustic scattering by various configurations of compliant tubes of rectangular cross sections.

Work has also been done for water waves, especially for gratings of cylindrical columns, either floating on the water surface or planted to the bottom of the water, used as breakwaters. Interested readers are referred to works by Miles (1983), Linton and Evans (1990), Evans and Linton (1991), Porter and Evans (1996, 1999), Maniar and Newman (1997), McIver (2000), and Ohl *et al.* (2001a, b), and the references cited therein.

Despite the difference in the types of waves, when a wave impinges onto an infinite grating, a portion of the wave becomes localized in the vicinity of the grating and the remainder propagates in the form of planar wave. This behavior itself suggests an obvious approximation commonly known as the *layered multiple-scattering method* for analyzing the scattering due to multiple gratings. This method treats periodically arranged scatterers as layers of gratings stacked together. Within each layer, the infinite grating problems can be used to approximate the wave transmission and reflection characteristics. Then, the multiple-scattering problem becomes analyzing the wave propagation in multilayered panels. Some of the aforementioned references have in fact endeavored into the method. This method has recently garnered renewed interest in the study of photonic and phononic band gap materials. These materials are artificial materials having internal periodically arranged scatterers. They prohibit waves in a certain frequency range from propagating through the material. The layered multiple-scattering method has been used in many analyses such as Esquivel-Sirvent and Cocoletzi (1994), Botten *et al.* (2000, 2004), Liu *et al.* (2000), Psarobas and Sigalas (2002), Platts *et al.* (2003a, b), Sainidou *et al.* (2005). However, so far this method has not been compared with exact solutions for validation and error qualification.

The main thrust of this paper is to evaluate the layered multiple-scattering method. The infinite grating problem is first revisited, with the use of *T*-matrix notation. This notation not only maintains the full mathematical rigor, but also accentuates the physics framed in an input-output perspective. More importantly, the matrix notation allows scatterers in the grating to be abstract. A set of planar wave expansion basis matrices is defined, based on the alternative representation of Schlömilch series obtained by Twersky (1961).The characteristics of the propagating modes and the evanescent modes are analyzed. The layered multiple-scattering analysis is then derived. The approximate solution is compared with the analytically exact solution obtained via the computational system constructed in the author's previous work (Cai and Williams, 1999b).

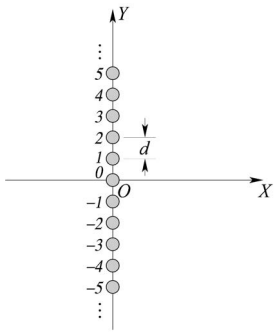


FIG. 1. Geometry of infinite grating problem.

II. INFINITE GRATING PROBLEM REVISITED

A. Problem description

Consider an infinite number of identical cylindrical scatterers equally spaced by a distance d apart along the y axis, as shown in Fig. 1. Scatterers are numbered from $-\infty$ to ∞: those with negative numbers are located in the lower half-plane, those with positive numbers in the upper half-plane, and scatterer 0 is located at the origin. Coordinates in the global coordinate system are referred to either by Cartesian coordinates (x,y) or by polar coordinates (r,θ).

It is assumed that the T matrix for each scatterer is known. The T matrix is defined in the corresponding single-scatterer problem and relates wave expansion coefficients of the scattered wave to those of the incident wave. When the incident and scattered waves are expressed as

$$\phi^{\text{inc}} = \{\boldsymbol{A}\}^T\{\boldsymbol{J}(r,\theta)\} \quad \text{and} \quad \phi^{\text{scattered}} = \{\boldsymbol{B}\}^T\{\boldsymbol{H}(r,\theta)\}, \quad (1)$$

where $\{\boldsymbol{A}\}$ and $\{\boldsymbol{B}\}$ are the wave expansion coefficients for the incident and scattered waves, respectively, and $\{\boldsymbol{J}(r,\theta)\}$ and $\{\boldsymbol{H}(r,\theta)\}$ are the regular and singular wave expansion bases, respectively, the linearity of the system requires

$$\{\boldsymbol{B}\} = [\boldsymbol{T}]\{\boldsymbol{A}\}, \quad (2)$$

where $[\boldsymbol{T}]$ is the so-called T matrix. Entries for the wave expansion bases at the nth row are

$$\{J(r,\theta)\}_n = J_n(kr)\mathrm{e}^{\hat{\imath}n\theta} \quad \text{and} \quad \{H(r,\theta)\}_n = H_n^{(1)}(kr)\mathrm{e}^{\hat{\imath}n\theta}, \quad (3)$$

where $\hat{\imath} = \sqrt{-1}$, $J_n(\cdot)$ and $H_n^{(1)}(\cdot)$ are Bessel and Hankel functions of the first kind, respectively, and n runs from $-\infty$ to ∞. Since the T matrix is defined in a specific coordinate system, the assumption about the T matrix implies that a set of local polar coordinate systems, denoted as (r_i,θ_i) for scatterer i, has been defined. All θ_i's are measured from the global x direction.

For generality, no further assumptions regarding the physical and geometrical compositions of the scatterers are made. This allows the broadest range of scatterers to be used to construct the grating, such as those generated by the so-called scatterer polymerization methodology (Cai and Williams, 1999a) or multilayered scatterers (Cai, 2004, 2005).

Considerations here are limited to elastic antiplane shear waves, which are commonly called the SH waves. In such cases, the only nontrivial displacement component is the z component. In the steady state, this displacement can be expressed as $z=\phi e^{\hat{\imath}\omega t}$, where ω is the circular frequency and ϕ is the complex amplitude of the displacement. It is noted that the mathematical description of an SH wave problem is essentially the same as that for acoustic waves.

B. Multiple-scattering solution

Multiple-scattering problems have been studied extensively in the past few decades. Twersky (1952a) established an early solution by tracking the wave-scatterer interactions. Waterman (1969) introduced the concept of the T matrix for analyzing the scattering from geometrically complicated scatterers. Varadan and colleagues obtained multiple-scattering solutions using the T-matrix method for a variety of problems [see Varadan *et al.* (1988), for a list of their work]. The author used the T-matrix concept to frame Twersky's ordered scattering in an input-output perspective and subsequently developed a multiple-scattering solution (Cai and Williams, 1999a), and the scatterer polymerization method (Cai and Williams, 1999a) and the multiple scattering in single scatterers method (Cai, 2004, 2005) for analyzing complicated scatterers. Note that none of the aforementioned multiple-scattering solutions consider the effects of possible rigid-body translations of the scatterers.

According to the multiple-scattering theory (Cai and Williams, 1999a), the total wave field ϕ^{total} consists of the incident wave and an infinite number of scattered waves, one from every scatterer; that is,

$$\phi^{\text{total}} = \phi^{\text{inc}} + \sum_{i=-\infty}^{\infty} \{\boldsymbol{C}_i\}^T \{\boldsymbol{H}(r_i, \theta_i)\}, \tag{4}$$

where $\{\boldsymbol{C}_i\}$ and $\{\boldsymbol{H}(r_i, \theta_i)\}$ are the wave expansion coefficient matrix and the singular wave expansion basis, respectively, for a scattered wave in scatterer i's local coordinate system.

The general solution for the multiple-scattering problem (Cai and Williams, 1999a) is

$$\{\boldsymbol{C}_i\} = [\boldsymbol{T}]\left(\{\boldsymbol{A}_i\} + \sum_{\substack{j=-\infty \\ j \neq i}}^{\infty} [\boldsymbol{R}_{ji}]^T \{\boldsymbol{C}_j\} \right), \tag{5}$$

where $\{\boldsymbol{A}_i\}$ is the wave expansion coefficient for the incident wave in scatterer i's local coordinate system and $[\boldsymbol{R}_{ji}]$ is the coordinate translation matrix between scatterer i and scatterer j's local coordinate systems, whose entry at the pth row and qth column is given by

$$[R_{ji}]_{pq} = e^{\hat{\imath}(p-q)\theta_{ji}} H^{(1)}_{p-q}(kd_{ji}), \tag{6}$$

where (d_{ji}, θ_{ji}) are the polar coordinates in scatterer j's local coordinate system for o_i, the origin of scatterer i's local coordinate system. Note that $d_{ji}=|i-j|d$, and $\theta_{ij}=\pi/2$ when $i>j$ and $\theta_{ij}=3\pi/2$ when $i<j$.

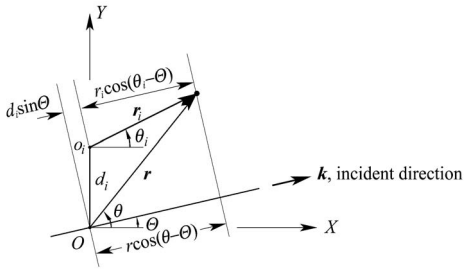


FIG. 2. Angled incident wave.

Assume a planar incident wave of unit amplitude propagates along a direction that forms an angle Θ with the X axis, that is,

$$\phi^{\text{inc}} = e^{\hat{\imath}kr\cos(\theta-\Theta)} = e^{\hat{\imath}k(x\cos\Theta + y\sin\Theta)}. \tag{7}$$

Without losing generality, the incident wave is expressible in the following wave expansion form:

$$\phi^{\text{inc}} = \{\boldsymbol{A}\}^T \{\boldsymbol{J}(r, \theta)\}. \tag{8}$$

The incident wave observed at o_i has a phase difference compared to that observed at the global origin O, as sketched in Fig. 2. This phase difference leads to the following relation between the wave expansion coefficients in local coordinate systems of two different scatterers i and j as

$$\{\boldsymbol{A}_j\} = e^{\hat{\imath}(j-i)kd\sin\Theta} \{\boldsymbol{A}_i\}. \tag{9}$$

Due to the infinity nature, all scatterers play the identical role, with the only difference being the different phase of the incident wave they are exposed to. This means that waves scattered by different scatterers only differ by the same phase difference, that is,

$$\{\boldsymbol{C}_j\} = e^{\hat{\imath}(j-i)kd\sin\Theta} \{\boldsymbol{C}_i\}. \tag{10}$$

Note that the global coordinate system is the same as scatterer 0's local coordinate system: $\{\boldsymbol{A}_0\}=\{\boldsymbol{A}\}$. Also, denote $\{\boldsymbol{C}_0\}=\{\boldsymbol{C}\}$. Expressing all waves in the global coordinate system, Eqs. (4) and (5) can be rewritten as

$$\phi^{\text{total}} = \{\boldsymbol{A}\}^T \{\boldsymbol{J}(r, \theta)\} + \{\boldsymbol{C}\}^T \sum_{i=-\infty}^{\infty} e^{\hat{\imath}ikd\sin\Theta} \{\boldsymbol{H}(r_i, \theta_i)\} \tag{11}$$

and

$$\{\boldsymbol{C}\} = [\boldsymbol{T}]\{\boldsymbol{A}\} + [\boldsymbol{T}]\left(\sum_{\substack{j=-\infty \\ j \neq 0}}^{\infty} [\boldsymbol{R}_{j0}]^T e^{\hat{\imath}jkd\sin\Theta} \right) \{\boldsymbol{C}\}, \tag{12}$$

where Eq. (10) specialized for the case $i=0$ has been used. Introducing the following matrix

$$[\boldsymbol{L}] = \sum_{\substack{j=-\infty \\ j \neq 0}}^{\infty} [\boldsymbol{R}_{j0}]^T e^{\hat{\imath}jkd\sin\Theta}, \tag{13}$$

which contains only the geometric information of the grating and is hence called the *lattice sum*, Eq. (12) can be solved formally as

$$\{\boldsymbol{C}\} = ([\boldsymbol{I}] - [\boldsymbol{T}][\boldsymbol{L}])^{-1} [\boldsymbol{T}] \{\boldsymbol{A}\}. \tag{14}$$

For each entry in the $[\boldsymbol{L}]$ matrix, the summation is performed in the way such that a pair of scatterers j and $-j$ is summed first, giving, for the pth row and qth column,

$$[L]_{pq} = \hat{\imath}^{p-q}\mathcal{H}_{p-q}(kd,\Theta), \tag{15}$$

where

$$\mathcal{H}_n(z,\Theta) = \sum_{i=1}^{\infty} H_n^{(1)}(iz)[(-1)^n e^{\hat{\imath}iz\sin\Theta} + e^{-\hat{\imath}iz\sin\Theta}], \tag{16}$$

and Eq. (6) has been used.

The summation in Eq. (16) is a *Schlömilch series* (Watson, 1966). In general, this series converges rather slowly, especially when z is small, which happens to be in the ranges where most of the interests in infinite gratings lie. Twersky (1961) developed an alternative representation of this series using elementary functions, that gives a better convergence. Twersky's representation and the author's experience in its computational implementation are summarized in Appendix A. In the context of the following discussions, it suffices to treat $\mathcal{H}_n(z,\Theta)$ as a special function, which becomes singular when $z(1\pm\sin\Theta)/(2\pi)$ is an integer.

C. The planar wave expansion basis

The total scattered wave, the second term on the right-hand side of Eq. (11), involves another Schlömilch series. Twersky (1962) also obtained an alternative expression using elementary functions for this series. Adapting Twersky's formula for the present notation is discussed in Appendix B. The resulting formula can be written as

$$\sum_{i=-\infty}^{\infty} e^{\hat{\imath}kd\sin\Theta} e^{\hat{\imath}n\theta_i} H_n(kr_i) = \frac{2}{kd}\sum_{m=-\infty}^{\infty} \frac{(\sin\alpha_m - \hat{\imath}(x/|x|)\cos\alpha_m)^n}{\cos\alpha_m} e^{\hat{\imath}k(|x|\cos\alpha_m + y\sin\alpha_m)}, \tag{17}$$

where the fraction $x/|x|$ simply denotes the sign of x,

$$\sin\alpha_m = \sin\Theta + m\frac{2\pi}{kd}, \tag{18}$$

and the corresponding $\cos\alpha_m$ is defined as

$$\cos\alpha_m = \begin{cases} \sqrt{1-\sin^2\alpha_m} & \text{when } |\sin\alpha_m| < 1, \\ \hat{\imath}\sqrt{\sin^2\alpha_m - 1} & \text{when } |\sin\alpha_m| > 1. \end{cases} \tag{19}$$

Similar to the Schlömilch series in Eq. (16), this series also diverges when $kd(1\pm\sin\Theta)/(2\pi)$ becomes an integer.

Define a column matrix $\{\boldsymbol{\mathcal{P}}(x,y)\}$ and a square matrix $[\boldsymbol{M}]$ whose entries are

$$\{\mathcal{P}(x,y)\}_m = e^{\hat{\imath}k(|x|\cos\alpha_m + y\sin\alpha_m)}, \tag{20}$$

and

$$[M]_{mn} = \frac{2}{kd}\frac{(\sin\alpha_m - \hat{\imath}(x/|x|)\cos\alpha_m)^n}{\cos\alpha_m}. \tag{21}$$

Then, the total wave in the field can be written as

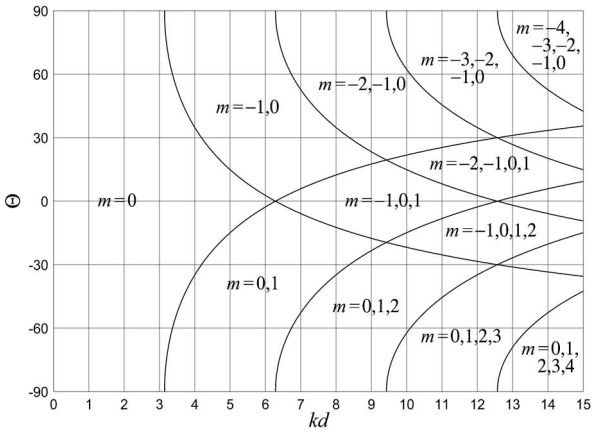


FIG. 3. Diagram of propagating modes in kd-Θ space. Each curve denotes the condition in which one of modes becomes propagating. Different modes that exist in areas partitioned by curves are shown.

$$\phi^{\text{total}} = e^{\hat{\imath}k(x\cos\Theta + y\sin\Theta)} + ([\boldsymbol{M}]\{\boldsymbol{C}\})^{\boldsymbol{T}}\{\boldsymbol{\mathcal{P}}(x,y)\}. \tag{22}$$

Matrix $\{\boldsymbol{\mathcal{P}}(x,y)\}$ has the appearance of planar waves and is hence called the *planar wave expansion basis*. Matrix $[\boldsymbol{M}]$ serves as the *mode converter* that converts cylindrical wave modes (in polar coordinate system) into planar wave modes (in Cartesian coordinate system).

1. The propagating modes

The planar wave modes that correspond to real values of α_m (or, $|\sin\alpha_m| < 1$) represent the *propagating modes* of the grating. In these modes, $-\pi/2 < \alpha_m < \pi/2$, and m is limited by

$$-m_- \le m \le m_+, \tag{23}$$

where

$$m_- = \left[\frac{kd(1+\sin\Theta)}{2\pi}\right], \quad m_+ = \left[\frac{kd(1-\sin\Theta)}{2\pi}\right], \tag{24}$$

and $[\cdot]$ denotes the largest integer no larger than the enclosed argument. In these modes, on the right half-plane ($x>0$), the transmitted wave propagates in a direction that forms an angle α_m with the x axis, and on the left half-plane ($x<0$), the reflected wave propagates in a direction that forms an angle $\pi-\alpha_m$ with the x axis. These propagation directions are independent of the material properties of the grating.

The number of propagating modes depends on the normalized frequency kd and the incident angle. The mode diagram in Fig. 3 shows the existence of different propagating modes on the $kd-\Theta$ plane. The delineating curves are determined by setting $\sin\alpha_m = 1$ in Eq. (18). Mode $m=0$ is always a propagating mode. The region in which $m=0$ is the only propagating mode is often called the *one-mode region*.

2. The evanescent modes

When m falls outside the range as defined in Eq. (23), α_m assumes a complex value. These modes are called the *evanescent modes* of the grating. For mode m, $\sin^2\alpha_m > 1$, the planar wave expansion basis in Eq. (20) becomes

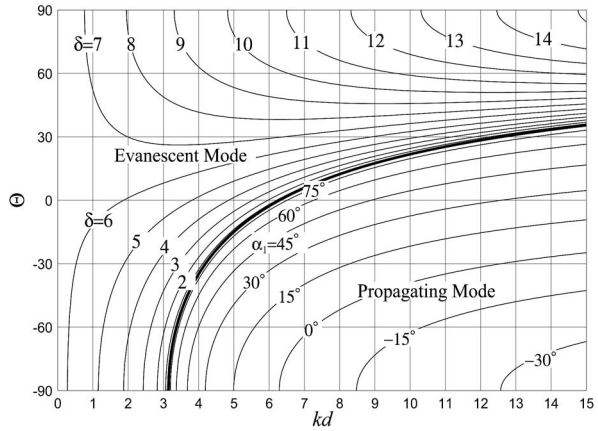


FIG. 4. Details of mode in $m=1$. The bold curve delineates propagating and evanescent regions. In the evanescent region, contours for δ, which characterizes spatial exponential decay of the mode in the function form $e^{-\delta|x|/d}$, are shown. In the propagating region, the contours for the direction of the transmitted wave are shown.

$$\{\mathcal{P}(x,y)\}_m = e^{-|x|\sqrt{(k\sin\Theta + 2m\pi/d)^2 - k^2}} e^{\hat{\imath}y(k\sin\Theta + 2m\pi/d)}. \quad (25)$$

Equation (25) suggests that, along the grating (the y direction), the mode shape is periodic. When the incident wave is normal to the grating, the spatial period is $d/|m|$. With an angled incident wave, the period varies with the frequency: it is shortened as the frequency increases for modes that propagate in directions that fall within the acute region between the incident direction and the grating; it is lengthened as the frequency increases in the remaining region.

More importantly, in the direction normal to the grating (the x direction), the mode shape decays exponentially as $|x|$ increases; higher modes decay faster than the lower modes. Define

$$\delta = \sqrt{(kd\sin\Theta + 2m\pi)^2 - (kd)^2}. \quad (26)$$

Then, the extent of an evanescent mode in the x direction can be quantitatively described by $e^{-\delta|x|/d}$, which means that, at a distance d from the grating, the evanescent mode has decayed to a fraction of $e^{-\delta}$.

For mode $m=1$, Fig. 4 shows the characteristics of this mode in the $kd-\Theta$ space. The bold curve, which appears in Fig. 3, delineates the propagating and evanescent regions. Contours for δ are shown in the evanescent region, and contours for the propagating direction α_1 are shown in the propagating region. For mode $m=-1$, the picture is Fig. 4 flipped about $\Theta=0°$ axis. For all other $m\neq 0$ modes, the picture is similar to mode $m=\pm 1$.

Figure 4 suggests that an evanescent mode has only limited presence within a small neighborhood near the delineating curve. An evanescent mode evolves with the increase of frequency as the following: the mode starts as an evanescent mode at extremely low frequencies. Its spatial extent increases as the frequency increases and eventually becomes a propagating mode traveling along the grating—the Wood's anomaly as noted by Twersky (1961). Afterwards, its propagation direction approaches to that of the incident wave as the frequency continues to increase.

D. Exact solution in planar wave expansion basis

Often, it is more convenient to split the unified notation in Eq. (22) into separated expressions for the transmitted and the reflected waves. Define the *forward planar wave expansion basis* $\{\boldsymbol{\mathcal{P}}_+(x,y)\}$ and the *backward planar wave expansion basis* $\{\boldsymbol{\mathcal{P}}_-(x,y)\}$ with entries

$$\{\mathcal{P}_\pm(x,y)\}_m = e^{\hat{\imath}k(\pm x\cos\alpha_m + y\sin\alpha_m)}. \quad (27)$$

Then, the transmitted and the reflected waves can be expressed as

$$\phi^{\text{transmitted}} = \{\boldsymbol{\mathcal{C}}_+\}^T\{\boldsymbol{\mathcal{P}}_+(x,y)\},$$

$$\phi^{\text{reflected}} = \{\boldsymbol{\mathcal{C}}_-\}^T\{\boldsymbol{\mathcal{P}}_-(x,y)\}, \quad (28)$$

where $\{\boldsymbol{\mathcal{C}}_+\}$ and $\{\boldsymbol{\mathcal{C}}_-\}$ are planar wave expansion coefficients for the transmitted and the reflected waves, respectively. They are determined by

$$\{\boldsymbol{\mathcal{C}}_+\} = [\boldsymbol{M}_+]\{\boldsymbol{C}\} + \{\boldsymbol{\mathcal{A}}\}, \quad \{\boldsymbol{\mathcal{C}}_-\} = [\boldsymbol{M}_-]\{\boldsymbol{C}\}, \quad (29)$$

where

$$[M_\pm]_{mn} = \frac{2}{kd}\frac{(\sin\alpha_m \mp \hat{\imath}\cos\alpha_m)^n}{\cos\alpha_m}, \quad (30)$$

and $\{\boldsymbol{\mathcal{A}}\}$ is the planar wave expansion coefficient of the incident wave. For the incident wave given in Eq. (8), $\{\mathcal{A}\}_0 = 1$ and $\{\mathcal{A}\}_m = 0$ for all other m.

Recall that both Schlömilch series become singular when $kd(1\pm\sin\Theta)/(2\pi)$ becomes an integer, in which case, $\cos\alpha_m = 0$. To observe the behavior of the exact solution near the singularity point, Eqs. (14) and (29) can be combined to write

$$\{\boldsymbol{\mathcal{C}}_\pm\} = [\boldsymbol{M}_\pm]([\boldsymbol{I}] - [\boldsymbol{T}][\boldsymbol{L}])^{-1}[\boldsymbol{T}]\{\boldsymbol{A}\} + \{\boldsymbol{\mathcal{A}}\}, \quad (31)$$

where $\{\boldsymbol{\mathcal{A}}\}$ appears only for the case with the + sign. Matrices $[\boldsymbol{M}_\pm]$ and $[\boldsymbol{L}]$ become singular simultaneously, due to the term $1/\cos\alpha_m$ in Eq. (30) for $[\boldsymbol{M}_\pm]$ and in Eq. (15) [and, in turn, Eqs. (A2), (A3), and (A6)] for $[\boldsymbol{L}]$. Near the singular point, as the near-singular terms become dominant, keeping only the dominant terms, Eq. (31) becomes

$$\lim_{\cos\alpha_m\to 0}\{\boldsymbol{\mathcal{C}}_\pm\} = -[\boldsymbol{M}_\pm][\boldsymbol{L}]^{-1}\{\boldsymbol{A}\} + \{\boldsymbol{\mathcal{A}}\}. \quad (32)$$

The common factor $1/\cos\alpha_m$ in the $[\boldsymbol{M}_\pm]$ and $[\boldsymbol{L}]$ matrices cancels each other due to the inversion of $[\boldsymbol{L}]$ matrix. As a result, there is no singularity in the final expressions for waves.

E. The approximate solution

Since there is only a limited number of planar propagating wave modes, it immediately becomes conceivable to approximate the wave fields by accounting for only these propagating modes. Such an approximation can be readily achieved by reducing the size of both the planar wave expansion coefficient matrices and the corresponding planar wave expansion bases so that the index m only runs from $-m_-$ to m_+. Performing such size reduction does not change the matrix expressions for the transmitted and reflected waves in Eq. (28).

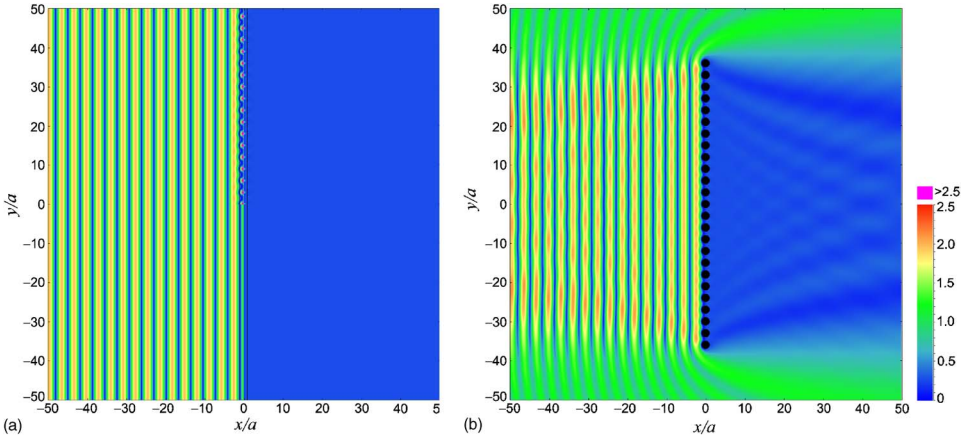


FIG. 5. (Color online) Comparison of the infinite grating solution (a) with the multiple-scattering solution for finite grating (b) when $kd=3$ ($ka=1$). In (a), the lower half-space shows the approximate solution that accounts only for propagating modes; the black lines denote the extent of grating itself.

F. Numerical examples

In the following, comparisons among the exact and approximate infinite grating solutions and the multiple-scattering solution for finite gratings are made through numerical examples. The gratings are comprised of circular steel cylinders embedded in epoxy host. The multiple-scattering solutions are obtained for a grating of 25 cylinders via a computational system that has previously been verified to be analytically exact (Cai and Williams, 1999a, b). The shear modulus is 1.17 GPa for the epoxy and 200 GPa for the steel. The mass density is 1100 kg/m^3 for the epoxy and 7800 kg/m^3 for the steel. Unless otherwise noted, the spacing between any two adjacent cylinders in a grating is $d=3a$, where a is the radius of the cylinders, and all materials are assumed to be lossless.

Figures 5–7 compare the displacement amplitude fields in the vicinity of the grating at normalized frequencies $kd=3,6$, and 9 ($ka=1$, 2, and 3), respectively. In the figures for the infinite grating solutions, the exact solutions are shown in the upper half-space and the approximation solutions are shown in the lower half-space; a pair of vertical thin black lines denotes the extent of the grating itself ($|x|=a$).

When $kd=3$ and 6, there is only one propagating mode. When $kd=3$, the approximate and the exact solutions are identical, except for a slight difference in a small region near the grating. By the first crest in the backward direction, which is located at approximately a half wavelength from the centerline, all the evanescent modes have diminished. In comparison, when $kd=6$, the evanescent modes extend into the fourth crest in the backward direction, which is located approximately at $|x|=2\lambda\approx 2.09d$. This is consistent with Fig. 4, which gives a nondimensional characteristic length $\delta\approx 2$ for the first evanescent mode $m=\pm 1$.

When $kd=9$, there are three propagating modes. Waves in these modes interfere, forming a weave pattern which does not give a clear indication of propagating directions of these modes. All evanescent modes have sufficiently decayed outside the region occupied by the grating, and the approximate solutions are identical to the exact solutions.

In the multiple-scattering solutions for the corresponding finite gratings, edge effects due to the finite size of the grating are prevalent. For the cases of one propagating mode ($kd=3$ and 6), the edge effects appear in the forward direction as multiple streaks at different angles with respect to the x axis. In the backward direction, these streaks are superimposed onto other waves to become lumps on crests. The edge effects are minimal in a small region $|x|/a<10$ and $|y|/a<25$, where the multiple-scattering solutions agree with the infinite grating solutions. For the cases of three propagating modes ($kd=9$), the region having minimal edge effects appears as a diamond-shaped region, with the extent in the x direction increasing as frequency increases.

Figure 8 compares the displacement transmission spectra as computed by the infinite grating solution and by the multiple-scattering solution. The infinite grating spectrum is determined by $|\{\mathcal{C}_+\}_0|$, which represents the amplitude of the

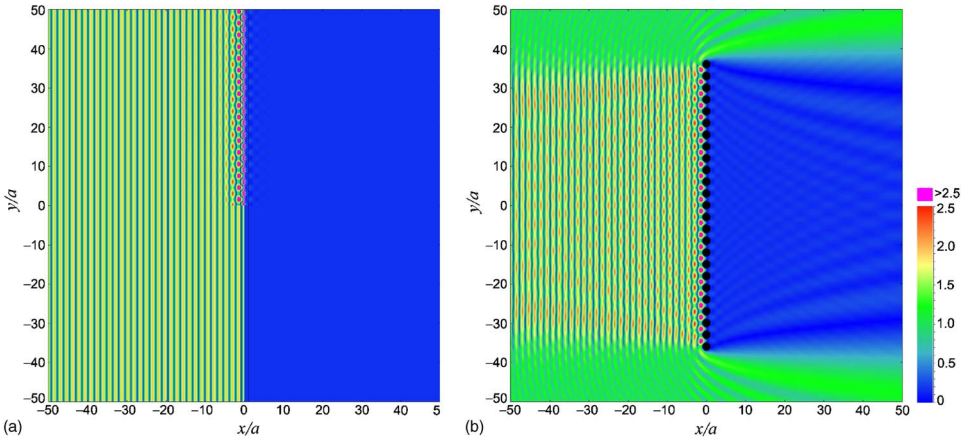


FIG. 6. (Color online) Comparison of the infinite grating solution (a) with the multiple-scattering solution for finite grating (b) when $kd=6$ ($ka=2$). In (a), the lower half-space shows the approximate solution that accounts only for propagating modes; the black lines denote the extent of grating itself.

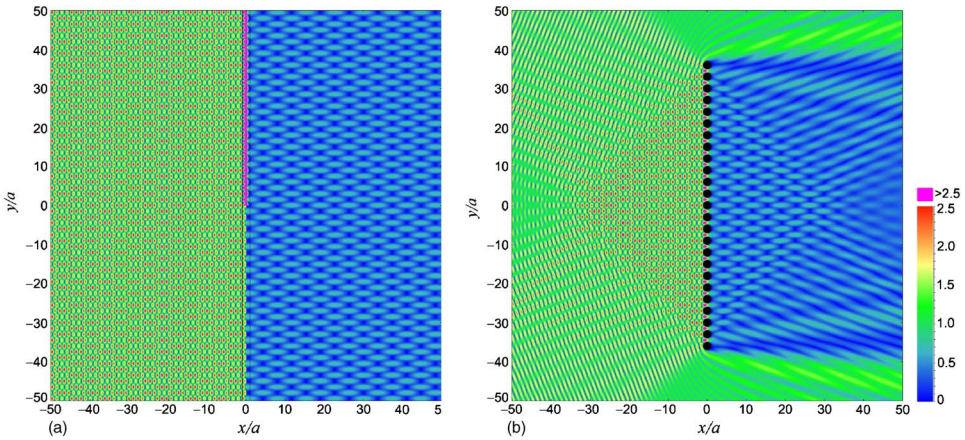


FIG. 7. (Color online) Comparison of the infinite grating solution (a) with the multiple-scattering solution for finite grating (b) when $kd=9$ ($ka=3$). In (a), the lower half-space shows the approximate solution that accounts only for propagating modes; the black lines denote the extent of grating itself.

propagating mode $m=0$. The spectrum for the multiple-scattering solution is obtained by averaging the displacement amplitude along $x/a=10$ in the range $y/d\in[0,10]$. The x and y are so chosen such that the evanescent modes have sufficiently decayed and the edge effects are minimal.

The two spectra shown in Fig. 8 are extremely close, since both solutions are exact for the respective problems. At certain very low frequencies ($kd<0.4$), the multiple-scattering spectrum exceeds unity. This is due to the edge effects, which are more prevalent in lower frequencies than in higher frequencies. For the infinite grating problem, there is no such effect and hence the transmission coefficient never exceeds unity.

When there is more than one propagating mode, a meaningful single transmission coefficient needs to account for the propagating directions of different modes. This is rather difficult to calculate for the multiple-scattering solution. Hence, the spectra shown in the range between $kd=2\pi$ and $kd=7.5$ ($ka=2\pi/3$ and $ka=2.5$) do not have much physical meaning, except to indicate, by the closeness of two curves, that the contributions from the propagating modes $m=\pm1$ are very small. This portion of spectra is included primarily to demonstrate that there is no computational difficulty near the singularity point of the Schlömilch series.

III. LAYERED MULTIPLE-SCATTERING METHOD

Using the approximate solution for the infinite grating problem, wave interactions with the grating can be viewed in simple terms such as transmission and reflection coefficients, analogous to those for a thin homogeneous panel separating an infinite host medium. Consequently, an array of regularly arranged scatterers can be viewed as comprising of multiple layers of infinite gratings, and the approximate solution for the infinite grating problem would enable the problem to be treated as multiple transmissions and reflections occurring in a multilayer panel. This approach, called the *layered multiple-scattering method*, would significantly simplify the analysis.

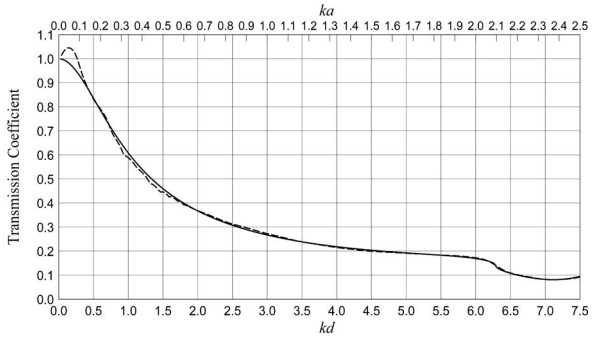


FIG. 8. Comparison of wave transmission spectrum as obtained by the infinite grating solution (solid curve) and the multiple-scattering solution (dashed curve).

A. Transmission and reflection coefficients in regularly arranged scatterer array

As an example to illustrate the layered multiple-scattering method, a recursive procedure for calculating the wave transmission and reflection coefficients is outlined in the following for a case in which all gratings are identical, and the distance between any two adjacent layers is b. The consideration is further limited to the one-mode region. A set of recursive formulas was obtained by Huang and Heckl (1993) for the case when all layers are aligned to form a rectangular grid.

A more general case is considered here. Layers are numbered from left to right. Each layer has its own local coordinate system, denoted as (x_p,y_p) for the pth layer. A global coordinate system (x,y) is defined to coincide with the first layer's. Layers are arranged such that their y positions are shifted with respect to the x axis, by an amount $\Delta_p=[1+(-1)^p]\Delta/2$ for the pth layer. That is, there is no shift for odd-numbered layers and Δ for even-numbered layers. It is further assumed that the transmission and reflection coefficients for individual gratings have been obtained. These coefficients are in fact wave expansion coefficients associated with individual local coordinate systems, and $T=\{\mathcal{C}_+\}_0$ and $R=\{\mathcal{C}_-\}_0$.

Assume the first $p-1$ layers have been analyzed, yielding the cumulative transmission and reflection coefficients $\mathcal{T}_{p-1}$ and $\mathcal{R}_{p-1}$, respectively. The task at hand is to obtain the cumulative transmission and reflection coefficients, $\mathcal{T}_p$ and $\mathcal{R}_p$, respectively, for the first p layers. All cumulative transmission and reflection coefficients are associated with the global coordinate system. Obviously, $\mathcal{T}_1=T$ and $\mathcal{R}_1=R$. Figure 9(a) shows the geometry of the problem, and Fig. 9(b)

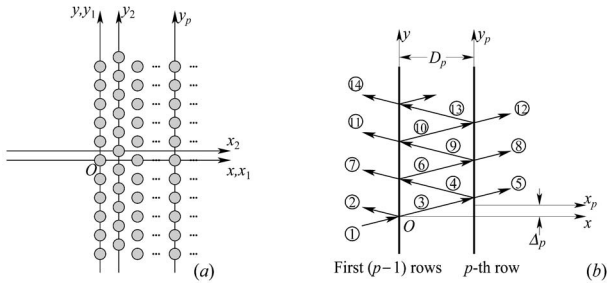


FIG. 9. Process for pth step recursive analysis for a multiple layer gratings. First $p-1$ layers have been analyzed and treated as one single layer. (a) Scatterer arrangement. (b) Schematic of analysis.

illustrates the multiple transmission and reflection processes between the first $p-1$ layers and the pth layer, as well as the relation between the two coordinate systems. Note that $x=x_p+D_p$, $y=y_p+\Delta_p$, and $D_p=(p-1)b$.

The incident wave 1 is assumed to be a planar wave as expressed in Eq. (8). Waves 2 and 3 are the reflected and transmitted waves by the first $(p-1)$ layers, which are expressible as

$$\phi_2 = \mathcal{R}_{p-1} e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)},$$

$$\phi_3 = \mathcal{T}_{p-1} e^{\hat{\imath}k(D_p\cos\Theta + \Delta_p\sin\Theta)} e^{\hat{\imath}k(x_p\cos\Theta + y_p\sin\Theta)},$$

where wave 3 is expressed in the (x_p, y_p) coordinate system, in which it incidents upon layer p, producing waves 4 and 5. Thus,

$$\phi_4 = R\mathcal{T}_{p-1} e^{\hat{\imath}k(D_p\cos\Theta + \Delta_p\sin\Theta)} e^{\hat{\imath}k(-x_p\cos\Theta + y_p\sin\Theta)}$$

$$= R\mathcal{T}_{p-1} e^{2\hat{\imath}kD_p\cos\Theta} e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}.$$

Note that Δ_p disappears when wave 4 is expressed in the global coordinate system. Afterwards, wave 4 incidents upon the first $p-1$ layers, producing waves 6 and 7. In this process, the incident wave encounters the right surface of the first $p-1$ layers. If $p-1$ is even, the first $p-1$ layers are skew symmetric about their geometric center. If $p-1$ is odd, the first $p-1$ layers are symmetric about the central layer. In both cases, the resulting waves can be identically expressed as

$$\phi_6 = \mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{2\hat{\imath}kb\cos\Theta} e^{\hat{\imath}k(x\cos\Theta + y\sin\Theta)},$$

$$\phi_7 = \mathcal{T}_{p-1} R\mathcal{T}_{p-1} e^{2\hat{\imath}kD_p\cos\Theta} e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}.$$

Expressions for all subsequent waves can be similarly obtained, as tabulated in Table I.

The total transmitted wave consists of waves 5, 8, 12, and higher order terms. Hence, adding the wave expansion coefficients as listed in Table I gives

$$\mathcal{T}_p = \frac{\mathcal{T}_{p-1} T}{1 - \mathcal{R}_{p-1} R e^{2\hat{\imath}kb\cos\Theta}}, \tag{33}$$

where the following Taylor expansion for $(1-x)^{-1} = 1 + x + x^2 + x^3 + \cdots$ has been used. Similarly the reflection coefficient for the first p layers can be found as

TABLE I. Expressions for waves depicted in Fig. 13.

Wave	Coefficient	Basis
1	1	$e^{\hat{\imath}k(x\cos\Theta + y\sin\Theta)}$
2	$\mathcal{R}_{p-1}$	$e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}$
3	$\mathcal{T}_{p-1} e^{\hat{\imath}kD_p\cos\Theta}$	$e^{\hat{\imath}k(x_p\cos\Theta + y_p\sin\Theta)}$
4	$R\mathcal{T}_{p-1} e^{\hat{\imath}kD_p\cos\Theta}$	$e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}$
5	$T\mathcal{T}_{p-1} e^{\hat{\imath}kD_p\cos\Theta}$	$e^{\hat{\imath}k(x_p\cos\Theta + y_p\sin\Theta)}$
6	$\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}k(D_p+2b)\cos\Theta}$	$e^{\hat{\imath}k(x_p\cos\Theta + y_p\sin\Theta)}$
7	$\mathcal{T}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}k2D_p\cos\Theta}$	$e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}$
8	$T\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}k(D_p+2b)\cos\Theta}$	$e^{\hat{\imath}k(x_p\cos\Theta + y_p\sin\Theta)}$
9	$R\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}2k(D_p+2b)\cos\Theta}$	$e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}$
10	$\mathcal{R}_{p-1} R\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}k(D_p+4b)\cos\Theta}$	$e^{\hat{\imath}k(x_p\cos\Theta + y_p\sin\Theta)}$
11	$\mathcal{T}_{p-1} R\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}2k(D_p+b)\cos\Theta}$	$e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}$
12	$T\mathcal{R}_{p-1} R\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}k(D_p+4b)\cos\Theta}$	$e^{\hat{\imath}k(x_p\cos\Theta + y_p\sin\Theta)}$
13	$R\mathcal{R}_{p-1} R\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}k(2D_p+4b)\cos\Theta}$	$e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}$
14	$\mathcal{T}_{p-1} R\mathcal{R}_{p-1} R\mathcal{R}_{p-1} R\mathcal{T}_{p-1} e^{\hat{\imath}k(2D_p+4b)\cos\Theta}$	$e^{\hat{\imath}k(-x\cos\Theta + y\sin\Theta)}$

$$\mathcal{R}_p = \mathcal{R}_{p-1} + \frac{\mathcal{T}_{p-1}^2 R e^{2\hat{\imath}(p-1)kb\cos\Theta}}{1 - \mathcal{R}_{p-1} R e^{2\hat{\imath}kb\cos\Theta}}. \tag{34}$$

Expressions in Eqs. (33) and (34) are equivalent to those obtained by Huang and Heckl (1993). However, the present derivation allows position shifts in y position. The fact that these position shifts Δ_p do not enter the final expressions suggests a possible further generalization, to the first approximation: that any shifts in the y positions, regular or irregular, would have no effects. With an approximation that takes only the propagating modes into account, an infinite grating becomes a mathematical interface with homogeneous properties. This means that the interlayer distance is the only parameter that determines the behavior of the scatterer array.

B. Examples

Consider two arrays of scatterers constructed from two and four layers of equally spaced infinite gratings, respectively, and $b=d$. The approximate solutions obtained by the layered multiple-scattering method are compared with analytically exact multiple-scattering solutions for the corresponding arrays comprised of finite gratings of 25 scatterers in each grating. No exact multiple-scattering solution for multiple infinite gratings is available for this comparison because the Schlömilch series are only applicable to a single infinite grating.

The displacement transmission spectra are shown in Figs. 10 and 11 for the cases of two and four gratings, respectively. The spectra for the exact multiple-scattering solution for the finite gratings are the averaged displacement amplitude measured at a distance $x/a=10$ from the last layer in the forward direction. Figures 10 and 11 show that the curves for the two solutions in general have excellent agreement. As expected, as the number of layers increases, a major stop band is formed, giving a quiescent field in the forward direction for $kd>1.5$ $(ka>0.5)$. At very low frequencies, the transmission coefficient computed by the exact multiple-scattering solution for the finite gratings again occasionally exceeds unity, due to the diffracted waves.

There are two distinctive features in these spectra that do not appear in the corresponding single-grating spectra and

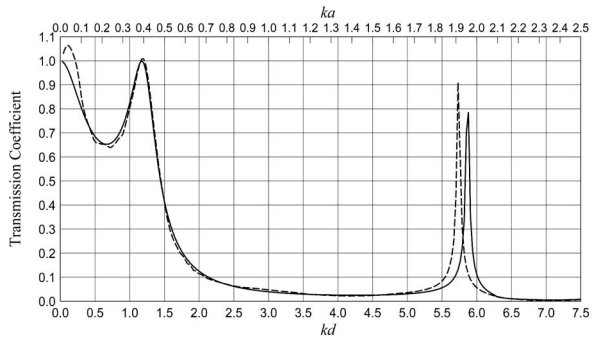


FIG. 10. Comparison of wave transmission spectrum for two layers of grating as obtained by the layered multiple-scattering solution for infinite gratings (solid curve) and the multiple-scattering solution for finite gratings (dashed curve).

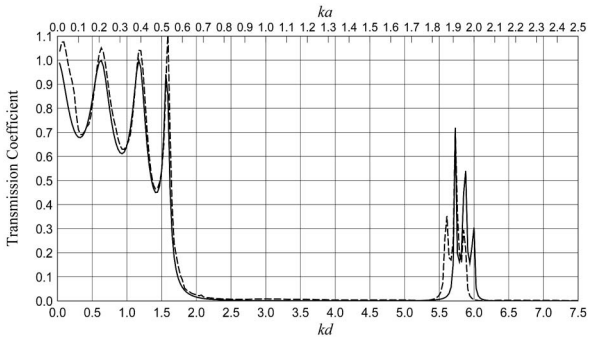


FIG. 11. Comparison of wave transmission spectrum for four layers of grating as obtained by the layered multiple-scattering solution for infinite gratings (solid curve) and the multiple-scattering solution for finite gratings (dashed curve).

thus are attributed entirely to the nature of multiple layers. The first is several peaks and valleys in the low-frequency range before the stop band's cutoff frequency. The number of these valleys and spikes equals the number of layers minus one. This has been observed by a number of similar studies, such as Kristiansen and Fahy (1972), Vovk *et al.* (1978), and Radlinski (1989). The second is the same number of closely spaced resonance spikes between $kd=5.5$ and 6.0, and there is a noticeable shift in peak frequencies between the two solutions.

Figure 12 compares the displacement amplitude distributions at $kd=6$ ($ka=2$) for the cases of two and four gratings. In each case, the multiple-scattering solution for finite gratings is shown in the upper half-space, and the layered multiple-scattering solution is shown in the lower half-space. This particular frequency is chosen because the evanescent mode $m=\pm 1$ has a large extent at this frequency, as observed earlier in Fig. 6.

Figure 12 shows that there is a slight difference between the exact and approximate solutions in displacement amplitude in the forward direction, which is also observable from the spectra in Figs. 10 and 11. This slight difference is completely overwhelmed by other more dominant effects in the backward direction. There are periodic lumps riding atop the crests in the backward direction in the exact multiple-scattering solution. However, a closer comparison with the corresponding single-grating case in Fig. 6 suggests that these lumps are mostly due to the edge effects, and the lumps due to the evanescent modes are again limited to the first three to four crests. In other words, the extent of the evanescent modes does not increase as the number of layers increases.

C. Effects of scatterer arrangement

The geometric parameters used in the layered multiple-scattering method include only the scatterer spacing within a grating d and the interlayer spacing b. This implies that scatterer arrangements such as the ones in Fig. 13, in which one is a rectangular arrangement and the other is a triangular arrangement but both have an interlayer spacing of $\sqrt{3}d/2$, would produce the same result. Note that the two configurations in Fig. 13 represent the extremes of possible variations in scatterer arrangement under the given constraints.

In Fig. 14, the transmission spectrum calculated by the layered multiple-scattering method is compared with the spectra obtained by the exact multiple-scattering solution for finite gratings for both arrangements. Figure 14 shows that the three spectra are indeed very close to each other. All the peaks appear to have shifted to higher frequencies when compared to the similar setting when $b=d$ as in Fig. 10. Specifically, comparing the peaks by the multiple-scattering solution, the peak at $kd\approx 5.73$ in Fig. 10 is shifted to $kd\approx 6.54$ in Fig. 14—a shift by a factor of 1.14; the peak at $kd\approx 1.20$ in Fig. 10 is shifted to $kd\approx 1.44$ in Fig. 14—a shift by a factor of 1.20. Both shift factors are very close to 1.155,

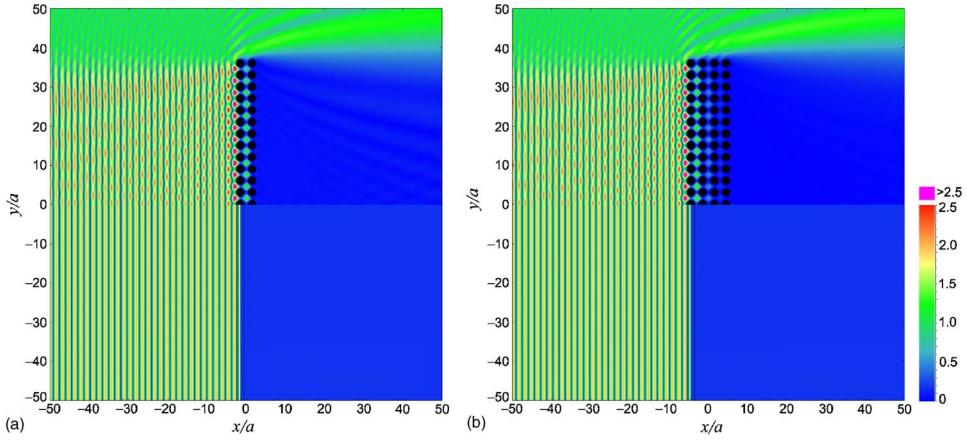


FIG. 12. (Color online) Comparison of the layered multiple-scattering solution with the exact multiple-scattering solution for (a) two layers of gratings and (b) four layers of gratings, when $kd=6$ ($ka=2$). The multiple-scattering solution is shown in the upper half-space and the layered multiple scattering solution is shown in lower half-space.

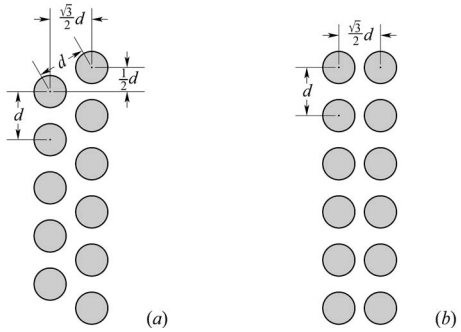


FIG. 13. Different arrangements with same interlayer distance of $\sqrt{3}d/2$: (a) triangular arrangement and (b) rectangular arrangement.

the ratio between the interlayer distances in the two cases. Therefore, the interlayer distance b is the overriding parameter, at least to the first approximation, that affects the features in the spectra due to the nature of multiple layers.

There are two noticeable differences among the three spectra. The first is the peak at $kd \approx 6.54$ only appears in the case of rectangular arrangement using multiple-scattering solution. This peak occurs in the frequency range where there are three propagating modes, for which the layered multiple-scattering solution is no longer valid. The second difference is the slight shift of the peak frequencies at $kd \approx 1.4$: all three spectra give slightly different peak frequencies.

D. Effects of evanescent modes

The evanescent modes are the only difference between the exact infinite grating solution and the approximate infinite grating solution. Between the exact multiple-scattering solution and the layered multiple-scattering solution, the differences include the edge effect due to the finite grating used in the former and the omission of the evanescent modes in the latter. Examples so far suggest that the shifts in various key frequencies appear to be the most prominent effect of the evanescent modes. Such a shift occurs in the case of two or more layers, and the shift does not seem to increase as the number of layers increases.

The simpler case of two gratings offers an elementary scenario for analyzing the mechanism of the frequency shift in the approximate solutions. Mathematically, a peak would occur when the denominator in Eq. (33) approaches zero or an extreme. Denote the single grating reflection coefficient as $R=|R|e^{i\psi}$, where ψ represents the phase in the reflected wave. For a normal incidence ($\Theta=0°$), a peak would occur if $|R|^2e^{2i(\psi+kb)}$ approaches to unity; this would require that $|R|$ be sufficiently close to unity, and that

$$\psi + kb = n\pi, \tag{35}$$

where n is any integer. Equation (35) agrees with an earlier observation that the interlayer spacing b is the dominant factor in determining the peak frequencies.

In the multiple layer setting, the presence of an evanescent mode would affect the way the two layers interact. The most likely effect of the evanescent mode is in the form of a small correction to the phase ψ. According to Eq. (35), this would cause the peak frequencies to shift slightly. Furthermore, this correction to ψ is independent of the number of layers in the scatterer arrangement, and hence the peak frequency shift is not affected by the number of layers.

The phase shift due to evanescent modes can also explain the peak frequency shift in the spectra for triangular and rectangular scatterer arrangements obtained via the multiple-scattering solution. In these two arrangements, the square arrangement can be viewed as the result of shifting the second layer in the triangular arrangement by a distance $d/2$ in the y direction. In the presence of the evanescent mode m, according to Eq. (25), the mode shape is periodic in the y direction, and the geometrical shift of the second layer would result in a phase shift of $e^{im\pi}$ in the expression for the evanescent mode, which would in turn result in another slight correction to the phase ψ.

It is also noted that the problem under consideration is an elastic SH wave problem. In such a problem, scatterers undergo pure shear deformation and no net volume change would occur. As reported by others (such as Vovk and Grinchenko, 1978; Radlinski, 1989), shell-like scatterers undergoing the breath-mode deformation might cause significant intergrating interactions.

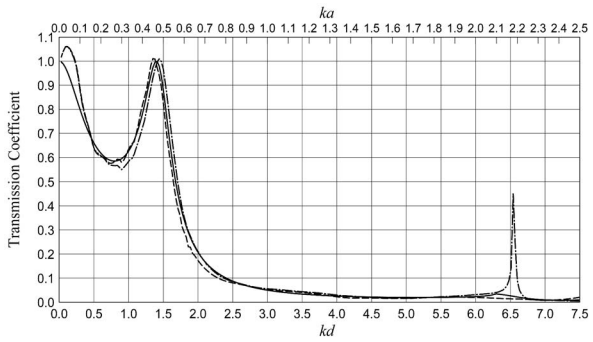


FIG. 14. Transmission spectra for different fiber arrangements with the same interlayer distance. Solid curve: layered multiple scattering solution; dashed curve: triangular arrangement; dot-dashed curve: rectangular arrangement.

IV. CONCLUSIONS

In this paper, the layered multiple-scattering method for analyzing the scattering from an array of regularly arranged scatterers is evaluated by comparing with exact solutions obtained by other means. In this method, each layer is an infinite grating, and the multiple-scattering process is treated as a multiple transmission-reflection process in a multilayer panel.

One of the exact solutions used in the evaluation is a general multiple-scattering theory. The theory uses a T matrix to represent a scatterer, which opens the possibility to use abstract scatterers for the analysis. The other exact solution is the exact solution for an infinite grating. This solution is obtained by combining the T-matrix formulation of the multiple scattering theory and the alternative expressions for Schlömilch series (Twersky, 1961). The infinity nature enables the scattered waves due to an planar incident wave to be expressed in a planar wave expansion basis using a Cartesian coordinate system. The resulting waves are further di-

vided into propagating and evanescent modes. Details of these modes are analyzed. Numerical examples suggest that only the first evanescent modes would have significant presence in a limited frequency range.

The layered multiple-scattering method only accounts for the propagating modes in each grating. This method not only provides an excellent approximation for most frequency ranges, but also significantly reduces the computational complexity. However, this method may give a slight shift in peak frequencies in the spectra. Numerical examples also show that, in a regular arrangement of scatterers, the interlayer distance is the dominant parameter that affects the features due to the multi-layer setting, such as the critical frequencies of a stopband.

ACKNOWLEDGMENTS

The author gratefully acknowledges the support of the National Science Foundation for this research under Grant No. CMS-0510940. The program manager is Dr. Y.-W. Chung. The author wishes to express his sincere appreciation to an anonymous reviewer for providing many insightful and enlightening comments, as well as several important references.

APPENDIX A: FIRST SCHLÖMILCH SERIES

The first Schlömilch series used in an infinite grating problem is in the computation of the wave expansion coefficients, as in Eq. (16);

$$\mathcal{H}_n(x,\Theta)=\sum_{s=1}^{\infty} H_n(sx)[(-1)^n e^{\hat{\imath}sx\sin\Theta}+e^{-\hat{\imath}sx\sin\Theta}]. \quad \text{(A1)}$$

Twersky (1961) obtained an alternative representation for the series. For $n>0$,

$$\mathcal{H}_{2n}(x,\Theta)=2\sum_{m=0}^{\infty}\frac{e^{-\hat{\imath}2n\alpha_m}}{x\cos\alpha_m}+2\sum_{m=-\infty}^{-1}\frac{e^{\hat{\imath}2n\alpha_m}}{x\cos\alpha_m}+\frac{\hat{\imath}}{n\pi}+\frac{\hat{\imath}}{\pi}\sum_{s=1}^{n}\frac{(-1)^m 2^{2m}(n+m-1)!\,B_{2m}(\Delta\sin\Theta)}{(2m)!\,(n-m)!\,\Delta^{2m}}, \quad \text{(A2)}$$

$$\mathcal{H}_{2n+1}(x,\Theta)=2\sum_{m=0}^{\infty}\frac{e^{-\hat{\imath}(2n+1)\alpha_m}}{x\cos\alpha_m}-2\sum_{m=-\infty}^{-1}\frac{e^{\hat{\imath}(2n+1)\alpha_m}}{x\cos\alpha_m}+\frac{2}{\pi}\sum_{s=0}^{n}\frac{(-1)^m 2^{2m}(n+m)!\,B_{2m+1}(\Delta\sin\Theta)}{(2m+1)!\,(n-m)!\,\Delta^{2m+1}}, \quad \text{(A3)}$$

where $\Delta=x/(2n)$, $B_n(x)$ is the Bernoulli polynomial, and α_m are defined by a pair of functions $\sin\alpha_m$ and $\cos\alpha_m$ in Eqs. (18) and (19).

Note that α_m become complex when $|\sin\alpha_m|>1$. For such cases, Twersky (1961) defined $\alpha_m=\pi/2-\hat{\imath}|\eta_m|$ when $m>0$ and $\alpha_m=-\pi/2+\hat{\imath}|\eta_m|$ when $m<0$, where $\cosh\eta_m=\sin\Theta+2m\pi/x$. In the implementation for this paper, it is found that, without introducing η_m, $\sin\alpha_m$ and $\cos\alpha_m$ in Eqs. (18) and (19) in conjunction with the following identify suffice:

$$e^{\pm\hat{\imath}n\alpha_m}=(e^{\pm\hat{\imath}\alpha_m})^n=(\cos\alpha_m\pm\hat{\imath}\sin\alpha_m)^n. \quad \text{(A4)}$$

For $n=0$, according to Twersky (1961), $\mathcal{J}_0(x,\Theta)$ and $\mathcal{Y}_0(x,\Theta)$ are obtained by Magnus and Oberhettinger as

$$\mathcal{H}_0(x,\Theta)=\mathcal{J}_0(x,\Theta)+\hat{\imath}\mathcal{Y}_0(x,\Theta). \quad \text{(A5)}$$

Again, without using η_m, $\mathcal{J}_0(x,\Theta)$ and $\mathcal{Y}_0(x,\Theta)$ can be equivalently written as

$$\mathcal{J}_0(x,\Theta)=\frac{2}{x}\sum_{m=-m_-}^{m^+}\frac{1}{\cos\alpha_m}-1, \quad \text{(A6)}$$

$$\mathcal{Y}_0(x,\Theta)=-\frac{2}{\pi}\left(\gamma+\ln\frac{x}{4\pi}\right)+\frac{1}{\pi}\left(\sum_{m=1}^{m_+}\frac{1}{m}+\sum_{m=1}^{m_-}\frac{1}{m}\right)-2\sum_{m=m_++1}^{\infty}\left(\frac{1}{x\sqrt{\sin^2\alpha_m-1}}-\frac{1}{2\pi m}\right)-2\sum_{m=-\infty}^{-m_--1}\left(\frac{1}{x\sqrt{\sin^2\alpha_m-1}}+\frac{1}{2\pi m}\right), \quad \text{(A7)}$$

where m_- and m_+ are defined in Eq. (24), and $\gamma=0.5772156649\ldots$ is the Euler constant. Although the above expression could be further simplified by noting that the summations over $1/m$ in fact cover the entire integer range, the split expression as shown in Eq. (A7) ensures a better convergence.

For $n<0$, the following relation is used:

$$\mathcal{H}_{-n}(x,\Theta)=(-1)^n\mathcal{H}_n(x,\Theta). \quad \text{(A8)}$$

In the implementation, it is found that when x is large (such as $x>50$), especially when $x\sin\Theta$ and n are also large (such as $x\sin\Theta>35$ and $n>50$), the performance of Twersky's series deteriorates. In such cases, using the following asymptotic expression for the Hankel functions generally produces better results:

$$H_n(z)\simeq\sqrt{\frac{2}{\pi z}}e^{\hat{\imath}(z-n\pi/2-\pi/4)}. \quad \text{(A9)}$$

Combining Eqs. (A9) and (A1) gives

$$\mathcal{H}_{2n}(x,\Theta)\simeq\frac{4e^{-\hat{\imath}(n+1/4)\pi}}{\sqrt{2\pi x}}\sum_{m=1}^{\infty}\frac{e^{\hat{\imath}mx}\cos(mx\sin\Theta)}{\sqrt{m}}, \quad \text{(A10)}$$

$$\mathcal{H}_{2n+1}(x,\Theta)\simeq-\frac{4e^{-\hat{\imath}(n+1/4)\pi}}{\sqrt{2\pi x}}\sum_{m=1}^{\infty}\frac{e^{\hat{\imath}mx}\sin(mx\sin\Theta)}{\sqrt{m}}. \quad \text{(A11)}$$

APPENDIX B: SECOND SCHLÖMILCH SERIES

Twersky (1962) derived the following formula for a Schlömilch series that often appears in infinite grating problems: for $x>0$,

$$\hat{\imath}^n \sum_{s=-\infty}^{\infty} e^{\hat{\imath} s d \sin\Theta} e^{-\hat{\imath} n \varphi_s} H_n(r_s) = 2 \sum_{m=-\infty}^{\infty} C_m e^{-\hat{\imath} n \alpha_m} e^{\hat{\imath} x \cos\alpha_m + \hat{\imath} y \sin\alpha_m}, \tag{B1}$$

where

$$C_m = \frac{1}{d \cos\alpha_m}, \tag{B2}$$

and all parameters are defined in the same way as in the present paper, except that the local polar coordinate systems are defined such that, for scatterer s, $\varphi_s = \tan^{-1}(y - sd)/x$ for $x > 0$ and $\pi - \varphi_s$ when $x < 0$, whereas, in the present paper, the polar coordinates are defined such that ϕ_s varies from 0 to 2π. The following relations are noted: in the first quadrant, $\phi_s = \varphi_s$; in the second and third quadrants, $\phi_s = \pi - \varphi_s$; and in the fourth quadrant, $\phi_s = 2\pi + \varphi_s$.

Replacing n by $-n$ and using the relation $H_{-n}^{(1)}(\cdot) = (-1)^n H_n^{(1)}(\cdot)$ give

$$\hat{\imath}^n \sum_{s=-\infty}^{\infty} e^{\hat{\imath} s d \sin\Theta} e^{\hat{\imath} n \phi_s} H_n(r_s) = 2 \sum_{m=-\infty}^{\infty} C_m e^{\hat{\imath} n \alpha_m} e^{\hat{\imath} x \cos\alpha_m + \hat{\imath} y \sin\alpha_m}. \tag{B3}$$

It can be readily verified that this expression is valid for both the first and fourth quadrants.

For $x < 0$, according to Twersky (1962), the relation is obtained when φ_s on the left-hand side of Eq. (B1) is replaced by $\pi - \varphi_s$ and α_m on the right-hand side is replaced by $\pi - \alpha_m$. It can be similarly shown that the relation becomes the following:

$$\hat{\imath}^n \sum_{s=-\infty}^{\infty} e^{\hat{\imath} s d \sin\Theta} e^{\hat{\imath} n \phi_s} H_n(r_s) = 2 \sum_{m=-\infty}^{\infty} C_m e^{\hat{\imath} n (\pi - \alpha_m)} e^{\hat{\imath} x \cos(\pi - \alpha_m) + \hat{\imath} y \sin(\pi - \alpha_m)}. \tag{B4}$$

Note the relations

$$\sin(\pi - \alpha_m) = \sin\alpha_m \quad \text{and} \quad \cos(\pi - \alpha_m) = -\cos\alpha_m, \tag{B5}$$

and, for $x < 0$, $-x = |x|$, the right-hand side of Eq. (B4) can be alternatively written as

$$\hat{\imath}^n \sum_{s=-\infty}^{\infty} e^{\hat{\imath} s d \sin\Theta} e^{\hat{\imath} n \phi_s} H_n(r_s) = 2 \sum_{m=-\infty}^{\infty} (-1)^n C_m e^{-\hat{\imath} n \alpha_m} e^{\hat{\imath} |x| \cos\alpha_m + \hat{\imath} y \sin\alpha_m}. \tag{B6}$$

There is a slight ambiguity in Twersky's original paper as to whether α_m in the expression for C_m in Eq. (B2) should be replaced by $\pi - \alpha_m$ when $x < 0$. Numerical computations using Eq. (B4) confirm that Eq. (B2) should remain unchanged.

Combining Eqs. (B3), (B6), and (A4), a unified expression that is valid for both $x > 0$ and $x < 0$ can be written as

$$\sum_{s=-\infty}^{\infty} e^{\hat{\imath} s d \sin\Theta} e^{\hat{\imath} n \phi_s} H_n(r_s) = \frac{2}{d} \sum_{m=-\infty}^{\infty} \frac{(\sin\alpha_m - \hat{\imath}(x/|x|)\cos\alpha_m)^n}{\cos\alpha_m} e^{\hat{\imath}(|x| \cos\alpha_m + y \sin\alpha_m)}, \tag{B7}$$

where the fraction $x/|x|$ simply denotes the sign of x.

Achenbach, J. D., and Li, Z. L. (**1986**). "Propagation of horizontally polarized transverse waves in a solid with a periodic distribution of cracks," Wave Motion **8**(4), 371–379.

Angel, Y. C., and Achenbach, J. D. (**1987**). "Harmonic waves in an elastic solid containing a doubly periodic array of cracks," Wave Motion **9**(5), 377–385.

Botten, L. C., Nicorovici, N. A. P., Asatryan, A. A., McPhedran, R. C., de Sterke, C. M., and Robinson, P. A. (**2000**). "Formulation for electromagnetic scattering and propagation through grating stacks of metallic and dielectric cylinders for photonic crystal calculations. Part I. Method," J. Opt. Soc. Am. A **17**(12), 2165–2176.

Botten, L. C., White, T. P., de Sterke, C. M., McPhedran, R. C., Asatryan, A. A., and Langtry, T. N. (**2004**). "Photonic crystal devices modelled as grating stacks: matrix generalizations of thin film optics," Opt. Express **12**(8), 1592–1604.

Brigham, G. A., Libuha, J. J., and Radlinski, R. P. (**1977**). "Analysis of scattering from large planar gratings of compliant cylindrical shells," J. Acoust. Soc. Am. **61**, 48–59.

Burke, J. E., and Twersky, V. (**1966**). "On scattering of waves by the infinite grating of elliptic cylinders," IEEE Trans. Antennas Propag. **AP-14**(4), 465–480.

Cai, L.-W. (**2004**). "Multiple scattering in single scatterers," J. Acoust. Soc. Am. **115**, 986–995.

Cai, L.-W. (**2005**). "Scattering of elastic anti-plane shear waves by multilayered eccentric scatterers," Q. J. Mech. Appl. Math. **58**(2), 165–183.

Cai, L.-W., and William, Jr., J. H. (**1999a**). "Large scale multiple scattering problems," Ultrasonics **37**(7), 453–462.

Cai, L.-W., and William, Jr., J. H. (**1999b**). "Full-scale simulations of elastic wave scattering in fiber reinforced composites," Ultrasonics **37**(7), 463–482.

Esquivel-Sirvent, R., and Cocoletzi, G. H. (**1994**). "Band structure for the propagation of elastic waves in Superlattices," J. Acoust. Soc. Am. **95**, 86–90.

Evans, D. V., and Linton, C. M. (**1991**). "Trapped modes in open channels," J. Fluid Mech. **225**, 153–175.

Heckl, M. A. (**1992**). "Sound propagation in bundles of periodically arranged cylindrical tubes," Acustica **77**(3), 143–152.

Heckl, M. A. (**1994**). "Oblique sound transmission through tube bundles and tube gratings," Ultrasonics **32**(4), 275–286.

Heckl, M. A., and Mulholland, L. S. (**1995**). "Some recent developments in the theory of acoustic transmission in tube bundles," J. Sound Vib. **179**(1), 37–62.

Huang, X. Y., and Heckl, M. A. (**1993**). "Transmission and dissipation of sound waves in tube bundles," Acustica **78**(4), 191–200.

Ivanov, V. P. (**1971**). "Plane-wave diffraction by an *N*-layer grating," Sov. Phys. Acoust. **17**(2), 202–207.

Kalhor, H. A., and Ilyas, M. (**1982**). "Scattering of plane electromagnetic waves by a grating of conducting cylinders embedded in a dielectric slab over a ground plane," IEEE Trans. Antennas Propag. **AP-30**(4), 576–579.

Kavaklioğlu, O. (**2000**). "Scattering of a plane wave by an infinite grating of circular dielectric cylinders at oblique incidence: E-polarization," Int. J. Electron. **87**(3), 315–336.

Kavaklioğlu, O. (**2001**). "On diffraction of waves by the infinite grating of circular dielectric cylinders at oblique incidence: Floquet representation," J. Mod. Opt. **48**(1), 125–142.

Kavaklioğlu, O. (**2002**). "On Schlömilch series representation for the transverse electric multiple scattering by an infinite grating of insulating dielectric circular cylinders at oblique incidence," J. Phys. A **35**(9), 2229–2248.

Klyukin, I. I., and Chabanov, V. E. (**1975**). "Sound diffraction by a plane grating of cylinders," Sov. Phys. Acoust. **20**(5), 519–523.

Kristiansen, U. R., and Fahy, F. J. (**1972**). "Scattering of acoustic waves by an N–layer periodic grating," J. Sound Vib. **24**(3), 315–335.

Lakhtakia, A., Varadan, V. V., and Varadan, V. K. (**1986a**). "Reflection characteristics of a dielectric slab containing dielectric or perfectly conducting cylindrical gratings," Appl. Opt. **25**(6), 887–894.
Lakhtakia, A., Varadan, V. V., and Varadan, V. K. (**1986b**). "Reflection characteristics of an elastic slab containing a periodic array of elastic cylinders: SH wave analysis," J. Acoust. Soc. Am. **80**, 311–316.
Lakhtakia, A., Varadan, V. V., and Varadan, V. K. (**1988**). "Reflection characteristics of an elastic slab containing a periodic array of elastic cylinders: P and SV wave analysis," J. Acoust. Soc. Am. **83**, 1267–1275.
Leiko, A. G., and Mayatskii, V. I. (**1974**). "Diffraction of plane sound waves by an infinite grating of perfectly compliant cylinders," Sov. Phys. Acoust. **20**(3), 256–258.
Leiko, A. G., and Mayatskii, V. I. (**1975**). "Diffraction of a plane sound wave by an infinite grating of perfectly rigid elliptical cylinders," Sov. Phys. Acoust. **20**(4), 389–390.
Linton, C. M., and Evans, D. V. (**1990**). "The interaction of waves with arrays of vertical circular cylinders," J. Fluid Mech. **215**, 549–569.
Liu, Z. Y., Chan, C. T., Sheng, P., and Goertzen, A. L. (**2000**). "Elastic wave scattering by periodic structures of spherical objects: Theory and experiment," Phys. Rev. B **62**(4), 2446–2457.
Lord Rayleigh (**1907**). "On the dynamical theory of gratings," Proc. R. Soc. London, Ser. A **17**, 399–416.
Maniar, H. D., and Newman, J. N. (**1997**). "Wave diffraction by a long array of cylinders," J. Fluid Mech. **339**, 309–330.
McIver, P. (**2000**). "Water-wave propagation through an infinite array of cylindrical structures," J. Fluid Mech. **424**, 101–125.
Miles, J. W. (**1982**). "On Rayleigh scattering by a grating," Wave Motion **4**(3), 285–292.
Miles, J. W. (**1983**). "Surface-wave diffraction by a periodic row of submerged ducts," J. Fluid Mech. **128**, 155–180.
Millar, R. F. (**1961a**). "Scattering by a grating I," Can. J. Phys. **39**(1), 81–103.
Millar, R. F. (**1961b**). "Scattering by a grating II," Can. J. Phys. **39**(1), 104–118.
Millar, R. F. (**1963a**). "Plane wave spectra in theory I. Scattering by a finite number of bodies," Can. J. Phys. **41**(12), 2106–2134.
Millar, R. F. (**1963b**). "Plane wave spectra in grating theory II. Scattering by an infinite grating of identical cylinders," Can. J. Phys. **41**(12), 2135–2154.
Millar, R. F. (**1964a**). "Plane wave spectra in grating theory III. Scattering by a semiinfinite grating of identical cylinders," Can. J. Phys. **42**(6), 1149–1184.
Millar, R. F. (**1964b**). "Plane wave spectra in grating theory IV. Scattering by a finite grating of identical cylinders," Can. J. Phys. **42**(12), 2395–2410.
Millar, R. F. (**1966**). "Plane wave spectra in grating theory V. Scattering by a semi-infinite grating of isotropic scatterers," Can. J. Phys. **44**(11), 2839–2874.
Mulholland, L. S., and Heckl, M. A. (**1994**). "Multi-directional sound wave propagation through a tube bundle," J. Sound Vib. **176**(3), 377–398.
Ohl, C. O. G., Taylor, R. E., Taylor, P. H., and Borthwick, A. G. L. (**2001**). "Water wave diffraction by a cylinder array. 1. Regular waves," J. Fluid Mech. **442**, 1–32.
Ohl, C. O. G., Taylor, R. E., Taylor, P. H., and Borthwick, A. G. L. (**2001b**). "Water wave diffraction by a cylinder wave. 2. Irregular waves," J. Fluid Mech. **442**, 33–66.
Platts, S. B., Movchan, N. V., McPhedran, R. C., and Movchan, A. B. (**2003a**). "Transmission and polarization of elastic waves in irregular structures," J. Eng. Mater. Technol. **125**(1), 2–6.
Platts, S. B., Movchan, N. V., McPhedran, R. C., and Movchan, A. B. (**2003b**). "Band gaps and elastic waves in disordered stacks: normal incidence," Proc. R. Soc. London, Ser. A **459**(2029), 221–240.
Porter, R., and Evans, D. V. (**1996**). "Wave scattering by periodic arrays of breakwaters," Wave Motion **23**(2), 95–120.
Porter, R., and Evans, D. V. (**1999**). "Rayleigh-Bloch surface waves along periodic gratings and their connection with trapped modes in waveguides," J. Fluid Mech. **386**, 233–258.
Psarobas, I. E., and Sigalas, M. (**2002**). "Elastic band gaps in a fcc lattice of mercury spheres in aluminum," Phys. Rev. B **66**(5), 052302.
Radlinski, R. P. (**1989**). "Scattering from multiple gratings of compliant tubes in a viscoelastic layer," J. Acoust. Soc. Am. **85**, 2301–2310.
Radlinski, R. P., and Simon, M. M. (**1982**). "Scattering by multiple gratings of compliant tubes," J. Acoust. Soc. Am. **72**(2), 607–614.
Radlinski, R. P., and Janus, R. S. (**1986**). "Scattering from two and three gratings of densely packed compliant tubes," J. Acoust. Soc. Am. **80**(6), 1803–1809.
Sainidou, R., Stefanou, N., Psarobas, I. E., and Modinos, A. (**2005**). "A layer-multiple-scattering method for phononic crystals and heterostructures of such," Comput. Phys. Commun. **166**, 197–240.
Scarpetta, E., and Sumbatyan, M. A. (**1995**). "Explicit analytical results for one-mode normal reflection and transmission by a periodic array of screens," J. Math. Anal. Appl. **195**(3), 736–749.
Scarpetta, E., and Sumbatyan, M. A. (**1997**). "On wave propagation in elastic solids with a doubly periodic array of cracks," Wave Motion **25**(1), 61–72.
Scarpetta, E., and Sumbatyan, M. A. (**2002**). "In-plane wave propagation through elastic solids with a periodic array of rectangular defects," J. Appl. Phys. **69**(1), 179–188.
Scarpetta, E., and Sumbatyan, M. A. (**2003**). "Wave propagation through elastic solids with a periodic array of arbitrarily shaped defects," Math. Comput. Modell. **37**(1), 19–28.
Twersky, V. (**1952a**). "Multiple scattering of radiation by an arbitrary configuration of parallel cylinders," J. Acoust. Soc. Am. **24**, 42–46.
Twersky, V. (**1952b**). "On a multiple scattering theory of the finite grating and the Wood anomalies," J. Appl. Phys. **23**(10) 1099–1118.
Twersky, V. (**1956**). "On the scattering of waves by an infinite grating," IRE Trans. Antennas Propag. **AP-4**(3), 330–345.
Twersky, V. (**1961**). "Elementary function representations of Schlomilch series," Arch. Ration. Mech. Anal. **8**(4), 323–332.
Twersky, V. (**1962**). "On scattering of waves by the infinite grating of circular cylinders," IRE Trans. Antennas Propag. **AP-10**(6), 737–765.
Varadan, V. V., Lakhtakia, A., and Varadan, V. K. (**1988**). "Comments on recent criticism of the *T*-matrix method," J. Acoust. Soc. Am. **84**(6), 2280–2284.
Vovk, I. V., and Grinchenko, V. T. (**1978**). "Diffraction of a plane wave by a double-layer grating of hollow elastic bars," Sov. Phys. Acoust. **24**(6), 480–482.
Vovk, I. V., Grinchenko, V. T., and Kononuchenko, L. A. (**1976**). "Diffraction of a sound wave by a plane grating formed by hollow elastic bars," Sov. Phys. Acoust. **22**(2), 113–115.
Waterman, P. C. (**1969**). "New formulation of acoustic scattering," J. Acoust. Soc. Am. **45**, 1417–1429.
Watson, G. N. (**1966**). *Theory of Bessel Functions*, 2nd ed. (Cambridge U.P., Cambridge).

The Wigner-Smith matrix in acoustic scattering: Application to fluid-loaded elastic plates

H. Franklin, P. Rembert,[a] and O. Lenoir
Laboratoire d'Acoustique Ultrasonore et d'Electronique (LAUE), CNRS UMR 6068, Université du Havre, Place Robert Schuman, 76610 Le Havre, France

(Received 9 November 2005; revised 2 May 2006; accepted 2 May 2006)

The Wigner-Smith matrix **Q** is built up by differentiation of the unitary condition of the scattering matrix **S**. The matrices **Q** and **S** both contain the same information but with different points of view. For structures with simple geometrical shapes such as plates or cavities, the acoustic scattering is a two channel scattering represented by a 2×2 **S** matrix. The elements of the **Q** matrix can be described: (i) by means of the phase derivatives of the elements of the matrix **S**, (ii) by means of the phase derivatives of the eigenvalues of the **S** matrix. The equivalence of these two descriptions allows one to express the phase derivatives of (i) in terms of the phase derivatives of (ii). The Wigner-Smith matrix concept enables one to unify and to improve both the phase gradient method and the eigenvalue method in the frame of the multichannel scattering. It obviously incorporates the resonance scattering theory. Approximate resonant formulas and numerical results are given for the case of fluid loaded elastic isotropic plates in order to check the validity of the method.
 [DOI: 10.1121/1.2207574]

PACS number(s): 43.20.Ks, 43.40.Dx, 43.40.Rj [RMW] Pages: 62–73

I. INTRODUCTION

The resonance scattering theory (RST) has been applied to various submerged structures such as plates or stacks of plates,[1,2] targets of separable geometry such as cylinders, tubes or spheres,[3,4] and also to cavities in an elastic medium.[5–7] Its main purpose was to analyze the resonance properties of these structures by finding the convenient approximations (called the Breit-Wigner formula) of the elements of the scattering matrix in the frequency domain (plates, cylinders, spheres, cavities) or in the angular domain (plates). Some difficulties have been encountered when applying the RST to cases where overlapping phenomena between resonances occur.[8,9] This fact, added to the need for fast methods easy to implement, have led to building the phase gradient method (PGM) applied to plates[10,11] and cylindrical shells.[12] The PGM consists in analyzing the phase derivatives of the elements of the scattering matrix. The peaks of the phase derivatives with respect to the frequency variable (or angular variable) give both the resonance positions (located at the values of the variable where sharp peak maxima or minima occur) and twice the inverse of the width (by the measure of the peak heights). The PGM then enables a more complete characterization of the resonance properties, the overlapping phenomenon[8,9] in the case of plates also being solved. As an extension of the method, the derivative variable can be one of the physical parameters of the structures: densities, wave velocities, etc. This enables one to study the sensitivity of the waves scattered by these structures, with respect to these parameters.

The scattering matrix needed for the analysis takes several forms.

For cavities in an elastic medium, polarization conversions between L (longitudinal) and T (transverse) waves arise in the scattering. As a consequence, the main diagonal of the scattering matrix is made up of 2×2 block matrices $\mathbf{S}_m$ Ref. 5 (m denotes the mode) while all the other blocks are equal to zero. The form of the matrices $\mathbf{S}_m$ is

$$\mathbf{S}_m = \begin{pmatrix} s_m^{LL} & s_m^{LT} \\ s_m^{LT} & s_m^{TT} \end{pmatrix}.$$

The superscript LL stands for the polarization conversion from the incoming L wave into an outgoing L wave while the superscript LT stands for the polarization conversion from the incoming L wave into an outgoing T wave.

For fluid-loaded plates or stacks of plates, the scattering matrix is a 2×2 matrix denoted **S**, whose elements are the reflection and transmission coefficients.[13] In the case of plates loaded by different fluids on the faces, one has

$$\mathbf{S} = \begin{pmatrix} r_1 & \chi t_1 \\ \chi t_1 & r_2 \end{pmatrix},$$

r_1, r_2, denoting the reflection coefficients of the plate, and χt_1 the transmission coefficient through the plate.

The dimension of the matrix $\mathbf{S}_m$—Ref. 7 (or $\mathbf{S}$[13]) being two, one is in the field of the two-channel scattering formalism. Each of the elements of these matrices is coupled with the three others and gives only a partial information about the scattering. In RST or PGM, the studies are generally led by calculating at first the elements one by one, and then by comparing them. However, the need for methods of fast evaluations of the resonance spectrum, avoiding both the study of the elements one by one and as much as possible the steps of the derivative processes led us to suggest an eigenvalues method (EM) exploiting the unitary properties of the matrix $\mathbf{S}_m$ (or **S**). It then remains to find the connection be-

[a]Electronic mail: pascal.rembert@univ-lehavre.fr

tween the PGM and the EM. Because, from a mathematical point of view, the analysis of the two matrices $\mathbf{S}_m$ and $\mathbf{S}$ follows nearly the same way, the physical approach will focus mainly on the case of plates in this paper.

Let $e^{2i\delta_\lambda}$ and $e^{2i\delta_\mu}$ denote the eigenvalues of the $\mathbf{S}$ matrix. The plots of the moduli of the two transition terms $T_\lambda = (e^{2i\delta_\lambda}-1)/2i$ and $T_\mu = (e^{2i\delta_\mu}-1)/2i$ give, although from a different point of view, the same results as the PGM (which uses the phase derivative of the reflexion coefficient). In the particular case of identical fluids loading the plate ($r_2 = r_1$), the connection between PGM and EM seems straightforward. As a proof, let us consider the derivative with respect to the frequency for instance, $\mathbf{S}'\mathbf{S}^\dagger + \mathbf{S}\mathbf{S}^{\dagger\prime} = 0$, of the unitary relation $\mathbf{S}\mathbf{S}^\dagger = \mathbf{I}$. The prime- and dagger denote the derivative and the adjoint operator, respectively, and $\mathbf{I}$ is the (2×2) identity matrix. The first term of the previous derivative multiplied by $-i$, i.e., $\mathbf{Q} = -i\mathbf{S}'\mathbf{S}^\dagger$, is known as the Wigner-Smith matrix or lifetime matrix, and was initially developed in the frame of quantum mechanics by Smith.[14] This formalism allows one to define rigorously and estimate the time spent by an asymptotically free particle within an interaction region. Here, the diagonal elements of the $\mathbf{Q}$ matrix exhibit the phase derivative of the reflection coefficient. But the $\mathbf{Q}$ matrix can also be expressed in terms of the phase derivatives of the eigenvalues of the $\mathbf{S}$ matrix. By equaling the trace (first scalar invariant) of these matrices, a simple formula relating the phase derivatives is obtained.

More precisely, two descriptions of the $\mathbf{Q}$ matrix are possible. At first, the (ϕ', φ') description, presented in Sec. II, deals with the derivatives of the reflection and transmission coefficient phases ϕ and φ. Its corresponding matrix is noted $\mathbf{Q}_{\phi\varphi}$. The second one, the $(\delta'_\lambda, \delta'_\mu)$ description, presented in Sec. III, deals with the phase derivatives of the eigenvalues $e^{2i\delta_\lambda}$ and $e^{2i\delta_\mu}$ of the $\mathbf{S}$ matrix. Its corresponding matrix is noted $\mathbf{Q}_{\lambda\mu}$. The requirement $\mathbf{Q}_{\phi\varphi} = \mathbf{Q}_{\lambda\mu}$ means, in particular, that the traces are equal, as well as the determinants (second scalar invariants). It also provide basis relations between the sets of phase derivatives (ϕ', φ') and $\delta'_\lambda, \delta'_\mu$ as shown in Sec. IV. The connection between PGM and EM can then be achieved by means of relations between phase derivatives.

Several analyses in acoustic scattering (by cylindrical or spherical structures) require the removal of the background component that masks the resonances from the scattered field. It is then shown, in Sec. V, how the results of the previous sections can account for the background and how they lead to express the scattering matrix as a sum of density matrices.[15] The concept of the density matrix is useful, in particular, for geometrical interpretations. In Sec. VI, the two-channel RST is incorporated. Approximate forms of the phase derivatives and of the elements of various matrices are obtained in the frequency variable. At first, the advantages of a use of the transition terms in place of the transmission coefficients are shown. In particular, around some given frequencies, the transition terms behave as if they exchange their properties. This fact is not encountered in the transmission coefficient. Then, the phase derivatives φ', δ'_λ, and δ'_μ are investigated. Contrary to φ', the use of the phase derivatives δ'_λ and δ'_μ enhances the contrast between great peaks due to plate resonances and small peaks (located at the same frequencies previously mentioned) caused by the asymmetry in fluid loading. Connection between the phase derivatives and the transition terms are explained. At last, the derivative of the mixing angle (e.g., the angle related to the orthogonal matrix that diagonalizes $\mathbf{S}$) is shown to be an adequate tool to indicate the asymmetry in fluid loading. The plots of the eigenvalues of the $\mathbf{Q}$ matrix show the spreading of the resonances between the two channels of the scattering. In Sec. VII, the work of the preceding section is carried out in the angular domain for the phase derivatives φ', δ'_λ, and δ'_μ.

II. THE (ϕ', φ') DESCRIPTION

Let us consider plates or stacks of plates loaded with different fluids (of indices 1 and 2) at each external face. If the incident wave comes from fluid 1 (respectively, fluid 2), let r_1 (respectively, r_2) be the reflection coefficient of the system and χt_1 be the transmission coefficient. These coefficients are given in Appendix A. They depend on a set of n variables (frequency f, sine of the incidence angle $\bar{k}_x$) and physical parameters (densities, sound velocities in fluids c_1 and c_2, longitudinal c_L and transverse c_T wave velocities in the solids, etc.). This set forms the n elements of a vector x. For instance, $n=4$ and $\boldsymbol{x} = (f, c_L, c_T, c_1)$ for an elastic plate loaded by the same fluid on its two faces.[10] The scattering matrix of the structure

$$\mathbf{S} = \begin{pmatrix} r_1 & \chi t_1 \\ \chi t_1 & r_2 \end{pmatrix}, \tag{1}$$

is a (2×2) complex, symmetric, and unitary matrix. In this last case, $\mathbf{S}\mathbf{S}^\dagger = \mathbf{S}^\dagger\mathbf{S} = \mathbf{I}$. The expansion of the first product leads to

$$\mathbf{S}\mathbf{S}^\dagger = \begin{pmatrix} r_1 r_1^* + (\chi t_1)(\chi t_1)^* & r_1(\chi t_1)^* + (\chi t_1) r_2^* \\ (\chi t_1) r_1^* + r_2(\chi t_1)^* & r_2 r_2^* + (\chi t_1)(\chi t_1)^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{2}$$

the asterisk denoting the complex conjugate. The derivative $(')$ of the unitary condition $\mathbf{S}\mathbf{S}^\dagger = \mathbf{I}$ with respect to one of the components of vector $\boldsymbol{x}$ gives the following expression:

$$\mathbf{S}'\mathbf{S}^\dagger + \mathbf{S}\mathbf{S}'^\dagger = 0. \tag{3}$$

From Eq. (3), a self-adjoint matrix $\mathbf{Q}_{\phi\varphi}$, also called Wigner-Smith matrix,[14] can be defined by

$$\mathbf{Q}_{\phi\varphi} = -i\mathbf{S}'\mathbf{S}^\dagger. \tag{4}$$

The subscripts ϕ and φ represent the phases of the reflection and the transmission coefficients written as follows

$$r_1 = |r_1| e^{i\phi}, \quad r_2 = |r_2| e^{i\psi}, \quad \chi t_1 = |\chi t_1| e^{i\varphi}, \tag{5}$$

with $|r_1| = |r_2|$. From these expressions, one of the unitary relations of the scattering matrix, Eq. (2):

$$r_1(\chi t_1)^* + (\chi t_1) r_2^* = 0 \tag{6}$$

gives

$$e^{i(\phi-\varphi)} = -e^{i(\varphi-\psi)}. \tag{7}$$

By differentiation, one has

$$\phi' - \varphi' = \varphi' - \psi'. \tag{8}$$

Hence, the elements q_{lm} $(l,m=1,2)$ of the $\mathbf{Q}_{\phi\varphi}$ matrix take the forms

$$q_{11} = \{\phi'|r_1|^2 + \varphi'|\chi t_1|^2\}, \tag{9a}$$

$$q_{12} = q_{21}^* = \{(\phi' - \varphi')|r_1||\chi t_1| + iW\}e^{i(\phi-\varphi)}, \tag{9b}$$

$$q_{22} = \{\psi'|r_1|^2 + \varphi'|\chi t_1|^2\}, \tag{9c}$$

Here, W indicates the Wronskian of $|r_1|$ and $|\chi t_1|$:

$$W = |r_1||\chi t_1|' - |r_1|'|\chi t_1|. \tag{10}$$

The first scalar invariant of $\mathbf{Q}_{\phi\varphi}$, i.e., the trace of the matrix, is

$$\mathrm{tr}(\mathbf{Q}_{\phi\varphi}) = 2\varphi'. \tag{11}$$

It does not depend on $|r_1|$ and $|\chi t_1|$. The second scalar invariant is

$$\det(\mathbf{Q}_{\phi\varphi}) = \phi'\psi'|r_1|^2 + \varphi'^2|\chi t_1|^2 - W^2. \tag{12}$$

The eigenvalues $\Lambda^{\pm}_{\phi\varphi}$ of the matrix $\mathbf{Q}_{\phi\varphi}$, according to the properties of Hermitian matrices, are real and given by

$$\Lambda^{\pm}_{\phi\varphi} = \varphi' \pm \sqrt{\varphi'^2 - \det(\mathbf{Q}_{\phi\varphi})} = \left(\frac{\phi' + \psi'}{2}\right) \pm \sqrt{\left(\frac{\phi' - \psi'}{2}\right)^2 |r_1|^2 + W^2}. \tag{13}$$

III. THE $(\delta'_\lambda, \delta'_\mu)$ DESCRIPTION

We examine now the second way to express the $\mathbf{Q}$ matrix by using the phase derivatives of the eigenvalues of the scattering matrix $\mathbf{S}$. The $\mathbf{S}$ matrix can be put into the diagonal form $\mathbf{S}_{\mathrm{eig}}$ by means of the transform $\mathbf{S}=\mathbf{R}\mathbf{S}_{\mathrm{eig}}\mathbf{R}^t$ (the superscript t stands for transposition), in which

$$\mathbf{S}_{\mathrm{eig}} = \begin{pmatrix} e^{2i\delta_\lambda} & 0 \\ 0 & e^{2i\delta_\mu} \end{pmatrix}, \tag{14}$$

and

$$\mathbf{R} = \begin{pmatrix} \cos\varsigma & -\sin\varsigma \\ \sin\varsigma & \cos\varsigma \end{pmatrix}. \tag{15}$$

The quantities $e^{2i\delta_\lambda}$ and $e^{2i\delta_\mu}$ represent[13] the eigenvalues of $\mathbf{S}$ and are such that $|e^{2i\delta_\lambda}| = |e^{2i\delta_\mu}| = 1$. $\mathbf{R}$ is a real and orthogonal matrix, $\mathbf{R}\mathbf{R}^t = \mathbf{R}^t\mathbf{R} = \mathbf{I}$, and ς is the rotation angle (also called the mixing angle). The first (respectively, second) column of R is the eigenvector associated with $e^{2i\delta_\lambda}$ (respectively, $e^{2i\delta_\mu}$). First, the use of the relationship $\mathbf{S} = \mathbf{R}\mathbf{S}_{\mathrm{eig}}\mathbf{R}^t$ leads to the following form for the Wigner-Smith matrix

$$\mathbf{Q}_{\lambda\mu} = \mathbf{R}\mathbf{M}\mathbf{R}^t, \tag{16}$$

with

$$\mathbf{M} = i\mathbf{R}'^t\mathbf{R} + \mathbf{Q}_{\mathrm{eig}} - i\mathbf{S}_{\mathrm{eig}}\mathbf{R}'^t\mathbf{R}\mathbf{S}^\dagger_{\mathrm{eig}}, \tag{17}$$

and

$$\mathbf{Q}_{\mathrm{eig}} = -i\mathbf{S}'_{\mathrm{eig}}\mathbf{S}^\dagger_{\mathrm{eig}} = \begin{pmatrix} 2\delta'_\lambda & 0 \\ 0 & 2\delta'_\mu \end{pmatrix}. \tag{18}$$

In a second time, the relation between the rotation matrix $\mathbf{R}$ and its derivative being such that

$$\mathbf{R}'^t\mathbf{R} = i\varsigma' \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{19}$$

the previous expression for $\mathbf{M}$ becomes

$$\mathbf{M} = \begin{pmatrix} 2\delta'_\lambda & A \\ A^* & 2\delta'_\mu \end{pmatrix}, \tag{20}$$

with

$$A = i\varsigma'[1 - e^{2i(\delta_\lambda - \delta_\mu)}]. \tag{21}$$

From these results, one obtains the elements $\bar{q}_{lm}$ $(l,m=1,2)$ of the $\mathbf{Q}_{\lambda\mu}$ matrix:

$$\bar{q}_{11} = 2\delta'_\lambda \cos^2\varsigma + 2\delta'_\mu \sin^2\varsigma - (2\,\mathrm{Re}\,A)\sin\varsigma\cos\varsigma, \tag{22a}$$

$$\bar{q}_{12} = \bar{q}^*_{21} = A\cos^2\varsigma - A^*\sin^2\varsigma + 2(\delta'_\lambda - \delta'_\mu)\sin\varsigma\cos\varsigma, \tag{22b}$$

$$\bar{q}_{22} = 2\delta'_\lambda \sin^2\varsigma + 2\delta'_\mu \cos^2\varsigma + (2\,\mathrm{Re}\,A)\sin\varsigma\cos\varsigma. \tag{22c}$$

Here, $\mathrm{Re}\,A$ means the real part of A. The first scalar invariant only depends on the sum of the phase derivatives δ'_λ and δ'_μ:

$$\mathrm{tr}(\mathbf{Q}_{\lambda\mu}) = 2(\delta'_\lambda + \delta'_\mu). \tag{23}$$

The second scalar invariant takes the form

$$\det(\mathbf{Q}_{\lambda\mu}) = 4\delta'_\lambda\delta'_\mu - |A|^2. \tag{24}$$

From these relations and from Eq. (13), the following forms of the $\mathbf{Q}_{\lambda\mu}$ matrix eigenvalues can be deduced

$$\Lambda^{\pm}_{\lambda,\mu} = (\delta'_\lambda + \delta'_\mu) \pm \sqrt{(\delta'_\lambda + \delta'_\mu)^2 - \det(\mathbf{Q}_{\lambda\mu})}. \tag{25}$$

It is easily verified that they are real quantities if $\det(\mathbf{Q}_{\lambda\mu})$ is replaced by its expression given in Eq. (24). It must also be noticed that $\mathrm{tr}(\mathbf{Q}_{\lambda\mu})$ and that $\det(\mathbf{Q}_{\lambda\mu}) = \det(\mathbf{M})$. So, the matrices $\mathbf{Q}_{\lambda\mu}$ and $\mathbf{M}$ have the same eigenvalues. If $\mathbf{U}$ is the unitary matrix transforming $\mathbf{M}$ into a diagonal one, i.e., $\mathbf{M}_{\mathrm{eig}}$, then $\mathbf{M} = \mathbf{U}\mathbf{M}_{\mathrm{eig}}\mathbf{U}^\dagger$ and finally

$$\mathbf{Q}_{\lambda\mu} = (\mathbf{R}\mathbf{U})\mathbf{M}_{\mathrm{eig}}(\mathbf{R}\mathbf{U})^\dagger. \tag{26}$$

Therefore, matrix $\mathbf{R}\mathbf{U}$ diagonalizes $\mathbf{Q}_{\lambda\mu}$.

IV. IMPLICATIONS OF THE REQUIREMENT $\mathbf{Q}_{\phi\varphi} = \mathbf{Q}_{\lambda\mu}$

This requirement implies that both the traces and the determinants in each description are equal:

$$\mathrm{tr}(\mathbf{Q}_{\lambda\mu}) = \mathrm{tr}(\mathbf{Q}_{\phi\varphi}), \tag{27}$$

and

$$\det(\mathbf{Q}_{\phi\varphi}) = \det(\mathbf{Q}_{\lambda\mu}). \tag{28}$$

In a more explicit form, Eq. (27) becomes

$$\delta'_\lambda + \delta'_\mu = \varphi'. \tag{29}$$

Equation (29) enables the connection between the PGM and the EM. It does not depend on the rotation angle and involves the phase of the transmission coefficient rather than the one of the reflection coefficients. In a sense, this is not surprising because the transmission coefficient is the coefficient that remains unchanged whatever the side of the plate impinged by the incident wave (the phase of the reflection coefficient changes while the transmission coefficient remains invariant). By using the relation between the phase derivatives in Eq. (8), it follows

$$\delta'_\lambda + \delta'_\mu = \frac{\phi' + \psi'}{2}. \tag{30}$$

While the phase derivatives δ'_λ and δ'_μ describe eigenmodes (each one describes a part of the resonance spectrum for the plate, according to the separation implied by the eigenvalues of the **S** matrix), the sum $\delta'_\lambda + \delta'_\mu$ describes an average of the phase derivatives ϕ' and ψ'. It must be noticed also, from Eqs. (29) and (30), that $\delta'_\lambda + \delta'_\mu = (\varphi' + \phi' + \psi')/3$.

To examine briefly one possible generalization of Eq. (29), let $\mathbf{Q}_{\phi\varphi j}$ and $\mathbf{Q}_{\lambda\mu j}$ be the Wigner-Smith matrices when the differentiation is done with respect to one of the components x_j ($j=1$ to n) of the $\boldsymbol{x}$ vector defined in Sec. II. Then, the linear combinations $\Sigma_j x_j \mathbf{Q}_{\phi\varphi j}$ and $\Sigma_j x_j \mathbf{Q}_{\lambda\mu j}$ on one hand, and the traces in each description on the other hand, are equal:

$$\sum_j x_j \mathbf{Q}_{\lambda\mu j} = \sum_j x_j \mathbf{Q}_{\phi\varphi j}, \tag{31}$$

and

$$\sum_j x_j \mathrm{tr}\, \mathbf{Q}_{\lambda\mu j} = \sum_j x_j \mathrm{tr}\, \mathbf{Q}_{\phi\varphi j}. \tag{32}$$

By introducing the notation $\boldsymbol{x} \cdot \nabla_x = \Sigma_j x_j \partial / \partial x_j$, one obtains from Eq. (32)

$$\boldsymbol{x} \cdot \nabla_x (\delta_\lambda + \delta_\mu) = \boldsymbol{x} \cdot \nabla_x \varphi. \tag{33}$$

The phases φ, δ_λ, and δ_μ are homogeneous functions of zero order of the components x_j of the x vector. For instance, in the case of identical fluid-loading, the right-hand side member is equal to zero.[10] One deduces the relations

$$\boldsymbol{x} \cdot \nabla_x \delta_\lambda = \boldsymbol{x} \cdot \nabla_x \delta_\mu = 0, \tag{34}$$

which constitute an extension of the PGM to the field of the multi-channel scattering formalism.

V. INTRODUCTION OF THE DENSITY MATRIX FORMALISM

In this section, one examines at first the form taken by the **Q** matrix when **S** is expressed as the product of two matrices. Such a product (or factorization) is encountered in problems where the background needs to be removed, for instance in the scattering by fluid cavities. It is also encountered when a factorization needs to be applied to the matrix $\mathbf{S}_{\mathrm{eig}}$, in view to decompose the two channel scattering as a product of single channel ones. In a second time, one introduces the density matrices related to the scattering (in this present case, by fluid-loaded plates). The concept of the density matrix has been summarized in Ref. 15 for quantum mechanics and statistical physics. It can also be used in acoustic scattering to quantify the spreading of the resonances between the outgoing channels.[7]

Let us assume that the scattering matrix **S** expresses as the product of two matrices $\mathbf{S}_{\mathrm{left}}$ and $\mathbf{S}_{\mathrm{right}}$:

$$\mathbf{S} = \mathbf{S}_{\mathrm{left}} \mathbf{S}_{\mathrm{right}}, \tag{35}$$

with $\mathbf{S}_{\mathrm{left}} \mathbf{S}^\dagger_{\mathrm{left}} = \mathbf{S}^\dagger_{\mathrm{left}} \mathbf{S}_{\mathrm{left}} = \mathbf{I}$ and $\mathbf{S}_{\mathrm{right}} \mathbf{S}^\dagger_{\mathrm{right}} = \mathbf{S}^\dagger_{\mathrm{right}} \mathbf{S}_{\mathrm{right}} = \mathbf{I}$. Once the matrices $\mathbf{Q} = -\mathrm{i}\mathbf{S}'\mathbf{S}^\dagger$, $\mathbf{Q}_{\mathrm{left}} = -\mathrm{i}\mathbf{S}'_{\mathrm{left}} \mathbf{S}^\dagger_{\mathrm{left}}$, and $\mathbf{Q}_{\mathrm{right}} = -\mathrm{i}\mathbf{S}'_{\mathrm{right}} \mathbf{S}^\dagger_{\mathrm{right}}$ have been defined, the following relation is easily deduced:

$$\mathbf{Q} = \mathbf{Q}_{\mathrm{left}} + \mathbf{S}_{\mathrm{left}} \mathbf{Q}_{\mathrm{right}} \mathbf{S}^\dagger_{\mathrm{left}}. \tag{36}$$

Since the trace of a product of several matrices is invariant by a cyclic permutation of these matrices, it follows:

$$\mathrm{tr}\, \mathbf{Q} = \mathrm{tr}\, \mathbf{Q}_{\mathrm{left}} + \mathrm{tr}\, \mathbf{Q}_{\mathrm{right}}. \tag{37}$$

As a general rule, the phases related to the background (represented by the $\mathbf{S}_{\mathrm{left}}$ matrix) vary very slowly in comparison with the phases of the resonant scattering[12] (represented by the $\mathbf{S}_{\mathrm{right}}$ matrix). It follows, in Eq. (37), that $\mathrm{tr}\, \mathbf{Q}_{\mathrm{left}}$ is very small in comparison with $\mathrm{tr}\, \mathbf{Q}_{\mathrm{right}}$.

Let us examine now another case where the matrix $\mathbf{S}_{\mathrm{eig}}$ can be factorized as follows:

$$\mathbf{S}_{\mathrm{eig}} = \mathbf{S}_{\mathrm{eig}(\lambda)} \mathbf{S}_{\mathrm{eig}(\mu)}, \tag{38}$$

with

$$\mathbf{S}_{\mathrm{eig}(\lambda)} = \begin{pmatrix} e^{2i\delta_\lambda} & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{S}_{\mathrm{eig}(\mu)} = \begin{pmatrix} 1 & 0 \\ 0 & e^{2i\delta_\mu} \end{pmatrix}. \tag{39}$$

Such a factorization enables one to distinguish, inside the two channel problem, two separate matrices in which one of the channels is closed (the corresponding element is equal to one). Then, the scattering matrix $\mathbf{S} = \mathbf{R}\mathbf{S}_{\mathrm{eig}}\mathbf{R}^t$ can be put into the form of Eq. (35) with

$$\mathbf{S}_{\mathrm{left}} = \mathbf{R}\mathbf{S}_{\mathrm{eig}(\lambda)}\mathbf{R}^t, \quad \mathbf{S}_{\mathrm{right}} = \mathbf{R}\mathbf{S}_{\mathrm{eig}(\mu)}\mathbf{R}^t. \tag{40}$$

Next, the introduction of the so-called density matrices,[15] $\mathbf{P}_\lambda$ and $\mathbf{P}_\mu$, defined by

$$\mathbf{P}_\lambda = \mathbf{R}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\mathbf{R}^t, \quad \mathbf{P}_\mu = \mathbf{R}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\mathbf{R}^t, \tag{41}$$

with $\mathbf{P}_\lambda \mathbf{P}_\mu = \mathbf{P}_\lambda \delta_{\lambda\mu}$ ($\delta_{\lambda\mu}$ represents here the Kronecker delta symbol) and the use of the transition terms $T_{\lambda,\mu} = (e^{2i\delta_{\lambda,\mu}} - 1)/2i$ leads to

$$\mathbf{S}_{\mathrm{left}} = \mathbf{I} + 2iT_\lambda \mathbf{P}_\lambda, \quad \mathbf{S}_{\mathrm{right}} = \mathbf{I} + 2iT_\mu \mathbf{P}_\mu. \tag{42}$$

Therefore, the **S** matrix can be expressed by means of the density matrices:

$$\mathbf{S} = \mathbf{I} + 2i[T_\lambda \mathbf{P}_\lambda + T_\mu \mathbf{P}_\mu]. \tag{43}$$

In terms of the pseudovector $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ whose components are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{44}$$

we have

$$\mathbf{P}_\lambda = \tfrac{1}{2}(\mathbf{I} + \vec{L}\cdot\vec{\sigma}), \quad \mathbf{P}_\mu = \tfrac{1}{2}(\mathbf{I} - \vec{L}\cdot\vec{\sigma}), \tag{45}$$

the relation $\mathbf{P}_\mu = \mathbf{I} - \mathbf{P}_\lambda$ being easily deduced and

$$\vec{L} = (\sin 2\varsigma, 0, \cos 2\varsigma). \tag{46}$$

It follows from Eqs. (43) and (45) that

$$\mathbf{S} = \mathbf{I} + \mathbf{i}[T_\lambda(\mathbf{I} - \vec{L}\cdot\vec{\sigma}) + T_\mu(\mathbf{I} + \vec{L}\cdot\vec{\sigma})]. \tag{47}$$

Though this form appears rather formal, its advantage is to exhibit the $\vec{L}$ vector allowing geometrical discussions. When the mixing angle ς varies, the extremity of the $\vec{L}$ vector describes, in the (u_1, u_3) plane of the space materialized by three orthogonal axes indexed u_1, u_2, u_3, a circle of radius one centered at the origin $(0,0,0)$. We recall that for a fluid cavity in an elastic medium, $\vec{L}$ describes an helical path on the unit sphere.[7] There are two reasons to such a difference of behavior. For the fluid cavity, (i) only one of the transition terms describes the whole scattering process, while the other one is null (only one eigenchannel is involved in the scattering, the other being closed); (ii) the rotation matrix $\mathbf{R}$ involves two independent angles instead of one for the plate.

By taking into account both the derivative of Eq. (43),

$$\mathbf{S}' = 2i[T'_\lambda \mathbf{P}_\lambda + T'_\mu \mathbf{P}_\mu + T_\lambda \mathbf{P}'_\lambda + T_\mu \mathbf{P}'_\mu], \tag{48}$$

and the Hermitian conjugate

$$\mathbf{S}^\dagger = \mathbf{I} - 2i(T^*_\lambda \mathbf{P}_\lambda + T^*_\mu \mathbf{P}_\mu), \tag{49}$$

and by using the relations $\mathbf{P}^t_\lambda = \mathbf{P}_\lambda$, $\mathbf{P}^t_\mu = \mathbf{P}_\mu$, $\mathbf{P}'_\mu \mathbf{P}_\lambda = -\mathbf{P}'_\lambda \mathbf{P}_\lambda$, $\mathbf{P}'_\lambda \mathbf{P}_\mu = -\mathbf{P}'_\mu \mathbf{P}_\mu$, and $\mathbf{P}_\lambda \mathbf{P}_\mu = \mathbf{P}_\lambda \delta_{\lambda\mu}$, the Wigner-Smith matrix $\mathbf{Q}_{\lambda\mu} = -i\mathbf{S}'\mathbf{S}^\dagger$ is obtained under the form

$$\begin{aligned}\mathbf{Q}_{\lambda\mu} = 2[&T'_\lambda(1 - 2iT^*_\lambda)\mathbf{P}_\lambda + T'_\mu(1 - 2iT^*_\mu)\mathbf{P}_\mu + T_\lambda \mathbf{P}'_\lambda \\ &+ T_\mu \mathbf{P}'_\mu - 2i(T_\lambda - T_\mu)(T^*_\lambda \mathbf{P}'_\lambda \mathbf{P}_\lambda - T^*_\mu \mathbf{P}'_\mu \mathbf{P}_\mu)].\end{aligned} \tag{50}$$

With the following identities:

$$\operatorname{tr} \mathbf{P}_\lambda = \operatorname{tr} \mathbf{P}_\mu = 1, \tag{51}$$

$$\operatorname{tr} \mathbf{P}'_\lambda = \operatorname{tr} \mathbf{P}'_\mu = 0, \tag{52}$$

$$\operatorname{tr} \mathbf{P}'_\lambda \mathbf{P}_\lambda = \operatorname{tr} \mathbf{P}'_\mu \mathbf{P}_\mu = 0, \tag{53}$$

it follows

$$\operatorname{tr} \mathbf{Q}_{\lambda\mu} = T'_\lambda(1 - 2iT^*_\lambda) + T'_\mu(1 - 2iT^*_\mu). \tag{54}$$

VI. APPLICATION: DERIVATIVES WITH RESPECT TO THE FREQUENCY

A. Connection with the two channel RST

It must be noticed, at first, that in the case of asymmetrically loaded plates, the modes in the λ channel (respectively, μ channel) tend to the antisymmetric (respectively, symmetric) modes when the fluids are identical.[13] For the fluids considered here, one would not see objection to use the words antisymmetric and symmetric. In the two channel theory, a resonance corresponds to the case where one of the eigenphases, δ_λ, varies strongly while the other eigenphase δ_μ remains constant. This case occurs when the resonances are separated enough to consider them as independent ones. Let us denote by X the dimensionless frequency variable. In these conditions, in the neighborhood of a resonance position X_p, one gets the Breit-Wigner approximations

$$e^{2i\delta_\lambda} \approx \frac{X - X_p - i\Gamma^+_p/4}{X - X_p + i\Gamma^+_p/4}, \quad e^{2i\delta_\mu} \simeq 1, \tag{55}$$

from the exact formulas (given in Appendix B in the particular case of a single plate). The term $X_p - i\Gamma^+_p/4$ represents the pole of the reflection (or the transmission) coefficient or the pole of $e^{2i\delta_\lambda}$. It is recalled that $\Gamma^+_p = \Gamma_{1,p} + \Gamma_{2,p}$ and $\Gamma^-_p = \Gamma_{1,p} - \Gamma_{2,p}$.[13] The terms $\Gamma_{1,p}$ and $\Gamma_{2,p}$ are the partial widths quantifying the contribution of each fluid, defined by[1]

$$\Gamma_{m,p} = +2\tau_{m,p} \Big/ \left(\frac{dCa}{dX}(X_p)\right), \quad m = 1,2. \tag{56}$$

They are related to the mixing angle ς by the formulas $\Gamma^+_p \cos^2 \varsigma = \Gamma_{1,p}$ and $\Gamma^+_p \sin^2 \varsigma = \Gamma_{2,p}$. From Eq. (55), it follows that

$$\delta_\lambda = 2 \arctan \frac{\Gamma^+_p}{X_p - X}, \quad \delta_\mu \approx p\pi (p \in \mathbb{Z}), \tag{57}$$

and that

$$\delta'_\lambda \approx \frac{\Gamma^+_p/4}{(X - X_p)^2 + (\Gamma^+_p/4)^2}, \quad \delta'_\mu \approx 0. \tag{58}$$

In the neighborhood of the resonance position X_p, the approximations for the moduli of the reflection and transmission coefficients are

$$|r_1| = \frac{[(X - X_p)^2 + (\Gamma^-_p/4)^2]^{1/2}}{[(X - X_p)^2 + (\Gamma^+_p/4)^2]^{1/2}}, \tag{59}$$

$$|\chi t_1| = \frac{(\Gamma_{1,p}\Gamma_{2,p}/4)^{1/2}}{[(X - X_p)^2 + (\Gamma^+_p/4)^2]^{1/2}}. \tag{60}$$

Those for the phases of the reflection coefficients are

$$\phi = -\arctan \frac{\Gamma^+_p/4}{X - X_p} - \arctan \frac{\Gamma^-_p/4}{X - X_p}, \tag{61}$$

$$\psi = -\arctan \frac{\Gamma^+_p/4}{X - X_p} + \arctan \frac{\Gamma^-_p/4}{X - X_p}, \tag{62}$$

while for the transmission coefficient

$$\varphi = -\arctan \frac{\Gamma^+_p/4}{X - X_p}. \tag{63}$$

The corresponding derivatives are

$$|r_1|' = \frac{(X - X_p)[(\Gamma^+_p/4)^2 - (\Gamma^-_p/4)^2]}{[(X - X_p)^2 + (\Gamma^+_p/4)^2]^{1/2}[(X - X_p)^2 + (\Gamma^+_p/4)^2]^{3/2}}, \tag{64}$$

$$|\chi t_1|' = -\frac{(X-X_p)(\Gamma_{1,p}\Gamma_{2,p}/4)^{1/2}}{[(X-X_p)^2+(\Gamma_p^+/4)^2]^{3/2}}, \tag{65}$$

$$\phi' = \frac{\Gamma_p^+/4}{(X-X_p)^2+(\Gamma_p^+/4)^2} + \frac{\Gamma_p^-/4}{(X-X_p)^2+(\Gamma_p^-/4)^2}, \tag{66}$$

$$\psi' = \frac{\Gamma_p^+/4}{(X-X_p)^2+(\Gamma_p^+/4)^2} - \frac{\Gamma_p^-/4}{(X-X_p)^2+(\Gamma_p^-/4)^2}, \tag{67}$$

and

$$\varphi' = \frac{\Gamma_p^+/4}{(X-X_p)^2+(\Gamma_p^+/4)^2}. \tag{68}$$

In particular, at the resonance frequency $X=X_p$, one has

$$\delta_\lambda' = \varphi' \simeq \frac{4}{\Gamma_p^+}, \tag{69}$$

and

$$\phi' \simeq \frac{4}{\Gamma_p^+} + \frac{4}{\Gamma_p^-}, \quad \psi' \simeq \frac{4}{\Gamma_p^+} - \frac{4}{\Gamma_p^-}. \tag{70}$$

Let us give now the form of the $\mathbf{Q}_{\phi\varphi}$ matrix at the resonance frequency $X=X_p$. Formulas (64) and (65) ensure that the Wronskian W given by Eq. (10) vanishes in the off diagonal elements q_{12} and q_{21}. After straightforward calculations, it follows

$$\mathbf{Q}_{\phi\varphi}^{X=X_p} = \frac{16}{\Gamma_p^{+2}} \begin{pmatrix} \frac{\Gamma_{1,p}}{2} & \frac{\mp\Gamma_{1,p}^{1/2}\Gamma_{2,p}^{1/2}}{2} \\ \frac{\mp\Gamma_{1,p}^{1/2}\Gamma_{2,p}^{1/2}}{2} & \frac{\Gamma_{2,p}}{2} \end{pmatrix}. \tag{71}$$

One can verify that $\mathbf{Q}_{\lambda\mu}^{X=X_p} = \mathbf{Q}_{\phi\varphi}^{X=X_p}$. At the resonance frequency X_p, the diagonal elements of the $Q_{\phi\varphi}$ matrix provide the partial widths $\Gamma_{1,p}$ and $\Gamma_{2,p}$, once the total width Γ_p^+ is determined. In other words, Eq. (71) shows that the diagonal elements of the modified matrix $(1/\delta_\lambda'^2)\mathbf{Q}_{\phi\varphi}$ give the half values of the partial widths at the resonance frequency, the trace of this matrix being equal to the half of the total width. In the same way, the density matrix $\mathbf{P}_\lambda$ and the $\vec{L}$ vector become

$$\mathbf{P}_\lambda^{X=X_p} = \begin{pmatrix} \frac{\Gamma_{1,p}}{\Gamma_p^+} & \frac{\mp\Gamma_{1,p}^{1/2}\Gamma_{2,p}^{1/2}}{\Gamma_p^+} \\ \frac{\mp\Gamma_{1,p}^{1/2}\Gamma_{2,p}^{1/2}}{\Gamma_p^+} & \frac{\Gamma_{2,p}}{\Gamma_p^+} \end{pmatrix}, \tag{72}$$

and

$$\vec{L}^{X=X_p} = \left(\frac{\Gamma_p^-}{\Gamma_p^+}, 0, \pm\frac{\Gamma_{1,p}^{1/2}\Gamma_{2,p}^{1/2}}{\Gamma_p^+}\right). \tag{73}$$

As expected, the trace of each density matrix is equal to one. The comparison of Eqs. (71) and (72) leads to $\mathbf{Q}_{\phi\varphi}^{X=X_p} = (8/\Gamma_p^+)\mathbf{P}_\lambda^{X=X_p}$, i.e., the elements of the two matrices differ by factor $8/\Gamma_p^+$, which represents twice the amplitude of the phase derivative φ' (or δ_λ') at the resonance frequency. For identical fluid loading, $\Gamma_p^-=0$. Therefore, the $\vec{L}^{X=X_p}$ vector is along the u_3 axis. In the case of a resonance in the μ channel, similar calculations lead to the density matrix

$$\mathbf{P}_\mu^{X=X_p} = \begin{pmatrix} \frac{\Gamma_{2,p}}{\Gamma_p^+} & \frac{\mp\Gamma_{1,p}^{1/2}\Gamma_{2,p}^{1/2}}{\Gamma_p^+} \\ \frac{\mp\Gamma_{1,p}^{1/2}\Gamma_{2,p}^{1/2}}{\Gamma_p^+} & \frac{\Gamma_{1,p}}{\Gamma_p^+} \end{pmatrix}, \tag{74}$$

which can also be deduced formally by means of the relation $\mathbf{P}_\mu = \mathbf{I} - \mathbf{P}_\lambda$ [Eq. (45)].

B. Numerical results

When the fluids are identical, the information provided by the EM and the PGM are identical. The only advantage of the EM is to provide automatically the separation between symmetric and antisymmetric modes. When the fluids are different, only the EM is convenient to demonstrate the asymmetry of the loading. As an obvious consequence, the use of the $\mathbf{Q}_{\lambda\mu}$ description is more efficient than $\mathbf{Q}_{\phi\varphi}$, as will be shown in the following.

Consider the case of an aluminum plate loaded with water on one face and with glycerine on the other face (see Appendix A, for the constants). All the curves are given for the incidence angle $\theta_1=5°$ and are plotted versus the frequency variable $X=k_1d$ (k_1 is the wave number in fluid F_1). Only the case of derivatives with respect to the frequency are considered in the following.

1. The transition terms

The case of an aluminum plate loaded with water on one face and with glycerine on the other face has been studied previously by means of the transition terms[13] defined as the elements T_λ and T_μ of the matrix $\mathbf{T}=(\mathbf{S}_{\text{eig}}-\mathbf{I})/2i$. We recall in Fig. 1(a) the curves of the squared modulus of the transition terms $|T_\lambda|^2$ (antisymmetric modes) and $|T_\mu|^2$ (symmetric modes). Generally, the peaks of $|\chi t_1|^2$ do not reach unity. As a consequence, these plots cannot provide the accurate estimation of the widths and the positions of the resonances. At the opposite, the $|T_\lambda|^2$ and $|T_\mu|^2$ curves, plotted together, look like that of the transmission coefficient but have peaks reaching unity. It is then possible, in principle, to evaluate from the corresponding curves the position and width of the resonances, even if, as can be seen in Fig. 1(a), several resonance peaks have not rigorously the symmetric behavior expected from the Breit-Wigner formula (at values of $|T_\lambda|^2$ and $|T_\mu|^2$ less than 0.25). Near the frequencies $X\simeq 3.40$, 6.65, 9.93, 13.37, 16.73, 19.97, a careful examination would show that the curves of the two transition terms are not smooth and exhibit sharp variations, as in Fig. 1(b). Around these frequencies, the transition terms behave as if they exchange their properties (from symmetric to antisymmetric and conversely). Obviously, when the fluids loading the plate are identical, the shape of the curves is different, i.e., it becomes smooth, and no exchange of property can be assumed. This phenomenon does not happen at all the incidence angles, for instance, at $\theta_1=20°$, it no longer exists.

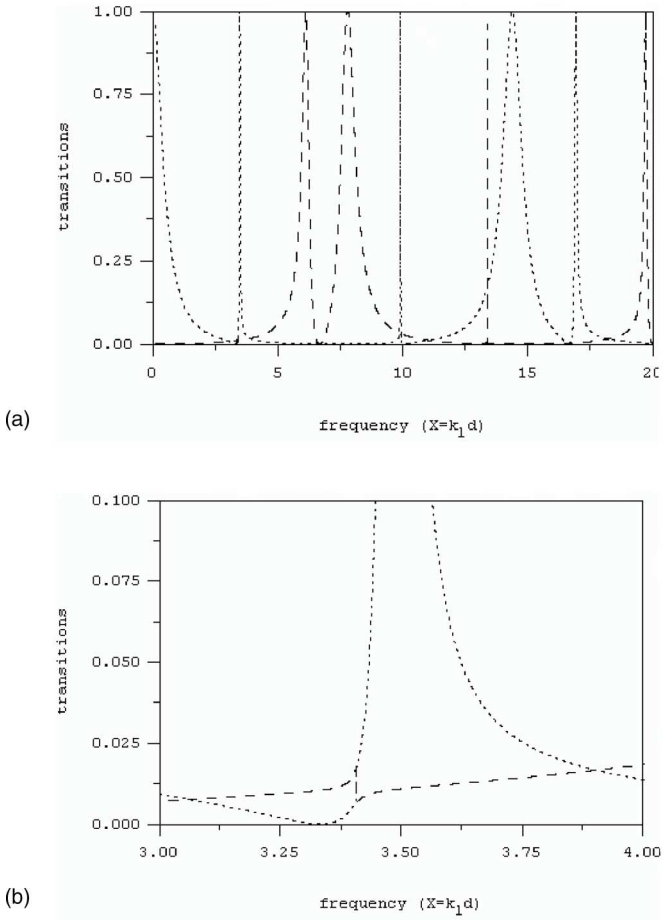


FIG. 1. (a) Squared moduli of the transition terms $|T_\lambda|^2$ (dotted) and $|T_\mu|^2$ (dashed) in the frequency variable $X=k_1d$ for the incidence angle $\theta_1=5°$ (reported from Ref. 13). (b) Zoom of (a) in the frequency range from $X=3$ to $X=4$ showing the discontinuity at $X=3.40$.

2. The phase derivatives

In Fig. 2, the curve of the phase derivative φ', truncated at a value of 25, shows peak maxima located at the same resonance frequencies as those found in Fig. 1(a). Both the resonance frequencies and peak heights are indicated in Tables I and II. According to the definition given by the PGM, the higher the peak, the smaller the resonance width; the height of one peak must be equal to four times the inverse of the corresponding resonance width. To verify this fact, the "exact" values of the heights of φ' reported from the curves and the "approximate" values calculated with Eq. (56) are reported in Tables I and II. The agreement is good between exact and approximate values. For the antisymmetrical modes, the values of the partial widths of the resonances are also presented.

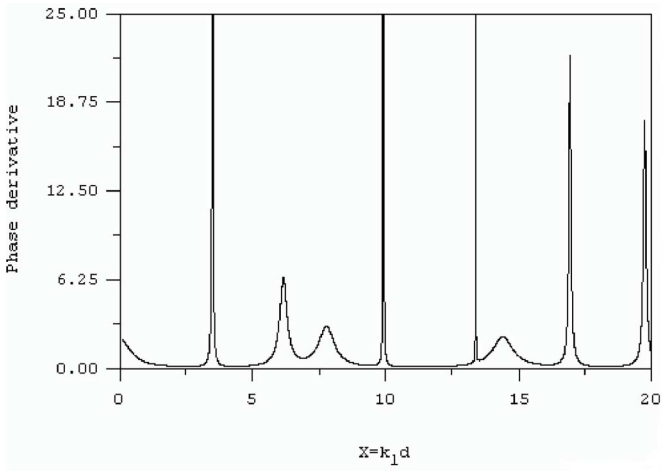


FIG. 2. The phase derivatives with respect to the frequency φ'. The peak heights have been truncated to a value of 25. Both the position and peaks heights are given in Tables I and II.

In Fig. 3(a), the measurement of the peak heights of the phase derivatives δ'_λ (dotted curve) and δ'_μ (dashed curve) would show that they are a little smaller than those of Fig. 2. This is a consequence of Eq. (29): for each frequency, one must add the contributions of both δ'_λ and δ'_μ to be able to find exactly the value of φ'. As a general rule, relatively far from the plate resonance frequencies, the two derivatives δ'_λ and δ'_μ are negligible. But around the frequencies $X\simeq 3.40$, 6.65, 9.93, 13.37, 16.73, 19.97 where the transition terms have a singular behavior, δ'_λ (or δ'_μ) shows small peaks, as in the zoom of Fig. 3(b) for δ'_μ. Both the dotted and the dashed curves have smooth variations everywhere, except in the vicinity of the frequency $X\simeq 3.40$ corresponding to the position of the small peak maximum. The presence of the small peaks is due to the asymmetrical fluid loading. When the fluids are identical, these peaks vanish. So, at this stage, it is possible to distinguish between the plate resonances (great peaks) and the asymmetry resonances (small peaks, second-order phenomena comparatively to the great peaks). For the small resonance peaks, it can be noted that no approximation formula has been found.

If one reports once more to Fig. 3(b), a new set of two smooth and composite peaks can be exhibited and interpreted as resulting from the exchange of properties already mentioned for the transition terms. For increasing frequencies, the plot of the small peak begins with a dotted part, and, after its maximum, becomes a dashed part. At the opposite, the plot of the great peak beginning with a dashed part, ends with a dotted part. The accurate examination of Figs. 3(a) and 1(a) shows that the frequency at which one goes from dot to dash (and conversely) corresponds exactly to the frequency at which the moduli of the transition terms have abrupt vertical variations, as in Fig. 1(b).

A comparison between Figs. 2 and 3(a) leads one to point out that the great peaks of δ'_λ and δ'_μ may be considered in place of φ'. With a good accuracy, they provide most of the resonance widths. It can be seen that $\delta'_\lambda\simeq\varphi'$ or $\delta'_\mu\simeq\varphi'$. For the same reasons that led one to use the transitions in place of the transmission coefficient, the eigenphase derivatives must be used in place of the phase derivative of the transmission coefficient, as they contain much more information. As a first conclusion, the use of the eigenphase derivatives presents at least two advantages:

(i) Contrary to φ', they enable one to distinguish between the two types of resonances for the plate: symmetric (dashed curve peaks) or antisymmetric (dotted curve peaks).

TABLE I. The first antisymmetric resonances at the incidence angle $\theta_1=5°$. The agreement is good between the "exact" resonance heights measured on the curve of φ', Fig. 2, and the approximate one calculated from Eq. (56).

Generalized Lamb mode A_p	Resonance frequency X_p	Resonance heights $4/\Gamma_p^+$ (exact)	Resonance heights $4/\Gamma_p^+$ (approximated)	Partial widths Γ_1 (approximated)	Partial widths Γ_2 (approximated)
A_1	3.485	56.7	57.5	2.64×10^{-2}	4.31×10^{-2}
A_2	9.913	153.4	152.6	9.95×10^{-3}	1.63×10^{-2}
A_3	14.395	2.19	2.19	6.93×10^{-1}	1.131
A_4	16.920	22.03	22.6	6.72×10^{-2}	1.097×10^{-1}

(ii) The small peaks, not easy to detect in φ', indicate the existence of asymmetrical loading by the fluids. They vanish for identical fluids.

One can also represent the curves corresponding to the projection of the peaks in the (X,θ_1) plane. Those of the Wigner-Smith matrix eigenvalue $\Lambda^+_{\phi\varphi}$, in Fig. 4, look like those of φ' (not represented in the paper). However, the computations show that the peak heights [i.e., the intensities in the (X,θ_1) plane] differ. The only exception concerns the mode A_0 which lies in the $\Lambda^-_{\phi\varphi}$ eigenvalue. At the same scale of representation in the (X,θ_1) plane, the peaks of $\Lambda^-_{\phi\varphi}$ would appear as second-order phenomena. These results show that the symmetric and antisymmetric characters of the resonances cannot be identified by means of the Wigner-Smith matrix eigenvalues.

3. The mixing angle

At first, let us illustrate the role played by the mixing angle ς in the relation that links $\mathbf{Q}_{\phi\varphi}$ and the frequency derivatives δ'_λ and δ'_μ. It is shown (see the end of Sec. VI A) that at the resonance frequency $X=X_p$ of the eigenvalue $e^{2i\delta_\lambda}$, one gets the remarkable relation $\mathbf{Q}^{X=X_p}_{\phi\varphi}=(8/\Gamma_p^+)\mathbf{P}^{X=X_p}_\lambda$. It is also shown in Appendix B, that for $X=X_p$, the first diagonal element $\cos^2\varsigma$ of the density matrix $\mathbf{P}^{X=X_p}_\lambda$ can be approximated by $\cos^2\varsigma\approx(\varepsilon^2+1)/4$ ($\varepsilon^2=\tau_1/\tau_2$ being nothing but the contrast between the fluids). It follows that the plot of the first diagonal element q_{11} of $\mathbf{Q}_{\phi\varphi}$ should coincide at $X=X_p$ with the plot of $2\delta'_\lambda$ weighted by $(\varepsilon^2+1)/4$. Indeed, in Fig. 5(a) plotted for $\theta_1=15°$, the heights of the q_{11} peaks centered at the values of X_p: 13.3, 28.1, 43.5, 59.1, and 74.7 equate the heights of the $2\delta'_\lambda[(\varepsilon^2+1)/4]$ peaks. Figure 5(b) illustrates the equation $\mathbf{Q}_{\phi\varphi}=(8/\Gamma_p^+)\mathbf{P}^{X=X_p}_\lambda$ at resonance frequencies $X=X_p$ of the eigenvalue $e^{2i\delta_\mu}$, by comparing q_{11} and $2\delta'_\mu[(\varepsilon^2+1)/4]$ at the same incidence angle $\theta_1=15°$. Once again, at the values of X_p: 19.6, 35.6, 51.3, and 66.9, the heights of the peaks in both plots remarkably coincide.

In Fig. 6, the peaks of the mixing angle derivative $|\varsigma'|$, plotted for $\theta_1=5°$, are located at the frequencies $X\simeq3.40$, 6.65, 9.93, 13.37, 16.73, 19.97 already mentioned for the transition terms. One effect of the mixing angle is to point out the frequencies where the transition terms show abrupt variations. It also informs on the character of the fluid-loading: identical fluids imply $|\varsigma'|=0$ whatever the frequency, while different fluids imply $|\varsigma'|\neq0$ in limited intervals as indicated in the figure. Therefore, it localizes what we have called the asymmetry resonances and seems to be an adequate tool to indicate the difference in fluid-loading.

TABLE II. The first symmetrical resonances at the incidence angle $\theta_1=5°$. The agreement is good between the "exact" resonance heights measured on the curve of φ', Fig. 2, and the approximate one calculated from Eq. (56).

Generalized Lamb mode S_p	Resonance frequency X_p	Resonance heights $4/\Gamma_p^+$ (exact)	Resonance heights $4/\Gamma_p^+$ (approximated)
S_1	6.155	6.36	6.64
S_2	7.776	2.93	2.95
S_3	13.374	920	1026.21
S_4	19.73	17.44	18.07

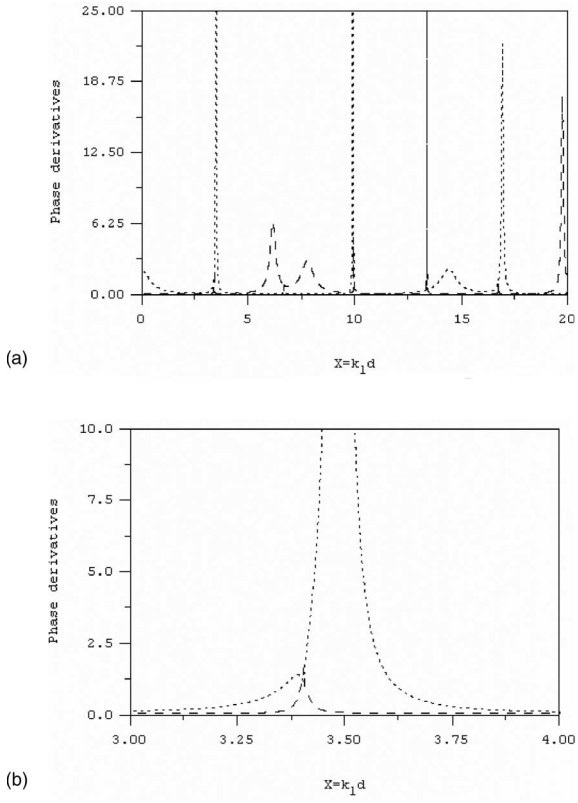


FIG. 3. (a) The eigenphase derivatives with respect to the frequency δ'_λ (dotted), δ'_μ (dashed). The peak heights have been truncated to a value of 25. (b) An enlargement of (a) at the neighborhood of the resonance frequency $X=3.40$. The contribution of δ'_μ (dashed) to the value of the peak height of δ'_λ (dotted) is negligible.

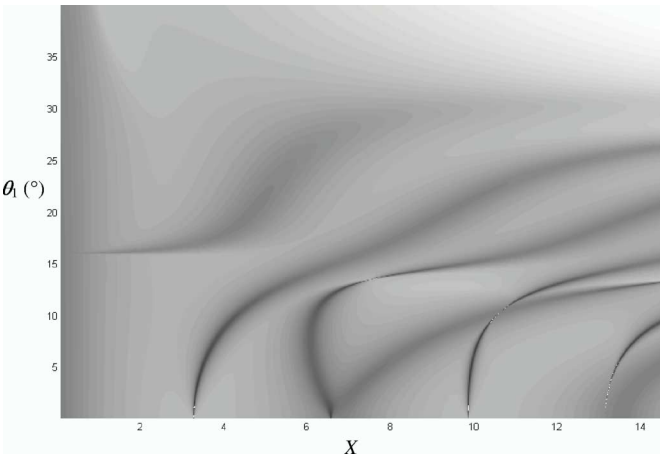


FIG. 4. Curves of the peaks of the eigenvalue $|\Lambda^+_{\phi\varphi}|$ in the (X,θ_1) plane (white=lowest value, black=highest value).

VII. APPLICATION: DERIVATIVES WITH RESPECT TO THE INCIDENCE ANGLE

In this section, the frequency is assumed to be fixed. The analysis of Sec. VI A can be adapted to the case where the incidence angle is the variable. For convenience, we report here only some of the formulas needed for the discussions. $\bar{k}_x=\sin\theta_1$ denotes the sine of the incidence angle (see Appendix A). In the vicinity of an antisymmetric resonance position $\bar{k}_{x,p}$, the approximations for the eigenvalues of the scattering matrix are[13]

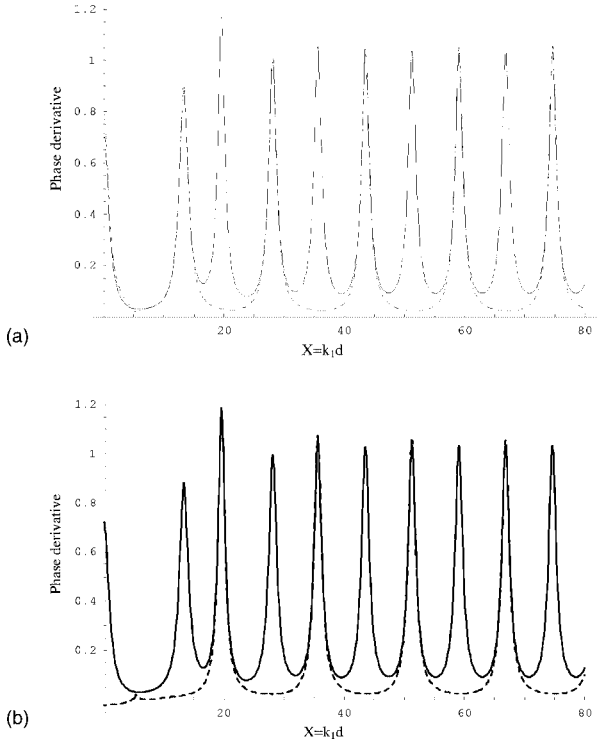


FIG. 5. (a) Plot of q_{11} component of $\mathbf{Q}_{\phi\varphi}$ (plain line) compared to $2\delta'_\lambda[(\varepsilon^2+1)/4]$ (dashed line) at $\theta_1=15°$. (b) Plot of q_{11} component of $\mathbf{Q}_{\phi\varphi}$ (plain line) compared to $2\delta'_\mu[(\varepsilon^2+1)/4]$ (dashed) at $\theta_1=15°$.

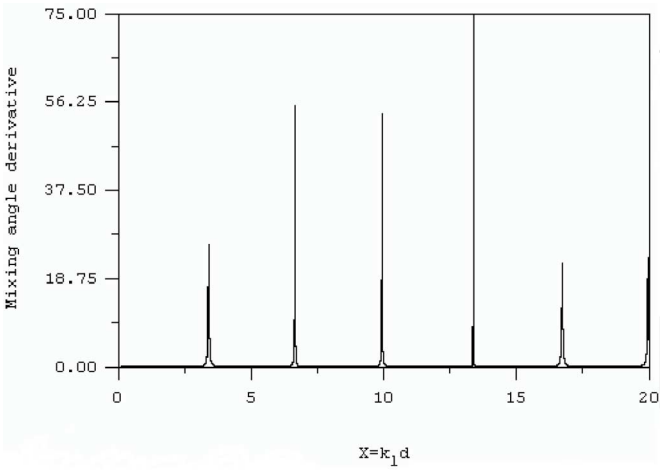


FIG. 6. The absolute value of the mixing angle derivative with respect to the frequency $|\varsigma'|$ plotted vs the frequency for the incidence angle $\theta_1=5°$. The peak located at $X\simeq 13.37$ culminates at 489.9.

$$e^{2i\delta_\lambda}\approx\frac{\bar{k}_x-\bar{k}_{x,p}+i\gamma_p^+/4}{\bar{k}_x-\bar{k}_{x,p}-i\gamma_p^+/4},\quad e^{2i\delta_\mu}\simeq 1. \tag{75}$$

The $\bar{k}_{x,p}+i\gamma_p^+/4$ term represents the pole of the reflection (or the transmission) coefficient or the pole of $e^{2i\delta_\lambda}$. It is recalled that $\gamma_p^+=\gamma_{1,p}+\gamma_{2,p}$ and $\gamma_p^-=\gamma_{1,p}-\gamma_{2,p}$. The terms $\gamma_{1,p}$ and $\gamma_{2,p}$ are the angular partial widths defined by

$$\gamma_{m,p}=-2\tau_{m,p}\Big/\left(\frac{dCa}{d\bar{k}_x}(\bar{k}_{x,p})\right),\quad m=1,2. \tag{76}$$

From the above-noted expressions, the phase derivatives (now with respect to $\bar{k}_x$) are

$$\delta'_\lambda\approx\frac{-\gamma_p^+/4}{(\bar{k}_x-\bar{k}_{x,p})^2+(\gamma_p^+/4)^2},\quad \delta'_\mu\approx 0. \tag{77}$$

It can be deduced, from the approximate expressions of the reflection and transmission coefficients, that $\varphi'\approx\delta'_\lambda$. In particular, at the resonance frequency $\bar{k}_x=\bar{k}_{x,p}$, one has

$$\delta'_\lambda=\varphi'\simeq\frac{-4}{\gamma_p^+}. \tag{78}$$

In order to examine briefly the properties of the phase derivatives with respect to the incidence angle, two values of the dimensionless frequency are considered. The value $X=4.23$, or in frequency-thickness units $fd\simeq 2$ Mhz mm, is chosen to plot the curves presented in Fig. 7 [see also Fig. (12) of Ref. 10]. The resonances are well separated. In addition, the two eigenphase derivatives δ'_λ and δ'_μ enable an easier distinction between antisymmetric (dotted curve) and symmetric (dashed curve) modes. The first resonance minimum corresponding to the A_1 mode is located at the angle $\bar{\theta}_p\simeq 9.70$ (or $\bar{k}_{x,p}\simeq 0.17$) with an amplitude $-4/\gamma_p^+=-158.90$. The resonance width is then $\gamma_p^+\simeq 0.0025$. For identical fluids (water), it has been found $\bar{\theta}_p\simeq 9.69$ and $-4/\gamma_p^+\simeq -209.55$. As expected, the effect of a heavier fluid on one of the faces is two enlarge the resonance width, with no significant modification of the position.

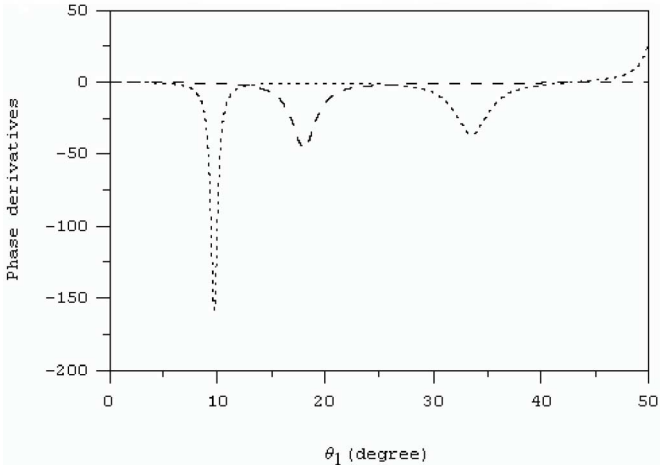


FIG. 7. The phase derivatives with respect to the sine of the angle. δ'_λ (dotted) and δ'_μ (dashed) plotted vs the incidence angle θ_1. Dimensionless frequency $X=4.23$.

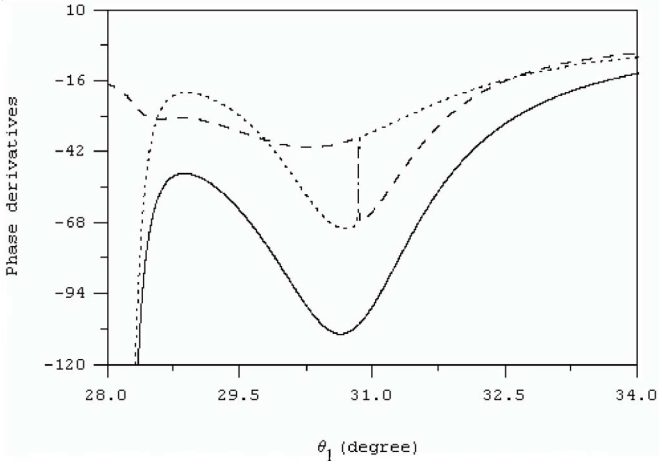


FIG. 9. The phase derivatives with respect to the sine of the angle. φ' (solid), δ'_λ (dotted), and δ'_μ (dashed) plotted vs the incidence angle θ_1. Dimensionless frequency $X=25$.

For the dimensionless frequency $X=25$, we give at first, in Fig. 8, the curves of the two transition terms in the angular domain $[28°–34°]$ that includes the Rayleigh angle. When $\theta_1>29°$, the modulus of the transmission coefficient is equal to zero, while each of the transition terms (dotted curve for $|T_\lambda|^2$, dashed curve for $|T_\mu|^2$) detects one resonance. The presence of two resonances is due to the two generalized Rayleigh waves that propagate along the water/aluminum and the glycerine/aluminum interfaces. The curves of the phase derivatives φ' (solid curve), δ'_λ (dotted curve), and δ'_μ (dashed curve), are given in Fig. 9. In the $[29°–34°]$ domain, the curve of φ' exhibits only one minimum, the amplitude of which provides an easy interpretation only when the fluids loading the plate are identical. One can verify, as stated by Eq. (29), that φ' is an average of the eigenphase derivatives δ'_λ and δ'_μ. At the scale of the figure, both curves of δ'_λ and δ'_μ present a jump at the Rayleigh angle $\theta_1 \simeq 30.84$. A zoom near this angle would show that the jumps are not truly vertical. If the curves of δ'_λ and δ'_μ are put together as in the figure, they draw two resonance minima, one located at $\theta_1 \simeq 30.69$ with an amplitude equal to -70.28, the other located at $\theta_1 \simeq 30.29$ with an amplitude equal to -40.45. As we are in the limit of very great values of the frequency where the interfaces are practically decoupled and the waves propagating along them connected to the fundamental modes A_0 and S_0, the widths correspond to partial widths. They must be computed with the help of the formula $-2/\gamma_{1,0}=-70.28$ and $-2/\gamma_{2,0}=-40.45$ rather than $-4/\gamma_0^+$, index 0 referring to the A_0 and S_0 modes. Finally, the values of $\bar{k}_{x,m}=\bar{k}_{x,0}+i\gamma_{m,0}/2$ $(m=1,2)$ that correspond to the minima are $\bar{k}_{x,1}\simeq 0.510+i0.014$ and $\bar{k}_{x,2}\simeq 0.504+i0.024$. They are in good agreement with those values obtained from the computation of the poles of the reflection coefficients at the water/aluminum ($\bar{k}_{x,1}\simeq 0.510+i0.015$) and the glycerine/aluminum ($\bar{k}_{x,2}\simeq 0.504+i0.027$) interfaces.

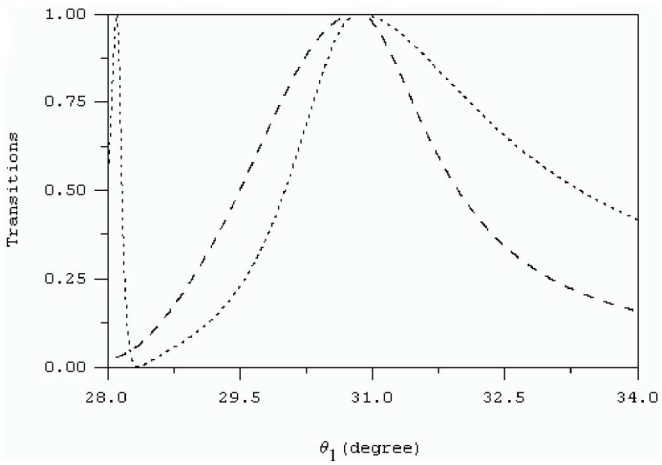


FIG. 8. Squared moduli of the transition terms $|T_\lambda|^2$ (dotted), and $|T_\mu|^2$ (dashed) plotted vs the incidence angle θ_1. Dimensionless frequency $X=25$.

VIII. CONCLUSION

The Wigner-Smith matrix enables the unification of three methods (RST, PGM, and EM) originally built up independently. It also provides a general frame useful to solve some scattering problems in acoustics. In this paper, the general properties of the Wigner-Smith matrix have been investigated in the case of structures with simple geometrical shape, with a special focus on plates. It has been shown that the study of the resonances can follow two ways: either by considering a description which uses the phases of the elements of the scattering matrix (e.g., the phases of the reflection and transmission coefficients), or by considering a description which uses the phases of the eigenvalues of the scattering matrix. The second description, although not straight, gives more results than the first one. The main results are as follows.

(1) The phase derivative of the transmission coefficient is equal to the sum of the eigenphase derivatives. Then, by a separate study of the eigenphases, it is possible to decompose and to understand a little more the physical content of the transmission coefficient.
(2) The description which uses the eigenphase derivatives allows one to build up density matrices useful to in-

troduce a geometrical interpretation for the scattering. It also allows one to interpret the effects of the mixing angle derivatives.

(3) The numerical computations led in the case of plates, show that the mixing angle derivative helps to distinguish between the effects due to the asymmetrical loading by the fluids and the resonances of the plate.

Although the numerical computations have been performed for plates, the method proposed here can be used for the study of the acoustic scattering by other submerged structures. For fluid-loaded stack of plates[16] as for single cavities embedded in an elastic medium, the results of Secs. II–IV need very little modifications. The Wigner-Smith matrix concept can be extended to cases of higher dimension for the scattering matrix. For instance, a stack of plates sandwiched between elastic semi-infinite media admits a 4×4 scattering matrix (four channel scattering). The four eigenphases derivatives can be helpful to analyze accurately the vibration modes. In multiple scattering by cylindrical or spherical cavities in an elastic medium, the scattering matrix is of infinite dimension. It is also made up of 2×2 block submatrices similar to those described in the Sec. I. The 2×2 blocks can be studied separately (even if they are not unitary) or the infinite matrix truncated to an $N\times N$ matrix under the constraint that the unitary condition is satisfied. The N eigenphase derivatives of such a matrix can serve to study the multiple scattering.

APPENDIX A

We recall here the reflection and transmission coefficients for fluid-loaded plates[13] (F1/S/F2 systems). Fluid Fi ($i=1,2$) has density ρ_i, sound wave speed c_i with, for time harmonic waves of angular frequency $\omega=2\pi f$, an associate wave number $k_i=\omega/c_i$. Solid S has density ρ. The velocity of the longitudinal L (respectively, transverse T) wave is c_L (respectively, c_T) with an associate wave number $k_L=\omega/c_L$ (respectively, $k_T=\omega/c_T$). The thickness of the plate is denoted by d. From the above-defined physical quantities, dimensionless parameters can be built, with reference to fluid F1, that help to express the antisymmetric Ca, symmetric Cs, and fluid loading τ_i functions as follows:

$$Ca = 4\bar{k}_x^2\bar{k}_{zL}\bar{k}_{zT}\tan(\bar{k}_{zT}X/2) + (2\bar{k}_x^2 - n_T^2)^2\tan(\bar{k}_{zL}X/2), \tag{A1}$$

$$Cs = 4\bar{k}_x^2\bar{k}_{zL}\bar{k}_{zT}\cot(\bar{k}_{zT}X/2) + (2\bar{k}_x^2 - n_T^2)^2\cot(\bar{k}_{zL}X/2), \tag{A2}$$

$$\tau_i = \frac{\rho_i}{\rho} n_T^4 \frac{\bar{k}_{zL}}{\bar{k}_{zi}}. \tag{A3}$$

In Eqs. (A1)–(A3),

$$\bar{k}_x = \sin\theta_1 = n_i\sin\theta_i = n_L\sin\theta_L = n_T\sin\theta_T, \tag{A4}$$

with $n_i=c_1/c_i$, $n_L=c_1/c_L$, $n_T=c_1/c_T$ and

$$\bar{k}_{zi} = (n_i^2 - \bar{k}_x^2)^{1/2}, \quad \bar{k}_{zL} = (n_L^2 - \bar{k}_x^2)^{1/2}, \quad \bar{k}_{zT} = (n_T^2 - \bar{k}_x^2)^{1/2}, \tag{A5}$$

while $X=k_1d$ (dimensionless frequency). The reflection coefficients r_1 (F1/S interface), r_2 (F2/S interface), and the transmission coefficient t_1 are (with $i=\sqrt{-1}$):

$$r_1 = \frac{(Ca - i\tau_1)(Cs - i\tau_2) + (Ca + i\tau_2)(Cs + i\tau_1)}{(Ca + i\tau_1)(Cs - i\tau_2) + (Ca + i\tau_2)(Cs - i\tau_1)}, \tag{A6}$$

$$r_2 = \frac{(Ca + i\tau_1)(Cs + i\tau_2) + (Ca - i\tau_2)(Cs - i\tau_1)}{(Ca + i\tau_1)(Cs - i\tau_2) + (Ca + i\tau_2)(Cs - i\tau_1)}, \tag{A7}$$

$$t_1 = \frac{2i\tau_1(Ca + Cs)\bar{k}_{z1}/\bar{k}_{z2}}{(Ca + i\tau_1)(Cs - i\tau_2) + (Ca + i\tau_2)(Cs - i\tau_1)}, \tag{A8}$$

The usual energy conservation law is $|r_1|^2+|\chi t_1|^2=1$ with $\chi = \sqrt{\rho_2 k_{z2}/\rho_1 k_{z1}}$.

The constants used in the computations are: $\rho = 2790$ kg/m^3, $c_L=6380$ m/s and: $c_T=3100$ m/s for aluminum; $\rho_1=1000$ kg/m^3 and $c_1=1485$ m/s for water; $\rho_2 = 1260$ kg/m^3 and $c_2=1920$ m/s for glycerine.

APPENDIX B

We give in the following the exact forms for the phase derivatives δ'_λ and δ'_μ. By using the formulas of Appendix A, a straightforward calculation gives

$$e^{2i\delta_\lambda} = \frac{P^*}{P}, \quad e^{2i\delta_\mu} = \frac{V^*}{V}, \tag{B1}$$

where

$$P = \frac{1}{2}\{2\sqrt{\tau_1\tau_2}(Ca + Cs)\sqrt{\Delta} - (\tau_1 + \tau_2)(Cs - Ca)\} + 2i\tau_1\tau_2, \tag{B2}$$

$$V = -\frac{1}{2}\{2\sqrt{\tau_1\tau_2}(Ca + Cs)\sqrt{\Delta} + (\tau_1 + \tau_2)(Cs - Ca)\} + 2i\tau_1\tau_2, \tag{B3}$$

with

$$\Delta = \left[\frac{(\tau_1 - \tau_2)(Ca - Cs)}{2\sqrt{\tau_1\tau_2}(Ca + Cs)}\right]^2 + 1. \tag{B4}$$

From expressions (B1)–(B4), the phase derivatives δ'_λ and δ'_μ are

$$\delta'_\lambda = \frac{\mathrm{Im}(PP'^*)}{|P|^2}, \quad \delta'_\mu = \frac{\mathrm{Im}(VV'^*)}{|V|^2}. \tag{B5}$$

In the particular case where identical fluids load the plate ($\tau_1=\tau_2$), one has

$$\delta'_\lambda = \frac{\tau_1 C'a - \tau'_1 Ca}{Ca^2 + \tau_1^2}, \quad \delta'_\mu = \frac{\tau_1 C's - \tau'_1 Cs}{Cs^2 + \tau_1^2}. \tag{B6}$$

The sum $\delta'_\lambda + \delta'_\mu$ gives Eq. (3) of Ref. 10.

Assuming that $\tau_1=\varepsilon\tau$ and $\tau_2=\tau/\varepsilon$, $\tau=\sqrt{\tau_1\tau_2}$ represents the logarithmic mean between τ_1 and τ_2, while $\epsilon^2=\tau_1/\tau_2$ can

be interpreted as the contrast factor between the acoustic impedances of fluids 1 and 2. From now on, our purpose is to demonstrate that the coupling angle ς, at the resonance of an eigenmode λ or μ and whatever the variable (frequency or angle), only depends on the contrast factor ε^2. The case of the frequency variable is considered in the following. At a resonance frequency $X=X_p$ of the eigenvalue $e^{2i\delta_\lambda}$, the resonance condition $\mathrm{Re}(P(X_p))\approx 0$ leads to

$$\frac{(Ca+Cs)}{(Cs-Ca)}\sqrt{\Delta}\approx\frac{(\tau_1+\tau_2)}{2\sqrt{\tau_1\tau_2}}=\frac{\varepsilon^2+1}{2\varepsilon}, \tag{B7}$$

and by applying the definition of $\sqrt{\Delta}$, Eq. (B4), at the same frequency,

$$\sqrt{\Delta}\approx\frac{\varepsilon^2+1}{2\varepsilon}. \tag{B8}$$

It can be concluded that $Ca\approx 0$. The first-order expansion of $P(X)$ near X_p,

$$P(X)\approx 2i\tau_1\tau_2+(X-X_p)\left\{\sqrt{\tau_1\tau_2}\sqrt{\Delta}\partial_X(Ca+Cs)-\sqrt{\tau_1\tau_2}\sqrt{\Delta}\frac{(Ca+Cs)}{(Cs-Ca)}\partial_X(Cs-Ca)\right\}_{X=X_p}, \tag{B9}$$

provides the resonance width Γ_p^+ in term of the partial width $\Gamma_{1,p}$:

$$\frac{\Gamma_p^+}{4}=\frac{2\tau_1\tau_2}{\sqrt{\tau_1\tau_2}\sqrt{\Delta}\left\{\partial_X(Ca+Cs)-\frac{(Ca+Cs)}{(Cs-Ca)}\partial_X(Cs-Ca)\right\}_{X=X_p}}=\frac{1}{\varepsilon^2+1}\Gamma_{1,p}. \tag{B10}$$

The comparison of this last relation with $\Gamma_p^+\cos^2\varsigma=\Gamma_{1,p}$ gives

$$\cos^2\varsigma=\frac{\varepsilon^2+1}{4}. \tag{B11}$$

Finally, the density matrix $\mathbf{P}_\lambda$, defined in Eq. (41), becomes

$$\mathbf{P}_\lambda^{X=X_p}=\begin{pmatrix}\frac{\varepsilon^2+1}{4} & \frac{\sqrt{(\varepsilon^2+1)(3-\varepsilon^2)}}{4}\\ \frac{\sqrt{(\varepsilon^2+1)(3-\varepsilon^2)}}{4} & \frac{3-\varepsilon^2}{4}\end{pmatrix}. \tag{B12}$$

[1] R. Fiorito, W. Madigosky, and H. Überall, "Resonance theory of acoustic waves interacting with an elastic plate," J. Acoust. Soc. Am. **66**, 1857–1866 (1979).
[2] P. D. Jackins and G. C. Gaunaurd, "Resonance acoustic scattering from stacks of bonded elastic plates," J. Acoust. Soc. Am. **80**, 1762–1776 (1986).
[3] L. Flax, L. R. Dragonette, and H. Überall, "Theory of elastic resonance excitation by sound scattering," J. Acoust. Soc. Am. **63**, 723–731 (1978).
[4] L. Flax, G. C. Gaunaurd, and H. Überall, "Theory of Resonance scattering," in *Physical Acoustics*, edited by P. Mason and R. N. Thurston (Academic, New York, 1981), Vol. **15**, pp. 191–294.
[5] S. G. Solomon, H. Überall, and K. B. Yoo, "Mode conversion and resonance scattering of elastic waves from a cylindrical fluid-filled cavity," Acustica **55**, 147–159 (1984).
[6] M. S. Choi and Y. M. Cheong, "Matrix theory of elastic resonance scattering and its application fluid-filled cavities," Acust. Acta Acust. 85, 170–180 (1999).
[7] P. Rembert, H. Franklin, and J.-M. Conoir, "Multichannel resonant scattering theory applied to fluid-filled cylindrical cavity in an elastic medium," Wave Motion **40**, 277–293 (2004).
[8] A. Freedman, "On the overlapping resonances: Concept of acoustic transmission through an elastic plate. I. An examination of properties," J. Sound Vib. **82**, 181–195 (1982).
[9] A. Freedman, "On the overlapping resonances: Concept of acoustic transmission through an elastic plate. II. Numerical examples and physical properties," J. Sound Vib. **82**, 197–213 (1982).
[10] O. Lenoir, J. Duclos, J.-M. Conoir, and J.-L. Izbicki, "Study of Lamb waves based upon the frequential and angular derivatives of the phase of the reflection coefficient," J. Acoust. Soc. Am. **94**, 330–343 (1993).
[11] S. Derible, P. Rembert, O. Lenoir, and J.-L. Izbicki, "Elastic plate: Experimental measurements of resonance widths," Acoust. Lett. **16**, 208–213 (1993).
[12] J.-M. Conoir, J.-L. Izbicki, and O. Lenoir, "Phase gradient method applied to scattering by an elastic shell," Ultrasonics **35**, 157–169 (1997).
[13] H. Franklin, E. B. Danila, and J.-M. Conoir, "S-matrix theory applied to acoustic scattering by asymmetrically fluid-loaded elastic isotropic plates," J. Acoust. Soc. Am. **110**, 243–253 (2001).
[14] F. T. Smith, "Lifetime matrix in Collision Theory," Phys. Rev. **118**, 349–356 (1960).
[15] D. Ter Haar, "Theory and applications of the density matrix," Rep. Prog. Phys. **24**, 304–362 (1961).
[16] H. Franklin, J.-L. Izbicki, T. Marie-Françoise, and P. Rembert, "Submerged plane layered isotropic media: Properties of the scattering matrix eigenvalues with application to bilayers," J. Acoust. Soc. Am. **116**, 1893–1896 (2004).

A stable boundary element method for modeling transient acoustic radiation

D. J. Chappell,[a] P. J. Harris, D. Henwood, and R. Chakrabarti
School of Computing and Mathematical Sciences, University of Brighton, Lewes Road, Brighton BN2 4GJ, United Kingdom

(Received 10 October 2005; revised 13 April 2006; accepted 17 April 2006)

Transient acoustic radiation from a closed axisymmetric three-dimensional object is modeled using the time domain boundary element method. The widely reported instability problems are overcome by reformulating the integral equation to obtain a Burton and Miller type equation in the time domain. The stability of such an approach is mathematically justified and supported by subsequent numerical results. The hypersingular integrals which arise are evaluated using a method valid for any surface discretization. Numerical results for the radiation of a spherical wave are presented and compared with an exact solution. The accuracy and stability of the results are verified for several geometrically different radiating objects. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2202909]

PACS number(s): 43.20.Px, 43.20.Bi [SFW] Pages: 74–80

I. INTRODUCTION

Time domain boundary integral methods have been used to solve wave propagation problems since the 1960s.[1,2] Since then increasing computer power has made numerical solutions possible over longer run times and so long-time instabilities in the time marching numerical solutions have become evident.[3–5] A number of methods have been suggested to resolve this such as time-averaging[6,7] and modified time-stepping.[4] Using an implicit formulation with high order interpolation and quadrature was also found to give stable results for all practical purposes.[8,9] Ha-Duong *et al.*[10] obtained stable results using a Galerkin approach and used an energy identity to prove stability of the Galerkin approximation. However, Galerkin methods are difficult and costly to implement and so have remained relatively unpopular despite their theoretical advantages.

The cause of these instabilities is discussed in Ref. 6 (in the context of the electric field integral equation) and is shown to be related to internal resonances of the scattering (or radiating) body. A similar argument is applied to the methods for acoustics problems later in this paper. In order to prevent these instabilities a time domain Burton-Miller type integral equation formulation like that originally applied to frequency domain problems[11] is used. A time domain formulation was proposed by Michielssen *et al.*[12] for acoustics scattering problems. It is shown that this formulation avoids the solution being corrupted by internal resonances and thus allows stability. The main difficulty in applying this method is evaluating the hypersingular integrals which are introduced. One method for doing this is to apply a similar limiting procedure to that of Terai[13] in the frequency domain as in Refs. 12 and 14. However, this has the disadvantage of being restricted to piecewise flat surface discretizations. The method employed here reformulates the hypersingular integrals into weakly singular ones using a Taylor series expansion and the identity of Meyer *et al.*[15]

II. THE INTEGRAL EQUATION FORMULATION

Let $\Omega \subset \mathbb{R}^3$ be a finite object with regular boundary surface Γ. Let $\Omega_+ = \mathbb{R}^3 \backslash \bar{\Omega}$ denote the unbounded exterior acoustic field and $\Omega_- = \Omega \backslash \Gamma$ denote the interior of Ω. Assume that Ω_+ is filled with a homogeneous compressible acoustic medium with speed of sound c. Let the radiated velocity potential be denoted by $\varphi : (\Omega_+ \cup \Gamma) \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ and consider the following initial-boundary value problem:

$$\nabla^2 \varphi(\underline{x},t) = \frac{1}{c^2} \frac{\partial^2 \varphi}{\partial t^2}(\underline{x},t) \quad \text{in } (\Omega_+ \cup \Gamma) \times \mathbb{R}_{\geqslant 0}, \tag{1}$$

$$\varphi(\underline{x},0) = 0 \quad \text{in } \Omega_+ \cup \Gamma, \tag{2}$$

$$\dot{\varphi}(\underline{x},0) = 0 \quad \text{in } \Omega_+ \cup \Gamma, \tag{3}$$

$$\frac{\partial \varphi}{\partial \hat{\underline{n}}_x}(\underline{x},t) = f(\underline{x},t) \quad \text{on } \Gamma \times \mathbb{R}_{\geqslant 0}, \tag{4}$$

where $\hat{\underline{n}}_x$ denotes the outward unit normal vector to Γ at $\underline{x}$, $\dot{\varphi}$ the time derivative and $f : \Gamma \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ is known. It is well known that this is a well-posed problem (see, for example, Ref. 10).

The initial-boundary value problem above may be represented by the following integral equation using Green's second identity[16]

$$\mathcal{D}(\varphi(\underline{x},t)) - \mathcal{S}\left(\frac{\partial \varphi}{\partial \hat{\underline{n}}_x}(\underline{x},t) \right) = \begin{cases} \frac{1}{2}\varphi(\underline{x},t) & \underline{x} \in \Gamma \\ \varphi(\underline{x},t) & \underline{x} \in \Omega_+, \end{cases} \tag{5}$$

where in the case $\underline{x} \in \Gamma$, it is assumed that Γ is locally differentiable at $\underline{x}$. The single-layer potential $\mathcal{S}$, and the double-layer potential $\mathcal{D}$ are defined as usual in terms of the fundamental solution of the wave equation

[a]Electronic mail: d.j.chappell@brighton.ac.uk

$$G(\underline{x},t) := \frac{1}{4\pi|\underline{x}|}\delta\left(t-\frac{|\underline{x}|}{c}\right),$$

where δ is the Dirac delta function. Explicitly, they are defined by

$$\mathcal{S}\varphi(\underline{x},t) := \int_0^{t^+}\int_\Gamma G(\underline{x}-\underline{\xi},t-s)\varphi(\underline{\xi},s)dS_{\underline{\xi}}ds$$

and

$$\mathcal{D}\varphi(\underline{x},t) := \int_0^{t^+}\int_\Gamma \frac{\partial G}{\partial \hat{\underline{n}}_{\underline{\xi}}}(\underline{x}-\underline{\xi},t-s)\varphi(\underline{\xi},s)dS_{\underline{\xi}}ds,$$

where $t^+=t+\varepsilon$ for arbitrarily small ε. This avoids taking the upper limit on the time integral at the singularity in δ. The classical jump relations in the form $[\mathcal{S}\varphi]=0$ and $[\mathcal{D}\varphi]=\varphi$ hold, where $[-]$ represents the jump across Γ. It is easy to show that the integral equation (5) is equivalent to the well-known Kirchoff or retarded potential integral equation:[17]

$$\frac{-1}{4\pi}\int_\Gamma\left\{\frac{\partial R}{\partial \hat{\underline{n}}_{\underline{\xi}}}\left(\frac{\varphi(\underline{\xi},\tau)}{R^2}+\frac{\dot{\varphi}(\underline{\xi},\tau)}{cR}\right)+\frac{1}{R}\frac{\partial\varphi}{\partial \hat{\underline{n}}_{\underline{\xi}}}(\underline{\xi},\tau)\right\}dS_{\underline{\xi}} = \begin{cases}\frac{1}{2}\varphi(\underline{x},t) & \underline{x}\in\Gamma\\ \varphi(\underline{x},t) & \underline{x}\in\Omega_+.\end{cases} \tag{6}$$

Here $R=|\underline{x}-\underline{\xi}|$ and $\tau=t-R/c$ is the retarded time.

A. Stability problems

As mentioned previously, it is well known that time marching solutions obtained using Eq. (6) suffer from poor stability. Since $\partial\varphi/\partial\hat{\underline{n}}_{\underline{\xi}}$ is known for all time with $\underline{x}\in\Gamma$, the term containing it will not affect the stability of the time marching solution. To study the stability properties it is therefore sufficient to consider the case where this term is zero, hence Eq. (6) becomes

$$\frac{1}{2}\varphi(\underline{x},t)=\frac{-1}{4\pi}\int_\Gamma\frac{\partial R}{\partial \hat{\underline{n}}_{\underline{\xi}}}\left(\frac{\varphi(\underline{\xi},\tau)}{R^2}+\frac{\dot{\varphi}(\underline{\xi},\tau)}{cR}\right)dS_{\underline{\xi}}. \tag{7}$$

The initial conditions mean the only time harmonic solutions that should be admitted by Eq. (7) are zero. However, supposing that $\varphi(\underline{x},t)=\hat{\varphi}(\underline{x})e^{-i\omega t}$ with $\omega\in\mathbb{R}_{\geqslant 0}$ is a time harmonic solution of Eq. (7) valid for $t\geqslant t'>0$, where t' is some arbitrary positive time gives

$$\frac{1}{2}\hat{\varphi}(\underline{x})e^{-i\omega t}=\frac{-1}{4\pi}\int_\Gamma\frac{\partial R}{\partial \hat{\underline{n}}_{\underline{\xi}}}\hat{\varphi}(\underline{\xi})e^{-i\omega t}\frac{e^{ikR}(1-ikR)}{R^2}dS_{\underline{\xi}}, \tag{8}$$

where $k:=\omega/c$ is the wavenumber. Hence the following equation for $\hat{\varphi}$ is obtained

$$\frac{1}{2}\hat{\varphi}(\underline{x})=\int_\Gamma\frac{\partial}{\partial \hat{\underline{n}}_{\underline{\xi}}}\left(\frac{e^{ikR}}{4\pi R}\right)\hat{\varphi}(\underline{\xi})dS_{\underline{\xi}}, \tag{9}$$

which is the well-known analogue of Eq. (7) in the frequency domain. This equation is known to admit nontrivial solutions for a set of discrete resonant wavenumbers.[11] Hence Eq. (7) must also admit these resonant solutions for $t\geqslant t'>0$.

The fact that these resonant time harmonic solutions are admitted by Eq. (7) is not in itself a cause of instability. However, the solution of the discretized version of Eq. (7) may be considered as a sum of discrete frequency components of the form $\hat{\varphi}(\underline{x})e^{\nu t}$ with $\nu\in\mathbb{C}$. Time harmonic solutions occur here when the sum is over a single value of ν lying on the imaginary axis. As in the case of the electric field integral equation studied in Ref. 6, the discretization leads to a loss of accuracy causing the values of ν to deviate from their theoretical values. This loss in accuracy is more severe for higher frequencies. For a time discretization with time step Δt, angular frequencies $\omega\geqslant\pi/\Delta t$ cannot be represented by the discrete model (Nyquist condition) and for $\omega\approx\pi/\Delta t$ the discrete model will be highly inaccurate and hence so will the values of ν in the calculation.[6] This is a problem since the resonances become more dense as ω increases[11] and so the smaller Δt is taken to be (necessary to approximate higher frequency radiation), the greater the chance that $\omega_{\text{res}}\approx\pi/\Delta t$ for some resonant angular frequency ω_{res}. If this occurs, then the value of ν corresponding to the resonance will be calculated inaccurately and may in general move off the imaginary axis into the right or left half planes. Any excited resonances for which ν moves into the right half plane give rise to an exponentially increasing oscillation or instability.

B. Burton-Miller-type time domain integral equation

In order to cure the above-described instability problems, a time domain analogue of the well-known Burton and Miller method in the frequency domain is used to obtain a solution for φ on Γ. A similar formulation to the one suggested in Ref. 12 is adapted to apply to radiation problems. This entails taking a linear combination of the derivative of Eq. (5) with respect to time and the normal derivative of Eq. (5). Explicitly this yields

$$(1-\alpha)\left\{\mathcal{D}(\dot{\varphi}(\underline{x},t))-\mathcal{S}\left(\left(\frac{\dot{\partial}\varphi}{\partial\hat{\underline{n}}_x}\right)(\underline{x},t)\right)-\frac{1}{2}\dot{\varphi}(\underline{x},t)\right\} + \alpha c\left\{\frac{\partial}{\partial\hat{\underline{n}}_x}\mathcal{D}(\varphi(\underline{x},t))-\frac{\partial}{\partial\hat{\underline{n}}_x}\mathcal{S}\left(\frac{\partial\varphi}{\partial\hat{\underline{n}}_x}(\underline{x},t)\right)-\frac{1}{2}\frac{\partial\varphi}{\partial\hat{\underline{n}}_x}(\underline{x},t)\right\} = 0, \tag{10}$$

where $\alpha\in(0,1)$ is a coupling parameter. The factor of c is required to balance the relative sizes of the terms in the two equations, i.e., so the power of c in the coefficients of like terms is the same in both equations. Note that the following easily proven identity has been used here:

$$\frac{\partial}{\partial t}(\mathcal{S}\text{ or }\mathcal{D})\varphi=(\mathcal{S}\text{ or }\mathcal{D})\dot{\varphi}. \tag{11}$$

Equation (10) may also be written in a similar form to Eq. (6). The upper (time derivative) term is like Eq. (6) for $\underline{x}\in\Gamma$, but with $\dot{\varphi}$ substituted for φ and the right-hand side subtracted. The lower (normal derivative) term is more complicated and contains hypersingular integrals.

The methods used to evaluate these will be discussed later. Note that in this integral equation formulation the val-

ues of $(\partial\dot{\varphi}/\partial\hat{\underline{n}})$ on $\Gamma\times\mathbb{R}_{\geqslant 0}$ are also required. However, the same initial-boundary value problem (1,2,3,4) is being solved and so these values may be computed from Eq. (4), assuming f is differentiable with respect to time. In the case where f is given by a closed expression this may be done directly by differentiating, as is done in the examples considered later. In the case where f is obtained from discrete measured values such as in loudspeaker modeling, the easiest way to calculate $\dot{f}$ is by using a finite difference approximation.

The use of the Burton-Miller type integral equation (10) to avoid the above-described stability problems is now considered in detail. It is clear from the previous section that the stability problems will be avoided if the integral equation arising when homogeneous boundary conditions are applied does not admit nontrivial solutions. The Burton-Miller-type equation with homogeneous boundary conditions is given by

$$(1-\alpha)\left\{\mathcal{D}(\dot{\varphi}(\underline{x},t))-\frac{1}{2}\dot{\varphi}(\underline{x},t)\right\}+\alpha c\left\{\frac{\partial}{\partial\hat{\underline{n}}_x}\mathcal{D}(\varphi(\underline{x},t))\right\}=0. \tag{12}$$

Assume that φ is a nontrivial solution of Eq. (12) and consider the double-layer potential

$$U(\underline{x},t)=\mathcal{D}\varphi(\underline{x},t)$$

for any $\underline{x}\in\mathbb{R}^3$. The jump relations together with Eqs. (11) and (12) give the following relation between interior boundary values of $\dot{U}$ and $\partial U/\partial\hat{\underline{n}}_x$:

$$(1-\alpha)\dot{U}_-+\alpha c\left(\frac{\partial U}{\partial\hat{\underline{n}}_x}\right)_-=0, \tag{13}$$

where the subscript "−" denotes the interior boundary values. Applying Green's first identity to U and $\dot{U}$ on Ω_- gives

$$\frac{1}{2}\int_{\Omega_-}\left\{|\nabla_{\underline{\xi}}U(\underline{x},t^+)|^2+\frac{|\dot{U}(\underline{x},t^+)|^2}{c^2}\right\}dV_{\underline{\xi}}=\int_0^{t^+}\int_\Gamma\dot{U}_-\left(\frac{\partial U}{\partial\hat{\underline{n}}_x}\right)_-dS_{\underline{\xi}}ds, \tag{14}$$

assuming φ is such that the volume integral is always finite. Substituting Eq. (13) into Eq. (14) and noting that $\alpha\in(0,1)$ yields

$$\begin{aligned}0&\leqslant\frac{1}{2}\int_{\Omega_-}\left\{|\nabla_{\underline{\xi}}U(\underline{x},t^+)|^2+\frac{|\dot{U}(\underline{x},t^+)|^2}{c^2}\right\}dV_{\underline{\xi}}\\&=\frac{-\alpha c}{(1-\alpha)}\int_0^{t^+}\int_\Gamma\left|\frac{\partial U}{\partial\hat{\underline{n}}_x}\right|_-^2dS_{\underline{\xi}}ds\\&=\frac{-(1-\alpha)}{\alpha c}\int_0^{t^+}\int_\Gamma|\dot{U}|_-^2dS_{\underline{\xi}}ds\leqslant 0.\end{aligned} \tag{15}$$

This means that both $\dot{U}_-$ and $(\partial U/\partial\hat{\underline{n}}_x)_-$ are zero. The jump relations give that

$$\dot{U}_+=\dot{U}_-+\dot{\varphi}=\dot{\varphi},\quad\left(\frac{\partial U}{\partial\hat{\underline{n}}_x}\right)_+=\left(\frac{\partial U}{\partial\hat{\underline{n}}_x}\right)_-=0, \tag{16}$$

where the subscript "+" denotes the exterior boundary values. This means that U satisfies the homogeneous Neumann boundary condition. By uniqueness of the exterior Neumann problem $U=0$ at any point in Ω_+. Clearly then $\dot{U}=0$ at any point in Ω_+ and so by Eq. (16), $\dot{\varphi}=0$ at any point in Ω_+. This means that φ is constant in time and so by the initial condition (2) must also be zero. It has therefore been shown that the Burton-Miller-type integral equation with homogeneous boundary condition (12) does not admit nontrivial solutions. Recall from the previous section that the stability problems inherent in the Kirchoff integral equation are related to its admission of resonant time harmonic solutions at some time when the initial conditions have vanished and the boundary sources are controlling the problem. The above-presented argument shows that the Burton-Miller-type integral equation (10) does not allow any resonant solutions and hence does not suffer from these stability problems.

III. THE SPACE-TIME DISCRETIZATION

Equation (10) is discretized in space and time in order to obtain a numerical solution via the collocation method. To do this, Γ is divided into n elements and the time axis into a regular grid $t_k=(k-1)\Delta t$, $k=1,\ldots,N$, where Δt is the time step. The solution $\varphi(\underline{x},t)$ is represented in terms of basis functions $\psi_j(\underline{x})$, $j=1,\ldots,n$ which interpolate φ in space, and $T_k(t)$, $k=1,\ldots,N$ which interpolate φ in time. Explicitly this yields

$$\varphi(\underline{x},t)=\sum_{k=1}^{N}\sum_{j=1}^{n}\varphi_j^k\psi_j(\underline{x})T_k(t), \tag{17}$$

where φ_j^k are unknown coefficients. Piecewise constant functions are chosen for ψ_j, $j=1,\ldots,n$ as these simplify the evaluation of the hypersingular integral later on. Cubic time interpolation functions as used in Ref. 12 are chosen for T_k, $k=1,\ldots,N$ and are given by $T_k(t)=T(t-t_k)$, where

$$T(t)=\begin{cases}1+\frac{11}{6}\left(\frac{t}{\Delta t}\right)+\left(\frac{t}{\Delta t}\right)^2+\frac{1}{6}\left(\frac{t}{\Delta t}\right)^3, & -\Delta t<t\leqslant 0\\ 1+\frac{1}{2}\left(\frac{t}{\Delta t}\right)-\left(\frac{t}{\Delta t}\right)^2-\frac{1}{2}\left(\frac{t}{\Delta t}\right)^3, & 0<t\leqslant\Delta t\\ 1-\frac{1}{2}\left(\frac{t}{\Delta t}\right)-\left(\frac{t}{\Delta t}\right)^2+\frac{1}{2}\left(\frac{t}{\Delta t}\right)^3, & \Delta t<t\leqslant 2\Delta t\\ 1-\frac{11}{6}\left(\frac{t}{\Delta t}\right)+\left(\frac{t}{\Delta t}\right)^2-\frac{1}{6}\left(\frac{t}{\Delta t}\right)^3, & 2\Delta t<t\leqslant 3\Delta t\\ 0 & \text{otherwise.}\end{cases} \tag{18}$$

Figure 1 gives a plot of the nonzero part of $T(t)$ and this choice ensures that all quantities evaluated in Eq. (10) are interpolated by at least piecewise linear functions in time.

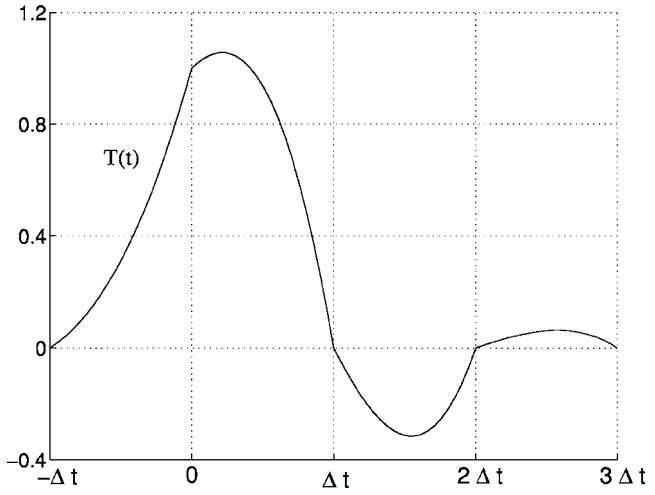


FIG. 1. The nonzero part of the cubic temporal basis function $T(t)$.

Expanding Eq. (10) out in the form of Eq. (6), substituting in Eq. (17) and evaluating at the kth time step yields an equation of the form

$$-\mathbf{A}^{(0)}\underline{\varphi}_k = \sum_{l=1}^{k-1} \mathbf{A}^{(l)}\underline{\varphi}_{k-1} + \underline{y}_k. \tag{19}$$

The terms in this matrix equation are given by

$$A_{i,j}^{(l)} = (1-\alpha)\left\{ \frac{-1}{4\pi}\int_{\Gamma_j} \frac{\partial R}{\partial \hat{\underline{n}}_\xi}\left[\frac{\dot{T}_{k-l}(\tau_k)}{R^2} + \frac{\ddot{T}_{k-l}(\tau_k)}{cR}\right] dS_\xi \right.$$
$$\left. - \frac{1}{2}\delta_{i,j}\dot{T}_{k-l}(t_k)\right\} + \frac{\alpha c}{4\pi}\left\{ \oint\!\!\!\!\!\!-_{\Gamma_j} \frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{1}{R}\right)\left[T_{k-l}(\tau_k) \right.\right.$$
$$\left.\left. + \frac{R}{c}\dot{T}_{k-l}(\tau_k)\right] dS_\xi + \int_{\Gamma_j} \frac{1}{c^2 R}\frac{\partial R}{\partial \hat{\underline{n}}_x}\frac{\partial R}{\partial \hat{\underline{n}}_\xi}\ddot{T}_{k-l}(\tau_k) dS_\xi \right\},$$

$$y_k = \frac{-(1-\alpha)}{4\pi}\int_\Gamma \frac{1}{R}\left(\frac{\partial \dot{\varphi}}{\partial \hat{\underline{n}}_\xi}\right)(\underline{\xi},\tau_k) dS_\xi$$
$$+ \alpha c\left\{ \frac{1}{4\pi}\int_\Gamma \frac{\partial R}{\partial \hat{\underline{n}}_x}\left[\frac{1}{R^2}\frac{\partial \varphi}{\partial \hat{\underline{n}}_\xi}(\underline{\xi},\tau_k) \right.\right.$$
$$\left.\left. + \frac{1}{cR}\left(\frac{\partial \dot{\varphi}}{\partial \hat{\underline{n}}_\xi}\right)(\underline{\xi},\tau_k)\right] dS_\xi - \frac{1}{2}\frac{\partial \varphi}{\partial \hat{\underline{n}}_x}(\underline{x},t_k)\right\},$$

$$\varphi_k = [\varphi_1^k, \ldots, \varphi_n^k]^T, \tag{20}$$

where $\oint\!\!\!\!\!\!-$ denotes the Hadamard finite part integral, $R=|\underline{x}_i - \underline{\xi}|$, $\tau_k = t_k - R/c$ and $\delta_{i,j}$ is the Kronecker delta.

The surface solution at any given time t_k may be generated using Eq. (19), by starting at the first time step and solving recursively (time marching) until reaching the desired time. The exterior solution at some point $\underline{x} \in \Omega_+$ may then be calculated using the discrete form of Eq. (6). The finite part integral is used since the integral here is hypersingular and so a method must be developed to evaluate its finite part numerically.

A. Evaluating the hypersingular integral

The hypersingular integral occurring in Eq. (10) can be seen in its discretized form as the Hadamard finite part integral appearing in terms of the matrix $\mathbf{A}$ in the previous section. Before discretization, this integral was of the form

$$\oint\!\!\!\!\!\!-_\Gamma \frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{1}{R}\right)\left[\varphi(\underline{\xi},\tau) + \frac{R}{c}\dot{\varphi}(\underline{\xi},\tau)\right] dS_\xi, \tag{21}$$

with notation as in Sec. II. As $R \to 0$, the first term in Eq. (21) is $O(R^{-3})$ and the second is $O(R^{-2})$. This can be reduced to the weakly singular case ($O(R^{-1})$) by applying Taylor's theorem as follows:

$$\varphi(\underline{\xi},\tau) = \varphi(\underline{x},t) + (\underline{\xi}-\underline{x})\cdot \nabla_\xi[\varphi(\underline{\xi},\tau)]|_{\xi=x} + O(R^2),$$
$$\dot{\varphi}(\underline{\xi},\tau) = \dot{\varphi}(\underline{x},t) + O(R), \tag{22}$$

which is valid if $\dot{\varphi}$ is differentiable in space on $\Gamma \times [0,\infty)$ and φ is twice differentiable in space on $\Gamma \times [0,\infty)$. Applying this to the hypersingular integral the following expression which is equivalent to Eq. (21) is obtained

$$\int_\Gamma \frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{1}{R}\right)[\varphi(\underline{\xi},\tau) - \varphi(\underline{x},t) - (\underline{\xi}-\underline{x})$$
$$\cdot \nabla_\xi[\varphi(\underline{\xi},\tau)]|_{\xi=x}] dS_\xi + \int_\Gamma \frac{R}{c}\frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{1}{R}\right)[\dot{\varphi}(\underline{\xi},\tau)$$
$$- \dot{\varphi}(\underline{x},t)] dS_\xi + \varphi(\underline{x},t)\oint\!\!\!\!\!\!-_\Gamma \frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{1}{R}\right) dS_\xi$$
$$+ \oint\!\!\!\!\!\!-_\Gamma \frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{1}{R}\right)(\underline{\xi}-x)\cdot \nabla_\xi[\varphi(\underline{\xi},\tau)]|_{\xi=x} dS_\xi$$
$$+ \dot{\varphi}(\underline{x},t)\oint\!\!\!\!\!\!-_\Gamma \frac{R}{c}\frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{1}{R}\right) dS_\xi. \tag{23}$$

This can now be evaluated since the first two integrals are weakly singular, the third may be evaluated using an identity from Meyer *et al.*[15] and it is easy to show that the last two integrals cancel in the case of piecewise constant spatial basis functions. The identity from Meyer gives that

$$\oint\!\!\!\!\!\!-_\Gamma \frac{\partial^2}{\partial \hat{\underline{n}}_x\, \partial \hat{\underline{n}}_\xi}\left(\frac{e^{ikR}}{R}\right) dS_\xi = k^2 \int_\Gamma (\hat{\underline{n}}_x \cdot \hat{\underline{n}}_\xi)\frac{e^{ikR}}{R} dS_\xi, \tag{24}$$

and so setting $k=0$, this can be used to evaluate the third integral in Eq. (23). A method for evaluating the hypersingular integral which is valid for any type of surface approximation has therefore been developed. This is important for applications such as loudspeakers where many elements would be required for flat elements to provide an accurate discretization.

IV. NUMERICAL RESULTS

The examples studied in this work are restricted to radiation from axisymmetric objects as this, simplifies the calculations. Figure 2 shows the generating curves for the three objects considered. The generating curve for the peanut is defined by three unit circles whose centers lie on an equilat-

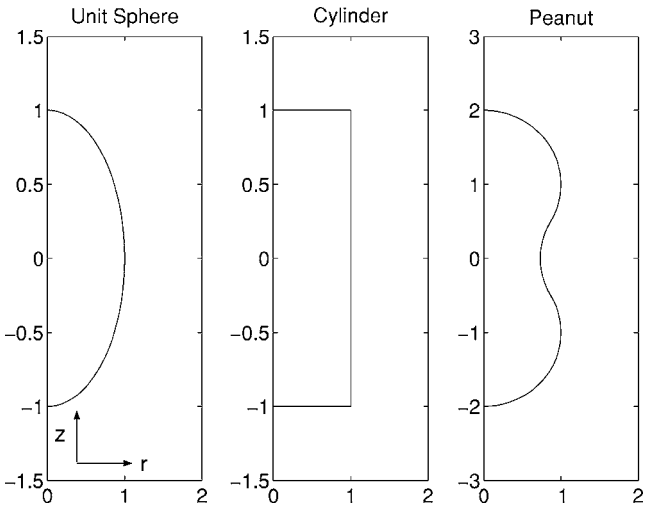


FIG. 2. Generating curves for the axisymmetric radiating objects considered.

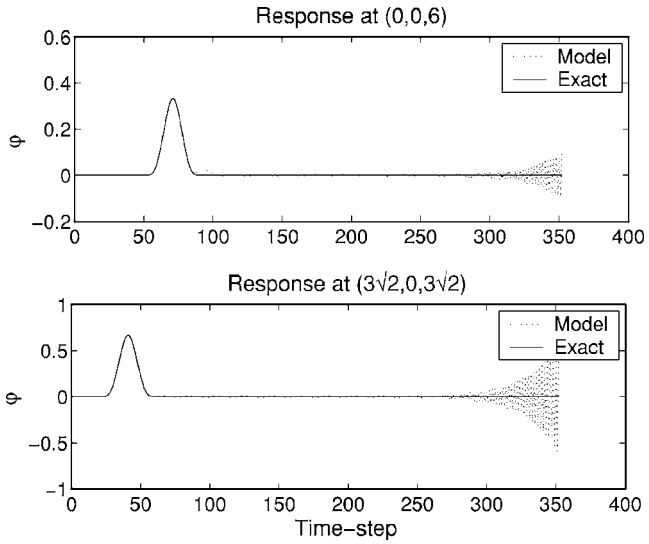


FIG. 3. The velocity potential calculated at two exterior points using the Kirchoff equation to calculate the surface solution on a unit sphere. Results are compared with an exact solution.

eral triangle. Starting from the point (0,2) at the top of the curve and moving along it in a clockwise direction, the first 2/5 of the arclength is formed by an arc from a circle centered at (0,1), the next 1/5 from a circle centered at $(\sqrt{3},0)$ and the final 2/5 from a circle centered at $(0,-1)$. All meshes used to approximate Γ are exact geometric representations defined in terms of arcs of circles or straight lines as appropriate.

In all cases the radiation of a spherically symmetric wave defined by

$$\varphi(R,t) = \frac{1}{R}\left(\frac{3}{4} - \cos\left(\frac{\pi(R-ct+3a)}{2a}\right) + \frac{1}{4}\cos\left(\frac{\pi(R-ct+3a)}{a}\right)\right)(H(R-ct+3a) - H(R-ct-a)) \tag{25}$$

is considered, which has the property that $(\partial\dot{\varphi}/\partial\hat{n})$ is continuous in time. Here a is the radius of some sphere $\mathbb{S}\subseteq\Omega$, R is the distance from the center of $\mathbb{S}$ to some point $\underline{x}\in\Omega_+$, and H is the Heaviside step function. It may be verified that Eq. (25) satisfies Eqs. (1)–(3) from the initial-boundary value problem. The boundary data required in Eq. (4) may be calculated from Eq. (25) using the chain rule and simplifications due to the axisymmetric geometry to give

$$\frac{\partial\varphi}{\partial r} = \frac{\partial\varphi}{\partial R}\sin\theta, \quad \frac{\partial\varphi}{\partial z} = \frac{\partial\varphi}{\partial R}\cos\theta,$$

where r and z are the coordinate axes shown in Fig. 2 and θ is the angle measured clockwise from the positive z axis. Hence the boundary data are given by

$$\frac{\partial\varphi}{\partial\hat{\underline{n}}} = n_r\frac{\partial\varphi}{\partial r} + n_z\frac{\partial\varphi}{\partial z} = \frac{\partial\varphi}{\partial R}[n_r\sin\theta + n_z\cos\theta], \tag{26}$$

where n_r and n_z are the components of $\hat{\underline{n}}$ in the r and z directions, respectively. It is also straightforward to compute $(\partial\dot{\varphi}/\partial\hat{\underline{n}})$ from the above. In the results that follow $\mathbb{S}$ is taken to be the sphere centered at the origin with the maximum possible radius a such that $\mathbb{S}\subseteq\Omega$ and the speed of sound has been normalized so that $c=1$. A Fourier transform shows that the wave defined by Eq. (25) is essentially band limited at frequency $f_{\max}$, which depends on the value of a. For accuracy, the time step is chosen as $\Delta t = 1/(10f_{\max})$ and the coupling parameter α is taken to be 0.5 as in Ref. 12.

The first example to be considered is that of a unit sphere. Since the radiated waves are spherically symmetric, the surface solution will be constant over Γ and so piecewise constant spatial basis functions will provide a good approximation of the analytic solution. Hence the relatively small value of $n=10$ boundary elements are used. The time step is taken to be 1/10 for the above-given reasons. This test problem is first solved using the simpler Kirchoff integral equation (6). Here the more widely used linear hat functions are applied for the time interpolation as in Ref. 5. Figure 3 shows the numerical results for φ compared with the exact solution at two different points in the exterior field. The results are clearly unstable, increasing exponentially after around 250 time steps. The results of repeating the same test problem, but using the Burton-Miller-type integral equation (10) are shown in Fig. 4. Here the results match the exact solution closely and so plots of the relative error at the two exterior points are given instead of the numerical and exact solutions. The relative error is Calculated from $(\varphi - \tilde{\varphi})/\max\varphi$, where φ and $\tilde{\varphi}$ denote the exact and numerical solutions, respectively. Note that the "max" used in the denominator is taken with respect to time and is used to avoid division by zero. The dotted vertical lines on the error plots in this paper are used to indicate the position of the radiated pulse shown in Fig. 3. The error is very small, peaking at 0.002 783 for the point (0,0,6) and at 0.003 293 for the point $(3\sqrt{2},0,3\sqrt{2})$. These peaks both occurred during the pulse. This shows that the results are accurate and stable over the 352 time steps shown.

Now the long-term stability of the method for calculating the surface solution is demonstrated. This is shown in terms of the characteristic ratio $cT/\mathrm{diam}(\Omega)$ as in Ref. 10, where T is the maximum time at which the solution is evalu-

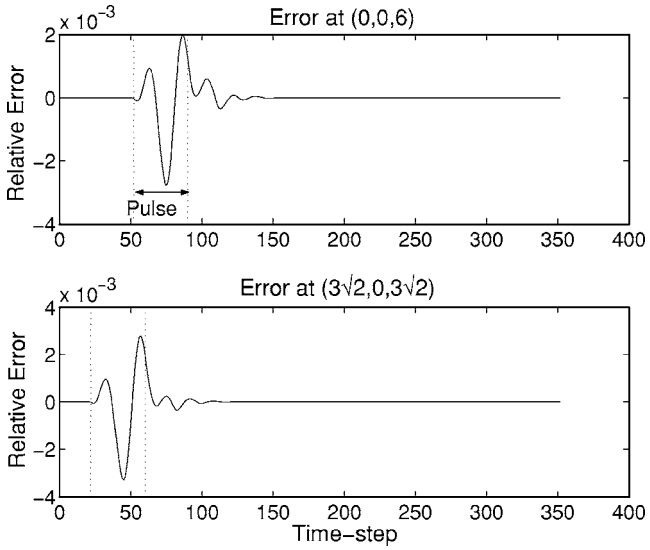


FIG. 4. The time variation of the relative error calculated at two exterior points using the Burton-Miller-type equation to calculate the surface solution on a unit sphere.

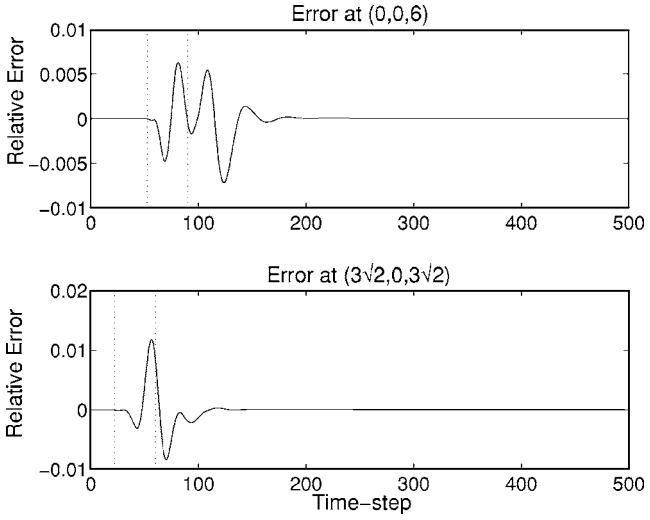


FIG. 6. The time variation of the relative error calculated at two exterior points using the Burton-Miller-type equation to calculate the surface solution on a cylinder.

ated and diam(Ω) is the diameter of Ω. In Ref. 10 long-term stability was shown with a characteristic ratio of greater than 500. Figure 5 shows the surface solution at the ten surface collocation points with a characteristic ratio of greater than 1000. This clearly shows the stability of the results over a long period of time (2002 s) and surpasses previous results.

Now the other examples of a cylinder and a peanut shaped mesh are considered. Here the spherical wave is not constant over Γ and so finer spatial discretizations are required for accurate results. More elements are also required since these objects have longer arclengths than the sphere. For the cylinder, $n=40$ is used and $\Delta t=1/10$ for the above-noted reasons. Figure 6 shows how the relative error varies in time at the same two points in the exterior field as before. The error is fairly small, peaking at 0.007 157 after the pulse for the point (0,0,6) and at 0.011 82 during the pulse for the point $(3\sqrt{2},0,3\sqrt{2})$. These peaks are greater than for the sphere, but this is not surprising given that the cylinder is not a regular surface as it has edges which are well known to have a detrimental effect on the accuracy of boundary element methods. However, the results are still reasonably accurate and stable over the 496 time steps shown.

For the peanut, $f_{\max}$ is larger than in the previous examples and so a smaller time step of 1/14 is used. This together with the fact that the peanut shaped mesh has the longest arclength of the examples considered means that the relatively large value of $n=60$ is taken. Figure 7 shows the results for the relative error in the exterior field, again at the same two points. The error is very small, peaking at 0.005 428 for the point (0,0,6) and at 0.003 742 for the point $(3\sqrt{2},0,3\sqrt{2})$. These peaks both occurred just before the end of the pulse. The errors shown in this case are smaller than for the cylinder but larger than for the sphere. This is the expected result since the peanut is a regular surface but the radiated wave is not constant over Γ like it is for the sphere. Therefore the results are accurate and stable over the 928 time steps shown.

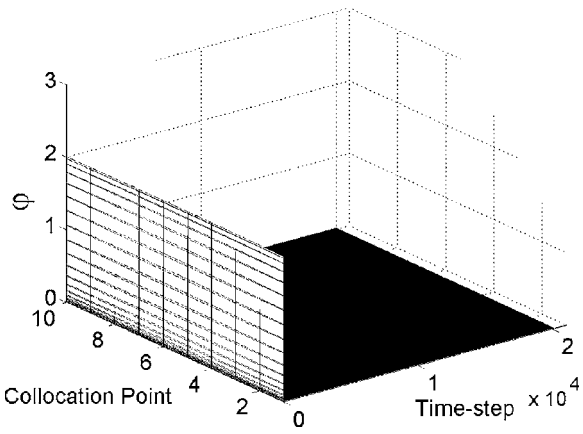


FIG. 5. The velocity potential calculated at the surface collocation points of the unit sphere, showing stability of the Burton-Miller-type equation for 20 020 time steps.

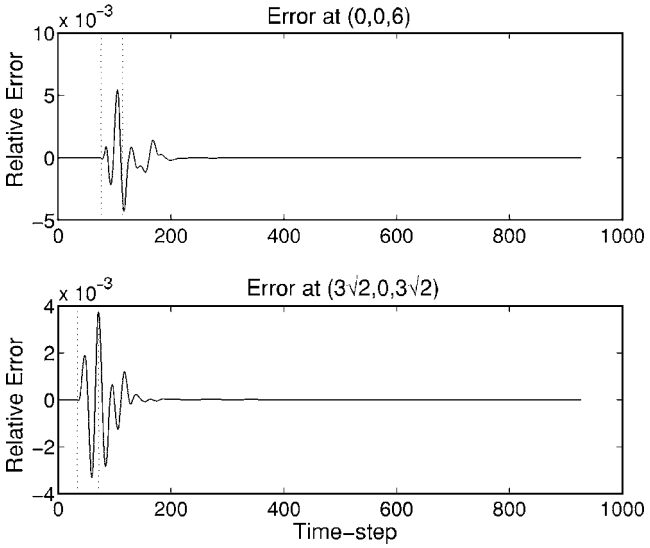


FIG. 7. The time variation of the relative error calculated at two exterior points using the Burton-Miller-type equation to calculate the surface solution on a peanut shaped axisymmetric object.

V. CONCLUSIONS

The methods presented accurately modeled transient acoustic radiation from various geometric objects. Furthermore, the Burton-Miller-type integral equation formulation employed was shown to be free of the long-time instabilities often suffered by such methods. The numerical results shown support these observations and a comparison with the more commonly used (but unstable) Kirchoff formulation was given. The hypersingular integrals arising in the Burton-Miller-type integral equation are reformulated as weakly singular ones using piecewise constant collocation and may be calculated for any choice of surface discretization.

ACKNOWLEDGMENT

The work carried out in this paper was partially funded by B&W Group Limited, Worthing, West Sussex, UK.

[1]M. B. Friedman and R. Shaw, "Diffraction of pulses by cylindrical objects of arbitrary cross section," J. Appl. Mech. **29**, 40–46 (1962).

[2]K. M. Mitzner, "Numerical solution for transient scattering from a hard surface of arbitrary shape-retarded potential technique," J. Acoust. Soc. Am. **42**, 391–397 (1967).

[3]B. P. Rynne, "Stability and convergence of time marching methods in scattering problems," IMA J. Appl. Math. **35**, 297–310 (1985).

[4]B. Birgisson, E. Siebrits, and A. P. Pierce, "Elastodynamic direct boundary element methods with enhanced numerical stability properties," Int. J. Numer. Methods Eng. **46**, 871–888 (1999).

[5]H. Wang, "Boundary integral modelling of transient wave propagation with application to acoustic radiation from loudspeaker," Ph.D. thesis, University of Brighton, 2004.

[6]B. P. Rynne and P. D. Smith, "Stability of time marching algorithms for the electric field integral equation," J. Electromagn. Waves Appl. **4**(12), 1181–1205 (1990).

[7]P. D. Smith, "Instabilities in time marching methods for scattering: Cause and rectification," Electromagnetics **10**, 439–451 (1990).

[8]M. J. Bluck and S. P. Walker, "Analysis of three-dimensional transient acoustic wave propagation using the boundary integral equation method," Int. J. Numer. Methods Eng. **39**, 1419–1431 (1996).

[9]S. J. Dodson, S. P. Walker, and M. J. Bluck, "Implicitness and stability of time domain integral equation scattering analysis," Appl. Comput. Electromagn. Soc. J. **13**(3), 291–301 (1998).

[10]T. Ha-Duong, B. Ludwig, and I. Terrasse, "A Galerkin BEM for transient acoustic scattering by an absorbing obstacle," Int. J. Numer. Methods Eng. **57**, 1845–1882 (2003).

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

[12]A. A. Ergin, B. Shanker, and E. Michielssen, "Analysis of transient wave scattering from rigid bodies using a Burton-Miller approach," J. Acoust. Soc. Am. **106**(5), 2396–2404 (1999).

[13]T. Terai, "On the calculation of sound fields around three dimensional objects by integral equation methods," J. Sound Vib. **69**(1), 71–100 (1980).

[14]Y. Kawai and T. Terai, "A numerical method for the calculation of transient acoustic scattering from thin rigid plates," J. Sound Vib. **141**(1), 83–96 (1990).

[15]W. L. Meyer, W. A. Bell, and B. T. Zinn, "Boundary integral solutions of three dimensional acoustic radiation problems," J. Sound Vib. **59**(2), 245–262 (1978).

[16]M. Costabel, "Time-dependent problems with the boundary integral equation method," *Encyclopedia of Comp. Mechanics*, edited by E. Stein, R. de Borst, and T. Hughes (Wiley, New York, 2004), Chap. 22.

[17]C. A. Coulson and A. Jeffrey, *Waves: A Mathematical Approach to the Common Types of Wave Motion* (Longman Group, London, 1977).

Simulation of ultrasonic fields radiated by a circular source through a layer with nonparallel boundaries

Elena Jasiūnienė,[a] Liudas Mažeika, and Rymantas Kažys
Ultrasound Institute, Kaunas University of Technology, Studentu 50, LT-51368 Kaunas, Lithuania

(Received 5 July 2005; revised 24 February 2006; accepted 12 April 2006)

Active elements of ultrasonic transducers are separated from the medium by thick protective layers, which may have nonparallel front and back surfaces. The objective of this work was to develop a simple method suitable for fast calculation of ultrasonic fields radiated through a layer with parallel or nonparallel boundaries and enabling one to take into account multiple reflections inside the layer. The main presumption of the proposed method is the following: after refraction at the boundary between two media, the ultrasonic field consists of plane and edge waves as before refraction. The proposed simulation method is based on transformation of a multi-layered medium into a virtual one without internal boundaries and equivalent to the actual medium from the point of a view of the times of flight of the direct plane and edge waves. The method enables simulation of radiated ultrasonic fields even after multiple reflections within the layer with parallel or nonparallel boundaries. The examples of simulated and experimentally measured ultrasonic fields of the transducer with a nonparallel front surface of protection layer, radiating into the water, are presented. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2202890]

PACS number(s): 43.20.Px, 43.20.El, 43.20.Bi [LLT] Pages: 81–89

I. INTRODUCTION

In the case of ultrasonic measurements in aggressive media—causing corrosion, possessing a high temperature, etc.—active elements of ultrasonic transducers are separated from the medium by thick protective layers. For separation usually a solid material such as stainless steel is used. The multiple reflections of ultrasonic waves take place in these layers due to mismatch of acoustic impedances. In order to reduce multiple reflections inside the layer, the protector is not with parallel boundaries, but with one boundary inclined with respect to another one, i.e., it looks like a wedge (Fig. 1). That enables us to reduce significantly the amplitude of the waves reflected repeatedly inside the layer, but the structure of the ultrasonic field radiated through a layer with nonparallel boundaries becomes complicated. If for radiation a circular piezoelectric element is used, the field transmitted through the layer is losing axial symmetry. This occurs due to the fact that the projection of the circular transducer on the inclined boundary of the layer produces an elliptical source of ultrasonic waves. Therefore, a simple method, suitable for fast simulation of ultrasonic fields, radiated through layers with nonparallel boundaries is needed.

Many authors have theoretically studied pressure waveforms radiated into different media by an idealized piston source.[1–3] The simplest method for simulation of ultrasonic fields is direct application of Huygens' principle. Experimental observations of the pulsed field of a circular ultrasonic transducer have shown that measured fields are in good agreement with theoretical results, calculated assuming an ideal piston behavior.[4,5] These studies have also demonstrated that the radiated fields consist of plane and edge waves.

Various investigations of ultrasonic fields in media with intermediate boundaries were also performed by many authors. Computational methods for determination of transmitted[5,6] and reflected ultrasonic waves through plane interfaces[7] and interfaces of a complex geometry,[8,9] using the Rayleigh integral and based on the Huygens principle, were proposed. A 3D computational method, based on the spatial impulse response and on the discrete representation computational concept, was used to simulate acoustic beams generated by arrays through interfaces[10] of arbitrary shapes. For an exact evaluation of the ultrasonic field in the presence of an interface, the angular spectrum method was proposed.[11,12] Comparisons of the Gauss-Hermite beam model and a boundary diffraction wave paraxial model were made for large angles of incidence.[13] The models for the computations of ultrasonic fields radiated by arbitrary shape transducers into objects under examination were also presented.[14,15] A multi-Gaussian-beam model enables modeling of the ultrasonic beam generated by a piston transducer radiating into complex geometries.[16] However, to our knowledge there are no modeling methods which take into account reflections inside the layer between nonparallel interfaces.

The objective of this work was to develop a simple approximate simulation method suitable for fast calculation of ultrasonic fields radiated through a layer with parallel or nonparallel boundaries in the case when the inclination angle of the layer boundary is relatively small ($<10°$), at the distances $z>2D$ from the transducer and enabling to take into account the multiple reflections inside the layer. For this purpose, the known model for calculation of the ultrasonic field of a single circular transducer in a homogeneous medium[17] and the extended model for the case of a multi-layered medium with parallel or nonparallel boundaries[18,19] were modified to take into account the multiple reflections inside the layer. The proposed simulation method is based on transfor-

[a]Electronic mail: elena.jasiuniene@ktu.lt

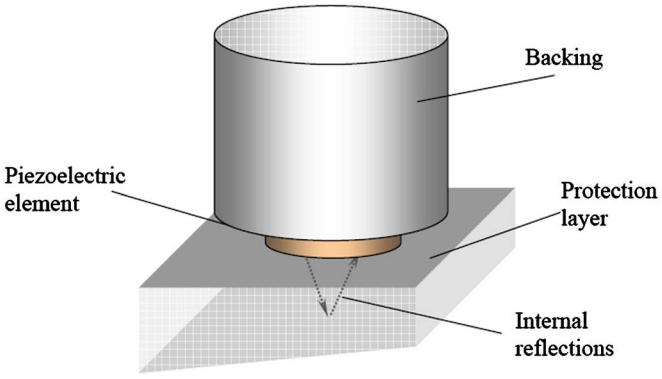


FIG. 1. (Color online) Ultrasonic transducer with nonparallel front surface of protection layer.

mation of a multi-layered medium into a virtual one without internal boundaries and equivalent to the actual medium from the point of view of the times of flight of direct plane and edge waves. The method proposed enables simulation of radiated ultrasonic fields even after multiple reflections within the layer with parallel or nonparallel boundaries.

II. THEORY AND MAIN STEPS OF THE PROPOSED METHOD

The method proposed is based on the assumption that an ultrasonic field in an inhomogeneous medium, e.g., after passing the inclined boundary of a protector, consists of plane and edge waves as before refraction. This means that, even after passing the boundary between two media, the shape of a spatial pulse response of the ultrasonic transducer remains the same, as in the case of a homogeneous medium; only the times of flight of plane and edge waves are different. However, when the surface of the circular piezoelectric element is inclined with respect to the boundary between the protector and medium, the structure is losing axial symmetry and the times of flight of the edge waves are not the same in the different planes containing the revolution axis of the transducer. The difference between the times of flight of the edge waves in this case will depend on the inclination angle of the boundary with respect to the surface plane of the piezoelectric element. However, as it will be shown later in this chapter, these differences in the case of relatively small inclination angles ($<10°$) are so small that they produce only the second-order effects and may be neglected. It allows for calculations of circular sources to exploit the mathematical model based on the spatial pulse response approach[20] and it is used for analysis of pulsed ultrasonic fields radiated by a circular transducer into a homogeneous medium.

In this case the spatial pressure pulse response of the circular transducer with radius R is given by the following expressions:

$$h_t(x,z,t)=\begin{cases} -\rho_0 c\,\delta(t-t_0), & x<R, t_0\leqslant t<t_1,\\ -\dfrac{\rho_0 c}{\pi}\dfrac{d\theta}{dt}, & x<R, t_1\leqslant t\leqslant t_2,\\ -\rho_0 c\left[\delta(t-t_0)+\dfrac{1}{\pi}\dfrac{d\theta_1}{dt}\right], & x=R, t_0\leqslant t\leqslant t_2,\\ -\dfrac{\rho_0 c}{\pi}\dfrac{d\theta}{dt}, & x>R, t_1\leqslant t\leqslant t_2, \end{cases} \tag{1}$$

where

$$\frac{d\theta}{dt}=\frac{1}{(c^2t^2-z^2)}\times\frac{-[c^2t(c^2t^2-z^2-x^2+R^2)]}{\sqrt{[2(c^2t^2-z^2)(x^2+R^2)-(c^2t^2-z^2)^2-(x^2-R^2)^2]}}, \tag{2}$$

$$\frac{d\theta_1}{dt}=-\frac{c^2t}{\sqrt{(c^2t^2-z^2)[4R^2-(c^2t^2-z^2)]}}, \tag{3}$$

$$t_0=z/c, \tag{4a}$$

$$t_1=\sqrt{(R-x)^2+z^2}/c, \tag{4b}$$

$$t_2=\sqrt{(R+x)^2+z^2}/c, \tag{4c}$$

t is the time, x,z are the spatial coordinates of the point, where the ultrasonic field is calculated, ρ_0 is the density of the homogeneous medium, R is the transducer radius, c is the ultrasound velocity, t_0 is the time of flight of the plane wave from the transducer surface to the point $P(x,z)$, and t_1 and t_2 are the times of flight of the edge waves from the nearest and farthest transducer edges to the point $P(x,z)$ accordingly.

Using various methods it was shown that the acoustic field of a circular transducer consists of plane and edge waves.[4,5] The whole surface of a piston generates a direct plane wave, which propagates in a cylindrical region having the piston at its base [Fig. 2(a)]. From the edge of the transducer the diffracted edge waves are radiated, which have toroidal shape. The edge waves are focused on the geometrical axis of the transducer. In the case of a circular transducer, the geometrical axis corresponds to equal distances from edges of the transducer.

The typical waveforms of the spatial pulse response are given in Fig. 2(b). It can be seen that a major part of the energy is concentrated in three pulses. The first pulse corresponds to the arrival time of a plane wave from the surface of a transducer. The second and third pulses correspond to the arrival times of edge waves from the nearest and farthest edges of a disk shaped transducer, respectively.

Simulation of the propagation of an ultrasonic wave through a solid layer with nonparallel surfaces is quite complicated, because on the boundary refraction and transformation of the waves take place. When an ultrasonic wave reaches the boundary between two media, part of the wave is transmitted into the second medium, and part of the wave is reflected back. Which part of the energy is transmitted, and which is reflected, depends on the type of transmitted wave, on the incidence angle, and on the acoustic impedances of both media.[21] The refraction angle of the wave is determined using Snell's law.

The various types of surface waves (surface, leaky waves, etc.) can be generated on the boundary between two

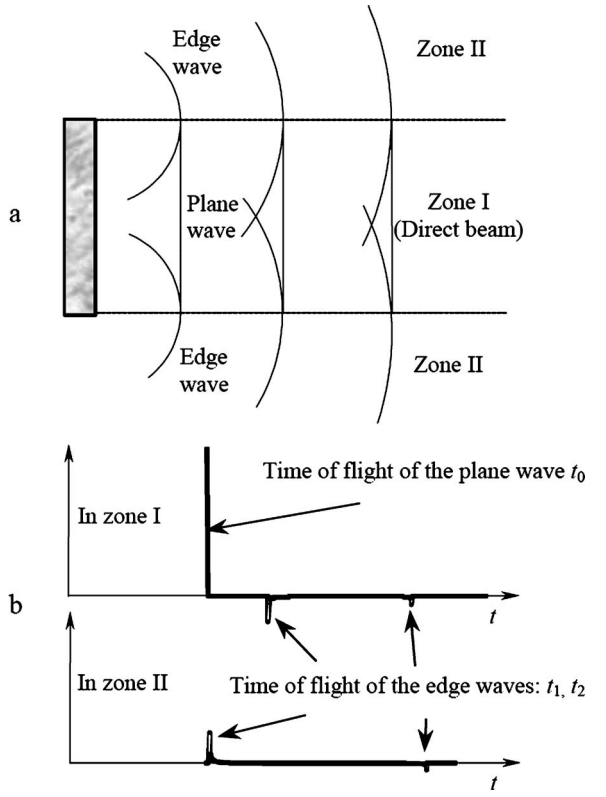


FIG. 2. (a) Direct plane and edge waves propagation zones. (b) Pulse response of the circular transducer in a homogeneous medium.

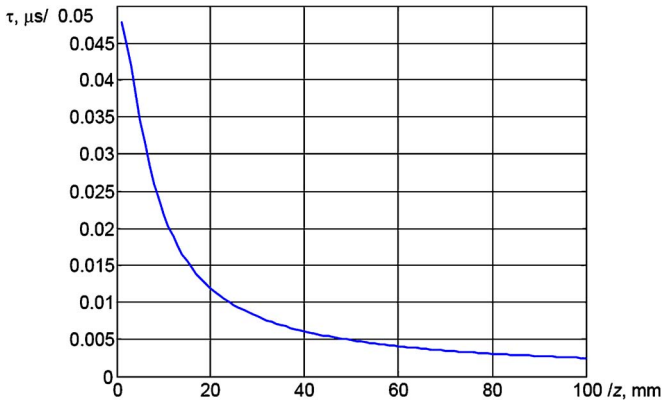


FIG. 3. (Color online) Delay time difference for the signals propagating in water from elliptical (maximal lateral dimension of the ellipse) and circular contours versus the distance from the transducer. The incidence angle in steel is 10°.

media depending on the type of incident wave, on the incidence angle, and on the acoustical properties of both media. However, influence of the surface waves is not essential in a far field.[22] So, using the method proposed, only longitudinal or shear waves inside a layer will be taken into account.

The structure of this field in a general case should depend on the reflection and transmission coefficients at the interfaces mainly due to the fact that the values of these coefficients depend on the angle of incidence. However, at the distances from the transducer $z>(2-4)D$, where D is the transducer diameter, the incidence angle varies in the range of 15° for different emitting points on the transducer surface and at longer distances is changing rather slowly. It means that in this range the values of the reflection and transmission coefficients change less than 10%. For longer distances this angle and corresponding changes are smaller and, consequently, the transmission and reflection coefficients for different rays will differ much less. Therefore in our case the influence of these coefficients on the structure of the radiated field will be rather small when the field is calculated not very close to the interface between a solid wedge and liquid medium, e.g., at the distances $z>(2-4)D$ from the transducer. So, performing calculations the transmission and reflection coefficients were taken as constant values not depending on the incidence angle, because the main interest of the proposed method is the structure of the radiated field.

As it was stated before, the main presumption of the proposed method is the following: after refraction at the boundary between two media, the ultrasonic field consists of plane and edge waves as before refraction. It means that after passing the boundary between two media the pulse response at an arbitrary selected point has the same character as at some point in a virtual homogeneous medium. Of course, for determination of the exact waveform of the pulse response the position of the point in a virtual homogeneous medium having the same times of flight of edge and plane waves must be found.

In the case of a circular transducer the times of flight of the edge waves are not the same in the planes containing the revolution axis of the transducer. The difference between the times of flight of the edge waves in this case will depend on the inclination angle of the boundary with respect to the surface plane of the piezoelectric element. Our investigation has shown that multiple reflections inside the protector are significantly reduced even when this angle and respectively the incidence angle do not exceed 6°–8°. In this case deviation of the circular transducer diameter from its projection on the inclined plane (ellipse) does not exceed 0.06–0.09 mm for the 10-mm-diam transducer. The time of flight difference of edge waves, which propagate from the axially symmetric source, and the elliptic source will depend on the incidence (or inclination) angle and the distance between the selected point in a working medium and the contour (edge) of the source. The refraction angle of ultrasonic waves in water in this range of incident waves does not exceed 1.5°–2°. The difference of the times of flight of edge waves for the above-mentioned transducer with a steel wedge is given in Fig. 3. From the results presented it follows that even in the worst case (the incidence angle to the steel-water boundary 10°) the difference of the times of flight at the distances longer than 20 mm does not exceed 5–10 ns. If the ultrasonic field in a liquid medium is calculated for 5-MHz frequency signal (the period $T=200$ ns), it means that this difference is smaller than $0.02-0.05T$, where T is the period of the ultrasonic signals. For lower frequency signals this normalized difference will be correspondingly smaller. Such a small difference in terms of the signal period may produce only the second-order effects in the simulated ultrasonic field. For example, when this difference is $0.02T$, the amplitude is reduced only by 0.12%. Please note that it is the worst possible estimation; usually this reduction is smaller. We think that

this example justifies the proposed approximate approach for fast calculation of ultrasonic fields radiated by a circular source through a solid layer with nonparallel boundaries in the case when the inclination angle of the layer boundary is relatively small, e.g., $<10°$.

Hence, it is possible to assume that a virtual homogeneous medium exists, in which the times of flight t_0, t_1, and t_2 are the same as in the given media with internal boundaries. Therefore, calculation of the ultrasonic field after transmission through a layer with parallel or nonparallel boundaries can be changed to the calculation of the field in a single virtual homogeneous medium.

According to the method proposed, the waveforms of ultrasonic signals within and after the layer with parallel or nonparallel boundaries are calculated in the following steps for each point in the field:

(1) The time of flight of the plane wave t_0 to the selected point $P(x,z)$ is found.
(2) The times of flight of the edge waves t_1 and t_2 to the same point $P(x,z)$ are calculated.
(3) An equivalent point in a virtual homogeneous medium, where the times of flight t_0, t_1, and t_2 are the same as in the media with boundaries, is found.
(4) The pressure pulse response for an equivalent point in the virtual homogeneous medium is calculated.
(5) The waveform of the ultrasonic signal is calculated using convolution of the spatial pulse response of the virtual medium and the driving signal.

III. CALCULATION OF THE ULTRASONIC FIELD, WHEN THE WAVE IS REFLECTED WITHIN THE LAYER WITH NONPARALLEL BOUNDARIES

A. Calculation of the time of flight of the plane wave

The wavefront of the plane wave is determined usually as the set of points in the space having the same propagation time from the source and, as a consequence, the same phase of the signal. In our case, when the transmitting transducer is a planar circular piezoelectric element, the cross section of the wave front in one plane is a line having a certain angle defined by Snell's law and crossing the point $P(x,z)$ (Fig. 4). So, calculation of the propagation time t_0 of the plane wave to the given point $P(x,z)$ can be separated into two tasks, the first of which is calculation of the plane wave propagation time from any point on the transducer surface to the point $P(x,z)$ in the wave front and the second one is determination if the point $P(x,z)$ is in the direct beam zone (plane wave propagation zone).

As the time of flight of the plane wave to the given point $P(x,z)$ is the same as the time of flight of the ray from the central point of the transducer to the same wavefront, the central point of the transducer was selected for the calculation of the plane wave propagation time $t_0=t_{0c}$ to the wave front, where the point $P(x,z)$ is given. The plane wave propagation time from the central point of the transducer to the given point is the sum of propagation times in the first (t_{01c}) and the second (t_{02c}) media (Fig. 4) and can be expressed by

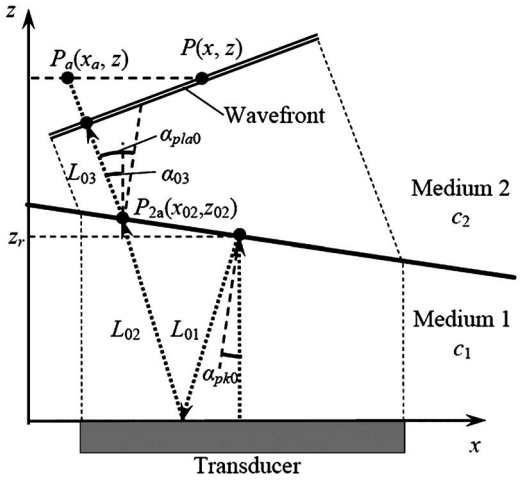


FIG. 4. Calculation of the time of flight of the plane wave, when the ultrasonic field after reflection within the layer with nonparallel boundary has to be calculated.

$$t_{0c} = t_{01c} + t_{02c} = \frac{z_r + L_{01} + L_{02}}{c_1} + \frac{L_{03}}{c_2}, \tag{5}$$

where c_1, c_2 are the ultrasound propagation velocities in the first and the second medium, z_r is the distance from the center of a transducer to the second medium, $L_{01}=z_r/\cos(2\alpha_{pk0})$ is the propagation distance of the reflected central ray from the boundary between two media back to the surface of the transducer, $L_{02}=L_{01}\cos(\alpha_{pk0})/\cos(3\alpha_{pk0})$ is the propagation distance of the reflected central ray from the plane of the transducer to the boundary between two media, $L_{03}=(z-z_{02})/\cos(\alpha_{03})-(x-x_a)\sin(\alpha_{03})$ is the propagation distance of the reflected central ray in the second medium, $x_a=x_{02}-(z-z_{02})\tan(\alpha_{03})$ is the x coordinate of the central ray of the transducer at the given distance z from the transducer, α_{pk0} is the inclination angle of the boundary between two media, $\alpha_{03}=\alpha_{pla0}-\alpha_{pk0}$ is the angle of the ray in the second medium with respect to the z axis, $\alpha_{pla0}=\arcsin[(c_2/c_1)\sin(3\alpha_{pk0})]$ is the refraction angle of the wave in the second medium once reflected in the first medium, and $P_{2a}(x_{02},z_{02})$ is the point on the boundary between the two media where the refraction takes place.

In order to be able to calculate the equivalent point in a virtual homogeneous medium, it has to be known if the point at which the ultrasonic field is calculated is in the region of the direct plane wave or outside it. Therefore, it must be checked whether the point $P(x,z)$ is in the direct beam zone using the following condition:

$$\begin{aligned} &\text{if } |x_a| \leqslant R, \text{ the point } P(x,z) \text{ is in the direct beam zone,} \\ &\text{if } |x_a| > R, \text{ the point } P(x,z) \text{ is out of the direct beam zone,} \end{aligned} \tag{6}$$

where R is the transducer radius.

B. Calculation of the times of flight of the edge waves

The propagation times of the edge waves t_1 and t_2 to the point $P(x,z)$ are given by

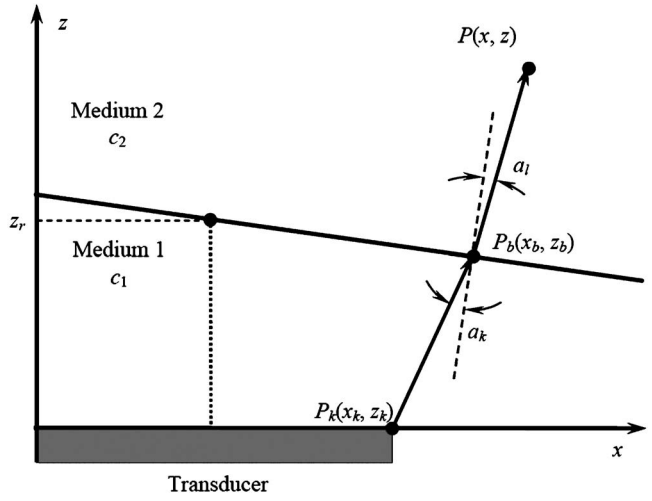


FIG. 5. Finding of the times of flight of directly transmitted edge waves.

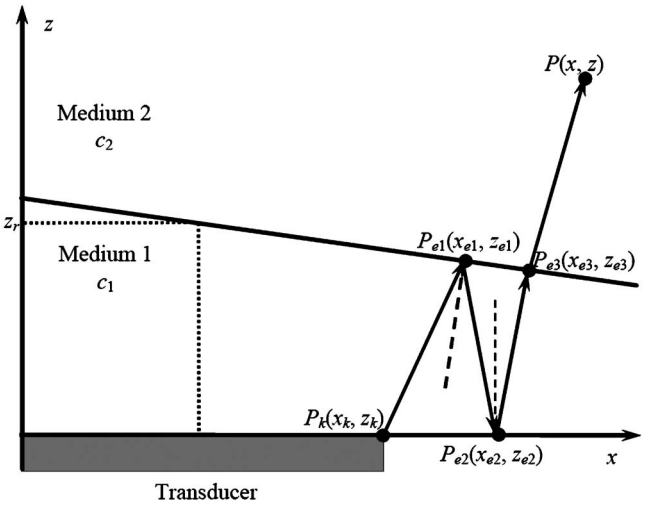


FIG. 6. Finding of the times of flight of the edge waves after reflections inside layer.

$$t_1 = \frac{r_{11}}{c_1} + \frac{r_{12}}{c_2}, \tag{7}$$

$$t_2 = \frac{r_{21}}{c_1} + \frac{r_{22}}{c_2}, \tag{8}$$

where r_{11}, r_{21} are the distances, which the edge waves propagate in the first medium, and r_{12}, r_{22} are the distances, which the edge waves propagate in the second medium.

In the case of direct propagation from the source to the target point (no reflection in the first medium) the distances (Fig. 5), which the ray has to propagate in the first and the second media from one edge, can be found from the system of equations:

$$\begin{aligned} &\frac{c_1}{\sin(\alpha_k)} = \frac{c_2}{\sin(\alpha_1)}, \\ &\alpha_k = \alpha_{pk0} + \arctan\left(\frac{x_b - x_k}{z_b - z_k}\right), \\ &\alpha_l = \alpha_{pk0} + \arctan\left(\frac{x - x_b}{z - z_b}\right), \\ &\tan(\alpha_{pk0}) = \frac{z_b - z_r}{x_b}, \end{aligned} \tag{9}$$

where α_k is the incidence angle of the ray from the edge of the transducer to the boundary between the two media, α_l is the refraction angle of the ray in the second medium, $P_k(x_k, z_k)$ is the point on the transducer edge, and $P_b(x_b, z_b)$ is the ray refraction point on the boundary between the two media.

This system of equations has no analytic solution and may be solved only using numerical methods. Moreover, this system of equations becomes more complicated for the case, when the reflection in the first medium takes place.

In this case, in order to find the distances that each ray of the edge wave has to pass in the first and the second media from the source to the target point $P(x,z)$, such points P_{e1}, P_{e2}, and P_{e3} have to be found. At the points $P_{e1}(x_{e1}, z_{e1})$ and $P_{e2}(x_{e2}, z_{e2})$ they would obey the reflection law and at the point $P_{e3}(x_{e3}, z_{e3})$—Snell's law (Fig. 6). Contrary to propagation of a plane wave, where the propagation direction from the planar source was defined in advance by the orientation of a transducer, the propagation direction of the edge wave is normal to the wave front. In the $x0z$ plane the sources of these waves are points corresponding to the edge of a circular transducer. So, the task is to find the ray paths radiated from these source points to the point $P(x,z)$ after reflections in the first medium and refraction on the boundary between the two media.

In order to simplify the task, the special "ray straightening" procedure has been proposed. This procedure is based on the fact that the beam path length of the ray reflected in the first medium is equivalent to the linear beam path in the virtual medium having triple angle $3\alpha_{pk0}$ with respect to the boundary between the first and the second media (Fig. 7).

The "ray straightening" procedure is implemented in two steps:

(i) At first, the rays S_{a2}, S_{a3} propagating in the first medium with the inclined boundary with the second medium and the transducer plane are projected

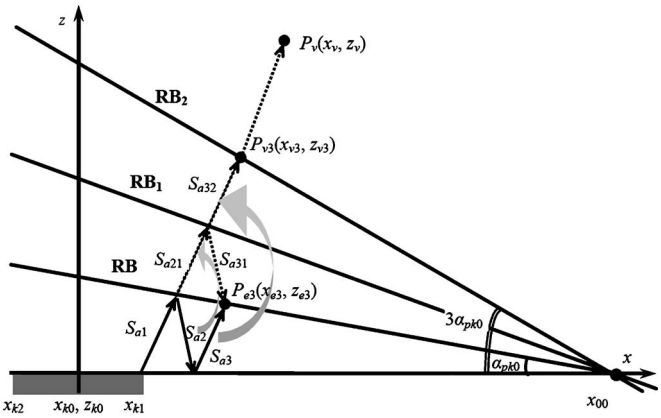


FIG. 7. Explanation of the "ray straightening" procedure.

specularly with respect to the boundary line. In this way the projections of the rays S_{a21}, S_{a31} and of the transducer plane RB_1 are obtained.

(ii) In the second step, the specular projection of the boundary RB and the ray S_{a31} are created. In Fig. 7, they correspond to RB_2 and S_{a32}.

It is possible to prove that the ray S_{a1} and the projected rays S_{a21}, S_{a32} are on a straight line. The incident angle of the ray S_{a3} with respect to the boundary between two media and the incident angle of the ray projection S_{a32} with respect to the projection of the boundary RB_2 also are equal. Moreover, if the target point $P(x,z)$ will be passed through this "ray straightening" procedure, the new virtual position $P_v(x_v,z_v)$ of the point $P(x,z)$ will be obtained also and the task with the reflection is transformed into the task without reflection.

The beam path (distances, which the ray has to propagate in the first and the second media) after the "ray straightening" procedure can be obtained from

$$\frac{c_1}{\sin(\alpha_k)} = \frac{c_2}{\sin(\alpha_l)},$$

$$\alpha_k = 3\alpha_{pk0} + \arctan\left(\frac{x_{v3} - x_k}{z_{v3} - z_k}\right),$$

$$\alpha_l = 3\alpha_{pk0} + \arctan\left(\frac{x_v - x_{v3}}{z_v - z_{v3}}\right), \tag{10}$$

$$\tan(3\alpha_{pk0}) = \frac{z_{v3} - z_{rv}}{x_{v3}},$$

where

$$z_{rv} = z_r \tan(3\alpha_{pk0})/\tan(\alpha_{pk0}),$$

$$x_v = [x - z_r/\tan(\alpha_{pk0})]\cos(2\alpha_{pk0}) = z\sin(2\alpha_{pk0}),$$

$$z_v = [x - z_r/\tan(\alpha_{pk0})]\sin(2\alpha_{pk0}) + z\cos(2\alpha_{pk0}).$$

Then the propagation distances of the waves propagating from the transducer edge in the first medium (r_{11} and r_{12}) and in the second (r_{21} and r_{22}) medium can be written as

$$r_{11} = \sqrt{(x_{v31} - x_{k1})^2 + (z_{v31} - z_{k1})^2},$$

$$r_{12} = \sqrt{(x_{v31} - x_v)^2 + (z_v - z_{v31})^2},$$

$$r_{21} = \sqrt{(x_{v32} - x_{k2})^2 + (z_{v32} - z_{k2})^2}, \tag{11}$$

$$r_{22} = \sqrt{(x_{v32} - x_v)^2 + (z_v - z_{v32})^2},$$

where (x_{k1},z_{k1}) and (x_{k2},z_{k2}) are the coordinates of the left and the right edges of the transducer, and $P_{v31}(x_{v31},z_{v31})$ and $P_{v32}(x_{v32},z_{v32})$ are the points of refraction on the boundary between the two media, corresponding to the rays propagating from the left and the right edges of the transducer.

The delay times of both rays t_1 and t_2 can be calculated using Eqs. (7) and (8).

C. Calculation of the equivalent point coordinates in a virtual homogeneous medium

As it was stated above, the proposed method is based on transformation of a multi-layered medium into a virtual one without internal boundaries and equivalent to the actual medium from the point of view of the times of flight of direct plane and edge waves. Such a transformation does not affect the source of ultrasonic waves, the piezoelectric element, because it is on the border of the transformed region. It means that all points on the surface of a piston-type source after the transformation vibrate in phase and the vibration profile on the surface is uniform.

It is assumed that an arbitrary point $P(x,z)$ in a multi-layered medium with planar parallel and nonparallel boundaries may be replaced by such an equivalent point $P_e(x_e,z_e)$ in a virtual homogeneous medium without boundaries that the times of flight t_0, t_1, and t_2 are the same as in the given medium with intermediate boundaries. The exact values of the times of flight of the plane wave t_0 and the edge waves t_1, t_2 in the case of the inclined layer are calculated according to the procedures defined in Secs. III A and III B. On the other hand, in homogeneous media these times of flight are defined by three equations [Eq. (4)]. In general they depend on four parameters: the transducer diameter D, the ultrasound velocity c, and the spatial coordinates x,z of the point under investigation. So, for given values of the times of flight t_0, t_1, t_2, one of these parameters can be selected arbitrarily; the other three should be calculated from the system of equations (4). In our case the predefined parameter is the transducer diameter and the parameters which must be calculated are the equivalent ultrasound velocity and the spatial coordinates $P_e(x_e,z_e)$ of the point in the virtual medium. It is necessary to point out that the equivalent ultrasound velocity c_e will be different for different points in the virtual homogeneous medium, e.g., it will be function of the spatial coordinates $c_e = c_e(x,y,z)$.

The ultrasound velocity in the virtual medium and the equivalent point coordinates are calculated differently inside the direct beam zone and outside it. The equivalent ultrasound velocity c_e in the virtual homogeneous medium is given by (Fig. 8)

$$c_e = \frac{2R}{\sqrt{t_2^2 - t_0^2} + \sqrt{t_1^2 - t_0^2}} \quad \text{in the direct beam zone,}$$

$$c_e = \frac{2R}{\sqrt{t_2^2 - t_0^2} - \sqrt{t_1^2 - t_0^2}} \quad \text{outside the direct beam zone.} \tag{12}$$

The coordinates of the equivalent point $P_e(x_e,z_e)$ in the $x0z$ plane of the virtual homogeneous medium can be written as

$$x_e = R - c_e\sqrt{t_2^2 - t_0^2}, z_e = t_0 c_e \quad \text{in the direct beam zone,}$$

$$x_e = R + c_e\sqrt{t_2^2 - t_0^2}, \tag{13}$$

$$z_e = t_0 c_e \quad \text{outside the direct beam zone.}$$

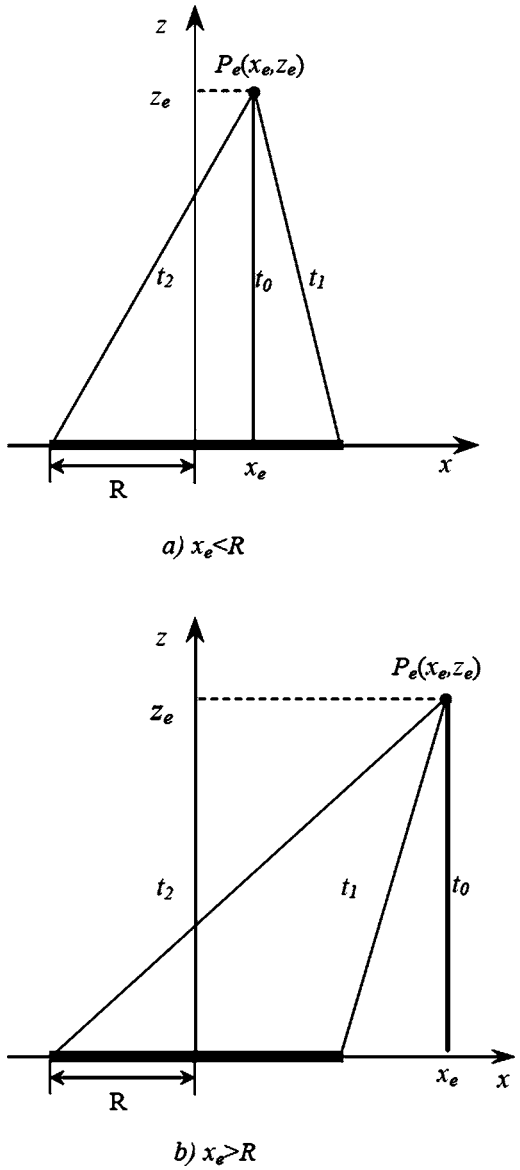


FIG. 8. Finding of the equivalent point in virtual homogeneous medium: (a) in the direct beam zone and (b) outside the direct beam zone.

D. Calculation of the spatial pulse response for an equivalent point in a virtual homogeneous medium

The spatial pulse response $h(x,y,z,t)$ for an equivalent point in a virtual homogeneous medium is found by means of the mixed analytic-numerical procedure.[17] This approach enables us to simulate an ultrasonic field in two media separated by an interface. The input parameters for this model are the transducer diameter, the ultrasound velocities c_1, c_2, and the densities ρ_1, ρ_2 corresponding to the first and second medium.

E. Convolution of the spatial pulse response and the driving signal

An acoustic pressure at an arbitrary point $P(x,y,z)$ is found as convolution of the driving pulse $u(t)$ and the spatial pulse response of the transducer $h(t,x,y,z)$:

$$p_a(t,x,y,z) = u(t) * h(t,x,y,z)k_c, \tag{14}$$

where $p_a(t,x,y,z)$ is the acoustical field of the transducer, $u(t)$ is the driving pulse, $*$ denotes the convolution, and $k_c = c_1/c_e$ is the constant factor.

The driving signal was approximated by

$$u(t) = e^{a(t-b)^2} \sin(2\pi ft), \tag{15}$$

where $a=k_a f\sqrt{(-2\ln 0.1)/p_s}$, $b=2p_s/3f$, p_s is the number of periods, k_a is the asymmetry factor, and f is the frequency.

Such a signal has the shape of a high-frequency ($f=5$ MHz) pulse with the Gaussian envelope. Steepness of the front and back slopes of the pulse can be set separately selecting a corresponding value of k_a. Simulation was carried out for a short pulse ($p_s=1.5$).

IV. SIMULATION AND EXPERIMENTAL RESULTS

Calculations of transient ultrasonic fields and waveforms in the time domain were performed for a disk-type transducer of radius $R=5$ mm and frequency $f=5$ MHz. It was assumed that the protection layer of the transducer was made of stainless steel and had the inclination angle of the front surface 2°. The thickness of the protection layer on the central axis was 13 mm. The ultrasound velocity in the protection layer was assumed to be $c_p=5800$ m/s and the density $\rho=7850$ kg/m^3. It was assumed that the transducer radiates waves into water ($c=1480$ m/s, $\rho=1000$ kg/m^3). The simulated pressure field only in the second medium (water) is presented in Fig. 9 as $p_{cs}(x,z)=\max_t|p(x,z,t)|$. For better understanding, the presented field is normalized with respect to the maximum value of the pressure in the second medium.

In Fig. 9 the simulated ultrasonic fields of the transducer in the second medium (water), after direct transmission through the nonparallel boundary stainless-steel–water (Fig. 9(a)), after single reflection inside the layer with the nonparallel boundary (Fig. 9(b)) and after double reflection inside the layer [Fig. 9(c)], are presented. Please note that the amplitudes of the fields shown in Figs. 9(a)–9(c) are normalized in each figure with respect to the maximal value of the particular field, e.g., it is impossible to compare amplitude relations in different plots. Normalization was performed with the purpose to display better the spatial structure of the ultrasonic fields. Actually, amplitudes of the signals reflected inside the protector are much lower than that of the directly transmitted signal and may be estimated from the waveforms of radiated ultrasonic pulses.

The waveforms of the ultrasonic pulses at the distance $z=50$ mm from the transducer are shown in Fig. 10. The transducer was driven by the signal described by Eq. (15). From the simulation results follows that ultrasonic beams, transmitted directly and after one or few reflections inside the layer, are propagating in water at slightly different directions. Therefore, each waveform was calculated also slightly at different positions ($x\neq 0$) where the maximal radiation of the particular beam is obtained. Please note that the amplitudes of the pulses shown in Figs. 10(a)–10(c) in each figure are normalized with respect to the maximal value of the directly transmitted pulse presented in Fig. 10(a).

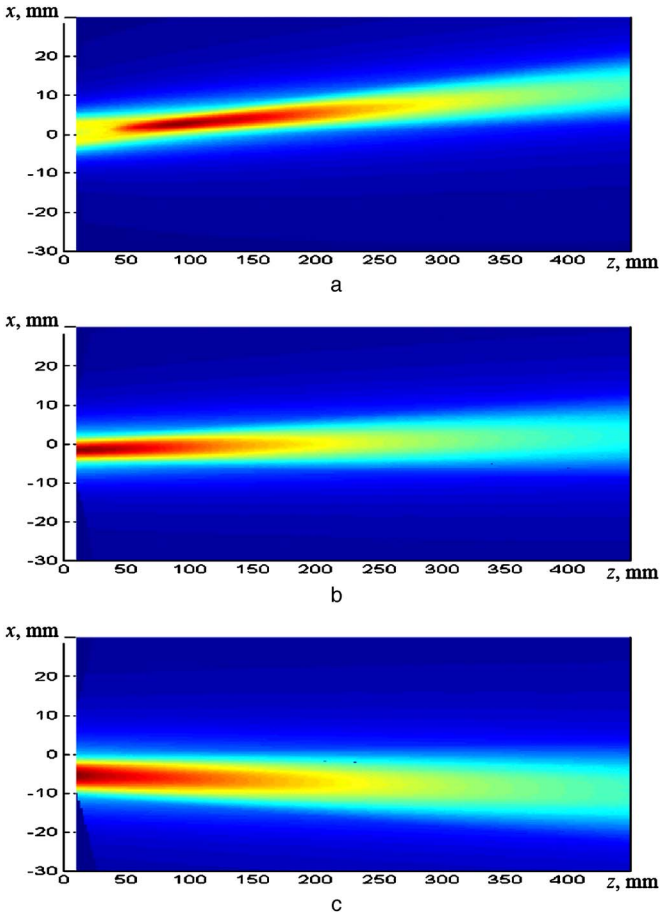


FIG. 9. (Color online) Simulated ultrasonic field of the transducer with a thick inclined protector (layer with nonparallel boundaries) in water: (a) directly transmitted signal, (b) signal after single reflection in the protector, and (c) signal after double reflection in the protector.

In order to test the validity of the computational method experimental investigations were performed. The measurements were carried out in water ($c=1500$ m/s, $\rho=1000$ kg/m^3) using the $f=5$ MHz circular ultrasonic transducer of radius $R=5$ mm inclined by 2° front surface. For

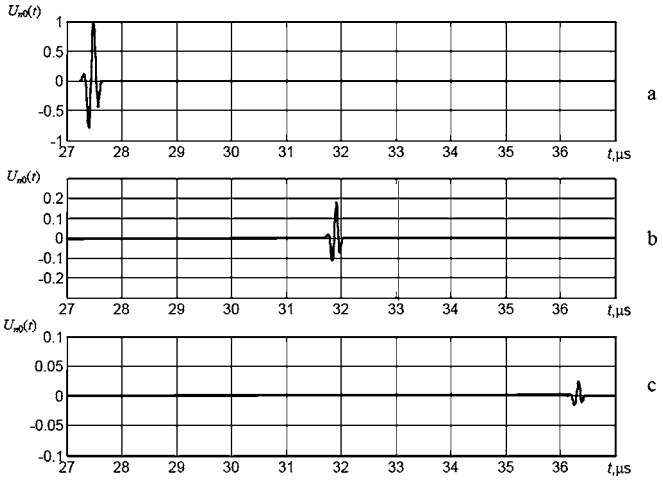


FIG. 10. Ultrasonic impulses at the distance $z=50$ mm from the transducer: (a) U_{n0}- directly transmitted impulse; (b) U_{n1}- impulse after one reflection inside the layer with the inclined surface; (c) U_{n2}- impulse after two reflections inside layer with the inclined surface.

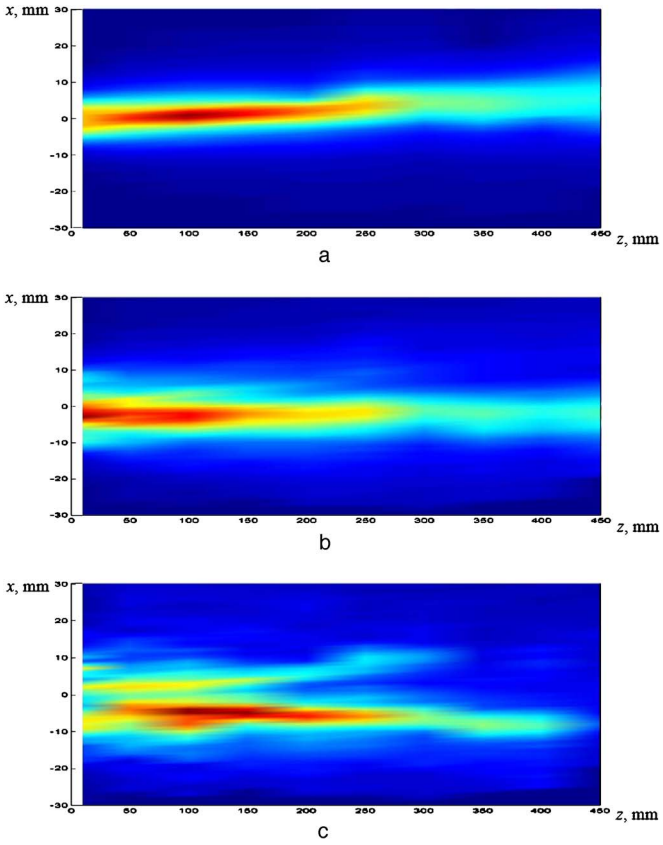


FIG. 11. (Color online) Experimentally measured ultrasonic field of the transducer with a thick inclined protector in water: (a) directly transmitted signal, (b) signal after single reflection in the protector, and (c) signal after double reflection in the protector.

experiments the water tank with a scanning system was used. The scanner has three independent axes, each of which can be positioned with a step of 9 μm. The absolute repeatability of the linear scanner is approximately 50 μm.

Ultrasonic fields were measured in a transmission mode. The measurements were performed in the water tank, using for reception of ultrasonic waves a wide-band (1–15 MHz) miniature piezoceramic probe (diameter 1 mm), which scanned along and across the direction of propagation of ultrasonic waves. The measurements were performed in the distance ranges 15–450 mm from the transducer. At each cross section, the measurements were carried out at 30 points. Data were collected by means of the computerized measuring system described above.

In Fig. 11, the measured ultrasonic field in water is presented. Here, as in Fig. 9, the measurement results of ultrasonic fields for three cases are presented. In Fig. 11(a) is shown the ultrasonic field in the second medium (water) after transmission through the nonparallel boundary stainless-steel–water; in Fig. 11(b), the ultrasonic field in the second medium (water) after reflection between the nonparallel boundary stainless-steel–water and the transducer surface; and in Fig. 11(c), before passing the nonparallel boundary the ultrasonic wave is twice reflected between the nonparallel boundary and the transducer surface.

As it was mentioned in the previous section, due to multiple reflections in the protector layer, the pulse response of

the transducer consists of a few decaying pulses. Therefore, the ultrasonic fields were measured separately for each pulse [Figs. 11(a)–11(c)]. Please note that the amplitudes of the fields shown in Figs. 11(a)–11(c), are normalized in each figure the same as in the case of simulated fields.

From the results presented in Figs. 9 and 11 it can be seen that the structures of the ultrasonic field in both cases (simulated and experimental) are similar. The comparison of such parameters of the field as beam width, inclination angle of the particular beam, and spatial positions of maximal signal values indicates that the proposed method gives accurate results.

The worse coincidence of theoretical and experimental results in the case of the twice-reflected beam in the solid layer can be explained by the fact that during reflection of the incident longitudinal wave at the inclined boundary a mode transformation takes place and, due to that, not only a longitudinal wave but also a shear wave is reflected back. This shear wave reflected once inside the protection layer is transformed on the boundary steel-layer–water into longitudinal wave and arrives to the measurement point in water almost at the same time as the twice-reflected longitudinal wave in the layer. So, the field presented in Fig. 11(c) after double reflection in the protector is the result of interference of these two waves. Please note that the mode conversion phenomenon inside the protector was not taken into account during simulations presented in Fig. 9.

From the results presented, it follows that a good correspondence between calculated and measured fields was obtained. That indicates that the proposed method may be used for prediction of the structure of ultrasonic fields in media with nonparallel boundaries with reasonable accuracy.

V. CONCLUSIONS

In this paper, the method and fast algorithm for simulation of ultrasonic fields excited by the ultrasonic disk-shaped transducers is presented. The known model for calculation of the ultrasonic field of a single circular transducer in a homogeneous medium was extended to the case of a multi-layered medium. The model enables calculation of ultrasonic fields in the medium with nonparallel boundaries, not only transmitted directly through nonparallel boundaries, but also after reflections in layers with nonparallel boundaries.

The simulated ultrasonic fields of the transducer with a nonparallel front surface of the protection layer, radiating into the water, were compared with the measurement results and a good correspondence between the calculated and the measured fields was obtained. Therefore, it is possible to conclude that the simulation results enable one to predict the structure of the field after passing a layer with nonparallel boundaries taking into account multiple reflections inside such a layer.

The application of the proposed modeling method for the analysis of ultrasonic fields radiated by a circular transducer with an inclined protective layer have shown that in this case not only reduction of the amplitude of multiple reflections in the layer may be achieved, but also the waves caused by these reflections are radiated in different directions. That is a very useful result for many practical applications because the total influence of these reflections on the directly transmitted signal is reduced essentially.

ACKNOWLEDGMENTS

The authors would like to thank R. Sliteris for performing the measurements and A. Voleisis for providing the ultrasonic transducer for the experiments.

[1]G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[2]J. N. Tjotta and S. Tjotta, "Near-field and far-field of pulsed acoustic radiators," J. Acoust. Soc. Am. **71**, 824–834 (1982).
[3]J. P. Weight, "Ultrasonic beam structures in fluid medium," J. Acoust. Soc. Am. **76**, 1184–1191 (1984).
[4]J. P. Weight and A. J. Hayman, "Observations of the propagation of very short ultrasonic pulses and their reflection by small targets," J. Acoust. Soc. Am. **63**, 396–404 (1978).
[5]A. J. Hayman and J. P. Weight, "Transmission and reception of short ultrasonic pulses by circular and square transducers," J. Acoust. Soc. Am. **66**, 945–951 (1979).
[6]M. El Amrani, P. Calmon, O. Roy, D. Royer, and O. Casula, "The ultrasonic field of transducers through a liquid-solid interface," Rev. Prog. Quant. Nondestr. Eval. **14**, 1075–1082 (1995).
[7]F. Buiochi, O. Martinez, L. G. Ullate, and F. Montero de Espinosa, "Computing reflection and transmission of ultrasonic beams through interfaces," *Proc. 8th ECNDT*, Barcelona, Spain, CD-ROM (2002), pp. 1–8.
[8]F. Buiochi, O. Martinez, L. G. Ullate, and F. Montero de Espinosa, "A computational method to predict the reflected and transmitted ultrasonic fields with interfaces of complex geometry," *IEEE Ultrasonic Symposium Proc.* CD-ROM – 537 (2002), pp. 1–4.
[9]F. Buiochi, O. Martinez, L. G. Ullate, and F. Montero de Espinosa, "A computational method to calculate the longitudinal wave evolution caused by interfaces between isotropic media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 181–192 (2004).
[10]F. Buiochi, O. Martinez, L. G. Ullate, and F. Montero de Espinosa, "3D computational method to study the focal laws of transducers arrays for NDE applications," Ultrasonics **42**, 871–876 (2004).
[11]L. W. Schmerr, T. P. Lerch, and A. Sedov, "Modeling the radiation of focused and unfocused ultrasonic transducers through planar interfaces," Rev. Prog. Quant. Nondestr. Eval. **14**, 1061–1066 (1995).
[12]D. Belgroune, J. F. de Belleval, and H. Djelouah, "Modeling of the ultrasonic field by the angular spectrum method in presence of interface," Ultrasonics **40**, 297–302 (2002).
[13]T. P. Lerch, L. W. Schmerr, and A. Sedov, "The paraxial approximation for radiation of a planar ultrasonic transducer at oblique incidence through an interface," Rev. Prog. Quant. Nondestr. Eval. **14**, 1067–1074 (1995).
[14]P. Calmon, A. Lhemery, I. Lecoeur-Taibi, R. Raillon, and L. Paradis, "Models for the computation of ultrasonic fields and their interaction with defects in realistic NDT configurations," Nucl. Eng. Des. **180**, 271–283 (1998).
[15]L. Odegaard, S. Holm, F. Teigen, and T. Klevelad, "Acoustic field simulation for arbitrarily shaped transducers in a stratified medium," Ultrasonics Symposium (1994), pp. 1535–1538.
[16]L. W. Schmerr, H.-J. Kim, R. Huang, and A. Sedov, "Multi-Gaussian ultrasonic beam modeling," WCU 2003, Paris (2003), pp. 93–99.
[17]E. Jasiūnienė and L. Mažeika, "The modified method for simulation of ultrasonic fields of disk shape transducer," Ultagarsas **33**, 33–37 (1999).
[18]R. Kažys, L. Mažeika, and E. Jasiūnienė, "Simulation of ultrasonic fields propagating through nonparallel boundaries," Lith. J. Phys. **44**, 359–366 (2004).
[19]R. Kažys, L. Mažeika, and E. Jasiūnienė, "Ultrasonic fields radiated through matching layers with nonparallel boundaries," Ultrasonics **42**, 267–271 (2004).
[20]D. E. Robinson, S. Lees, and L. Bess, "Near field transient radiation patterns for circular pistons," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-22**, 395–405 (1974).
[21]D. Esminger, *Ultrasonics. Fundamentals, Technology, Applications* (Marcel Dekker, New York, 1998).
[22]L. W. Schmerr and A. Sedov, "An elastodynamic model for compressional and shear wave transducers," J. Acoust. Soc. Am. **86**, 1988–1999 (1989).

On the sound field of a resilient disk in an infinite baffle

Tim Mellow
Nokia UK Ltd., Farnborough, Hants GU14 0NG, England

(Received 8 December 2005; revised 26 March 2006; accepted 27 April 2006)

The Rayleigh integral describing the near-field pressure of an axisymmetric planar monopole source with an arbitrary velocity distribution is solved with a method similar to that used by Mast and Yu [J. Acoust. Soc. Am. **118**(6), 3457–3464 (2005)] for a rigid disk in an infinite baffle. The closed-form solution is in the form of a double expansion, which is valid for distances from the observation point to the center of the source that are greater than its radius. However, for the remaining immediate near field, the King integral is solved using a combination of Gegenbauer's summation theorem and the Lommel expansion, resulting in a solution which is in the form of a triple expansion, reducing to a double expansion along the source's axis of symmetry. These relatively compact solutions in analytic form do not require numerical integration and therefore present no numerical difficulties except for a singularity at the rim. As an example of a monopole source with an arbitrary velocity distribution, equations describing the radiation characteristics of a resilient disk in an infinite baffle are derived. Using Babinet's principle, the pressure field of a plane wave passing through the complementary hole in an infinite rigid screen is calculated.
 [DOI: 10.1121/1.2206513]

PACS number(s): 43.20.Rz, 43.20.El, 43.20.Tb, 43.20.Wd [LLT] Pages: 90–101

I. INTRODUCTION

Despite the advances in acoustical simulation tools in recent years, it is still useful to have some canonical forms which can be rigorously calculated in a direct manner without any need for numerical integration, iteration, or least-squares minimization. Although such solutions are generally restricted to simple geometries (usually axisymmetric), they provide useful benchmarks for simulation. Most canonical forms, such as the spherical polar cap on a sphere,[1] are based upon spherical geometry.

Recently, Mast and Yu[2] derived an elegant pair of expansions for the pressure field of a rigid disk in an infinite baffle, based upon solutions to the Rayleigh[3] integral. The first expansion, or "outer" solution, provides a fast converging series for distances from the center of the disk to the observation point greater than the disk's radius, and can be regarded as an improved version of Stenzel's solution[4] using functions commonly found in text books as opposed to Stenzel's bespoke ones, which were based upon Lommel's polynomials.

Stenzel also derived what could be termed an "inner" solution for distances from the center of the disk to the observation point less than the disk's radius. This expansion can also be updated using hypergeometric function solutions to the radial integrals, which the author has already tried. However, although it converges in a similar fashion to the outer expansion, this is not the best solution available. In all of the aforementioned solutions, the origin of the spherical coordinate system is located at the center of the disk.

Hasegawa *et al.*[5] provided an expansion, together with recursion formulas, based upon an ingenious spherical coordinate system, the origin of which being located on the disk's axis of symmetry some distance in front of it. Mast and Yu[2] have developed this by locking the origin of the coordinate system to the same axial distance from the disk as the observation point, while keeping it on the axis of symmetry. This leads to their second expansion or "paraxial" solution for which the region of convergence looks like a funnel, falling just within the disk's perimeter at its surface and then spreading out to form a cone covering an angle of 45° either side of the axis of symmetry with the apex located at the disk's center. The limit of this expansion decreases to a single term on the axis of symmetry.

Hence, the Rayleigh integral has been shown to provide simple single expansion solutions for a rigid disk. In this paper, it is also shown to provide a similar outer solution for a monopole source with an arbitrary velocity distribution. However, for an inner or paraxial solution this does not appear to be the case and Stenzel[6] tackled the problem in a somewhat formidable analysis.

Surprisingly little attention has been paid to the King integral,[7–9] which could be solved by means of a double expansion. Greenspan[10] calculated it numerically to illustrate some special cases, such as the on-axis pressure, radiation force, and power for various monopole velocity distributions, as well as the transient response. Using two Lommel expansions, Williams[11] recast the King integral in a mathematically beautiful form, which he then used to illustrate some of its properties. Unfortunately, it can be shown that the integral in this expanded form yields a converging solution for only part of the near-field space. When the reverse Lommel expansions are applied, it is not obvious how to calculate the subsequent expression numerically and no results are provided.

In the current paper, the King integral is expanded using a combination of the Lommel expansion and the Gegenbauer summation theorem in order to ensure convergence. This yields a paraxial solution, which together with the outer solution, exhibits similar convergence characteristics to the expansions of Mast and Yu for the rigid disk.

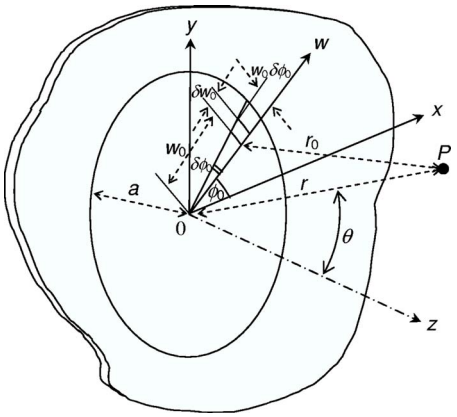


FIG. 1. Geometry of the disk in an infinite baffle.

A general aim of this paper is to present the most direct solutions possible for calculating the sound radiation characteristics of axisymmetric *monopole* sources, especially when the *velocity* distribution is unknown. In this respect, it is intended to complement a recent paper by Kärkkäinen *et al.*,[12] in which a solution was presented for calculating the radiation characteristics of *dipole* sources with unknown *pressure* distributions. The surface pressure distribution was described using a trial function consisting of a power series for which the coefficients were calculated by numerically solving a set of simultaneous equations. The simultaneous equations, in turn, had to be derived by analytically solving the integrals in the dipole part of the Kirchhoff-Helmholtz boundary integral formula (using the Green's function in cylindrical coordinates). In this paper, a similar procedure is followed, using a trial function for the surface velocity distribution and then solving the monopole part of the Kirchhoff-Helmholtz boundary integral formula.

Furthermore, the methods derived here can also be applied to nonrigid sources with fluid-structure coupling, such as plates and membranes. For example, similar formulation appears in a study on membranes by Kärkkäinen *et al.*[13]

II. RESILIENT DISK IN AN INFINITE BAFFLE

A. Boundary conditions

The infinitesimally thin rigid disk shown in Fig. 1 lies in the w plane with its center at the origin, where w is the radial ordinate of the cylindrical coordinate system and z is the axial ordinate. Due to axial symmetry, the polar ordinate ϕ of the spherical coordinate system can be ignored, so that the observation point P is defined simply in terms of the radial and azimuthal ordinates r and θ, respectively. The infinitesimally thin rigid baffle extends from the perimeter of the disk to $w=\infty$. On the baffle, the velocity is zero and therefore so is the normal pressure gradient. The infinitesimally thin membrane-like resilient disk is assumed to be perfectly flexible, has zero mass, and is free at its perimeter. It is driven by a uniformly distributed harmonically varying pressure $\tilde{p}_0$ and thus radiates sound from both sides into a homogeneous loss-free acoustic medium. In fact, there need not be a disk present at all and instead the driving pressure could be acting upon the air particles directly. However, for expedience, the area over which this driving pressure is applied shall be referred to as a disk from here onwards. On the surface of the disk and baffle, the following boundary conditions apply:

$$\frac{\partial}{\partial z_0}\tilde{p}(w_0,z_0)|_{z_0=0+}=\begin{cases}-ik\rho c\tilde{u}(w_0), & 0\leqslant w_0\leqslant a,\\ 0, & w_0>a,\end{cases} \tag{1}$$

where

$$\tilde{u}_0(w_0)=\sum_{m=0}^{\infty}\tilde{A}_m\left(1-\frac{w_0^2}{a^2}\right)^{m-(1/2)}, \tag{2}$$

and k is the wave number given by

$$k=\frac{\omega}{c}=\frac{2\pi}{\lambda}, \tag{3}$$

where ω is the angular frequency of excitation, ρ is the density of the surrounding medium, c is the speed of sound in that medium, and λ is the wavelength. The annotation $\tilde{}$ denotes a harmonically time-varying quantity and replaces the factor $e^{i\omega t}$. It is worth noting that the index of the first term of the expansion ($m=0$) is equal to $-1/2$, in order to satisfy the boundary condition of infinite velocity at the perimeter, as determined by Rayleigh.[3] The same expansion can be applied to any velocity distribution, providing the velocity is either infinite or zero at the perimeter. For example, in the case of a circular membrane with a clamped rim, the index of the first term would be equal to $+1/2$.

On the front and rear surfaces of the disk, the pressures are $\tilde{p}_+$ and $\tilde{p}_-$, respectively, which are given by

$$\tilde{p}_+=-\tilde{p}_-=\frac{\tilde{p}_0}{2}. \tag{4}$$

B. Solution of the free-space wave equation

Using the King integral, the pressure distribution is defined by

$$\tilde{p}(w,z)=2\int_0^{2\pi}\int_0^a g(w,z|w_0,z_0)\times\frac{\partial}{\partial z_0}\tilde{p}(w_0,z_0)|_{z_0=0+}w_0dw_0d\phi_0, \tag{5}$$

where the Green's function[14] is defined, in cylindrical coordinates, by the Lamb[15] or Sommerfeld[16] integral,

$$g(w,z|w_0,z_0)=\frac{i}{4\pi}\int_0^{\infty}J_0(\mu w)J_0(\mu w_0)\frac{\mu}{\sigma}e^{-i\sigma|z-z_0|}d\mu, \tag{6}$$

where

$$\sigma=\sqrt{k^2-\mu^2}. \tag{7}$$

Substituting Eqs. (1), (2), and (6), in Eq. (5) and integrating over the surface of the disk yields

$$\tilde{p}(w,z) = ka\rho c \sum_{m=0}^{\infty} \tilde{A}_m \Gamma\left(m+\frac{1}{2}\right) \int_0^{\infty} \left(\frac{2}{a\mu}\right)^{m-(1/2)} \times J_0(w\mu) J_{m+(1/2)}(a\mu) \frac{e^{-i\sigma z}}{\sigma} d\mu, \tag{8}$$

where Sonine's integral[17] has been used as follows:

$$\int_0^a \left(1-\frac{w_0^2}{a^2}\right)^{m-(1/2)} J_0(\mu w_0) w_0 dw_0 = \frac{a^2}{2}\Gamma\left(m+\frac{1}{2}\right) \times \left(\frac{2}{a\mu}\right)^{m+(1/2)} J_{m+(1/2)}(a\mu). \tag{9}$$

Applying the boundary condition of Eq. (4) leads to

$$\frac{\tilde{p}_0}{2} = ka\rho c \sum_{m=0}^{\infty} \tilde{A}_m \Gamma\left(m+\frac{1}{2}\right) \int_0^{\infty} \left(\frac{2}{a\mu}\right)^{m-(1/2)} \times J_0(w\mu) J_{m+(1/2)}(a\mu) \frac{1}{\sigma} d\mu. \tag{10}$$

C. Formulation of the coupled problem

Equation (10) can be written more simply as

$$\sum_{m=0}^{\infty} \tau_m I_m(w,k) = 1, \tag{11}$$

which is to be solved for the normalized power series coefficients τ_m as defined by

$$\tau_m = \frac{2\rho c \tilde{A}_m}{(m+1/2)\tilde{p}_0}. \tag{12}$$

The integral $I_m(w,k)$ can be split into two parts,

$$I_m(w,k) = I_{mR}(w,k) - iI_{mI}(w,k), \tag{13}$$

where the real part is given by

$$I_{mR}(w,k) = ka\Gamma\left(m+\frac{3}{2}\right) \int_0^k \left(\frac{2}{\mu a}\right)^{m-(1/2)} J_{m+(1/2)} \times (\mu a) J_0(\mu w) \frac{1}{\sqrt{k^2-\mu^2}} d\mu, \tag{14}$$

and the imaginary part is given by

$$I_{mI}(w,k) = ka\Gamma\left(m+\frac{3}{2}\right) \int_k^{\infty} \left(\frac{2}{\mu a}\right)^{m-(1/2)} J_{m+(1/2)} \times (\mu a) J_0(\mu w) \frac{1}{\sqrt{\mu^2-k^2}} d\mu. \tag{15}$$

D. Solution of the real integral

Substitution of $\mu = k \sin \vartheta$ in Eq. (14) gives

$$I_{mR}(w,k) = ka\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(1/2)} \int_0^{\pi/2} (\sin \vartheta)^{(1/2)-m} \times J_{m+(1/2)}(ka \sin \vartheta) J_0(kw \sin \vartheta) d\vartheta. \tag{16}$$

The Bessel functions in Eq. (16) are defined by[17]

$$J_0(kw \sin \vartheta) = \sum_{q=0}^{Q} \left(\frac{kw}{2}\right)^{2q} \frac{(-1)^q (\sin \vartheta)^{2q}}{(q!)^2}, \tag{17}$$

$$J_{m+(1/2)}(ka \sin \vartheta) = \sum_{r=0}^{R} \left(\frac{ka}{2}\right)^{2r+m+(1/2)} \times \frac{(-1)^r (\sin \vartheta)^{2r+m+(1/2)}}{r!\Gamma(r+m+3/2)}, \tag{18}$$

so that

$$I_{mR}(w,k) = 2\sum_{q=0}^{Q} \sum_{r=0}^{R} \frac{(-1)^{q+r}\Gamma(m+3/2)}{(q!)^2 r!\Gamma(r+m+3/2)} \left(\frac{ka}{2}\right)^{2(q+r+1)} \times \left(\frac{w}{a}\right)^{2q} \int_0^{\pi/2} (\sin \vartheta)^{2(q+r)+1} d\vartheta. \tag{19}$$

Solution of the integral in Eq. (19) is enabled by use of the following identity:[18]

$$\int_0^{\pi/2} (\sin \vartheta)^{2(q+r)+1} d\vartheta = \frac{\sqrt{\pi}\Gamma(q+r+1)}{2\Gamma(q+r+3/2)}. \tag{20}$$

Evaluating the integral over ϑ yields

$$I_{mR}(w,k) = \sqrt{\pi}\sum_{q=0}^{Q} \sum_{r=0}^{R} \frac{(-1)^{q+r}\Gamma(m+3/2)\Gamma(q+r+1)}{(q!)^2 r!\Gamma(r+m+3/2)\Gamma(q+r+3/2)} \times \left(\frac{ka}{2}\right)^{2(q+r+1)} \left(\frac{w}{a}\right)^{2q}. \tag{21}$$

E. Solution of the imaginary integral

1. Transformation of the integral into complex form

The following procedure converts the infinite limit of the integral in Eq. (15) into a finite one. First, the integral is converted into a form which can be integrated in the complex plane. The Bessel function $J_n(x)$ can be written in terms of the following pair of complex conjugate Hankel functions:[17]

$$J_n(x) = \frac{H_n^{(1)}(x) + H_n^{(2)}(x)}{2}, \tag{22}$$

which can now be used to separate I_{mI} into two complex conjugate integrals as follows:

$$I_{mI}(w,k) = \frac{I_{mI}^{(1)} + I_{mI}^{(2)}}{2}, \tag{23}$$

where, after substituting $\mu = kt$, the complex conjugate integrals are given by

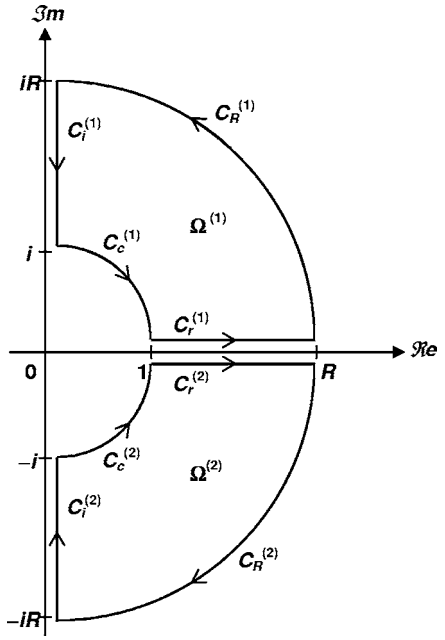


FIG. 2. Infinite integration contours in the complex t plane.

$$I_{ml}^{(1)} = 2\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(3/2)}\int_1^\infty J_0(kwt)H_{m+(1/2)}^{(1)} \times(kat)\frac{t^{(1/2)-m}}{\sqrt{t^2-1}}dt, \tag{24}$$

$$I_{ml}^{(2)} = 2\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(3/2)}\int_1^\infty J_0(kwt)H_{m+(1/2)}^{(2)} \times(kat)\frac{t^{(1/2)-m}}{\sqrt{t^2-1}}dt. \tag{25}$$

Referring to the complex t plane of Fig. 2, the integrals $I_{ml}^{(1)}$ and $I_{ml}^{(2)}$ can now be evaluated along contours $\Omega^{(1)}$ and $\Omega^{(2)}$, respectively. The contours are defined by

$$\begin{aligned}
&\Omega^{(1)} = t \in (C_r^{(1)} \cup C_R^{(1)} \cup C_i^{(1)} \cup C_c^{(1)}),\\
&C_r^{(1)} = [1,R],\\
&C_R^{(1)} = [Re^{i\vartheta}|0 \leqslant \vartheta \leqslant \pi/2],\\
&C_i^{(1)} = [iR,i],\\
&C_c^{(1)} = [e^{i\vartheta}|\pi/2 \geqslant \vartheta \geqslant 0],\\
&R \to \infty,\\
&\Omega^{(2)} \text{ symmetric to } \Omega^{(1)} \text{ with respect to the real axis.}
\end{aligned} \tag{26}$$

2. Contribution of $C_R^{(1)}$ and $C_R^{(2)}$

The contributions along $C_R^{(1)}$ and $C_R^{(2)}$ vanish for $R\to\infty$ due to the behavior of $H_{m+(1/2)}(t)$ as $|t|\to\infty$.

3. Contribution of $C_i^{(1)}$ and $C_i^{(2)}$

Noting that $\sqrt{t^2-1}=i\sqrt{1-t^2}$, the integral along $C_i^{(1)}$ can be written

$$2\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(3/2)}\int_{i\infty}^{i} J_0(kwt)H_{m+(1/2)}^{(1)}(kat)\frac{t^{(1/2)-m}}{i\sqrt{1-t^2}}dt, \tag{27}$$

which can be converted into an integral with real limits by substituting $t=is$ as follows:

$$-2i^{(1/2)-m}\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(3/2)}\int_1^\infty J_0(ikws)H_{m+(1/2)}^{(1)} \times(ikas)\frac{s^{(1/2)-m}}{\sqrt{1+s^2}}ds. \tag{28}$$

With help from the following identities:[17]

$$I_0(x) = J_0(ix), \tag{29}$$

$$K_\nu(x) = i^{\nu+1}\frac{\pi}{2}H_\nu^{(1)}(ix), \tag{30}$$

the integral can be written as

$$\frac{4i}{\pi}(-1)^m\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(3/2)}\int_1^\infty I_0(kws)K_{m+(1/2)} \times(kas)\frac{s^{(1/2)-m}}{\sqrt{1+s^2}}ds. \tag{31}$$

This integral is purely imaginary whereas the original is real valued. Hence, there is zero net contribution along $C_i^{(1)}$. The same is true for the contribution along $C_i^{(2)}$ where $\sqrt{t^2-1}=-i\sqrt{1-t^2}$.

4. Contribution of $C_c^{(1)}$ and $C_c^{(2)}$

Finally, the contributions along the unity quarter circle segments $C_c^{(1)}$ and $C_c^{(2)}$ can be calculated by using the substitution $t=e^{i\vartheta}$, so that the contribution along $C_c^{(1)}$ becomes

$$\Re\left(2i\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(3/2)}\int_{\pi/2}^{0} J_0(kwe^{i\vartheta})H_{m+(1/2)}^{(1)} \times(kae^{i\vartheta})\frac{e^{i((3/2)-m)\vartheta}}{\sqrt{e^{2i\vartheta}-1}}d\vartheta\right), \tag{32}$$

and likewise the contribution along $C_c^{(2)}$ becomes

$$\Re\left(2i\Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka}\right)^{m-(3/2)}\int_{-\pi/2}^{0} J_0(kwe^{i\vartheta})H_{m+(1/2)}^{(2)} \times(kae^{i\vartheta})\frac{e^{i((3/2)-m)\vartheta}}{\sqrt{e^{2i\vartheta}-1}}d\vartheta\right), \tag{33}$$

which is equal to Eq. (32). As there are no poles or zeros within the contours $\Omega^{(1)}$ or $\Omega^{(2)}$, it can be stated that, according to the residue theorem, the sum of the integrals around each of these contours is equal to zero. Therefore, I_{ml} can be written as

$$I_{mI}(w,k)=\frac{I_{mI}^{(1)}(w,k)+I_{mI}^{(2)}(w,k)}{2}=-2\Gamma\left(m+\frac{3}{2}\right)\times\left(\frac{2}{ka}\right)^{m-(3/2)}\Re\left(i\int_0^{\pi/2}J_0(kwe^{i\vartheta})\frac{e^{i((3/2)-m)\vartheta}}{\sqrt{e^{2i\vartheta}-1}}\times\{J_{m+(1/2)}(kae^{i\vartheta})+iY_{m+(1/2)}(kae^{i\vartheta})\}d\vartheta\right). \tag{34}$$

5. Expansion of the Bessel functions and final solution of the imaginary integral

A series expansion[17] of the Neumann function in Eq. (34) is given by

$$Y_{m+1/2}(kae^{i\vartheta})=\sum_{r=0}^{R}\left(\frac{ka}{2}\right)^{2r-m-(1/2)}\frac{(-1)^{r+m}e^{i\vartheta(2r-m-(1/2))}}{r!\Gamma(r-m+1/2)}. \tag{35}$$

Letting $\sin\vartheta=e^{i\vartheta}$ in Eqs. (17) and (18) and substituting these together with Eq. (35) in Eq. (34) gives

$$I_{mI}(w,k)=-2\sum_{q=0}^{Q}\sum_{r=0}^{R}\Re\left(\left(\frac{w}{a}\right)^{2q}\frac{(-1)^{q+r}\Gamma(m+3/2)}{(q!)^2r!\Gamma(r+m+3/2)}\times\left(\frac{ka}{2}\right)^{2(q+r+1)}i\int_0^{\pi/2}\frac{e^{2i\vartheta(q+r+1)}}{\sqrt{e^{2i\vartheta}-1}}d\vartheta-\frac{(-1)^{q+r+m}\Gamma(m+3/2)}{(q!)^2r!\Gamma(r-m+1/2)}\left(\frac{w}{a}\right)^{2q}\times\left(\frac{ka}{2}\right)^{2(q+r-m)}\int_0^{\pi/2}\frac{e^{2i\vartheta(q+r-m+(1/2))}}{\sqrt{e^{2i\vartheta}-1}}d\vartheta\right). \tag{36}$$

Solution of the integrals in Eq. (36) is enabled by use of the following identity:[18]

$$\int_0^{\pi/2}\frac{e^{2i\gamma\vartheta}}{\sqrt{e^{2i\vartheta}-1}}d\vartheta=\frac{1}{2\gamma}\left(\frac{\sqrt{\pi}\Gamma(\gamma+1)}{\Gamma(\gamma+1/2)}-{}_2F_1\left(\frac{1}{2},\gamma;\gamma+1;-1\right)e^{i\pi\gamma}\right). \tag{37}$$

Evaluating the integral over ϑ yields

$$I_{mI}(w,k)=2\Gamma\left(m+\frac{3}{2}\right)\Re\left(\sum_{q=0}^{Q}\sum_{r=0}^{R}F_Y(q,r,m)\times\left(\frac{ka}{2}\right)^{2(q+r-m)+1}\left(\frac{w}{a}\right)^{2q}-iF_J(q,r,m)\times\left(\frac{ka}{2}\right)^{2(q+r+1)}\left(\frac{w}{a}\right)^{2q}\right), \tag{38}$$

where the subfunctions F_Y and F_J are given by

$$F_Y(q,r,m)=(-1)^{q+r+m}\left(\frac{{}_2F_1(1/2,\alpha;\alpha+1;-1)e^{i\pi\alpha}}{2\alpha(q!)^2r!\Gamma(r-m+1/2)}-\frac{\sqrt{\pi}\Gamma(\alpha)}{2(q!)^2r!\Gamma(r-m+1/2)\Gamma(\alpha+1/2)}\right), \tag{39}$$

$$F_J(q,r,m)=(-1)^{q+r}\left(\frac{{}_2F_1(1/2,\beta;\beta+1;-1)e^{i\pi\beta}}{2\beta(q!)^2r!\Gamma(r+m+3/2)}-\frac{\sqrt{\pi}\Gamma(\beta)}{2(q!)^2r!\Gamma(r+m+3/2)\Gamma(\beta+1/2)}\right), \tag{40}$$

where

$$\alpha=q+r-m+1/2, \tag{41}$$

$$\beta=q+r+1. \tag{42}$$

However, for integer values of q and r, $iF_J(q,r,m)$ is purely imaginary and therefore makes no contribution to the real part of $I_{mI}(w,k)$. Similarly, the $e^{i\pi(q+r-m+1/2)}$ term of $F_Y(q,r,m)$ is also purely imaginary and can therefore be excluded. Thus, the final result can be written

$$I_{mI}(w,k)=-\sqrt{\pi}\sum_{q=0}^{Q}\sum_{r=0}^{R}\times\frac{(-1)^{q+r+m}\Gamma(m+3/2)\Gamma(q+r-m+1/2)}{(q!)^2r!\Gamma(r-m+1/2)\Gamma(q+r-m+1)}\times\left(\frac{ka}{2}\right)^{2(q+r-m)+1}\left(\frac{w}{a}\right)^{2q}. \tag{43}$$

F. Calculation of the power series coefficients (final set of simultaneous equations)

Truncating the infinite power series in (11) to order M and equating the coefficients of $(w/a)^{2q}$ yields the final set of $M+1$ simultaneous equations in τ_m as follows:

$$\sum_{m=0}^{M}({}_m\mathbf{P}_q(ka)+i\,{}_m\mathbf{T}_q(ka))\tau_m=\delta_{q0}, \tag{44}$$

where $\mathbf{P}$ shall be named the *Spence* function as defined by

$${}_m\mathbf{P}_q(ka)=\sqrt{\pi}\sum_{r=0}^{M}\frac{(-1)^{q+r}(ka/2)^{2(q+r+1)}}{(q!)^2r!(m+3/2)_r(q+r+1)_{1/2}}, \tag{45}$$

and $\mathbf{T}$ shall be named the *Stenzel* function as defined by

$${}_m\mathbf{T}_q(ka)=\sqrt{\pi}\sum_{r=0}^{M}\frac{(-1)^{q+r+m}(ka/2)^{2(q+r-m)+1}}{(q!)^2r!(m+3/2)_{r-2m-1}(q+r-m+1/2)_{1/2}}, \tag{46}$$

δ_{q0} is the Kronecker delta function and $(x)_n$ the Pochhammer symbol.[19] $\mathbf{P}$ and $\mathbf{T}$ are the monopole counterparts to the dipole cylindrical wave functions[12] $\mathbf{B}$ and $\mathbf{S}$. These equations are then solved for $q=0,1,2,\ldots,M-1,M$.

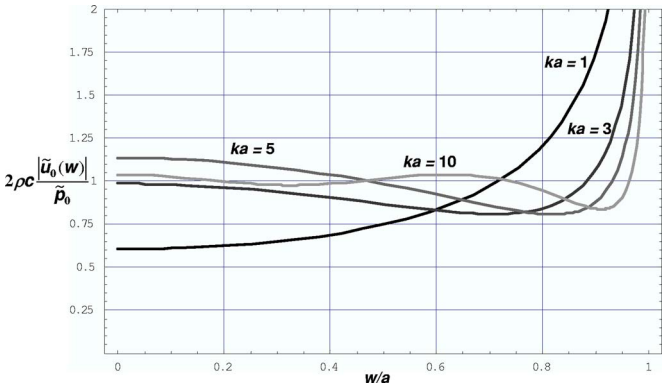


FIG. 3. Normalized surface velocity magnitude of the resilient disk.

G. Disk velocity

From Eq. (12) it follows that

$$\tilde{A}_m = \frac{\tau_m (m+1/2)\tilde{p}_0}{2\rho c}. \tag{47}$$

After substituting this in Eq. (2), the normalized disk velocity can be written as

$$\frac{2\rho c \tilde{u}_0(w_0)}{\tilde{p}_0} = \sum_{m=0}^{M} \tau_m \left(m+\frac{1}{2}\right)\left(1-\frac{w_0^2}{a^2}\right)^{m-(1/2)}, \quad 0 \leqslant w_0 \leqslant a. \tag{48}$$

The magnitude and phase of the normalized velocity are shown in Figs. 3 and 4, respectively, for various values of ka.

H. Radiation admittance

The total volume velocity $\tilde{U}_0$ produced by the disk can be found by integrating the disk velocity from (48) over the surface of the disk as follows:

$$\tilde{U}_0 = \int_0^{2\pi}\int_0^a \tilde{u}_0(w_0) w_0 dw_0 d\phi_0 = \frac{S\tilde{p}_0}{2\rho c}\sum_{m=0}^{M} \tau_m, \tag{49}$$

where S is the area of the disk given by $S=\pi a^2$. The acoustic radiation admittance is then given by

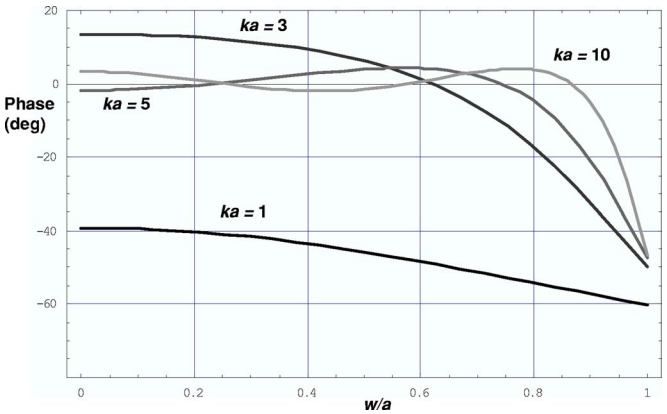


FIG. 4. Normalized surface velocity phase of the resilient disk.

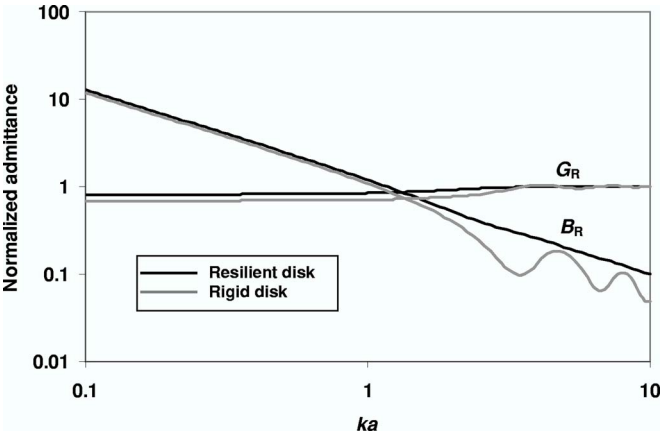


FIG. 5. Normalized radiation admittances of the rigid and resilient disk.

$$y_{ar} = \frac{\tilde{U}_0}{\tilde{p}_0} = \frac{S\tilde{u}_0}{\tilde{p}_0} = \frac{S}{2\rho c}(G_R + iB_R), \tag{50}$$

where G_R is the normalized *conductance* given by

$$G_R = \sum_{m=0}^{M} \Re(\tau_m) \approx \frac{8}{\pi^2}, \quad ka < 0.5, \tag{51}$$

and B_R is the normalized *susceptance* given by

$$B_R = \sum_{m=0}^{M} \Im(\tau_m) \approx \frac{4}{\pi ka}, \quad ka < 0.5. \tag{52}$$

The real and imaginary admittances G_R and B_R are plotted in Fig. 5 along with the actual admittance of a rigid disk for comparison.

I. Far-field pressure response

In the case of the far-field response, it is more convenient to use spherical coordinates so that the far-field polar responses can be obtained directly. Rayleigh's far-field approximation[14] is ideal for this purpose,

$$g(r,\theta,\phi|w_0,\phi_0,z_0) = \frac{1}{4\pi r} e^{-ik[r - w_0 \sin\theta\cos(\phi-\phi_0) - z_0\cos\theta]}, \tag{53}$$

which is inserted, together with Eqs. (1) and (2), in the following monopole part of the Kirchhoff-Helmholtz boundary integral formula in spherical coordinates:

$$\tilde{p}(r,\theta,\phi) = 2\int_0^{2\pi}\int_0^a g(r,\theta,\phi|w_0,\phi_0,z_0)|_{z_0=0+} \times \frac{\partial}{\partial z_0}\tilde{p}(w_0,\phi_0,z_0)|_{z_0=0+} w_0 dw_0 d\phi_0. \tag{54}$$

After integrating over the surface of the disk [while letting $\phi=\pi/2$ so that $\cos(\phi-\phi_0)=\sin\phi_0$], the far-field pressure is given by

$$\tilde{p}(r,\theta) = -\frac{ia\tilde{p}_0}{4r} e^{-ikr} D(\theta), \tag{55}$$

where the following identities have been used:[17]

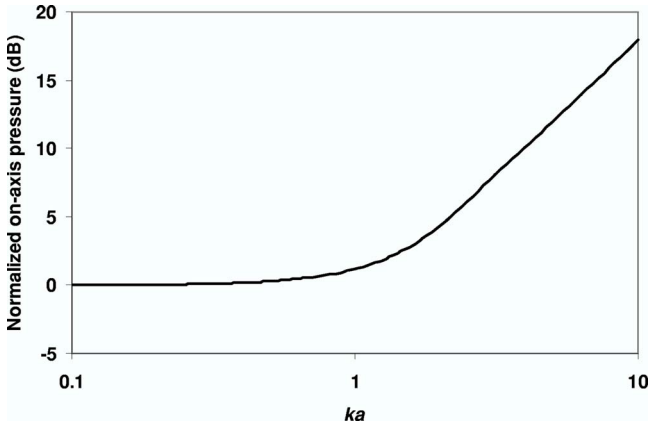


FIG. 6. Normalized far-field on-axis response of the resilient disk.

$$\frac{1}{2\pi}\int_0^{2\pi} e^{ikw_0 \sin\theta \sin\phi_0} d\phi_0 = J_0(kw_0 \sin\theta), \tag{56}$$

together with Eq. (9), where $\mu = k \sin\theta$. The directivity function $D(\theta)$ is given by

$$D(\theta) = ka\sum_{m=0}^{M} \tau_m \Gamma\left(m+\frac{3}{2}\right)\left(\frac{2}{ka\sin\theta}\right)^{m+(1/2)} J_{m+(1/2)} \times (ka\sin\theta). \tag{57}$$

The on-axis pressure is obtained by setting $\theta = 0$ in Eq. (53) before inserting it into Eq. (54). This results in an integral which is similar to the one for the radiation admittance in Eq. (49). Hence,

$$D(0) = ka\sum_{m=0}^{M} \tau_m \approx \begin{cases} 4i/\pi, & ka < 0.5 \\ ka, & ka > 2. \end{cases} \tag{58}$$

It is worth noting that the on-axis response is related to the radiation admittance by $D(0) = ka(G_R + iB_R)$. The asymptotic expression for low-frequency on-axis pressure is then simply

$$\tilde{p}(r,0) \approx \frac{a}{\pi r}\tilde{p}_0 e^{-ikr}, \quad ka < 0.5, \tag{59}$$

and likewise at high frequencies

$$\tilde{p}(r,0) \approx i\frac{ka^2}{4r}\tilde{p}_0 e^{-ikr}, \quad ka > 2, \tag{60}$$

which is the same as for a resilient disk in free space at all frequencies. The on-axis response is shown in Fig. 6, calculated from the magnitude of $D(0)$. The normalized directivity function $20\log_{10}(|D(\theta)|/|D(0)|)$ is plotted in Fig. 7 for various values of ka.

J. Near-field pressure when the distance from the center of the disk to the observation point is greater than the diaphragm's radius

Using the monopole part of the Kirchhoff-Helmholtz boundary integral formula,[14] the sound pressure at the observation point P can be written as

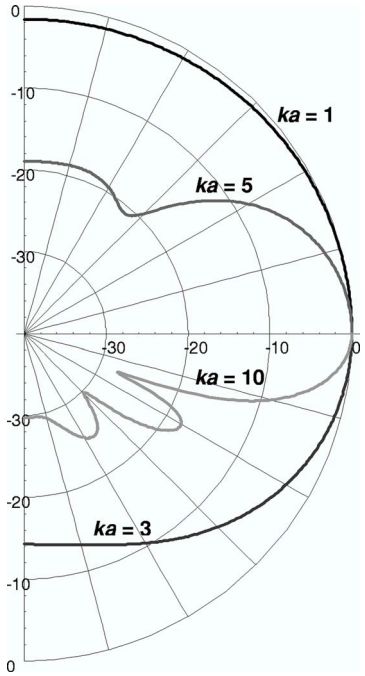


FIG. 7. Normalized far-field directivity function of the resilient disk.

$$\tilde{p}(r,\theta) = 2\int_0^{2\pi}\int_0^a g(r,\theta|w_0,\phi_0) \times \frac{\partial}{\partial z_0}\tilde{p}(w_0,z_0)|_{z_0=0+} w_0 dw_0 d\phi_0, \tag{61}$$

where a is the radius of the disk and $g(r,\theta|_{w_0,\phi})$ is the Green's function. The Green's function is defined by

$$g(r,\theta|w_0,\phi_0) = \frac{e^{-ikr_0}}{4\pi r_0}, \tag{62}$$

where

$$r_0 = \sqrt{r^2 + w_0^2 - 2rw_0\cos\phi_0\sin\theta}, \tag{63}$$

so that Eq. (61) becomes the Rayleigh integral. The Green's function of Eq. (62) can be expanded using the following formula, which is a special case of Gegenbauer's addition theorem:[17]

$$g(r,\theta|w_0,\phi_0) = -\frac{ik}{4\pi}\sum_{p=0}^{\infty}(2p+1)h_p^{(2)} \times (kr)j_p(kw_0)P_p(\cos\phi_0\sin\theta), \tag{64}$$

where $h_p^{(2)}$ is the spherical Hankel function.[19,14] Substituting Eqs. (64), (1), (2), and (47) in Eq. (61) enables the integrals in Eq. (61) to be separated as follows:

$$p(r,\theta) = -\frac{k^2\tilde{p}_0}{2\pi}\sum_{p=0}^{\infty}\left(p+\frac{1}{2}\right)h_p^{(2)}(kr)\sum_{m=0}^{\infty}\tau_m(m+1/2) \times \int_0^a\left(1-\frac{w_0^2}{a^2}\right)^{m-(1/2)} j_p(kw_0)w_0 dw_0 \times \int_0^{2\pi} P_p(\cos\phi_0\sin\theta)d\phi_0. \tag{65}$$

The Legendre function P_p can be expanded using the following addition theorem[17] (after setting one of the three angles in the original formula to $\pi/2$):

$$P_p(\cos\phi_0 \sin\theta) = P_p(0)P_p(\cos\theta) + 2\sum_{q=1}^{\infty}(-1)^q P_p^{-q}(0) \times P_p^q(\cos\theta)\cos q\phi_0, \tag{66}$$

which leads to the identity

$$\int_0^{2\pi} P_p(\cos\phi_0 \sin\theta)d\phi_0 = 2\pi P_p(0)P_p(\cos\theta) + 2\sum_{q=1}^{\infty}(-1)^q P_p^{-q}(0)P_p^q(\cos\theta) \times \frac{\sin 2\pi q}{q} = 2\pi P_p(0)P_p(\cos\theta), \quad \text{for integral } q. \tag{67}$$

It is also noted that[17]

$$P_{2p}(0) = \frac{\sqrt{\pi}}{p!\Gamma((1/2)-p)} = \frac{(-1)^p\Gamma(p+(1/2))}{\sqrt{\pi}p!}, \tag{68}$$

and

$$P_{2p+1}(0) = 0, \tag{69}$$

so that, after substituting Eqs. (67)–(69) in Eq. (65) and excluding the odd terms, only the radial integral remains as follows:

$$p(r,\theta) = -\sqrt{\pi}k^2\tilde{p}_0\sum_{p=0}^{\infty}\frac{(2p+(1/2))h_{2p}^{(2)}(kr)}{p!\Gamma((1/2)-p)} \times \sum_{m=0}^{\infty}\tau_m(m+1/2) \times \int_0^a\left(1-\frac{w_0^2}{a^2}\right)^{m-(1/2)} \times j_{2p}(kw_0)w_0dw_0P_{2p}(\cos\theta). \tag{70}$$

With the help of the following identity:[18]

$$\int_0^a\left(1-\frac{w_0^2}{a^2}\right)^{m-(1/2)} j_{2p}(kw_0)w_0dw_0 = \frac{\sqrt{\pi}}{k^2(m+(1/2))_{p+1}} \times \frac{p!}{\Gamma(2p+(3/2))}\left(\frac{ka}{2}\right)^{2p+2} \times {}_1F_2\left(p+1;p+m+\frac{3}{2},2p+\frac{3}{2};-\frac{k^2a^2}{4}\right), \tag{71}$$

the solution is given by

$$p(r,\theta) = -\tilde{p}_0\sum_{m=0}^{M}\tau_m\sum_{p=0}^{P}\frac{(-1)^p\Gamma(p+(1/2))h_{2p}^{(2)}(kr)P_{2p}(\cos\theta)}{\Gamma(2p+(1/2))(m+(3/2))_p} \times\left(\frac{ka}{2}\right)^{2p+2}{}_1F_2\left(p+1;p+m+\frac{3}{2},2p+\frac{3}{2};-\frac{k^2a^2}{4}\right). \tag{72}$$

This expansion converges providing $r \geqslant a$. Let an error function be defined by

$$\varepsilon_P(r,\theta) = \frac{|p_{2P}(r,\theta) - p_P(r,\theta)|}{|p_P(r,\theta)|}, \tag{73}$$

so that the pressure obtained with an expansion limit $2P$ is used as a reference. The calculations were performed using 30-digit precision with $P \approx 2ka$ and $M=P$. This produced values of ε typically on the order of 10^{-8} except at $r=a$ where it was on the order of 0.01. At $r=a$ and $\theta=\pi/2$ (i.e., the disk rim), there is a singularity which was avoided by the use of a small offset. For $ka<1$, values $M=P=2$ were used.

K. Near-field pressure when the distance from the center of the disk to the observation point is less than the diaphragm's radius

1. The near-field pressure as an integral expression

The simplest way to derive an expression for the immediate near-field pressure is to use the King integral. Applying the expression for $\tilde{A}_m$ in Eq. (47) to Eq. (8), the near-field pressure can be written

$$\tilde{p}(w,z) = \tilde{p}_0\sum_{m=0}^{\infty}\tau_m\Gamma\left(m+\frac{3}{2}\right)(I_{\text{Fin}}(m,w,z) - iI_{\text{Inf}}(m,w,z)), \tag{74}$$

where

$$I_{\text{Fin}}(m,w,z) = \frac{ka}{2}\int_0^k\left(\frac{2}{a\mu}\right)^{m-(1/2)}J_{m+(1/2)}(a\mu) \times J_0(w\mu)\frac{e^{-iz\sqrt{k^2-\mu^2}}}{\sqrt{k^2-\mu^2}}d\mu \tag{75}$$

and

$$I_{\text{Inf}}(m,w,z) = \frac{ka}{2}\int_k^{\infty}\left(\frac{2}{a\mu}\right)^{m-(1/2)}J_{m+(1/2)}(a\mu) \times J_0(w\mu)\frac{e^{-z\sqrt{\mu^2-k^2}}}{\sqrt{\mu^2-k^2}}d\mu. \tag{76}$$

2. Solution of the finite integral

Substituting $\mu=k\sqrt{1-t^2}$ in Eq. (75) in order to simplify the exponent yields

$$I_{\text{Fin}}(m,w,z) = \left(\frac{2}{ka}\right)^{m-(3/2)} \times \int_0^1\frac{J_{m+(1/2)}(ka\sqrt{1-t^2})J_0(kw\sqrt{1-t^2})}{(1-t^2)^{(m/2)+(1/4)}} \times e^{-ikzt}d\mu. \tag{77}$$

The Bessel functions in Eq. (77) can then be expanded using the following Lommel expansion:[20]

$$\frac{J_n(ka\sqrt{1-t^2})}{(1-t^2)^{n/2}} = \sum_{m=0}^{\infty}\left(\frac{ka}{2}\right)^m\frac{t^{2m}}{m!}J_{n+m}(ka), \tag{78}$$

which leads to

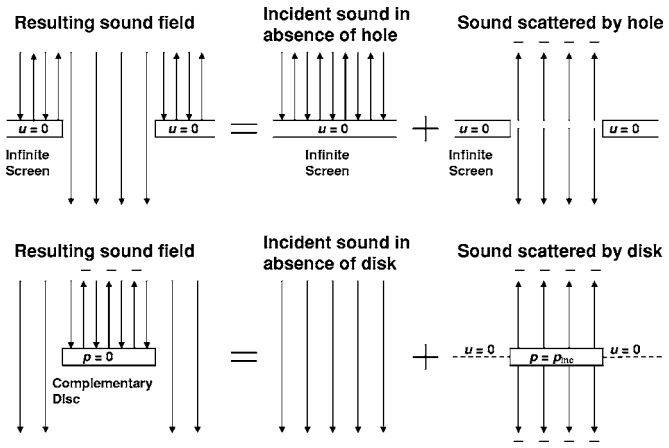


FIG. 8. Babinet's principle.

$$I_{\mathrm{Fin}}(m,w,z)=\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\left(\frac{ka}{2}\right)^{p-m+(3/2)}\times\left(\frac{kw}{2}\right)^{q}\frac{J_{p+m+(1/2)}(ka)J_q(kw)}{p!q!}\times\int_0^1 e^{-ikzt}t^{2(p+q)}d\mu. \tag{79}$$

The integral in Eq. (79) can be solved using the identity[17]

$$\int_0^1 e^{-ikzt}t^{2(p+q)}dt=\frac{\gamma(2p+2q+1,ikz)}{(ikz)^{2(p+q)+1}}, \tag{80}$$

where γ is the incomplete gamma function. Inserting Eq. (80) in Eq. (79) and truncating the summation limits gives the final solution to Eq. (75) as follows:

$$I_{\mathrm{Fin}}(m,w,z)=-\sum_{p=0}^{P}\sum_{q=0}^{Q}\frac{1}{p!q!(ikz)^{2(p+q)+1}}\left(\frac{ka}{2}\right)^{p-m+(3/2)}\times\left(\frac{kw}{2}\right)^{q}\times J_{p+m+(1/2)}(ka)J_q(kw)\gamma(2p+2q+1,ikz). \tag{81}$$

This solution converges for all $w>0$ and $z>0$, and was used for the region $0<w^2+z^2<a^2$, with a small offset at $z=0$. On the axis of symmetry ($w=0$), only the zeroth term of the expansion in q remains and the solution reduces to a single expansion,

$$I_{\mathrm{Fin}}(m,0,z)=-\sum_{p=0}^{P}\frac{1}{p!(ikz)^{2p+1}}\left(\frac{ka}{2}\right)^{p-m+(3/2)}J_{p+m+(1/2)}\times(ka)\gamma(2p+1,ikz), \tag{82}$$

which converges for $z>0$.

3. Solution of the infinite integral

For the finite integral, it was sufficient to expand both Bessel functions with the Lommel expansion. In the case of the infinite integral, this would lead to a solution which only converges for $z\geqslant w+a$. Therefore, in order to ensure convergence, one of the Bessel functions is to be expanded using Gegenbauer's summation theorem[17] as follows:

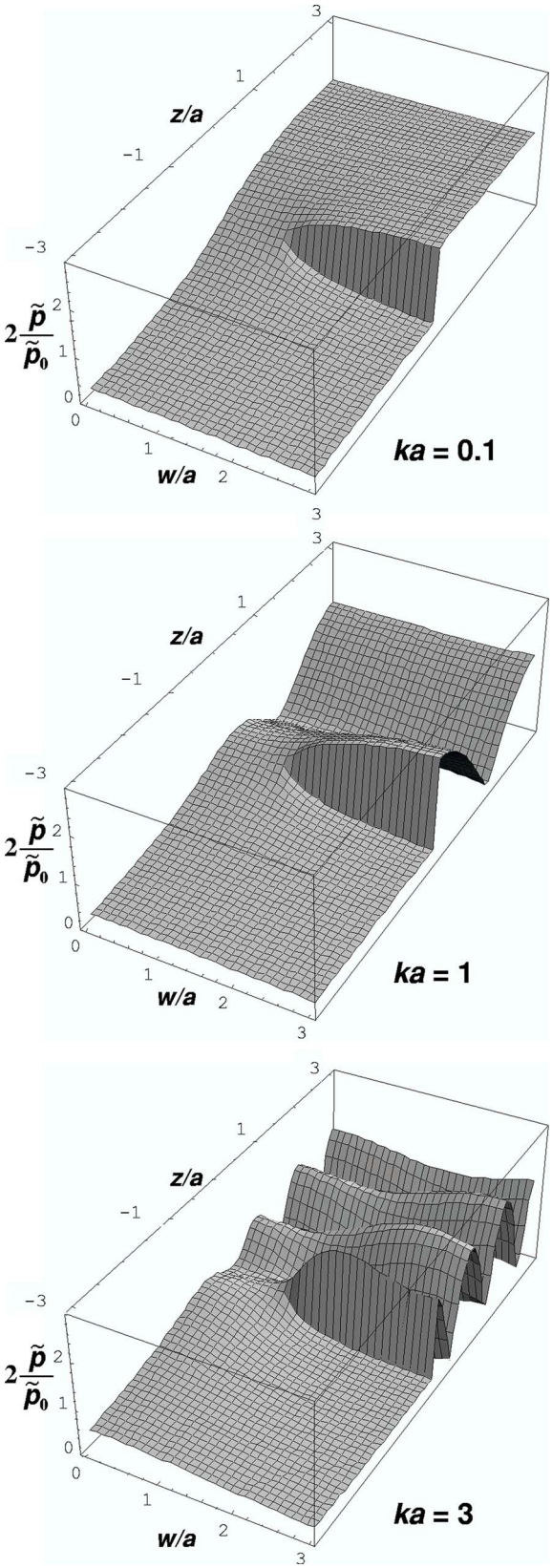


FIG. 9. Near-field pressure of a plane wave passing through a hole in an infinite screen for $ka=0.1$, 1, and 3.

$$\frac{J_n(kb\sqrt{t^2+1})}{(t^2+1)^{n/2}} = \Gamma(n)\left(\frac{2}{kbt}\right)^n \sum_{m=0}^{\infty} (m+n) J_{m+n}(kbt) \times J_{m+n}(kb) C_m^n(0), \quad (83)$$

where C_m^n is the Gegenbauer[17] polynomial given by

$$C_m^n(0) = \frac{\cos(m\pi/2)\Gamma(n+m/2)}{(m/2)!\Gamma(n)}, \quad (84)$$

where n is a positive real nonzero integer. Inserting Eq. (84) in Eq. (83) and noting that, due to the cosine term, all odd terms of the Gegenbauer polynomial are zero yields

$$\frac{J_n(ka\sqrt{t^2+1})}{(t^2+1)^{n/2}} = \left(\frac{2}{kat}\right)^n \sum_{m=0}^{\infty} \frac{(-1)^m}{m!}(2m+n)\Gamma(m+n) \times J_{2m+n}(ka) J_{2m+n}(kat). \quad (85)$$

Substituting $\mu = k\sqrt{t^2+1}$ in Eq. (76) in order to simplify the exponent yields

$$I_{\mathrm{Inf}}(m,w,z) = \left(\frac{2}{ka}\right)^{m-(3/2)} \int_0^{\infty} \frac{J_{m+(1/2)}(ka\sqrt{t^2+1})}{(t^2+1)^{(m/2)+(1/4)}} \times J_0(kw\sqrt{t^2+1}) e^{-kzt} dt. \quad (86)$$

Expanding $J_0(kw\sqrt{t^2+1})$ in Eq. (86) with Eq. (78) and $J_{m+(1/2)}[ka\sqrt{t^2+1})]$ with Eq. (85) yields

$$I_{\mathrm{Inf}}(m,w,z) = \left(\frac{2}{ka}\right)^{2m-1} \times \sum_{p=0}^{\infty}\sum_{q=0}^{\infty} \left(\frac{kw}{2}\right)^q \frac{(-1)^{p+q}(2p+m+(1/2))}{p!q!} \times \Gamma(p+m+(1/2)) J_{2q+m+(1/2)}(ka) J_q(kw) \times \int_0^{\infty} J_{2q+m+(1/2)}(kat) e^{-kzt} t^{2q-m-(1/2)} dt. \quad (87)$$

The integral in Eq. (87) can be solved using the following identity:[17]

$$\int_0^{\infty} e^{-kzt} J_{2p+m+(1/2)}(kat) t^{2q-m-(1/2)} dt = \frac{\Gamma(2p+2q+1)}{(k^2z^2+k^2a^2)^{q-(m/2)+(1/4)}} P_{2q-m-(1/2)}^{-2p-m-(1/2)}\left(\frac{z}{\sqrt{z^2+a^2}}\right), \quad (88)$$

so that, after truncating the summation limits, the final solution is given by

$$I_{\mathrm{Inf}}(m,w,z) = \left(\frac{2}{ka}\right)^{2m-1} \sum_{p=0}^{P}\sum_{q=0}^{Q} \frac{(-1)^{p+q}(2p+m+1/2)}{q!(k^2z^2+k^2a^2)^{q-(m/2)+(1/4)}} \times (p+1)_{m-(1/2)} \Gamma(2p+2q+1) \left(\frac{kw}{2}\right)^q \times J_{2p+m+(1/2)}(ka) J_q(kw) P_{2q-m-(1/2)}^{-2p-m-(1/2)} \times \left(\frac{z}{\sqrt{z^2+a^2}}\right). \quad (89)$$

This "paraxial" expression converges for $w^2 < a^2 + z^2$ and was used for the region $0 < w^2 + z^2 < a^2$, with a small offset at $z=0$. At $w \to a$ and $z \to 0$ (i.e., the disk rim), there is a singularity. On the axis of symmetry ($w=0$), only the zeroth term of the expansion in q remains and the solution reduces to a single expansion,

$$I_{\mathrm{Inf}}(m,0,z) = \left(\frac{2}{ka}\right)^{2m-1} \sum_{p=0}^{P} \frac{(-1)^p(2p+m+1/2)}{(k^2z^2+k^2a^2)^{-(m/2)+(1/4)}} \times (p+1)_{m-(1/2)} \Gamma(2p+1) \times J_{2p+m+(1/2)} \times (ka) P_{-m-(1/2)}^{-2p-m-(1/2)}\left(\frac{z}{\sqrt{z^2+a^2}}\right), \quad (90)$$

which converges for $z > 0$. The calculations for Eqs. (74), (81), and (89) were performed using 30-digit precision with $P \approx 2ka$, $Q \approx 4kw$, and $M = P$. This produced values of ε typically of the order of 10^{-8}. For $ka < 1$, values $M = P = 2$ and $Q = 4$ were used.

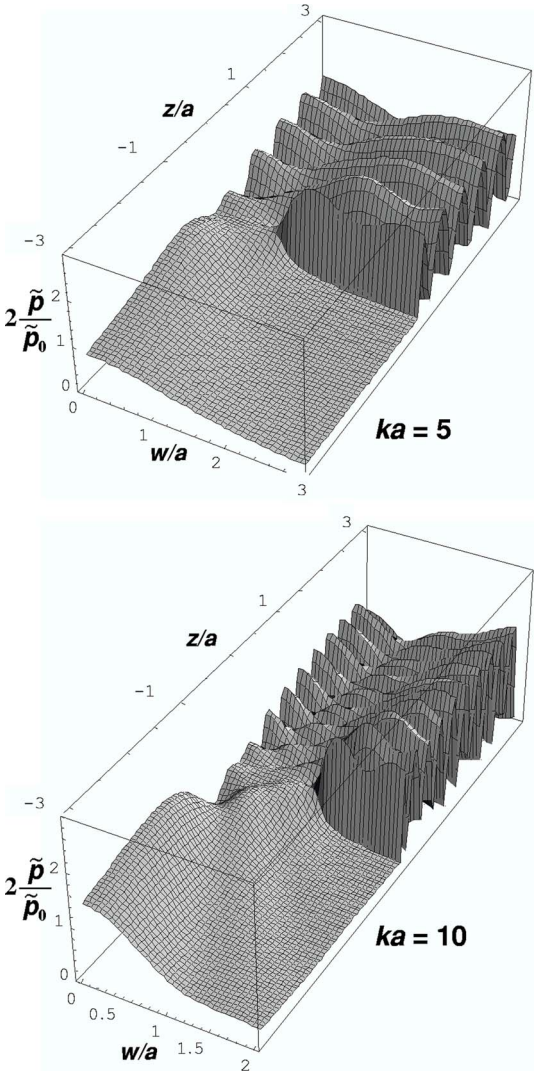


FIG. 10. Near-field pressure of a plane wave passing through a hole in an infinite screen for $ka=5$ and 10.

III. BABINET'S PRINCIPLE

Babinet's principle,[21] as developed by Bouwkamp,[22] states that the diffraction pattern resulting from the transmission of a plane wave through a hole in an infinite rigid screen (i.e., with infinite surface impedance) is equivalent to that produced by the scattering of the same incident wave by the complementary resilient disk. Furthermore, the scattered wave is identical to that produced if the disk itself were radiating in an infinite baffle, providing the surface pressure of the disk is equal to the pressure of the incident wave at the surface of the disk or hole in the absence of any obstacle. This is illustrated in Fig. 8. For clarity, the diagram portrays the scattering of a sound wave at some very high frequency where there is minimal diffraction. However, the principle applies at all frequencies. The resulting sound field is given by

$$\tilde{p}(\mathbf{r}) = \tilde{p}_{\mathrm{Inc}}(\mathbf{r}) + \tilde{p}_{\mathrm{Scat}}(\mathbf{r}), \tag{91}$$

where $\tilde{p}_{\mathrm{Inc}}(\mathbf{r})$ is the incident sound field in the absence of a hole or disk given in terms of the velocity potential $\tilde{\Psi}$ by

$$\tilde{p}_{\mathrm{Inc}}(z) = \begin{cases} -ik\rho c\tilde{\Psi}(e^{ikz} + e^{-ikz}), & \text{bright side of screen} \\ 0, & \text{dark side of screen} \\ -ik\rho c\tilde{\Psi}e^{ikz}, & \text{without disk (or screen)} \end{cases} \tag{92}$$

using Eq. (74) for $\tilde{p}_{\mathrm{Scat}}(w,z)$ (immediate near field), or Eq. (72) for $\tilde{p}_{\mathrm{Scat}}(r,\theta)$ (intermediate near field), or Eqs. (55) and (57) for $\tilde{p}_{\mathrm{Scat}}(r,\theta)$ (far field). Also, it can be stated that the volume velocity $\tilde{U}$ at the disk due to an incident wave is given by

$$\tilde{U} = y_{\mathrm{ar}}\tilde{p}_{\mathrm{Inc}}. \tag{93}$$

The radiation admittance y_{ar} is given by Eqs. (50)–(52). The results are shown in Figs. 9–11 for various values of ka. The pressure is plotted against the axisymmetric cylindrical ordinates w and z using

$$r = \sqrt{w^2 + z^2}, \tag{94}$$

$$\theta = \arctan w/z, \tag{95}$$

and the parameter P_{norm} is given by

$$P_{\mathrm{norm}} = \left| \frac{\tilde{p}(w,z)}{\rho c\tilde{u}_0} \right|. \tag{96}$$

IV. CONCLUSIONS

A set of solutions has been obtained for axisymmetric sources in infinite baffles which appear to be relatively compact and can be calculated fairly easily without numerical problems. As an example, the radiation characteristics of a resilient disk have been calculated. By applying Babinet's principle, this solution has also been used to calculate the pressure field of a plane wave passing through a circular hole in an infinite screen.

The radiation conductance (a.k.a. transmission coefficient), velocity magnitude and phase, and far-field directivity

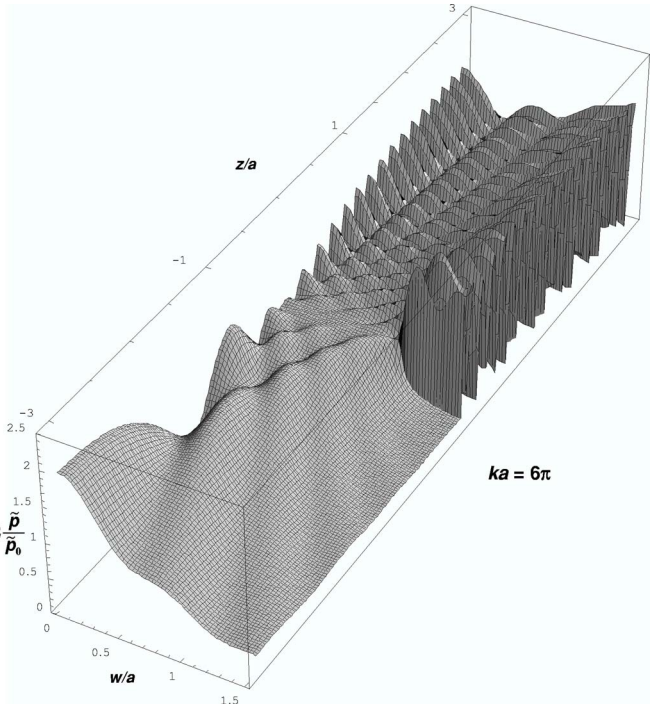


FIG. 11. Near-field pressure of a plane wave passing through a hole in an infinite screen for $ka=6\pi$.

function of the resilient disk in Figs. 5, 3, 4, and 7 respectively, show good agreement with the calculations of Spence[23,24] and the formulas provided here are intended to provide a simple alternative calculation method. For instance, Eqs. (44)–(46) and (50)–(52) for the radiation admittance appear to be relatively compact, thus eliminating the need for complicated spheroidal wave functions.

Although these derivations can be applied to membranes, plates, or shallow shells, to do so here would result in an overly long text. However, it would be interesting to see the cylindrical wave functions and pressure field expansions applied to fluid-structure coupled problems such as loudspeaker diaphragms, for example.

ACKNOWLEDGMENTS

The author would like to express his gratitude to N. Lobo for his invaluable advice in numerical matters and also to L. M. Kärkkäinen for his many useful suggestions.

[1] R. New, R. I. Becker, and P. Wilhelmij, "A limiting form for the near field of the baffled piston," J. Acoust. Soc. Am. **70**(5), 1518–1526 (1981).
[2] T. D. Mast and F. Yu, "Simplified expansions for radiation from a baffled circular piston," J. Acoust. Soc. Am. **118**(6), 3457–3464 (2005).
[3] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. **II**, pp. 107, 139, and 162.
[4] H. Stenzel, "Über die berechnung des schallfeldes einer kreisförmigen kolbenmembran (On the calculation of the sound field of a circular piston diaphragm)," E.N.T. **12**, 16–30 (1935).
[5] T. Hasegawa, N. Inoue, and K. Matsuzawa, "A new rigorous expansion for the velocity potential of a circular piston source," J. Acoust. Soc. Am. **74**(3), 1044–1047 (1983).
[6] H. Stenzel, "Über die Berechnung des Schallfeldes von kreisförmigen Membranen in starrer Wand (On the calculation of the sound field of circular membranes in rigid walls)," Ann. Phys. **4**, 303–324 (1949).
[7] L. V. King, "On the acoustic radiation field of the piezoelectric oscillator and the effect of viscosity on the transmission," Can. J. Res. **11**, 135–146

(1934).
[8]C. J. Bouwkamp, "A contribution to the theory of acoustic radiation," Philips Res. Rep. **1**, 251–277 (1945).
[9]A. Sommerfeld, "Die frei schwingende Kolbenmembran (The freely oscillating piston membrane)," Ann. Phys. **5**(42), 389–420 (1942/3).
[10]M. Greenspan, "Piston radiator: Some extensions of the theory," J. Acoust. Soc. Am. **65**(3), 608–621 (1979).
[11]A. O. Williams, "Acoustic field of a circular plane piston," J. Acoust. Soc. Am. **36**(12), 2408–2410 (1964).
[12]T. J. Mellow and L. M. Kärkkäinen, "On the sound field of an oscillating disk in an open and closed circular baffle," J. Acoust. Soc. Am. **118**(3), 1–15 (2005).
[13]T. J. Mellow and L. M. Kärkkäinen, "On the sound field of a membrane in an infinite baffle," on the CD ROM: *Audio Engineering Society Convention Papers, 118th Convention, Barcelona, 2005 May 28–31*, available from Audio Engineering Society Inc., Headquarters Office: 60 East 42nd Street, Room 2520, New York, NY 10165-2520.
[14]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 364–366, pp. 389.
[15]H. Lamb, "On the propagation of tremors over the surface of an elastic solid," Philos. Trans. R. Soc. London, Ser. A **203**, 1–42 (1904).
[16]A. Sommerfeld, "Über die Ausbreitung der Wellen in der drahtlosen Telegraphie (On the propagation of waves in wireless telegraphy)," Ann. Phys. **4**(28), 665–736 (1909).
[17]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed., edited by A. Jeffrey (Academic, New York, 2000), p. 671, Eq. (6.567.1); p. 900, Eq. (8.402); p. 901, Eqs. (8.405.1) and (8.405.2); p. 901, Eqs. (8.406.3) and (8.407.1); p. 900, Eq. (8.403); p. 902, Eq. (8.411.1); p. 930, Eq. (8.533.2); p. 963, Eq. (8.794.1); p. 959, Eq. (8.756.1); p. 887, Eq. (8.334.2); p. 342, Eq. (3.381.1); p. 930, Eq. (8.532.1); p. 980, Eqs. (8.930.1)–(8.930.7), p. 691, Eq. (6.621.1).
[18]S. Wolfram, *The MATHEMATICA BOOK*, 5th ed. (Wolfram Media, Champaign, IL, 2003). Symbolic computation by MATHEMATICA.
[19]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964), pp. 256, Eq. (6.1.22); pp. 437, Eq. (10.1.1).
[20]G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge University Press, London, 1944), pp. 141, Sec. 5.22, Eq. (5).
[21]J. Babinet, "Mémoires d'optique météorologique (Memoirs on meteorological optics)," C. R. Acad. Sci. **4**, 638–648 (1837).
[22]C. J. Bouwkamp, "Theoretical and numerical treatment of diffraction through a circular aperture," IEEE Trans. Antennas Propag. **AP18-2**, 152–176 (1970).
[23]R. D. Spence, "The diffraction of sound by circular disks and apertures," J. Acoust. Soc. Am. **20**(4), 380–386 (1948).
[24]R. D. Spence, "A note on the Kirchhoff approximation in diffraction theory," J. Acoust. Soc. Am. **21**(2), 98–100 (1949).

On the linewidth of the ultrasonic Larsen effect in a reverberant body

Richard L. Weaver[a] and Oleg I. Lobkis
Department of Theoretical and Applied Mechanics, University of Illinois at Urbana—Champaign, 216 Talbot Lab, 104 S. Wright Street, Urbana, Illinois 61801

(Received 24 March 2006; revised 20 April 2006; accepted 21 April 2006)

A simple ultrasonic feedback circuit is applied to a high-Q elastic body. At sufficient gain, an ultrasonic howl—or Larsen effect, ensues. It is a pure tone with an extraordinarily narrow spectrum. Theoretical estimates are constructed that predict linewidths proportional to the ratio between the spectral power density of the background noise and the intensity of the howl. In these experiments, this is of the order of nano-Hz and is unmeasurable. By augmenting the absorption and adding noise, the width is brought into a measurable regime, and the theory's prediction of a wider line is confirmed. It is speculated and demonstrated that these narrow lines permit high precision measurements of tiny changes in structures. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2205128]

PACS number(s): 43.25.Ts, 43.35.Yb, 43.38.Ew, 43.40.Le [TDM] Pages: 102–109

I. INTRODUCTION, THE LARSEN EFFECT

Unstable acoustic feedback is a familiar phenomenon. Danish physicist and electrical engineer Soren Absalon Larsen[1] is credited with first analyzing the effect. The usual audible experience is of a single tone, which appears when gain is above a critical level, waxes as the amplifier gain is adjusted above that threshold, and wanes or disappears when the gain is decreased. The tone varies in frequency as speaker-microphone distance is adjusted. Much attention has been focused on controlling and eliminating unstable feedback in systems of audio speakers and microphones,[2] and hearing aids, as well as structural control sensors and actuators.[3] There appears to have been little attention paid to the steady-state squeal (perhaps in part owing to its unpleasantness), or its amplitude and the steadiness of the tone. That steadiness is certainly affected by variations in ambient conditions and drift of electronics, but it is perhaps unclear what the limits to steadiness might be after those are controlled. Larsen himself observed stability to within 0.2% per hour.[1] Furduev[4] observed a comparable steadiness in an undersea environment. Here, we introduce an ultrasonic version of the Larsen effect for a reverberant solid and focus attention on the spectral width of the resulting single tone. We develop a theory for that width which predicts a dependence on background noise and a value, in our experiments, of the order of nanoHz. By artificially augmenting the noise, we bring the predicted width into a measurable regime and qualitatively confirm the theory. Applications are discussed.

II. ULTRASONIC FEEDBACK IN A REVERBERANT BODY

Consider the ultrasonic system pictured in Fig. 1(a). A piezoelectric sensor with useful sensitivity in the range from 100 to 2000 kHz, and a diameter of 3 mm is placed at an arbitrary position S on an irregular elastic body. We typically use[5,6] aluminum (shear and longitudinal wave speeds 3.1 and 6.15 mm/μs, respectively) of generic ray-chaotic shape and volume between 50 and 5000 cm^3. Relevant wavelengths lie in the range between 30 and 1.5 mm. The signal from the sensor at S is amplified by 40 dB, passed (optionally) through a filter to control the frequency range available, sampled by a digitizer, passed through an adjustable voltage divider, and amplified by a further 40 dB before being passed to a piezoelectric actuator at A that provides the ultrasonic source. A third transducer monitors the acoustic state of the elastic block at R. Turning on the amplifiers leads to a time-domain signal, such as that seen in Figs. 1(b) and 1(c), in which a signal (usually single tone) grows until the amplifiers saturate.

We define $\tilde{H}(\omega)$ to be the (Fourier transform of the) response of the transducer at A when an impulse is applied to the transducer at S in the absence of feedback. Figure 2(a) shows the spectrum $|\tilde{H}(\omega)|$ of the passive system. The broad peaks around 700, 1000, and 1500 kHz are characteristic of these transducers. The fact that the feedback tone chooses a frequency near 700 kHz is thus not surprising; this is in one of the main pass bands of the transducers. Sample steady-state spectra are shown in Fig. 2(b). We observe sharp lines (the frequencies of auto-oscillation) within the broad peaks of the ordinary passive spectrum. At modest gains, we typically observe a single line, but on occasion, especially at larger gain, the feedback shows many isolated frequencies. The precise position of the line(s) depends on the position of the actuator and sensor, the strength of gain, the temperature, and, sometimes hysteretically, the history of these parameters. At most levels of gain, there are many unstable frequencies, we nevertheless usually observe only one in the steady state. We suppose that it is the most rapidly growing

[a]Author to whom correspondence should be addressed. Electronic mail: r-weaver@uiuc.edu

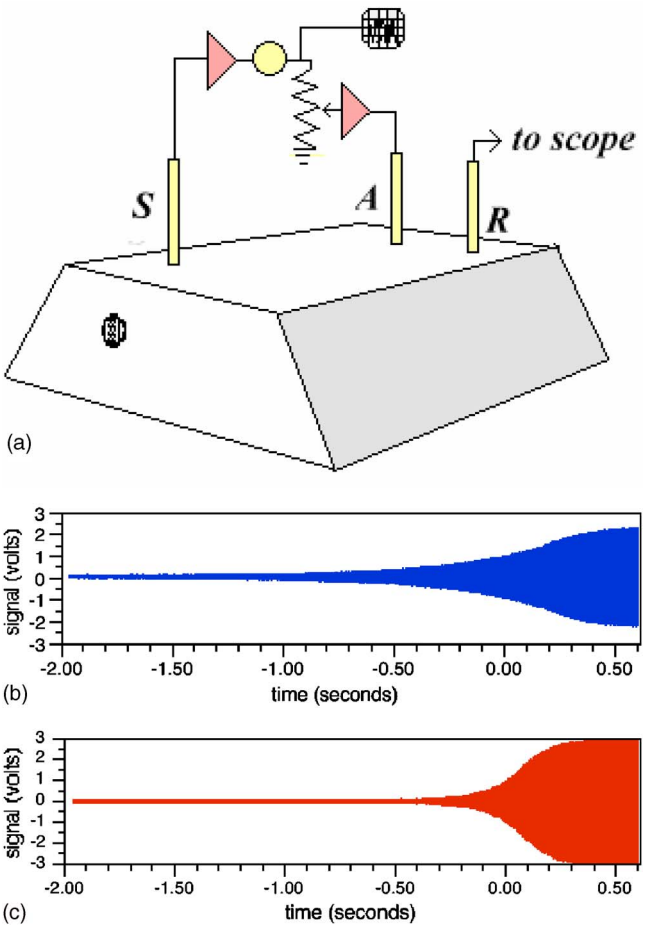


FIG. 1. (Color online) A piezoelectric transducer senses elastic waves at *S*. These signals are amplified, optionally filtered, optionally monitored by an oscilloscope, voltage-divided, and amplified again before being fed to a piezoelectric actuator at *A*. A third transducer monitors system response at *R*. At sufficient gain we observe (b) an unstable growth of a single frequency auto-oscillation at about 700 kHz. At slightly higher gain (c) the growth is faster.

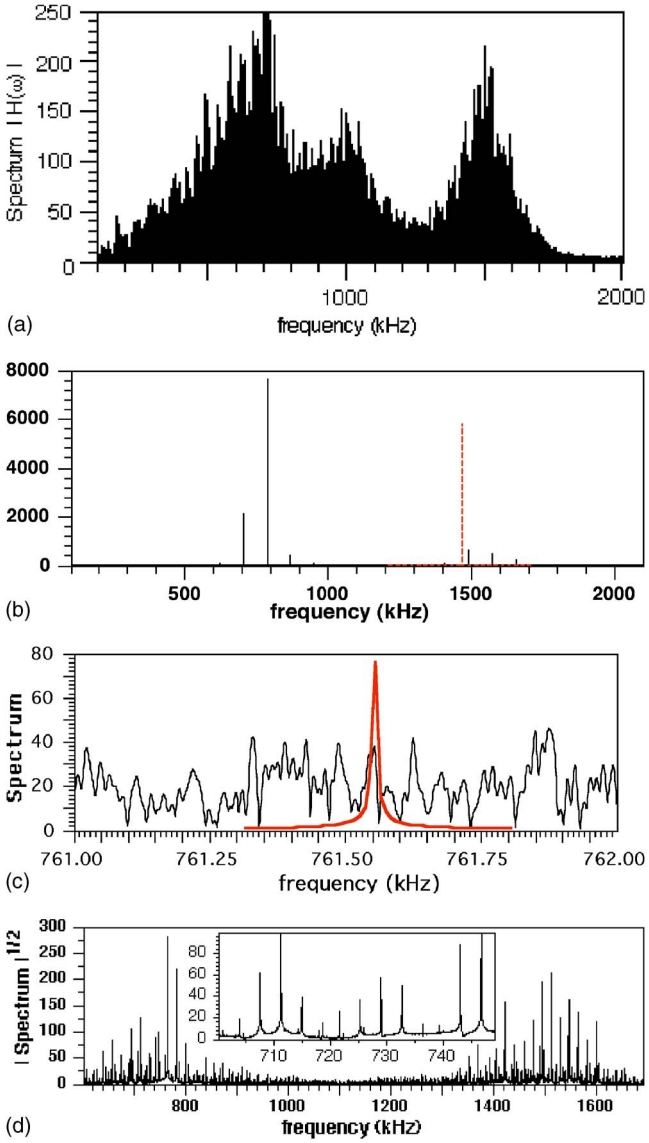


FIG. 2. (Color online) Qualitative behaviors of the feedback. Panel (a) shows the ordinary passive spectrum of a typical system like that of Fig. 1, obtained by Fourier analyzing the transient response of the transducer at *S* to an impulsive forcing at *A*. The broad peaks at 700, 1000, and 1500 kHz are characteristic of the transducers, as are the finer structures with spacings of the order of 43 kHz (related to a 23 microsecond round trip travel time up and down the transducer barrel). The finest structures (on a sub-kHz scale and not visible here) are due to the elastic block and correspond to a speck-lelike diffuse reverberant random response in this ray-chaotic body. Panel (b) shows typical simple feedback spectra consisting of one (dashed) or a few (solid) sharp lines. Panel (c) shows a fine resolution comparison of the ordinary spectrum and the corresponding feedback spectrum. It can be seen that the line (apparent width 10 Hz is artificially dictated by the amount of time captured) occurs near a local maximum of the ordinary spectrum. Panel (d) shows the (square root of) the spectrum of the auto oscillation for a case of high gain, a case with a multitude of lines. The inset shows the spectrum at finer resolution. These line spacings are dominated, but not exclusively, by a value of 3.7 kHz, suggesting a near-periodicity of 270 microseconds.

modes which survive. Linewidths are extraordinarily narrow, much narrower than would be expected based on the quality factor, Q (30,000), of the elastic body at these frequencies. We monitor the frequency of the auto-oscillation with a five-digit counter (a 0.1 s integration time at these frequencies). Line positions are stable to within our resolution (10 Hz) and constant over periods of minutes, depending only on temperature.

Figure 2(c) shows a fine scale overlay of the feedback spectrum and the ordinary passive spectrum $|\tilde{H}(\omega)|$. On this scale, the ordinary spectrum reveals features related to the reverberant cavity. At these frequencies, the modes of the passive cavity are closely spaced ("overlapped") and cannot be resolved, thus the features in $|H(\omega)|$ are those of a complex Gaussian process. In nuclear physics, they are called Ericson fluctuations;[7] they are also discussed in acoustics.[7] It may be seen that auto-oscillation tends to occur at a local maximum of the ordinary spectrum, but not precisely. Also, the line is not necessarily near the strongest of the peaks.

Figure 2(d) shows a spectrum obtained at high gain. The many lines (all narrow to within our resolution of 10 Hz) are perhaps ascribable to nonlinear mixings between at least two primary lines near 750 and 1500 kHz. At these acoustic amplitudes, nonlinearity is confined to the amplifiers and their saturation; it is not in the wave propagation.

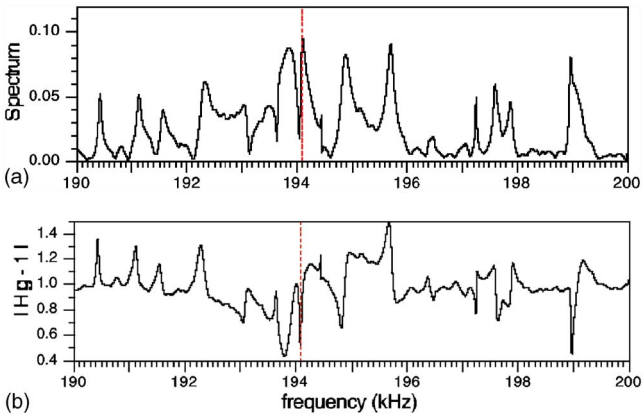


FIG. 3. (Color online) The ordinary spectrum $|H|$, and feedback tone at 194.1 kHz, (a). The Larsen frequency is close to a maximum of the ordinary spectrum, i.e., close to an eigenmode of the specimen. The line position is compared (b) with measured $|Hg-1|$, showing that the tone occurs near a zero of $Hg-1$.

The theory for the low amplitude linear dynamics of the structure below saturation may be described by

$$V_S(t) = X_S(t) \otimes G(S,A,t) \otimes X_A(t) \otimes V_A(t),$$

$$H_{SA}(t) = X_S(t) \otimes G(S,A,t) \otimes X_A(t), \tag{1}$$

$$V_A(t) = g(t) \otimes V_S(t) + N(t),$$

where $V_S(t)$ is the signal out of the sensor at S, and $V_A(t)$ is the signal into the actuator at A. They differ by the gain $g(t)$ and the amplifier noise $N(t)$. The symbol $\otimes$ represents a temporal convolution. X_S and X_A are transducer transfer functions and G is the elastodynamic Greens function between points A and S.[5,6,8]

Neglecting the noise $N(t)$ and seeking $V(t)$ in the form $\exp\{i\lambda t\}$, we find that freely decaying or growing oscillations correspond to the solutions λ to

$$\tilde{H}_{SA}(\lambda) = \int H_{SA}(t)\exp\{-i\lambda t\}dt = \tilde{X}_S(\lambda)\tilde{G}_{SA}(\lambda)\tilde{X}_A(\lambda)$$

$$= 1/\tilde{g}(\lambda). \tag{2}$$

The tilde represents a Fourier transform. At zero gain, Eq. (2) implies solutions λ wherever $\tilde{H}(\lambda)$ is infinite. These λ have a positive imaginary part, and correspond to freely decaying natural oscillations. For great enough gain, the fastest growing mode is that solution λ, which has the most negative imaginary part.

In Figs. 3, we plot measured $|\tilde{H}(\omega)|$ and, based on separate and awkward measurements of g, $|\tilde{H}(\omega)\tilde{g}(\omega)-1|$. The plot is for a gain setting g just above the minimum for instability, and for low frequencies as assured by the use of a low-pass filter. Modal density is about 1 mode/kHz in this specimen at 200 kHz; thus, the peaks in $|\tilde{H}(\omega)|$ correspond to eigenmodes of the body. When gain is at the threshold, $\mathrm{Im}\,\lambda=0$ and the free vibration condition (2) reduces to $\tilde{H}(\omega)\tilde{g}(\omega)=1$. Figure 3(a) shows $|\tilde{H}(\omega)|$ itself, and shows the peak centered at 194.106 kHz that appears to be related to the line at 194.095 kHz. (At these low frequencies, our frequency resolution is better; it is governed by the 1 s capture time of our 500,000 word digitizer when capturing at 0.5 MSa/s). Figure 3(b) shows $|\tilde{H}(\omega)\tilde{g}(\omega)-1|$, which has a sharp minimum at the frequency of the feedback instability. This behavior is also observed at high frequency, where modal overlap is large and the fluctuations in the ordinary spectrum are not modes but Ericson fluctuations.[7]

Theory (2) predicts $|\tilde{H}(\omega)\tilde{g}(\omega)-1|$ to be zero at the feedback line. That it only approaches zero in the plot is perhaps due to the gain being somewhat greater or less than threshold (by about 4 dB), or to an inaccurate measurement of g. These results are consistent with the simple linear instability theory, Eq. (2), but thorough confirmation will require a more careful control and measurement of g. A nonlinear model that included saturation would be more complex and would require a deeper understanding of the amplifiers; it is not attempted here.

These behaviors are found to occur also in disordered multiply scattering structures and in lower-Q systems. We have examined two-dimensional (2D) multiply scattering diffusing and localizing systems,[9] and a simple two-room system[10] that localizes at low frequency, and a block of Plexiglas; behaviors are similar. These structures have a higher absorption than do the ballistic aluminum bodies and correspondingly require higher gain before they autooscillate. In all of these systems, the lines remain more narrow than we can resolve.

III. THEORY FOR THE LINEWIDTH

The extremely narrow lines that are observed can be explained by appealing to an argument similar to that advanced by Schawlow and Townes[11] for the linewidth of a maser or laser. In this theory, the linewidth is governed by the strength of the background noise. (For a laser, the effective noise is related to the strength of the spontaneous emissions.) In our case, and as confirmed in separate measurements, the noise is dominated by the noise from the amplifier.

When the sensor S of Fig. 1(a) is de-attached from the acoustic system, the system is then excited at A chiefly by noise $N(t)$ from the amplifier. The signal received at R is then given by

$$V_R^{noS}(t) = X_R(t) \otimes G_{RA}(t) \otimes X_A(t) \otimes N(t) = H_{RA} \otimes N(t), \tag{3}$$

where G_{RA} is the (passive medium) Green's function between the points A and R. We take the noise to have autocorrelation

$$R_N(\tau) = \langle N(t)N(t+\tau)\rangle, \tag{4}$$

and spectral power density

$$\tilde{R}_N(\omega) = \int R_N(\tau)\exp\{-i\omega\tau\}d\tau. \tag{5}$$

This leads to an expression for the spectral power density of the signal at R

$$\tilde{R}_R^{noS}(\omega) = \int R_R^{noS}(\tau)\exp\{-i\omega\tau\}d\tau = |\tilde{H}_{RA}(\omega)|^2\tilde{R}_N(\omega), \tag{6}$$

where

$$R_R^{noS}(\tau) = \langle V_R^{noS}(t)V_R^{noS}(t+\tau)\rangle \tag{7}$$

When the sensor is attached to the structure the signal detected at S is amplified by gain g and passed to the actuator at A. The signal V_A applied to the actuator is then given, not by N as in Eq. (3), but self-consistently by

$$V_A^{withS}(t) = g(t) \otimes H_{SA}(t) \otimes V_A^{withS}(t) + N(t) \tag{8}$$

In the frequency domain, the solution is

$$\tilde{V}_A^{withS}(\omega) = [1 - \tilde{g}(\omega)\tilde{H}_{SA}(\omega)]^{-1}\tilde{N}(\omega). \tag{9}$$

The signal received at R is then

$$\tilde{V}_R^{withS}(\omega) = \tilde{H}_{RA}(\omega)[1 - \tilde{g}(\omega)\tilde{H}_{SA}(\omega)]^{-1}\tilde{N}(\omega) \tag{10}$$

The autocorrelation of $V_R(t)$ is

$$R_R^{withS}(\tau) = \langle V_R^{withS}(t)V_R^{withS}(t+\tau)\rangle, \tag{11}$$

and its spectral power density is

$$\tilde{R}_R^{withS}(\omega) = \int R_R^{withS}(\tau)\exp\{-i\omega\tau\}d\tau$$

$$= |\tilde{H}_{RA}(\omega)|^2\tilde{R}_N(\omega)/|[1 - \tilde{g}(\omega)\tilde{H}_{SA}(\omega)]|^2. \tag{12}$$

We now take g to be adjusted so that the system is on the verge of instability, but still linear. In this case, $g(\omega)H_{SA}(\omega)$ is unity at the Larsen frequency $\omega=\omega^*$, and varies near there like $gH \approx 1+(\omega-\omega^*+i\delta)T^*$. We remark that δ is the (half width at half power) width of the feedback line in radians per unit time, and is tiny; we also remark that ω^* is not necessarily one of the natural frequencies of the passive body. T^* is an as-yet undetermined quantity with dimensions of time; it may be complex.

The intensity of the screech is

$$I = \langle V_R^{withS}(t)^2\rangle = R_R^{withS}(\tau=0) = \frac{1}{2\pi}\int \tilde{R}_R^{withS}(\omega)d\omega,$$

$$= \frac{1}{2\pi}|\tilde{H}_{RA}(\omega^*)|^2\tilde{R}_N(\omega^*)|T^*|^{-2}\int d\omega/|\omega-\omega^*+i\delta|^2, \tag{13}$$

$$= \frac{1}{2}|\tilde{H}_{RA}(\omega^*)|^2\tilde{R}_N(\omega^*)|T^*|^{-2}/\delta.$$

Thus the Larsen linewidth is

$$\delta = \tilde{R}_R^{noS}(\omega^*)/2I|T^*|^2. \tag{14}$$

It is proportional to the spectral density of noise in the absence of feedback, inversely proportional to the intensity of the feedback, and inversely proportional to the square of $|T^*|$.

We may estimate the as yet undetermined quantity T^* by recognizing

$$T^* = d[\tilde{g}(\omega)\tilde{H}_{SA}(\omega)]/d\omega|_{\omega^*}. \tag{15}$$

Inasmuch as the frequency dependence in $\tilde{g}(\omega)\tilde{H}_{SA}(\omega)$ is dominated by that of H (time delays in the amplifier are modest compared to those in the acoustics), we may treat g as approximately constant, and equal to $1/\tilde{H}_{SA}(\omega^*)$. Thus,

$$T^* = d\{\ln \tilde{H}_{SA}(\omega)\}/d\omega|_{\omega^*}. \tag{16}$$

This quantity could be determined from laboratory measurements of H. A theoretical estimate may be constructed by considering the mean acoustic delay[12] of a passive signal between S and A in the absence of feedback,

$$t_{\text{delay}} \equiv \int_0^\infty tH_{SA}(t)^2dt \Big/ \int_0^\infty H_{SA}(t)^2dt. \tag{17}$$

In a reverberant room, where the mean energy H^2 is distributed uniformly in space and decays in time, such as $\exp\{-\omega t/Q\}$,[13,5,6] this is the absorption time, $t_{\text{delay}}=Q/\omega$. In an unbounded homogeneous system, it is the propagation time between sensor and actuator. In terms of the Fourier transform of H, it is

$$t_{\text{delay}} = -i\int_0^\infty \tilde{H}_{SA}^*(\omega)\left\{\frac{d\tilde{H}_{SA}(\omega)}{d\omega}\right\}d\omega \Big/ \int_0^\infty |\tilde{H}_{SA}(\omega)|^2d\omega,$$

$$= -i\int_0^\infty |\tilde{H}_{SA}(\omega)|^2\left\{\frac{d\ln\tilde{H}_{SA}(\omega)}{d\omega}\right\}d\omega \Big/ \int_0^\infty |\tilde{H}_{SA}(\omega)|^2d\omega. \tag{18}$$

The delay time is a $|H|^2$-weighted frequency average of the quantity $d\ln H/d\omega$, whose value at ω^* is T^*. We therefore make the estimate

$$T^* = it_{\text{delay}}. \tag{19}$$

In a reverberant body, this simplifies to $|T^*|^2=Q^2/\omega^{*2}$.

The Larsen frequency is presumably a special frequency, and T^* may be poorly approximated by its average. We therefore take estimate (19) as tentative.[14]

The estimate for the Larsen line width in a reverberant body is then

$$\delta = \omega^{*2}\tilde{R}_R^{noS}(\omega^*)/2IQ^2. \tag{20}$$

In an unbounded homogeneous system, where c is wave speed and r is distance between sensor and actuator, the estimate is

$$\delta = c^2\tilde{R}_R^{noS}(\omega^*)/2Ir^2. \tag{21}$$

The above estimates apply to linewidth at gains just below the nonlinear threshold. For higher gains, the amplifier will saturate, and we may expect some modification. Inasmuch as the saturation eliminates amplitude fluctuations but not phase noise, δ would be reduced by an additional factor of $(1/2)$.[15] The laser literature also commonly introduces an additional, "Petermann"[15] factor related to complexity of the normal modes and the different source of noise in a laser.

Prediction (20) is extraordinary. Consider systems like those we observe in the lab, with noise spectral density $\tilde{R}_R^{noS}$ of the order of 10^{-5} V^2 per 2π Mrad/s, or 1.6×10^{-11} V^2 s.

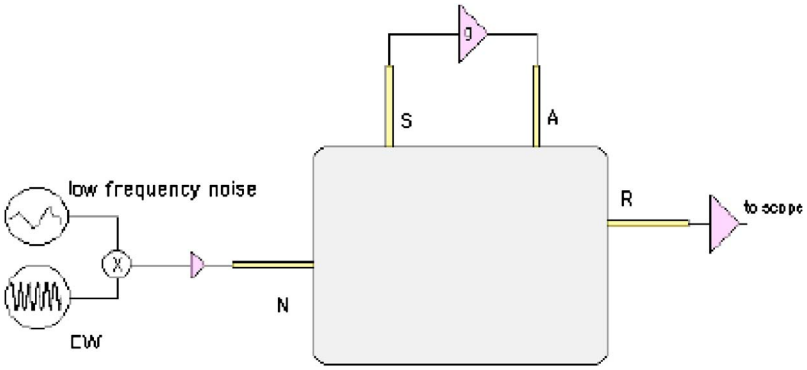


FIG. 4. (Color online) The ultrasonic feedback circuit of Fig. 1(a) is modified by adding narrow-band noise at a fourth transducer *N*.

This level corresponds to a noise spectrum distributed uniformly over a MHz with a root-mean-square (rms) amplitude of 3.2 mV. We take the intensity of feedback to be 5 V^2, at a frequency of the order of MHz, and in a body with a Q of 32,000. The theory then predicts a linewidth $\delta/2\pi = 0.01\ \mu$Hz. That is three cycles per century. This may be compared with a natural mode width $\omega/2\pi Q \approx 31$ cycles per second.

Predictions are less striking in an open system, or a system such as a large room that may be dominated by direct propagation, or a system with lower Q, but linewidth can nonetheless be far less than one might have guessed.

IV. LABORATORY MEASUREMENTS OF LINEWIDTH

We attempt to confirm this picture by lowering the quality factor Q and adding enough noise to significantly widen the Larsen line. A fourth transducer is driven by a narrow band noisy signal, centered on a frequency we set close to that of the feedback, see Fig. 4. This does not precisely correspond to the above theory where the noise entered at *A*, but it will suffice for qualitative comparison with theory. A steel bar with tape attached to many of its surfaces is used instead of the polished aluminum block of Sec. II, thus assuring a lower Q. Q was measured to be 500.

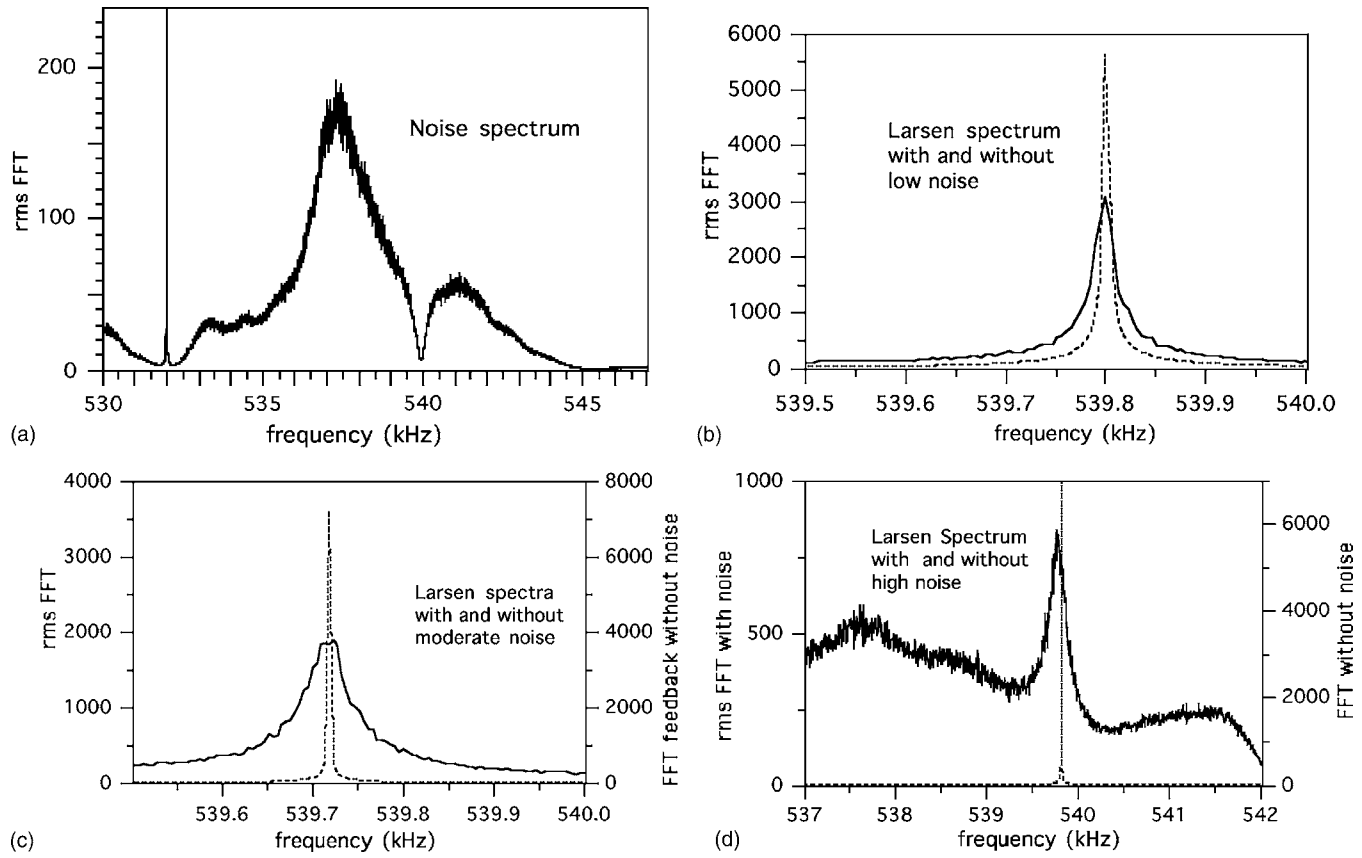


FIG. 5. (a) The noise spectrum obtained when the feedback is turned off. The line at 532 kHz is due to the continuous-wave source; it has leaked through the multiplier and is not part of the noise. The null at 540 kHz is due to a near zero of the transfer function $H_{NR}(\omega)$; the feedback circuit itself does not see the null. (b) The Larsen spectrum in the low-Q steel sample in absence of noise (dashed curve). has a finer width than can be resolved with the 200 ms signal capture length. In the presence of weak noise at the level illustrated in (a) the full-width-at-half-power Larsen linewidth is about 20 Hz (solid curve). (c) A higher noise level than that of (b) shows a further increase in Larsen line width. Full width at half power is now about 50 Hz. (d) At the highest noise level examined, the Larsen tone has a width of about 200 Hz. Formula (20) predicts, if we apply it even though it erroneously presumes the noise is from the amplifier, a width $\delta/2\pi = 1$ kHz.

Figure 5 shows the rms Larsen and noise spectra obtained at R for various levels of applied noise. Each spectrum was obtained by collecting a hundred 200 ms waveforms, Fourier transforming, squaring, averaging, and taking the square root. Figure 5(a) shows the noise process at R obtained by mixing a 532 kHz continuous wave form with audio-frequency noise, introducing the result at N, and disconnecting the feedback circuit. Figure 5(b) shows the spectrum of the signal at R with feedback, with and without the low noise level illustrated in Fig. 5(a). Figures 5(c) and 5(d) show the Larsen spectra with and without higher levels of noise. As predicted by theory, the spectrum of the feedback now has a finite measurable width. As predicted, the Larsen tone widens as the noise is increased.

Detailed quantitative investigation of prediction (20) is deemed outside the scope of the present work. Such an effort would require attention to uncertainties in the factor $|T^*|$, and modification of the experiment so as to introduce the noise at A rather than N. Nevertheless, after ignoring these caveats, we observe [Fig. 5(d)] that formula (20) is correct within a factor of 5.

V. APPLICATIONS

In an attempt to illustrate a potential application for this kind of measurement, we investigate the sensitivity of the Larsen tone to minor changes in the structure. Furduev[4] has made a similar suggestion for monitoring changes in the ocean. The Cramer-Rao bounds[16] (see Appendix), for the accuracy with which frequency may be determined, lie well below the nominal spectral precision $\Delta f = 1/T$ obtainable with a record length T. It should therefore be possible to discern frequency with much more accuracy than $1/T$. In order to determine changes in line position with high precision, we propose measuring that frequency, not by means of the peak in the spectrum, but rather by a nonlinear least-squares fit to a model of a single sinusoid plus noise. Our numerical simulations have shown that the frequency of a single sinusoid may be determined with high accuracy by this method. The approach here may be contrasted with that of Furduev's "regenerative monitoring," or with phase locked loops,[17] coda wave interferometry,[23,24,18] diffusing wave spectroscopy,[25] or ultrasonic sing around,[26] all useful for monitoring tiny changes in structures with high precision.

Figure 6 shows Larsen frequencies as determined by nonlinear least squares from a succession of 100 records over a period of 4 min. Each record had 200 ms length, and was captured at 2.5 MSa/s. The system of Fig. 1(a) consisting of cables, transducers, and aluminum block was placed under a Plexiglas dome to minimize air currents and corresponding temperature fluctuations. In a period of 4 min, the frequency dropped smoothly [Fig. 6(a)] from 537.20895 kHz to 537.20765 kHz. This slow change is ascribed to a slow increase in temperature.[18] The residual of its fit to a quadratic function of time suggests, Fig. 6(b), that our measurement has a limiting precision of about 1 mHz, but is also corrupted by some short term (and as yet not understood) irregular fluctuations of order 5 mHz.

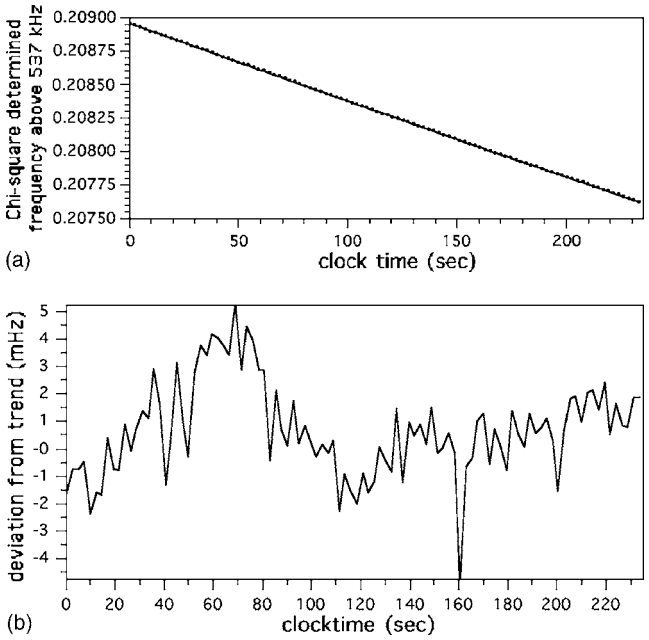


FIG. 6. Chi-square-determined frequencies (isolated symbols) fit well to a quadratic function of time (solid curve). Over 235 s, the frequency changes by 1.27 Hz, corresponding to 0.000248%, and a 0.0088 °C increase in temperature.[18] (b) Difference between the frequency as determined by chi-square minimization and its trend as seen in (a). Residuals are of order 5 mHz; fluctuations are of order 1 mHz. The Cramer-Rao bound of 20 μHz is not achieved.

In Fig. 7, we show similar data taken before during and after addition of a small drop (0.02 gm) of water to the 2.0 kg aluminum block. The water drop at $t = 44$ s is evident in the Larsen frequency, which changes there by a bit less than a part in 10^5. Changes attributable to slow temperature variations are observed here also. After regressing out those trends, we observe short-term fluctuations of the order of 50 mHz, barely discernable in the plot but far greater than

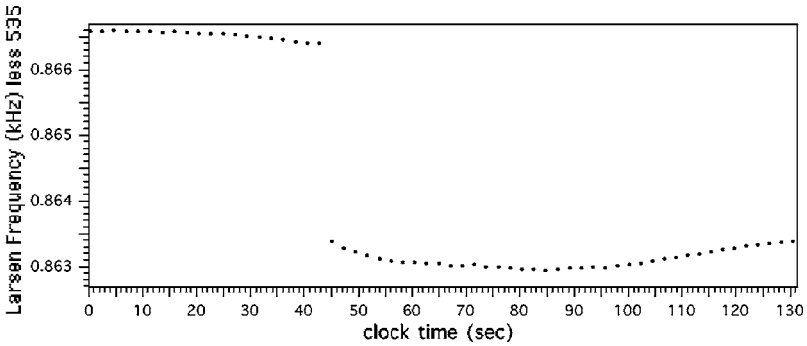


FIG. 7. Larsen frequency as determined by chi-square minimization before during and after a 20 mg drop of water is added to the aluminum block. The adding of the water drop is evident. Short-term trends, beyond those attributable to slow temperature changes and the water drop, are of the order of 50 mHz and barely distinguishable in this plot. The Larsen frequency returns to the vicinity of 535.866 kHz after the water drop is removed.

those seen in Fig. 6(b). The greater fluctuations here are not understood, but are possibly due to our abandonment of the Plexiglas dome in order to be free to add the water drop. Additional studies (not shown) confirm the hypothesis that the dome reduces the fluctuations. Remaining fluctuations are uncorrelated and of the order of mHz. Removal of the dome generates short-term fluctuations of 10's of mHz ($\sim 10^{-4}$ °C) over periods of tens of seconds Their origin is not known, but is possibly due to random air currents.

We conclude that Larsen frequencies can be monitored with high precision, and that changes may be associated with changes in acoustic properties. There is substantial interest in ascertaining the elastic properties of modern materials, for example, thin films,[19] or high-temperature superconductors.[20] Resonant ultrasound spectroscopy[19,20] is proposed for that purpose, as are other ultrasonic probes.[21] Similarly, we note wide interest in structural health monitoring by measurement of natural frequencies[22] or long dwell time diffuse wave fields,[23] and correlation of changes with damage. It may be that this kind of ultrasonic circuit would be a useful complement to ultrasonic nondestructive evaluation, or if scaled down in frequency to structural health monitoring.

VI. SUMMARY

A laboratory ultrasonic version of unstable acoustic feedback familiar from audio acoustics has been shown to exhibit extraordinary stable narrow tones. A theory for that linewidth suggests that it is chiefly governed by the quality factor of the ultrasonic body and the noise in the circuit; we calculate our tone's width to be of the order of nanoHz. By augmenting the noise and lowering the Q, we increase the width and qualitatively confirm the predictions of the theory. We measure line positions to a precision of the order of a few mHz (a part in 10^8), and demonstrate that they may be used to monitor tiny changes in a structure. It is intriguing to speculate on the ultimate limits with which small changes might be detectable using a dedicated system consisting of a well-stabilized acoustic environment and low noise stable electronics.

ACKNOWLEDGMENTS

This work was supported by the NSF CMS 05-28096. We thank Doug Mast for bringing the Cramer-Rao bounds to our attention, referees for helpful comments and for bringing Ref. 4 to our attention, and Finn Jacobsen and Alexey Yamilov for discussions.

APPENDIX, CRAMER-RAO BOUNDS

The limiting precision with which a Larsen frequency could be determined is given by a Cramer-Rao (CR) bound. CR bounds are given in terms of the Fisher information matrix:

$$F_{ij} = \frac{1}{\sigma^2}\sum_{n=1}^{N} \frac{\partial s_n}{\partial p_i}\frac{\partial s_n}{\partial p_j}, \tag{A1}$$

which is written in terms of the dependence of the model $s_n(p_i)$ for the nth datum on the parameters p to be estimated. In our case, $s_n = A\cos\omega t_n + B\sin\omega t_n$ and there are three parameters A, B, and ω to be estimated. Here, σ^2 is the variance of the (assumed uncorrelated Gaussian random) noise that corrupts the measurement of the nth datum. The CR bound, Δp_j, on the precision with which a parameter p_j may be estimated, is given in terms of its variance

$$\{\Delta p_j\}^2 = \text{var}\, p_j = [F^{-1}]_{jj}. \tag{A2}$$

In our case, this becomes

$$\text{var}\,\omega = 12\sigma^2/NIT^2, \tag{A3}$$

where N is the number of data points in the uniformly spaced record, T is its length, and $I = (1/2)(A^2 + B^2)$ is the signal intensity; σ^2/I is therefore a signal-to-noise ratio. The assumed uncorrelated Gaussian noise statistics correspond to a white-noise spectral power density $R_{\text{noise}}(\omega) = \sigma^2 T/N$. This allows comparison to the Larsen tone's intrinsic width δ of Eq. (20):

$$\Delta\omega = (\text{var}\,\omega)^{1/2} = [6\delta/T]^{1/2}[Q/\omega T]. \tag{A4}$$

We conclude that the CR bound for measurement of a Larsen frequency is the product of two quantities. One is the geometric mean of its intrinsic width δ and a quantity $6/T$ comparable to nominal spectral precision $2\pi/T$. The other is $Q/\omega T$ (typically between 10^{-3} and 10^{-1} in our measurements) equal to the ratio of the decay time Q/ω and the Larsen signal capture length T. Typical values for $\Delta\omega/2\pi$ in the aluminum samples are of the order of 20 μHz. Thus, an estimate of the Larsen frequency will be much less precise than its intrinsic width δ, but much more precise than the nominal spectral precision $2\pi/T$.

Achieving the CR bound in practice is unlikely, as the assumption of uncorrelated Gaussian noise is not necessarily met. It is also not clear that our clock has sufficient precision.

[1] A. Larsen, "Ein akustischer Wechselstromerzeuger mit regulierbarer Periodenzahl für schwache Ströme," {"An acoustic alternating-current generator for weak currents, with adjustable frequency"} Elektrotech. Z., ETZ **32**, 284–285 (1911); A. Kjerbye Nielsen, "'Larsen-effekten' og den fØrste elektriske tonegenerator baseret herpa," {"The 'Larsen effect' and the first electrical pure-tone generator based on this effects."} Teleteknik **3**, 140–148 (1984); D. Barbaro, "Self-starting acoustic oscillations in closed spaces," Alta Freq. **27**, 472–485 (1958).

[2] J. L. Nielsen and U. P. Svensson, "Performance of some linear time-varying systems in control of acoustic feedback," J. Acoust. Soc. Am. **106**, 240–254 (1999); M. Romanin and C. Trestino, "An acoustic feedback canceller," WSEAS Transactions on Circuits and Systems **2**, 851–857 (2003).

[3] J. M. Kates, "Constrained adaptation for feedback cancellation in hearing aids," J. Acoust. Soc. Am. **106**, 1010–1019 (1999); S. M. Kim, S. J. Elliott, and M. J. Brennan, "Decentralized control for multichannel active vibration isolation," IEEE Trans. Control Syst. Technol. **9**, 93–100 (2001).

[4] A. Furduev, "Acoustic monitoring of the sea medium variability: experimental testing of new methods," Acoust. Phys. **47**, 361–268 (2001).

[5] R. L. Weaver, "Wave chaos in elastodynamics," in *Waves and Imaging through Complex Media*, edited by P. Sebbah, Proceedings of the International Physics School on Waves and Imaging through Complex Media,

Cargese France (Kluwer Academic, 2001), pp. 141–186.
[6]U. Kuhl, H.-J. Stockmann, and R. Weaver, "Classical wave experiments on chaotic scattering," J. Phys. A **38**, 10433–10463 (2005).
[7]T. Ericson, Ann. Phys. **23**, 390 (1963); R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," J. Acoust. Soc. Am. **45**, 546 (1969).
[8]K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1975); J. D. Achenbach, *Wave propagation in Elastic Solids* (North–Holland/Elsevier, Amsterdam, 1973).
[9]R. Weaver, "Anderson localization of ultrasound," Wave Motion **12**, 129–142 (1990).
[10]R. L. Weaver and O. I. Lobkis, "Anderson localization in coupled reverberation rooms," J. Sound Vib. **231**, 1111–1134 (2000).
[11]A. L. Schawlow and C. H. Townes, "Infrared and optical masers," Phys. Rev. **112**, 1940 (1958).
[12]P. Sebbah, O. Legrand, and A. Genack, "Fluctuations in photon delay time and their relation to phase spectra in random media," Phys. Rev. E **59**, 2406 (1999).
[13]H. Kuttruff, "Energetic sound propagation in rooms," Acustica **83**, 622 (1997).
[14]In the special circumstance that the feedback tone occurs at the peak of an isolated natural mode, we indeed find such difference: $|T^*|$, there is *twice* the absorption time of that mode. This is seen by noting, for a Larsen frequency at the peak of an isolated mode with width η, $\tilde{H} \sim (\omega - \omega^* + i\eta)^{-1}$; $H^2(t) \sim \exp(-2\eta t)$; $t_{\text{delay}} = 1/2\eta$; $T^* = d \ln \tilde{H}/d\omega|_{\omega^*} = -1/i\eta = 2it_{\text{delay}}$.
[15]A. E. Siegman, *Lasers* (University Science Books, Mill Valley, CA1986); M. Patra, H. Schomerus, and C. W. J. Beenakker, "Quantum-limited linewidth of a chaotic laser cavity," Phys. Rev. A **61**, 023810 (2000).
[16]C. R. Rao, *Linear Statistical Inference and its Applications* (Wiley, New York, 1973); L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* (Addison–Wesley, Reading, MA, 1991).
[17]W. T. Yost, B. R. Macias, P. Cao, A. R. Hargens, and T. Ueno, "System for determination of ultrasonic wave speeds and their temperature dependence in liquids and *in vitro* tissues," J. Acoust. Soc. Am. **117**, 646 (2005).
[18]O. Lobkis and R. Weaver, "Coda wave interferometry in finite solids, recovery of the P to S Conversion rate in an elastodynamic billiard," Phys. Rev. Lett. **90**, 254302 (2003).
[19]Thin film characterization using resonant ultrasound spectroscopy, J. R. Gladden, J. H. So, R. Pradhan, and J. D. Maynard, J. Acoust. Soc. Am. **111**, 2399 (2002).
[20]M. Lei, J. L. Sarrao, W. M. Visscher, T. M. Bell, J. D. Thompson, A. Migliori, U. W. Welp, and B. W. Veal, "Elastic constants of a monocrystal of superconducting $YBa_2Cu_3O_{7-d}$," Phys. Rev. B **47**, 6154–6156 (1993).
[21]A. Briggs, *Acoustic Microscopy* (Clarendon Press, Oxford, 1992); *Advances in Acoustic Microscopy*, edited by A. Briggs and W. Arnold (Plenum Press, New York, 1995); A. G. Every, "Measurement of the near-surface elastic properties of solids and thin supported films," Meas. Sci. Technol. **13**, R21–R39 (2002).
[22]C. R. Farrar, S. W. Doeblíng, D. A. Nix, "Vibration-based structural damage identification," Philos. Trans. R. Soc. London, Ser. A **359**, 131–149 (2001).
[23]Y. Lu and J. E. Michaels, "A methodology for structural health monitoring with diffuse ultrasonic waves in the presence of temperature variations," Ultrasonics **43**, 717–731 (2005).
[24]R. Snieder, "The theory of coda wave interferometry," Pure Appl. Geophys. **163**, 455–473 (2006).
[25]M. L. Cowan, I. P. Jones, J. H. Page, and D. A. Weitz, "Diffusing acoustic wave spectroscopy," Phys. Rev. E **65**, 066605 (2002); J. De Rosny and P. Roux, "Multiple scattering in a reflecting cavity: Application to fish counting in a tank," J. Acoust. Soc. Am. **109**, 2587–2597 (2001).
[26]J. D. Aindow and R. C. Chivers, "A narrow-band sing-around ultrasonic velocity measurement system," J. Phys. E **15**, 1027–1031 (1982).

Outdoor sound propagation modeling in realistic environments: Application of coupled parabolic and atmospheric models

Bertrand Lihoreau, Benoit Gauvreau,[a] and Michel Bérengier
Laboratoire Central des Ponts et Chaussées, Section Acoustique Routière et Urbaine, B.P. 4129, 44341 Bouguenais cedex, France

Philippe Blanc-Benon
Laboratoire de Mécanique des Fluides et d'Acoustique de I'Ecole Centrale de Lyon, UMR CNRS 5509, 69314 Ecully Cedex, France

Isabelle Calmet
Laboratoire de Mécanique des Fluides de I'Ecole Centrale de Nantes, UMR CNRS 6598, B.P. 92101, 44321 Nantes cedex 03, France

(Received 11 March 2005; revised 14 April 2006; accepted 21 April 2006)

Predicting long-range sound propagation over a nonurban site with complex propagation media requires the knowledge of micrometeorological fields in the lower part of the atmospheric boundary layer, and more precisely its characteristics varying in both space and time with respect to local ("small-scale") and average ("long-term") conditions, respectively. Thus in this study, a mean-wind wide-angle parabolic equation (MW-WAPE) code is coupled with a dedicated micrometeorological code (SUBMESO) which simulates wind and temperature fields over moderately complex terrain with high resolution. Its output data are used as input data for the MW-WAPE code, which can also deal with different boundary conditions, such as the introduction of impedance jumps, thin screens or complex topography. Both codes are presented in the present paper. Comparisons between numerical predictions, and experimental data are also presented and discussed. Finally, we present an example of such a coupling method (MW-WAPE/SUBMESO) for the estimation of sound pressure levels at almost any site ("local scale"), for mean propagation conditions representative of long-term atmospheric conditions. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2204455]

PACS number(s): 43.28.Js, 43.50.Vt, 43.28.En, 43.28.Gq [DKW] Pages: 110–119

I. INTRODUCTION

In the framework of road traffic noise characterization and particularly for engineering applications (impact studies), there is a need for reliable sound pressure level (SPL) predictions for specific source-receiver configuration and propagation conditions, which must be representative of time ("long term") and space ("small scale" and site effects) characteristics of the acoustic situation. This is carried out in the present study, using a coupling approach with reference models for both acoustic and micrometeorological calculations. This paper focuses on this coupling method, which can give us access to SPL fluctuations due to both time and space variations of input data. Our micrometeorological code can use large-scale input data which can be chosen as representative of long-term (several years average) meteorological data, leading to small-scale (or "local-scale") output data, representative of long-term micrometeorological conditions at a considered site. At last, those data are used as input data for the acoustic model, leading to SPL predictions for small-scale *and* long-term representative propagation conditions.

Predicting long-range sound propagation over a nonurban site implies taking into account the mixed influence of ground characteristics (topography, obstacles, impedance, etc.) and atmospheric conditions (refraction and turbulence). These phenomena have been separately studied in the literature.[1] In recent years, several authors have developed numerical simulations of sound propagation in the atmosphere taking into account atmospheric models. To model sound propagation in the atmospheric boundary layer, the basic idea recently introduced is to use a mesoscale atmospheric model to simulate local wind and temperature profiles in an area above a terrain having a complex topography. This atmospheric model is coupled with an appropriate model for sound propagation. A first approach has been considered by Heimann and Gross[2] to simulate the temporal behavior of the sound pressure level across a narrow valley. In their work, a numerical sound particle model based on ray-tracing technique is coupled with a meteorological mesoscale model. Hole and Hauge[3] successfully applied the coupling method to describe the influence on sound propagation on a local scale of a morning air temperature inversion above a complex terrain. In their numerical simulations, the authors used a mesoscale atmospheric model (MM5) to provide input data for their acoustical predicting model based on a fast field program. In this model, the local sound speed in each vertical layer is calculated as the sum of the adiabatic sound speed and the wind component in the direction of the acoustic wave propagation. Recently, a different

[a]Author to whom correspondence should be addressed; electronic mail: benoit.gauvreau@lcpc.fr

approach has been considered to improve the modeling of sound propagation in an inhomogeneous moving atmosphere. These new numerical simulations are based on time-domain calculations performed with linearized equations of fluid dynamics.[4–6] The interest in these finite-difference time-domain techniques is their ability to deal with complicated phenomena in outdoor sound propagation, such as scattering by turbulence, three-dimensional (3D) effects by buildings, and topography. However, a high computational effort is necessary to run these solvers, and this approach is not yet appropriate to deal with long-range sound propagation problems.

The numerical predictions from our parabolic equation (PE) code have been previously quantitatively compared to numerical, analytical, and experimental results already published for gradually more complex situations: homogeneous, heterogeneous, and/or screened ground,[7,8] homogeneous, stratified, and/or turbulent atmosphere.[9,10] Concerning uneven ground, the model has been validated through a comparison with results obtained from a method using conformal mapping.[11,12] Comparative results show very good agreement for frequencies from 100 Hz to 5000 Hz.[13–15] Before presenting new computing results from our mean-wind wide-angle PE (MW-WAPE) code, this acoustic code is briefly presented in Sec. II. Numerical simulations have been developed using the paraxial approximation of the wave equation in bidimensional configurations with a split-step Padé marching algorithm. Afterward, PE predictions are compared to outdoor measurements especially carried out at the Laboratoire Central des Ponts et Chaussées (LCPC) experimental site (Sec. III). Comparative results for different geometrical configurations and atmospheric conditions are presented and discussed. Some improvements are obtained using the MW-WAPE instead of the classical PE (WAPE) regarding the convection effect of the wind on SPL.

Since wind and temperature profiles are required as input data for the acoustic code, they can be either experimentally determined or numerically synthesized by a micrometeorological model suited to the study of atmospheric flows at submesoscales and is called SUBMESO. This atmospheric model is presented in Sec. II B. It is a 3D nonhydrostatic compressible model derived from the advanced regional prediction system (ARPS) model.[16] The predictions are performed using large-eddy simulation (LES), which gives access to all 3D wind components as well as air temperature at each point of the mesh (50 m $\times$ 50 m) for different heights. Comparisons between SUBMESO predictions and micrometeorological data from the LCPC experimental site are shown and discussed in the next section. Finally, output data from SUBMESO is used as input data for PE predictions, leading to a coupling method (MW-WAPE/SUBMESO) to estimate the long-term SPL at a fixed site. An example of such a coupling approach is presented and discussed in Sec. IV.

II. THEORETICAL BACKGROUND

The principle of our coupling approach is to use *reference models* (i.e., reliable and validated models) for acoustic and micrometeorological calculations, in order to have access to reliable SPL predictions for a given situation in space (local scale) and time (long term). Thus, synthetic temperature and wind fields from SUBMESO are used as input data for PE calculations. In this section, we briefly present the MW-WAPE and SUBMESO codes, respectively.

A. Parabolic equation

The PE-based methods seem to be appropriate to solve the problem of acoustic propagation above a mixed ground with topographical irregularities in a refractive and turbulent atmosphere (see Sec. I). For numerical simulations of outdoor sound propagation, PEs have been derived using the approximation of the effective sound speed to take into account the convection effect of the wind. In this conventional approach, the real moving atmosphere is replaced by hypothetical motionless medium with the effective sound speed $c_{\text{eff}}=c+v_x$, where v_x is the wind velocity component along the direction of sound propagation between the source and receiver, and c is the adiabatic sound speed. This approach is convenient because both the source and receiver are close to the ground and the preferred direction of sound is nearly horizontal. However, in many problems of atmospheric acoustics, refracted sound waves propagate in directions which may significantly differ from the direction of propagation.[17–19] We use a specific PE developed by Ostashev *et al.*[20] and Dallois *et al.*[21] which does maintain the vector properties of the velocity medium. We consider bidimensional (x,z) propagation of a monochromatic acoustic wave in a homogeneous and moving medium. If the length scale of the medium L is much greater than the acoustic length scale, $\lambda \ll L$, an exact wave equation for this situation in the frequency domain is given by Ostashev *et al.*:[20]

$$\left[\Delta + k^2(1+\epsilon) - \sqrt{1+\epsilon}\frac{2ik}{c} v\cdot\nabla + \frac{v_x v_z}{c^2}\frac{\partial^2}{\partial_x \partial_z}\right] p(r) = 0, \quad (1)$$

where p is the acoustic pressure, $k=\omega/c$, c is the sound speed, $\omega=2\pi f$, f is the frequency, $\epsilon=(c_0/c(r))^2-1$ is the variation of the standard refraction index, c_0 is a reference sound speed, x and z are, respectively, the horizontal and vertical directions, and v stands for the velocity of the medium. When $v=0$, Eq. (1) is reduced to the Helmholtz equation:

$$[\Delta + k_0^2(1+\epsilon)]p(r) = 0. \quad (2)$$

The additional terms in Eq. (1), compared to Eq. (2), contain the effects of the moving medium. Ostashev *et al.*[20] and Dallois *et al.*[21] reduced Eq. (1) to WAPE. The first step is to write the two-dimensional equation for forward propagation:

$$\left[\frac{\partial}{\partial x} - ik\sqrt{Q}\right] p(r) = 0. \quad (3)$$

From here, the pseudo-operator $\sqrt{Q}$ is simplified using a Padé approximation to yield:

$$\sqrt{Q} = \frac{1+pL}{1+qL}, \quad (4)$$

where $L=Q-1$, $p=3/4$, and $q=1/4$. Considering the envelope of the pressure field defined as $\phi(r)=p(r)\exp(-ik_x)$, the

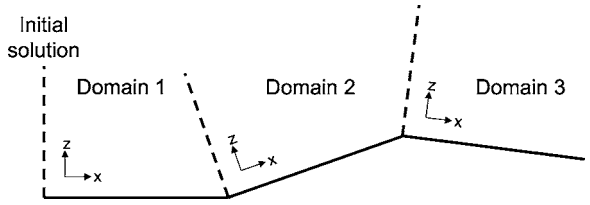


FIG. 1. Definition of the computational domains.

parabolic equation transforms to the MW-WAPE:

$$[1+qF_1-ipkM_1-qk^2M_1^2]\frac{\partial\phi}{\partial_x}=ik[(p-q)F_1 + ik(p-q)M_1-ikM_1F_1+qk^2M_1^2]\phi, \tag{5}$$

where

$$F_1=\frac{1}{c^2-v_x^2}\left[c_0^2+2ic_0\frac{v_z}{k}\frac{\partial}{\partial z}+\frac{c^2-v_z^2}{k^2}\frac{\partial^2}{\partial z^2}\right]-1,$$

and

$$M_1=\frac{2v_x}{k(c^2-v_x^2)}\left[ic_0-\frac{v_z}{k}\frac{\partial}{\partial z}\right].$$

If all velocities in Eq. (5) are set to zero, this equation is reduced to the classical Padé (1.1) WAPE derived from the Helmholtz equation [Eq. (2)]:

$$[I+qL]\frac{\partial\phi(r)}{\partial x}=ik_0[(p-q)L]\phi(r), \tag{6}$$

where

$$L=\varepsilon_{\text{eff}}+\frac{1}{k}\frac{\partial^2}{\partial z^2}, \tag{7}$$

with

$$\varepsilon_{\text{eff}}=n_{\text{eff}}^2-1=c_0^2/c_{\text{eff}}^2-1. \tag{8}$$

Equations (3) and (5) are discretized on a uniform mesh $(i\Delta x, j\Delta z)$ using a standard finite difference method. Vertical $(z-)$ derivatives are evaluated by centered difference approximations, and Crank-Nicholson scheme is implemented as a marching algorithm which takes the following form:

$$A\phi(x+\Delta x,z)=B\phi(x,z), \tag{9}$$

where A and B are pentadiagonal (MW-WAPE) or trigonal (WAPE) matrices. In our computations, the ground is modeled as a locally reacting surface with finite complex impedances calculated using the one parameter approximation from Delany and Bazley.[22] (see Sec. IV). Reflections at the top of the numerical grid are controlled by introducing a thin artificial absorption layer in the upper part of the computation domain.

The uneven ground is treated as a succession of flat domains.[11–15] After each flat domain, the coordinate system (x,z) is rotated so that the x-axis remains parallel to the ground (Fig. 1). The calculation above each domain needs an initial solution. The values of the initial solution for the domain $n+1$ are obtained from the interpolated values of the pressure field of the domain n, except for the first domain where the source is initialized by a Gaussian starter which has an adjustable width and takes into account the image source weighed by a complex reflection coefficient.

There are several ways of modeling the vertical wind and temperature, profiles near the ground: Linear, logarithmic, multilinear, linear logarithmic, hybrid, etc. As first-order approximations, temperature and wind profiles are set constant with distance (nonrange dependent) on each flat domain. Likewise, the profiles, are slightly rotated with each corresponding domain, since the angles between the rotated systems of coordinates are very small (inferior to 5°). Moreover, following Panofsky and Dutton[23] and Gilbert and White,[24] vertical temperature and wind profiles are assumed to be logarithmically shaped and expressed, respectively, as:

$$T(z)=T(z_0+d)+a_T\ln\left(\frac{z-d}{z_0}\right)\ (\text{K}),$$
$$v(z)=a_v\ln\left(\frac{z-d}{z_0}\right)\ (\text{m/s}), \tag{10}$$

where d is the displacement length, z_0 is the roughness parameter, and a_T and a_v are refraction parameters related to temperature and wind, respectively. The effective sound speed c_{eff} used in the classical Padé (1,1) PE [Eq. (6)] is defined from wind and temperature fields as:

$$c_{\text{eff}}(z)=c_0\sqrt{1+\frac{T(z)}{273.15}}+v(z)\cos(\theta)\ (\text{m/s}), \tag{11}$$

and next:

$$c_{\text{eff}}(z)\approx c_0\left(1+\frac{1}{2}\frac{T_0}{273.15}\right)+\left(\frac{1}{2}\frac{c_0}{273.15}a_T+\cos(\theta)a_v\right)\times\ln\left(\frac{z-d}{z_0}\right)\ (\text{m/s}), \tag{12}$$

where T_0(K) is a reference temperature (e.g., $T_0=293.15$ K), c_0 (m/s) is a reference sound speed for the same temperature (e.g., $c_0\approx 344$ m/s for $T_0=293.15$ K), and θ is the angle between wind direction and the direction of sound propagation. We can define an effective refraction parameter a_{eff} as follows:

$$a_{\text{eff}}=\frac{1}{2}\frac{c_0}{273.15}a_T+\cos(\theta)a_v\ (\text{m/s}). \tag{13}$$

This effective refraction parameter is used with WAPE while temperature and wind refraction parameters are used with MW-WAPE. In this paper, atmospheric turbulence is not considered. However, MW-WAPE can deal: with isotropic and homogeneous turbulence (only due to temperature fluctuations), where the random temperature field is modeled by a set of realizations (typically 50 realizations) which are generated by a superposition of discrete random Fourier modes.[25]

B. Atmospheric model

Our MW-WAPE code needs accurate propagation conditions as input data for calculations. Vertical sound speed

profiles are determined from wind and temperature profiles through Eq. (10) after data postprocessing. Those profiles can be also numerically synthesized by a micrometeorological model. This is carried out in the following section using an atmospheric code, called SUBMESO, whose theory is presented below. Next, the code is validated through comparison with experimental data from LCPC monitoring site located at Saint-Berthevin (France).

In this study, all the simulations are performed with the SUBMESO atmospheric model, which is derived from the ARPS.[16] It is a nonhydrostatic compressible model suited to the study of atmospheric flows at submesoscales. The equations are written in the so-called Gal-Chen, or terrain-following coordinate system. An option for the stretching of the mesh in the vertical z direction is available, while the grid needs to be regularly spaced in the horizontal x and y directions. The equations are discretized on the staggered grid of a Cartesian computational domain and are then transformed in the physical domain by means of the Jacobian method. Second-order accurate finite difference schemes are used to evaluate the derivatives. The solution is advanced in time using a time-splitting method[26] where all the terms in governing equations are computed explicitly. The simulations are performed using the LES method, which gives access to instantaneous fields. The flow is initialized using analytical profiles built with a meteorological preprocessor, the parameters of which are deduced from measurements: Roughness length, velocity friction, surface heat fluxes, etc. From these profiles, one can deduce the corresponding values of large-scale wind (at the highest level) that are constant during the simulation in the Rayleigh damping layer. The periodic computation on the flat coarse grid provides time-dependent turbulent inflow at the boundaries of the uneven nested grids. The subgrid fluxes are evaluated according to the Smagorinsky's model[27] modified by Lilly.[28]

The accuracy of the solution obtained by resolving the discrete Navier-Stokes equations depends directly on the mesh size. Because of the numerous time and space scales involved in the evolution of atmospheric flows, local refinement techniques are of great interest for meteorological simulations. The refinement technique used in this study is the nesting method in the horizontal direction, which consists in superposing fine grids covering small areas on a coarse grid covering the whole domain. The nesting procedure is managed as externally as possible—that is without modifying the core of the code—by means of the technical module adaptative grid refinement in FORTRAN (AGRIF) which was coupled to the atmospheric model.[29,30] Although the AGRIF module is designed to manage adaptive nesting, only fixed nested grids, are used in this study to achieve high resolution in the vicinity of particular sites. Particular care was given to the formulation of the nested grid boundary conditions, which should avoid spurious noise at the interface. Here, boundary points where the flow is entering—"inflow" boundary—are distinguished from boundary points where the flow is exiting—"outflow" boundary. At inflow boundaries, information should naturally arrive from outside, that is, from the coarse grid: The boundary values are specified through a Dirichlet-type condition. At outflow boundaries, the flow is mostly determined from the inner part of the fine grid and should be able to pass freely across the interface without boundary reflection. The method retained for outflow boundaries is the "radiative-nesting condition" applied to the velocity components and to the temperature while a Dirichlet condition is used for pressure.[31–33]

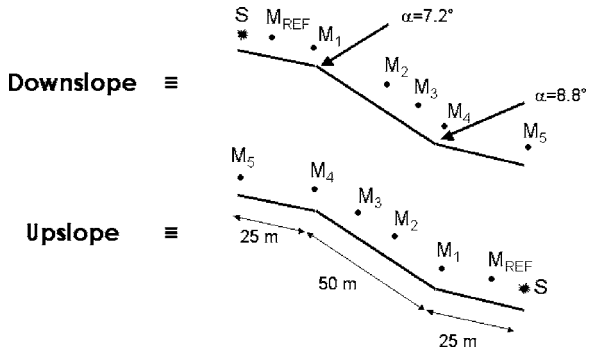


FIG. 2. Schematic illustration of the experimental setup at LCPC experimental site.

III. QUANTITATIVE RESULTS FROM MW-WAPE PREDICTIONS

A dedicated experimental campaign has been carried out at the LCPC monitoring site near Saint-Berthevin, whose protocol is briefly presented below. New results from comparison with MW-WAPE predictions are next discussed, focusing on the influence of the following parameters: Wind speed and direction, slope of the terrain, and receiver height.

A. Experimental setup

This specific campaign has been carried out on the most complex part of the LCPC experimental site: Uneven ground and very heterogeneous ground impedance.[13] For the calculations, the ground is modeled as a succession of three flat domains (Fig. 2): The first is 25 m long and has a slope of 10.21°, the second is 50 m long and has a slope of 17.42°, and the third is 25 m long and has a slope of 8.61°. The relative angle between the first and the second domain is $\alpha = 7.2°$ and $\alpha = 8.8°$ between the second and the third domain. Acoustic data have been collected using an impulsive and omnidirectional sound source (gun shots) and five microphones spatially distributed between 25 and 100 m from the source (M1 to M5 in Fig. 2). The reference microphone was located 10 m from the source (M_{ref}).

The ground impedance Z has been experimentally determined at several points through a "best-fit" calculation using the one-parameter formula from Delany and Bazley:[22]

$$Z = \rho_0 c_0 \left(1 + 0.0571 \left(\frac{\rho_0 f}{\sigma} \right)^{-0.754} + i 0.087 \left(\frac{\rho_0 f}{\sigma} \right)^{-0.732} \right), \tag{14}$$

where c_0 is the reference sound speed, ρ_0 is the air density, and σ is the airflow resistivity. The parameter σ is experimentally determined using a method developed by Bérengier *et al.*[34] The results of the fitting procedures for downslope propagation are (see Fig. 2): $\sigma = 600$ kPa s m^{-2} around the

TABLE I. Micrometeorological parameters deduced from experimental data: θ is the angle between wind direction and the direction of sound propagation, a_T and a_v are refraction parameters related to temperature and wind respectively, and a_{eff} is the effective refraction parameter (see Sec. II A).

Measurement	a_v(m/s)	θ(°)	a_T(K)	a_{eff} (m/s)
4	0.30	60	0.20	0.27
7	0.65	20	0.20	0.73
9	0.30	70	0.30	0.28
11	0.00	⋯	−0.30	−0.18

source, $\sigma=90$ kPa s m^{-2} around M1, $\sigma=160$ kPa s m^{-2} around M2, and $\sigma=200$ kPa s m^{-2} around M4.

In order to evaluate the micrometeorological conditions, we used an equipped tower located on the slope but far enough from the measurement line not to disturb acoustic propagation. The tower is equipped with ventilated air thermometers; (Young 41342VC) and accurate wind direction and wind speed sensors (Young 05305AQ), using a Young 26700 station, the accuracy is about 0.1 K, 2°, and 0.1 m/s respectively. The sampling rates of the temperature and wind measurements are too low as to derive turbulence parameters. These sensors are located at three different heights: 1, 3, and 10 m. Temperature and wind profiles are modeled following Eq. (10), where a_T and a_v are deduced from micrometeorological measurements (10 min average). For each acoustical measurement, the signal has been averaged over ten gun shots, which is a sufficient number to determine a reliable average value for acoustical measurements.[35] Results are given in terms of relative SPL coming from the difference between the spectrum at microphone M1, M2, M3, M4, or M5 and the spectrum at the reference microphone M_{ref}. Tables I and II, respectively, summarize micrometeorological and geometrical parameters deduced from data postprocessing, and used as input data for PE calculations.

B. Comparison between experimental results and numerical predictions

Figure 3 shows results from Measurement No. 7 performed for downslope propagation and for $h_s=h_M=h_{\text{ref}}=2m$ (see Table II). The corresponding micrometeorological parameters are summarized in Table I. The wind is moderately strong and almost directed in the source-receiver direction, which leads to a downward refracting atmosphere. Therefore, Fig. 3 allows us to compare acoustic results, either experimental (solid line) or calculated with the standard PE (WAPE – dotted line) and with the new PE (MW-WAPE – dashed line).

TABLE II. Geometrical parameters of experimental setup and numerical predictions: Source height h_s, reference receiver height h_{ref}, and microphone height h_{M} (see Sec. III A).

Measurement No.	Slope	h_s(m)	h_{ref}(m)	h_{M}(m)
4	Down	2	2	2
7	Down	2	2	2
9	Down	2	0.6	0.6
11	Up	2	2	2

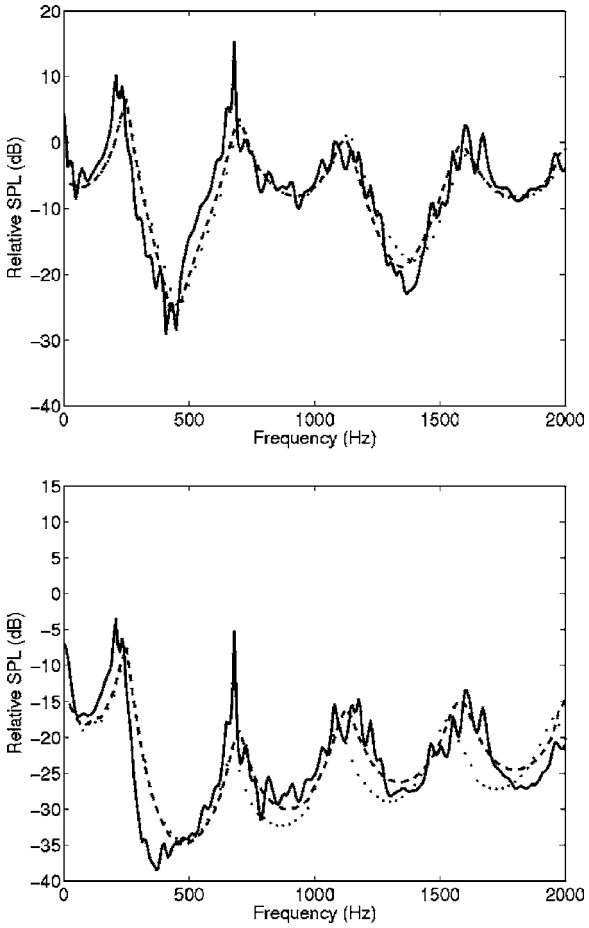


FIG. 3. Measurement No. 7. Relative SPL as a function of frequency: Comparison between experimental results (solid line) and PE predictions (WAPE in dotted line and MW-WAPE in dashed line). (a) M1 (25 m) and (b) M4 (75 m).

Numerical predictions are in very good agreement with experimental data, especially those given by the new PE (MW-WAPE), which gives a best localization of interference fringes for highest frequencies. This difference increases with the distance of propagation. In terms of geometric acoustics, this means that the receiver can be reached following two different paths: A direct and a reflected ray. The sound speed $(c+v\cos\theta)$ varies on each ray with θ, and the use of the effective sound speed $(c+v_x)$ introduces a cumulative phase error in standard parabolic equation. This error increases with receiver height, distance of propagation, and wind speed.

Additional calculations for different geometrical configurations, slopes, and atmospheric conditions have confirmed that there is no significant difference between the results from WAPE and MW-WAPE when the wind is very low. On the contrary, as far as the wind is moderately strong, cross-wind effects are always better taken into account using new MW-WAPE instead of standard WAPE. From now on, the numerical results further presented in this paper will be issued from the MW-WAPE code. Above 2000 Hz, wavelength and ground roughness are of the same order. This

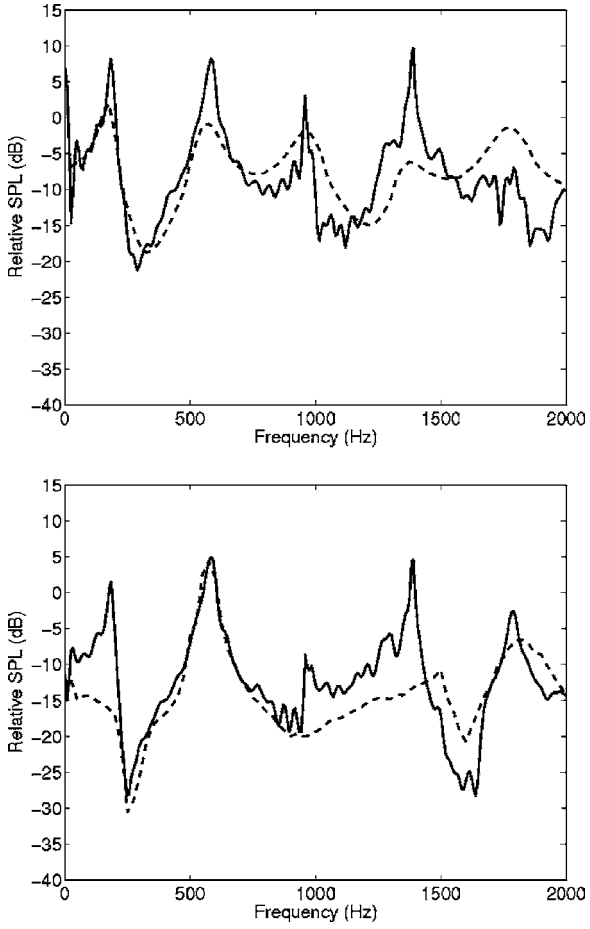


FIG. 4. Measurement No. 11. Relative SPL as a function of frequency: Comparison between experimental results (solid line) and MW-WAPE predictions (dashed line). (a) M1 (25 m) and (b) M4 (75 m).

leads to an additional diffraction which is not modeled by the acoustic code. That is the reason why comparative results above 2000 Hz are not presented. Regarding acoustic propagation for upslope cases, Fig. 4 shows results, from Measurement No. 11 ($h_s=h_{\mathrm{ref}}=h_{\mathrm{M}}=2m$, see Table II). The propagation conditions are slightly upward refracting (see Table I).

Numerical results show very good agreement with the experimental results at frequencies below 1000 Hz. Above 1000 Hz, numerical predictions are not as close to the experimental data. These discrepancies come from different causes. First, particular attention has to be paid to the numerical models sensibility, with respect to the spatial location of source and receivers (h_s, h_{ref}, and h_{M}, see Sec. III A), which can generate large uncertainties in MW-WAPE predictions when a lack of precision occurs in the *in situ* measurement of those geometrical input data. Second, the experimental terrain is more irregular on the bottom than on the top of the site. Thus, it is possible that discrepancies between numerical and experimental approaches result from 3D effects. Third, the Delany and Bazley's approach [Eq. (14)] gives good approximations, but remains limited. A more complex model of the ground, such as Attenborough's,[36] should improve our PE predictions. This model includes, more physical factors: Air flow resistivity σ, porosity Ω, grain shape factor n', and pore shape factor ratio s_f. Nevertheless, its implementation remains difficult because these

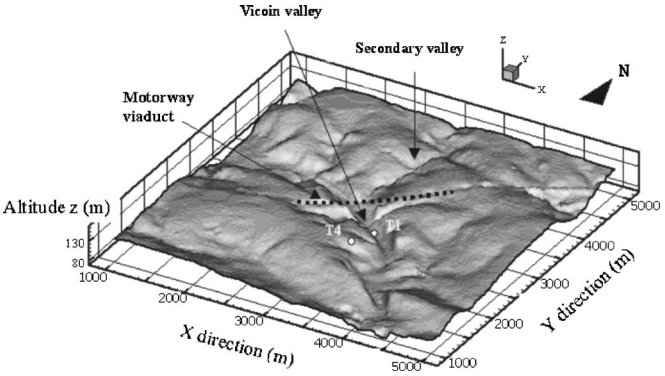


FIG. 5. Topography of the LCPC experimental site located at Saint-Berthevin. The resolution is 25 m. The vertical direction is stretched by a factor of 5 in order to make the topography more noticeable.

four factors (σ, Ω, n', and s_f) cannot be easily characterized by *in situ* measurements. Last, but not least, the differences between experimental and numerical SPL for the highest frequencies (above 1000 Hz) can also be explained by the effects of atmospheric turbulence,[37–41] which have not been taken into account in those MW-WAPE calculations for central processing unit time reasons.

IV. PRACTICAL EXAMPLES OF MW-WAPE/SUBMESO COUPLING

This section shows how SUBMESO gives us access to synthetic local (site-scale) wind and temperature vertical profiles from global ones (regional scale), which can be chosen next as representative values of "long-term" (e.g., a 30 years average) atmospheric conditions. Those last conditions are given by national meteorological stations (Météo-France, for instance), which record hourly data over dozens of years. Thus, SUBMESO can provide wind and temperature profiles for the characterization of a specific situation both in *space* (local scale) and *time* (long term). Then, output data from SUBMESO can be used as micrometerological input data for PE calculations between the source and receiver. Therefore, our MW-WAPE/SUBMESO coupling method can provide SPL representative of *local* and *long-term* atmospheric conditions between the source and receiver. This procedure is briefly presented below, before showing some calculations from the coupled MW-WAPE/SUBMESO code.

LES of the atmospheric flow above the LCPC experimental site located at Saint-Berthevin have been performed to assess the terrain-induced modifications of the mean flow and turbulence characteristics over hilly surfaces. This study is part of a research plan directed by the LCPC for controlling long-range noise pollution in the surroundings of motorways.

The studied area is centered on the permanent source of noise pollution, which is a motorway viaduct crossing the valley of the river Vicoin. This valley, which has an approximately average depth of 35 m and an average width of 200 m, crosses the domain from north-west to south-east. A tributary stream coming from the north in a smaller valley joins the river Vicoin near the center of the site. The smooth topography of the 16 km^2 area is displayed in Fig. 5, based

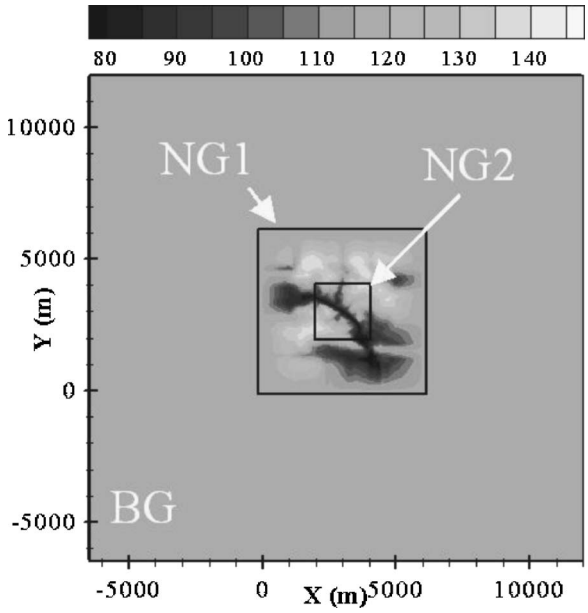


FIG. 6. BG (flat) and the two nested grids NG1 and NG2 (with topography), for the site of Saint-Berthevin.

on a digital terrain model (DTM) with a resolution of 25 m provided by the French National Geographical Institute (IGN). Total relief amplitude is 67 m ranging from 73 m in the lower part of the valley (near the south boundary) to 140 m at the northern edge of the domain.

Three levels of grids are used, which are displayed in Fig. 6. At the first level, a base grid (BG) has a horizontal resolution of 450 m with 43×43 points (18.9 km $\times$18.9 km). The terrain is flat and its altitude is set to the terrain average altitude $z_a=115.2$ m. At the second level, the nested grid (NG1) has a horizontal resolution of 150 m, with 44×44 points. The domain covered by this grid (6.15 km $\times$6.15 km) includes the whole 16 km^2 area defined in DTM. The area is extended in the horizontal directions so that the terrain is flat at the boundaries of NG1, insuring a proper connection at the BG/NG1 interface at the altitude z_a. Finally, a third grid (NG2) is nested into the grid NG1, centred on the motorway viaduct. It has a horizontal resolution of 50 m, with 44×44 points (2.15 km$\times$2.15 km). Note that the topography on the grid NG2 is thus more detailed than on the grid NG1. For the three grids, 32 layers are distributed in the vertical direction, following a geometric series (with a common ratio of 1.2). The thinnest mesh layer is 10 m deep, at the ground. A Rayleigh damping layer extends from $z=4000$ m to the top of the domain (7500 m). Periodic conditions are imposed at the boundaries of the BG, providing time-dependent turbulent inflow at the boundaries of the uneven area NG1.

A period of neutral atmospheric stratification (i.e., without thermal effects)—May 21, from 1:45 to 3:15—was selected among the data available from the experimental campaign conducted in May 2000. Initially, a uniform wind profile is set in the larger domain and the flow is forced by a constant large-scale wind ($U=3.3$ m/s in the west-east direction and $V=2.7$ m/s in the south-north direction). The initial wind direction is 230° clockwise from the north.

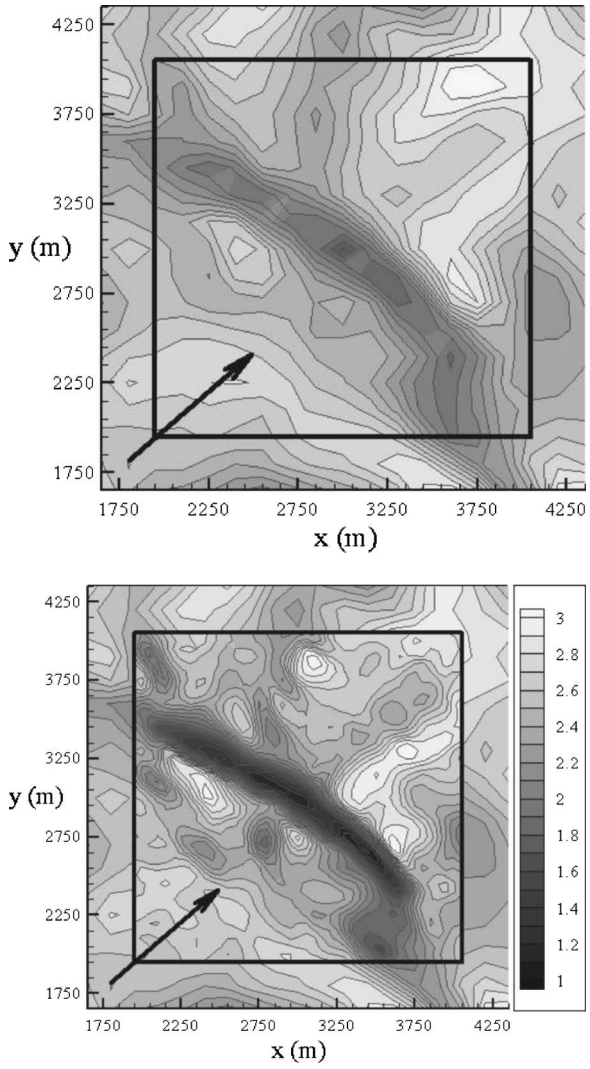


FIG. 7. Partial x-y view of the simulated wind field W_1 (m/s) at Saint-Berthevin, at the first grid level (5 m) above the ground (a) on NG1 (the black square represents the location of NG2) and (b) on NG2 overlapping NG1. The black arrow symbolizes the mean wind direction.

The surface temperature is uniform at the ground and kept constant during the whole simulation. As a first approach, the air is considered to be dry. After 1400 s, the wind profile calculated on NG2 locally matches quite well with the experimental profile measured at the tower (T4) (see Fig. 5 for the location of this tower). A time average is calculated between $t=1400$ s and $t=1500$ s with an output time step of 1 s for NG2. From this procedure, we get smoothed local fields hereafter designated by the index I. We compare these fields with their value averaged over the entire domain, hereafter designated by the index m. In the following discussion, we pay a particular attention to the deformation of the total wind.

The first important result is the improved quality of the simulated flow provided by the grid refinement technique. This conclusion is obvious from Fig. 7, which displays the simulated wind field at 5 m above the ground on both NG1 and NG2. The main features—slowdown and speedup effects in valleys and over hills—are visible at both resolution levels, but they are strongly diffused and smoothed on NG1. In

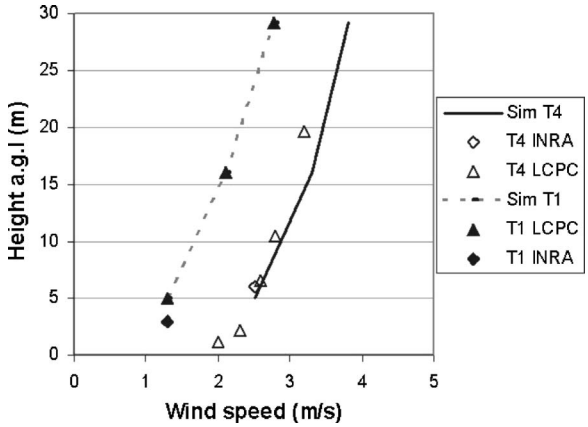


FIG. 8. Wind vertical profiles at Saint-Berthevin, from measurements ("T4/T1 INRA", "T4/T1 LCPC") and from simulation ("Sim T4" and "Sim T1"), at two locations T4 (3050 m, 2650 m) and T1 (3150 m, 2900 m). Note that the ground level is not the same for T1 to T4.

particular, we notice a great difference in the estimation of the slowdown in the Vicoin valley, which is nearly perpendicular to the mean wind direction: With a 50 m resolution, the wind speed is found to be reduced of up to 58% compared to its mean value W_m, whereas it is reduced of a maximum of 31% only with a 150 m resolution.

These differences can be explained by two factors. First, the highest resolution on NG2 naturally restrains the numerical diffusion of the solution. Second, the terrain features are much more accurately described on NG2 than on NG1, for which the topography is artificially smoothed. It is interesting to note that the value of the mean wind W_m on NG2 is between 2 and 3% lower than the value on NG1 in the first 300 m above the ground. This is probably due to more accurate calculation of the valley-induced deceleration when the horizontal resolution is improved from 150 m to 50 m.

The second important result is the absence of significant perturbations at the NG1/NG2 interface. The wind field is continuous from one grid to the other at inflow boundaries, as imposed by the Dirichlet-type condition. At outflow boundaries, we can naturally see slight "jumps" from NG2 to NG1 as the solution is less forced, but no induced numerical oscillations are visible. The radiative-nesting boundary condition appears to work very well in this case, by both forcing the NG2 solution to be consistent with the NG1 solution and avoiding spurious reflections at the interface. In the following analysis of the impact of relief on the flow, only results obtained on the high-resolution grid NG2 will be considered.

The vertical profiles of wind speed measured *in situ* and simulated are displayed in Fig. 8, at two different locations: T4 and T1 (see Fig. 5). T1 is located at the bottom of the valley ($z=82$ m), whereas T4 is located on a small plateau dominating the valley ($z=117$ m). Two simultaneous series of measurements (LCPC, permanent tower and Institut National de Recherche Agronomique (INRA), additional tower) are available during the considered period. From the measurements, "LCPC"—the main deformation into the valley—leads to an important decrease in the wind speed of about 1.3 m/s between T4 and T1. The deceleration of the flow into the valley is well reproduced by the model, despite the relatively low vertical resolution close to the ground, which does not allow us to predict the observed strong gradients of wind within the first meters above the ground. Note that at the same time, the wind is deflected to the north in the valley, which is in agreement with measurements.

Therefore, SUBMESO code provides wind field (and air temperature, not shown here because gradients are very weak in these conditions) at each point of its mesh, the first nodes of which are located 5 m above the ground. Wind speed values issued from SUBMESO (three components) are next expressed in terms of horizontal wind speed and direction in order to be adapted to PE calculations. Below 5 m, temperature and wind vertical profiles are assumed to be logarithmically shaped. Thus the refraction parameters a_T and a_v [Eq. (10)] can be fitted using the first point (5 m high) provided by SUBMESO. The corresponding vertical profile of sound speed [Eq. (12)] is assumed to be constant, but can be chosen as representative of the studied source-receiver configuration for medium-range propagation: valley, plateau, downslope, upslope, etc. It must be noticed that since mesh adaptation and data interpolation from SUBMESO to PE is not automatic yet, vertical profiles are still range independent. Further work is currently in progress in order to take into account the "exact" wind and temperature values at each point of the PE grid, directly interpolated from SUBMESO output data, leading to range dependent vertical profiles of sound speed on such irregular terrain.

Thus SUBMESO code provides synthetic *local* (site scale) wind and temperature vertical profiles from input data collected at larger scales (e.g., "Météo-France" national stations). Those large-scale input data can be chosen as average (e.g., a 30 year average) data, leading to SUBMESO output data representative of long-term (e.g., a 30 year average) atmospheric conditions at a local site. Therefore, output data from SUBMESO can be used as input data for the PE code, leading to SPL predictions for various micrometeorological conditions, including site effect (space sensitivity, e.g., topography) *and* long-term mean conditions (time sensitivity, e.g., a 30 year average). Those long-term averages can be chosen either for day, evening night, day and night, etc., periods, or for a more global period of 24 h.

The next part of this study show a comparison between experimental acoustic data extracted from a specific campaign carried out on the LCPC monitoring site located at Saint-Berthevin and MW-WAPE predictions (Fig. 9). During this experimental campaign, acoustic data have been measured for typical atmospheric conditions, i.e., for *local* micrometeorological conditions *directly* collected at the same site *and* at the same time using *synchronized* meteorological sensors from Saint-Berthevin equipped towers. Actually, after a scanning of the micrometeorological data measured during this experimental campaign, we chose a period (sample) which was the most representative of a "long-term" (average) period in terms of meteorological parameters predicted by SUBMESO (Table I—Measurement No. 4). Thus, it became possible to compare PE long-term predictions with the corresponding acoustic experimental data. Figures 9(a) and 9(b) give some examples of such long-term (a 30 year average of 24 h periods) PE predictions for downslope propaga-

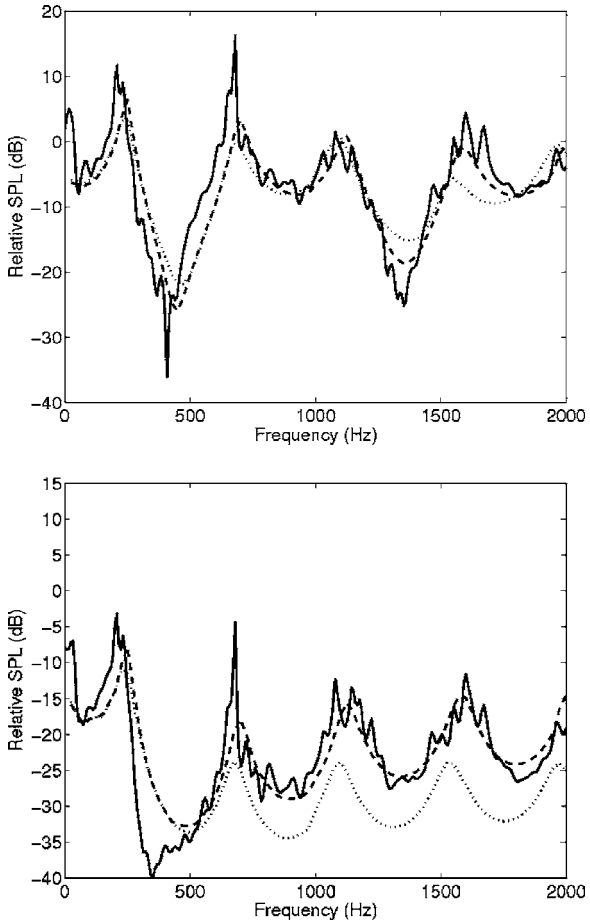


FIG. 9. Examples of results provided by the coupling of the SUBMESO micrometeorological code with the MW-WAPE acoustic code. Relative SPL as a function of frequency: Comparison between experimental results (solid line) and MW-WAPE predictions for long-term (dashed line) and for homogeneous (dotted line) conditions. (a) M1 (25 m) and (b) M4 (75 m).

tion and using the following geometrical parameters values: $h_s=h_M=h_{ref}=2m$. (Table II—Measurement No. 4). Moreover, in order to identify the real effect of long-term average meteorological parameters, we also calculate PE predictions in the same geometric configuration but for homogeneous conditions (dotted line). Micrometeorological parameters are deduced from SUBMESO calculations as presented above. As mentioned above, Figures 9(a) and 9(b) also show acoustic data (solid line) collected during the experimental campaign (Tables I and II—Measurement No. 4), which appear to be very similar to MW-WAPE predictions when considering "long term" atmospheric conditions instead of homogeneous conditions.

The SPL difference between homogeneous and long-term conditions, of course, increases with distance from source, but is already perceptible from medium range propagation [see Fig. 9(b), M4 at 75 m]. This difference can be also quantified, e.g., in one-third octave bands, which can be useful for engineering applications, such as noise impact studies. Figures 9(a) and 9(b) also show that discrepancies between PE predictions and experimental data can be greatly reduced if we can have access to "true" wind and temperature fields on the studied site during the corresponding period. Unfortunately, for engineering or operational situations, such experimental data cannot be always collected due to financial cost and/or *in situ* constraints. This study showed that SUBMESO can provide those wind and temperature profiles at almost any site (local scale), which can be chosen as mean values representative of a chosen period (short term or long term, e.g., 30 years).

Consequently, the MW-WAPE/SUBMESO coupled code can be used for estimating long-term SPL representative of mean (e.g., a 30 year average) atmospheric conditions at a local site, even on irregular terrain. It must be mentioned that this coupling approach do not yet take into account nonlinear effects of both acoustical and micrometeorological models. These first numerical results are presented as a first stage of a research program which is still in progress, and which will be carried out in the next few years, partially based on the exploitation of the database from the LCPC monitoring site located at Saint-Berthevin.

V. CONCLUSION

This study takes place in the framework of road traffic noise propagation. Our coupling approach gives us access to reliable SPL predictions for a specific source-receiver configuration (local scale) and propagation conditions (long term). This is carried out through the use of reference models for both acoustic (MW-WAPE) and micrometeorological (SUBMESO) calculations, which have been presented, discussed, and validated by comparison with experimental data.

The MW-WAPE code takes into account the convection properties of cross-wind effects on acoustic propagation. Moreover, in spite of some approximations in the ground impedance model, it can deal with realistic environments (uneven ground, impedance jumps, absorbent barriers, etc.) and complex propagation conditions (range dependent refraction profiles and atmospheric turbulence).

The SUBMESO micrometeorological code provides synthetic local (small scale) wind and temperature vertical profiles from global ones (regional scale). For the input profiles to the SUBMESO code, one can use long-term average atmospheric averages. In our study, we used 30 year averages, which for practical purposes can be interpreted as long-term averages. Then, output data from SUBMESO have been used as micrometeorological input data for MW-WAPE calculations between the source and receiver, in order to calculate SPL representative of *local* and *long-term* atmospheric conditions between the source and receiver. Examples of MW-WAPE/SUBMESO coupling have been presented and discussed. They can be considered as very promising results regarding both the complementary nature and the reliability of the coupled models, although some refinements still should be done on the automatic coupling procedure. Currently, further research is in progress in order to take into account range dependent profiles, line sources, 3D and nonlinear effects, and various propagation conditions.

ACKNOWLEDGMENTS

The authors are very grateful to Keith Wilson (Associate Editor) and reviewers for their insightful comments that have led to the improvement of this manuscript.

[1]Embleton, T. F. W., "Tutorial on sound propagation outdoors," J. Acoust. Soc. Am. **100**, 31–48 (1996).
[2]Heimann, D., and Gross, G., "Coupled simulation of meteorological parameters and sound level in a narrow valley," Appl. Acoust. **56**, 73–100 (1999).
[3]Hole, L. R., and Hauge, G., "Simulation of a morning air temperature inversion break-up in complex terrain and the influence on sound propagation on a local scale," Appl. Acoust. **64**, 401–414 (2003).
[4]Blumrich, R., and Heimann, D., "A linearized Eulerian sound propagation model for studies of complex meteorological effects," J. Acoust. Soc. Am. **112**, 446–455 (2002).
[5]Ostashev, V. E., Wilson, D. K., Liu, L., Aldridge, D. F., Symons, N. P., and Martin, D., "Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementations," J. Acoust. Soc. Am. **117**, 503–517 (2005).
[6]Salomons, E. M., Blumrich, R., and Heimann, D., "Eulerian time-domain model for sound propagation over a finite-impedance ground surface. Comparison with frequency-domain models," Acta. Acust. Acust. **88**, 483–492 (2002).
[7]Craddock, J. M., and White, M. J., "Sound propagation over a surface with varying impedance: A parabolic equation approach," J. Acoust. Soc. Am. **91**, 3184–3191 (1992).
[8]Rasmussen, K. B., and Galindo Arranz, M., "The insertion loss of screens under the influence of wind," J. Acoust. Soc. Am. **104**, 2692–2698 (1998).
[9]Galindo, M., "Approximations in the PE method. Phase and level errors in a downward refracting atmosphere," *Int. Symp. on Long Range Sound Propagation*, Lyon, France, 235–255 (1996).
[10]Salomons, E. M., "Diffraction by a screen in downwind sound propagation: A parabolic equation approach," J. Acoust. Soc. Am. **95**, 3109–3117 (1994).
[11]Blairon, N., Blanc-Benon, P., Bérengier, M., and Juvé, D., "Calculation of sound propagation over uneven terrain using parabolic equation," Invited Paper to 17th ICA, Proceedings Volume III, CD-Rom ISBN 88-88387-02-1, Rome (2001).
[12]Blairon, N., Blanc-Benon, P., Bérengier, M., and Juvé, D., "Outdoor sound propagation in complex environment: experimental validation of a PE approach," *10th Int. Symp. on Long Range Sound Propagation*, Grenoble, France, 114–128 (2002).
[13]Blairon, N., "Topography effect on acoustic propagation in the atmosphere: Numerical modelization using parabolic equation and validation by comparison with outdoor experiments," (in French) Ph. D. thesis, École Centrale de Lyon, France (2002).
[14]Gauvreau, B., Bérengier, M., Blanc-Benon, P., and Depollier, C., "Traffic noise prediction with the parabolic method: Validation, of a split-step Padé approach in complex environments," J. Acoust. Soc. Am. **112**, 2680–2687 (2002).
[15]Blanc-Benon, P., Lihoreau, B., Pénelon, T., Gauvreau, B., Calmet, I., and Bérengier, M., "Outdoor sound propagation in complex environments using the parabolic equation: A new PE code coupled with a micrometeorological code," *11th Int. Symp. on Long Range Sound Propagation*, Lake Morey, USA (2004).
[16]Xue, M., Droegemeier, K. K., and Wong, V., "The advanced regional prediction system (ARPS)—A multiscale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification," Meteorology and Atmos. Phys. **75**, 161–193 (2000).
[17]Godin, O. A., "Wide-angle parabolic equation for sound in 3D inhomogeneous moving medium," Dokl. Phys. **47**, 643–646 (2002).
[18]Lingevitch, J. F., Collins, M. D., Dacol, D. K., and Drob, D. P., "A wide angle and high Mach number parabolic equation," J. Acoust. Soc. Am. **111**, 729–734 (2002).
[19]Blanc-Benon, P., Dallois, L., and Juvé, D., "Long-range sound propagation in a turbulent atmosphere within parabolic equation," Acta. Acust. Acust. **87**, 659–669 (2001).
[20]Ostashev, V. E., Juvé, D., and Blanc-Benon, P., "Derivation of a wide-angle parabolic equation for sound waves in inhomogeneous moving media," Acta. Acust. Acust. **83**, 455–460 (1997).
[21]Dallois, L., Blanc-Benon, P., and Juvé, D., "A wide-angle parabolic equation for acoustic waves in homogeneous moving media: Application to atmospheric sound propagation," J. Comput. Acoust. **9**, 477–494 (2001).
[22]Delany, M. E., and Bazley, E. N., "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[23]Panofsky, H. A., and Dutton, J. A., *Atmospheric Turbulence* (Wiley, New York, 1984).
[24]Gilbert, K. E., and White, M. J., "Application of the parabolic equation to sound propagation in a refracting atmosphere," J. Acoust. Soc. Am. **85**, 630–637 (1989).
[25]Juvé, D., Blanc-Benon, P., and Chevret, P., "Numerical simulation of sound propagation through a turbulent atmosphere," *Proceedings of the 5th Int. Symp. on Long Range Sound Propagation*, Milton Keynes, UK 282–286 (1992).
[26]Klemp, J. B., and Wilhelmson, R. B., "The simulation of three-dimensional convective storm dynamics," J. Atmos. Sci. **35**, 1070–1096 (1978).
[27]Smagorinsky, J., "General circulation experiments with the primitive equations: I,"Mon. Weather Rev. **91**, 99–164 (1963).
[28]Lilly, J. B., "The representation of small-scale turbulence in numerical simulation experiments," *Proc. IBM Sci. Comput. Symp. on Env. Sci., N. Y.*, IBM Form 3201951, 195–210 (1967).
[29]Blayo, E., and Debreu, L., "Adaptive mesh refinement for finite-difference ocean models: First experiments," J. Phys. Oceanogr. **29**, 1239–1250 (1999).
[30]Pénelon, T., "Meteorological simulations of a rural site with a nonplan topography using the nesting technique with SUBMESO," (in French), Ph.D thesis, École Centrale de Nantes – Université de Nantes, France (2002).
[31]Carpenter, K. M., "Note on the paper: Radiation conditions for the lateral boundaries of limited-area numerical models by M. J. Miller and A. J. Thorpe (Q.J., 107, 605–628)," Q. J. R. Meteorol. Soc. **108**, 717–719 (1982).
[32]Chen, C., "A nested grid, nonhydrostatic, elastic model using a terrain-following coordinate transformation: The radiative-nesting boundary conditions," Mon. Weather Rev. **119**, 2852–2869 (1991).
[33]DeCroix, D. S., "Large-eddy simulations of convective and evening transition planetary boundary layers," Ph. D. thesis, Department of MEAS, NCSU, USA (2001).
[34]Bérengier, M., and Garai, M., "A state-of-the-art of in situ measurement of the sound absorption coefficient of road pavements," *17th Int. Cong. Acous.*, Rome, Italy (2001).
[35]Legeay, V., and Seznec, R., "On the determination of acoustic characteristics of absorbent materials," (in French), Acustica **53**, 171–192 (1983).
[36]Attenborough, K., "Acoustical impedance models for outdoor ground surface," J. Sound Vib. **99**, 521–544 (1985).
[37]Daigle, G. A., "Effects of atmospheric turbulence on the interference of sound waves near a hard boundary," J. Acoust. Soc. Am. **64**, 622–630 (1978).
[38]Daigle, G. A., "Effects of atmospheric turbulence on the interference of sound waves above a finite impedance boundary," J. Acoust. Soc. Am. **65**, 45–49 (1979).
[39]Chevret, P., Blanc-Benon, P., and Juvé, D., "A numerical model for sound propagation through a turbulent atmosphere near the ground," J. Acoust. Soc. Am. **100**, 3587–3599 (1996).
[40]Wilson, D. K., Brasseur, J. G., and Gilbert, K. E., "Acoustic scattering and the spectrum of atmospheric turbulence," J. Acoust. Soc. Am. **105**, 30–34 (1999).
[41]Karweit, M., Blanc-Benon, P., Juvé, D., and Comte-Bellot, G., "Simulation of the propagation of an acoustic wave through a turbulent velocity field: A study of phase variance," J. Acoust. Soc. Am. **89**, 52–62 (1991).

Rytov approximation of tomographic receptions in weakly range-dependent ocean environments

G. S. Piperakis[a] and E. K. Skarsoulis
Institute of Applied and Computational Mathematics, Foundation for Research and Technology Hellas, GR-711 10 Heraklion, Crete, Greece

G. N. Makrakis
Department of Applied Mathematics, University of Crete, GR-714 09 Heraklion, Crete, Greece and Institute of Applied and Computational Mathematics, Foundation for Research and Technology Hellas, GR-711 10 Heraklion, Crete, Greece

(Received 9 November 2005; revised 13 April 2006; accepted 13 April 2006)

An approximation method based on the Born and Rytov approximation is presented for the wave-theoretic prediction of acoustic arrival patterns associated with long-range pulse propagation in weakly range-dependent ocean environments. The environment is considered as a perturbation of a range-independent background state, and normal-mode theory is used, for the representation of the background Green's function. Using the Born and Rytov approximations, the perturbed Green's function corresponding to the range-dependent environment is expressed for each frequency within the source bandwidth in terms of the background Green's function and the medium (sound-speed) perturbation. The actual arrival pattern in the time domain is then computed through the inverse Fourier transform. Using the normal-mode representation, closed-form expressions for the first and second Born and Rytov approximations are derived, generalizing previous range-independent results, and indicating that the effect of range dependence on the acoustic field in the case of adiabatic perturbations is of second order. To cope with the multimodal nature of ocean acoustic propagation, a variation of the standard Rytov method is applied, proposed by Keller, according to which each modal component must be treated independently. A number of numerical examples demonstrate an advantage of the Rytov approximation (over the Born approximation) for time-domain calculations. [DOI: 10.1121/1.2202891]

PACS number(s): 43.30.Bp, 43.30.Dr, 43.30.Ft, 43.30.Qd [AIT] Pages: 120–134

I. INTRODUCTION

Ocean acoustic travel-time tomography[1] seeks to infer the ocean state from travel-time measurements of pulsed acoustic signals traveling through the water from broadband sources to distant receivers. Arrival times depend on sound speed, and sound speed is related with water temperature;[2] thus, from travel-time measurements temperature estimates can be obtained by inversion.[3] The spatial variability of sound speed (temperature) gives rise to refraction which in turn causes multipath propagation. Each path traverses different water masses with different temperature and sound-speed values. Thus a pulsed acoustic signal emitted by the source will reach the receiver at different time instants depending on the path it is traveling along. This leads to a sequence of arrivals at the receiver in the time domain conveying information about different water layers.

In long-range, deep-water propagation early arrivals at the receiver can be sufficiently described in terms of rays, in particular steep rays with a relatively small number of turning points, whereas late arrivals can be described by a limited set of low-order modes.[4] Still, the identification of individual modes in the late arrival pattern is not always possible due to mode interference and temporal overlapping, especially in cases of strong dispersion. A way to retrieve information about modes in such cases is by mode filtering using a vertical receiving array.[5,6] The other alternative for the analysis of late arrival patterns is full wave form inversion,[7–9] which, however, is associated with a large number of forward calculations and thus with a heavy computational burden. Several methods have been developed for accelerating the computation of wave-theoretic arrival patterns, based on approximations with respect to frequency, such as narrowband normal-mode approximations relying on Taylor expansions of eigenvalues and eigenfunctions,[10–12] broadband stationary-phase approximations,[13] and frequency-interpolation methods associated with normal modes[14] and the parabolic approximation.[15]

An approximation method for the wave-theoretic calculation of time-domain tomographic receptions, of the late arrival pattern in particular, in weakly range-dependent ocean environments is proposed here based on the Born and Rytov approximations of the frequency-domain Green's function. The method exploits the fact that many range-dependent ocean features such as internal waves,[16–18] mesoscale eddies, and large-scale variations[19–22] can be considered as range-dependent perturbations of a range-independent background state. In this connection, it makes use of the Born[23] and Rytov[24] second-order approximations for the perturbation of the Green's function caused by range-dependent

[a]Electronic mail: piperak@iacm.forth.gr

 0001-4966/2006/120(1)/120/15/$22.50

perturbations about a range-independent reference state. The corresponding arrival pattern is then calculated from the perturbed Green's function through the inverse Fourier transform. The range-independent background allows for the use of normal-mode theory as the basis for propagation modeling, leading to closed-form expressions for the perturbed Green's function in terms of background quantities and range-dependent sound-speed perturbations, offering a computationally efficient alternative to the exact range-dependent calculations at each frequency.[25]

A similar approach based on medium perturbations but in a range-independent framework was presented recently[26] relying upon perturbations of the vertical eigenvalue problem with respect to the sound speed. The present approach is more-general and allows for arbitrary sound-speed perturbations, either range independent or range dependent. The latter are assumed to be weak in the adiabatic sense, i.e., with large horizontal scales compared with the double-loop length of the corresponding eigenrays.[27]

The contents of the work are organized as follows: Section II addresses the Green's function and its perturbations in the Born and Rytov approximations up to second order, as well as the relations between the two approximations. In Sec. III using the normal-mode representation for the background acoustic field closed-form expressions for the first and second Born and Rytov approximations are derived. Section IV presents some numerical results from the application of the various approximations for the calculation of time-domain arrival patterns in range-dependent ocean environments, as well as comparisons with exact adiabatic and coupled-mode results. Finally, Sec. V contains a discussion of results and main conclusions from this work.

II. THE GREEN'S FUNCTION

The Green's function $G(\mathbf{x}|\mathbf{x}_s;\omega;c)$ of an ocean acoustic waveguide in the frequency domain describes the acoustic field of a harmonic point source of unit strength and satisfies the following inhomogeneous Helmholtz equation:

$$\left[\nabla^2 + \frac{\omega^2}{c^2(\mathbf{x})}\right] G(\mathbf{x}|\mathbf{x}_s;\omega;c) = -\delta(\mathbf{x}-\mathbf{x}_s), \tag{1}$$

where $\mathbf{x}$ is the space vector, ω the circular frequency of the source, and $\mathbf{x}_s$ its location, $c(\mathbf{x})$ the sound-speed distribution, and δ the Dirac delta function. The Laplacian operator is denoted by ∇^2 (the symbol Δ is reserved to denote variations in the following).

Equation (1) is supplemented by boundary and interface conditions according to which G vanishes at the sea surface whereas pressure and normal velocity are continuous across interfaces, as well as by a radiation condition according to which the field decays away from the source and consists of a system of outgoing waves.[25]

The acoustic field P of a source distribution $S(\mathbf{x};\omega)$, satisfying the inhomogeneous Helmholtz equation

$$\left[\nabla^2 + \frac{\omega^2}{c^2(\mathbf{x})}\right] P(\mathbf{x};\omega) = S(\mathbf{x};\omega) \tag{2}$$

and the same boundary/interface/radiation conditions as before, can be represented through the Green's function by the integral[28]

$$P(\mathbf{x},\omega) = -\int\int\int_V G(\mathbf{x}|\mathbf{x}';\omega;c) S(\mathbf{x}';\omega)\, dV(\mathbf{x}'), \tag{3}$$

i.e., it is a superposition of the acoustic fields of point sources distributed over the support of $S(\mathbf{x};\omega)$.

The acoustic pressure field in the time domain can be expressed through the inverse Fourier transform

$$p(\mathbf{x},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P(\mathbf{x},\omega) e^{j\omega t} d\omega. \tag{4}$$

In particular, the acoustic field of a point source in the time domain can be expressed through the inverse Fourier transform in terms of the source signal $P_s(\omega)$ in the frequency domain, i.e., $S(\mathbf{x};\omega) = -P_s(\omega)\delta(\mathbf{x}-\mathbf{x}_s)$, and the frequency-domain Green's function

$$p(\mathbf{x},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\mathbf{x}|\mathbf{x}_s;\omega;c) P_s(\omega) e^{j\omega t} d\omega. \tag{5}$$

Due to multipath propagation, the pressure amplitude at a fixed receiver location $\mathbf{x}=\mathbf{x}_r$ in the time domain consists in general of a number of peaks, the acoustic arrivals, whose shape and temporal locations are dependent on the source characteristics, the source/receiver locations, and the sound-speed distribution within the water column. In this connection, the function $a(t)=|p(\mathbf{x}_r,t)|$ is called arrival pattern at the receiver.

A. Green's function perturbations—Born approximation

Since the Green's function is dependent on the sound-speed distribution, perturbations of the latter will give rise to changes in the Green's function. Expressions relating the perturbations of the Green's function to the underlying sound-speed perturbations will be derived in the following.

Let a background (reference) state be characterized by a sound-speed distribution $c_0(\mathbf{x})$ with corresponding Green's function $G_0(\mathbf{x}|\mathbf{x}_s) = G(\mathbf{x}|\mathbf{x}_S;\omega;c_0)$ satisfying the inhomogeneous Helmholtz equation

$$\left[\nabla^2 + \frac{\omega^2}{c_0^2(\mathbf{x})}\right] G_0(\mathbf{x}|\mathbf{x}_s) = -\delta(\mathbf{x}-\mathbf{x}_s), \tag{6}$$

and the above-mentioned boundary/interface/radiation conditions.

A perturbation of the reference sound-speed by $\epsilon\Delta c$, where ϵ is a small parameter,[29] will cause a perturbation ΔG in the Green's function. The perturbed Green's function $G = G_0 + \Delta G$ satisfies the inhomogeneous Helmholtz equation

$$\left[\nabla^2+\frac{\omega^2}{[c_0(\mathbf{x})+\epsilon\Delta c(\mathbf{x})]^2}\right][G_0(\mathbf{x}|\mathbf{x}_s)+\Delta G(\mathbf{x}|\mathbf{x}_s)]$$
$$=-\delta(\mathbf{x}-\mathbf{x}_s), \tag{7}$$

and the same additional conditions as before. By subtracting Eq. (6) from Eq. (7) and adding the term $\omega^2\Delta G/c^2$ to both sides, the following equation is obtained:

$$\left[\nabla^2+\frac{\omega^2}{c_0^2(\mathbf{x})}\right]\Delta G(\mathbf{x}|\mathbf{x}_s)=-\left[\frac{\omega^2}{[c_0(\mathbf{x})+\epsilon\Delta c(\mathbf{x})]^2}-\frac{\omega^2}{c_0^2(\mathbf{x})}\right]\times[G_0(\mathbf{x}|\mathbf{x}_s)+\Delta G(\mathbf{x}|\mathbf{x}_s)]. \tag{8}$$

The perturbation ΔG satisfies the same boundary, interface, and radiation conditions as the unperturbed Green's function G_0, whereas the operators on the left-hand side of Eqs. (6) and (8) are identical ($\nabla^2+\omega^2/c_0^2$). In this connection, by considering the right-hand side of Eq. (8), as a function of $\mathbf{x}$, to be a distributed source term, the integral representation (3) can be used to express the solution of Eq. (8) as follows:

$$\Delta G(\mathbf{x}|\mathbf{x}_s)=\int\int\int_V G_0(\mathbf{x}|\mathbf{x}')\times\left[\frac{\omega^2}{[c_0(\mathbf{x}')+\epsilon\Delta c(\mathbf{x}')]^2}-\frac{\omega^2}{c_0^2(\mathbf{x}')}\right]\times[G_0(\mathbf{x}'|\mathbf{x}_s)+\Delta G(\mathbf{x}'|\mathbf{x}_s)]dV(\mathbf{x}'). \tag{9}$$

This is an integral equation for the perturbation ΔG of the Green's function. Expanding the expression in brackets up to the second order with respect to ϵ

$$\frac{\omega^2}{[c_0+\epsilon\Delta c]^2}-\frac{\omega^2}{c_0^2}=\omega^2\left(-\frac{2\epsilon}{c_0^3}\Delta c+\frac{3\epsilon^2}{c_0^4}\Delta c^2+O(\epsilon^3)\right),$$

and using an expansion of the Green's function perturbation ΔG with respect to ϵ,

$$\Delta G(\mathbf{x}|\mathbf{x}_s)=\epsilon\Delta G_1(\mathbf{x}|\mathbf{x}_s)+\epsilon^2\Delta G_2(\mathbf{x}|\mathbf{x}_s)+O(\epsilon^3), \tag{10}$$

Eq. (9) can be written in the form (up to second order):

$$\epsilon\Delta G_1(\mathbf{x}|\mathbf{x}_s)+\epsilon^2\Delta G_2(\mathbf{x}|\mathbf{x}_s)=\omega^2\int\int\int_V G_0(\mathbf{x}|\mathbf{x}')\times\left[-\frac{2\Delta c(\mathbf{x}')}{c_0^3(\mathbf{x}')}\epsilon+\frac{3\Delta c^2(\mathbf{x}')}{c_0^4(\mathbf{x}')}\epsilon^2\right]\times[G_0(\mathbf{x}'|\mathbf{x}_s)+\epsilon\Delta G_1(\mathbf{x}'|\mathbf{x}_s)+\epsilon^2\Delta G_2(\mathbf{x}'|\mathbf{x}_s)]dV(\mathbf{x}'). \tag{11}$$

Equating terms of equal order, expressions can be obtained for the terms in the expansion (10) of the Green's function perturbation.

First order (ϵ):

$$\Delta G_1(\mathbf{x}|\mathbf{x}_s)=-2\omega^2\int\int\int_V G_0(\mathbf{x}'|\mathbf{x}_s)G_0(\mathbf{x}|\mathbf{x}')\frac{\Delta c(\mathbf{x}')}{c_0^3(\mathbf{x}')}dV(\mathbf{x}'). \tag{12}$$

This is the first Born approximation[30,31] expressing the first-order perturbation ΔG_1 of the Green's function as linear functional of the underlying sound-speed perturbation Δc. The kernel $G_0(\mathbf{x}'|\mathbf{x}_s)G_0(\mathbf{x}|\mathbf{x}')$ represents a single scattering mechanism, in which a scatterer (sound-speed perturbation) at the position $\mathbf{x}'$, stimulated by the source at position $\mathbf{x}_s$, with stimulation magnitude $G_0(\mathbf{x}'|\mathbf{x}_s)$ acts as a secondary source whose acoustic field $G_0(\cdot|\mathbf{x}')$ is observed at the point $\mathbf{x}$. In this connection the first Born approximation is also called single-scattering approximation. The approximation (12) represents efficiently the perturbations caused by very weak scatterers and due to this it is alternatively called weak-scattering approximation. The volume V in Eq. (12) spans the support of the sound-speed perturbation δc.

Second order (ϵ^2):

$$\Delta G_2(\mathbf{x}|\mathbf{x}_s)=-2\omega^2\int\int\int_V\Delta G_1(\mathbf{x}'|\mathbf{x}_s)G_0(\mathbf{x}|\mathbf{x}')\times\frac{\Delta c(\mathbf{x}')}{c_0^3(\mathbf{x}')}dV(\mathbf{x}')+3\omega^2\int\int\int_V G_0(\mathbf{x}'|\mathbf{x}_s)G_0(\mathbf{x}|\mathbf{x}')\frac{\Delta c^2(\mathbf{x}')}{c_0^4(\mathbf{x}')}dV(\mathbf{x}'). \tag{13}$$

This is the second Born approximation[30] expressing the second-order perturbation ΔG_2 as quadratic functional of the underlying sound-speed perturbation. While the second integral represents a single-scattering mechanism applying on Δc^2, the kernel of the first integral

$$\Delta G_1(\mathbf{x}'|\mathbf{x}_s)G_0(\mathbf{x}|\mathbf{x}')=-2\omega^2\int\int\int_V G_0(\mathbf{x}''|\mathbf{x}_s)G_0(\mathbf{x}'|\mathbf{x}'')G_0(\mathbf{x}|\mathbf{x}')\times\frac{\Delta c(\mathbf{x}'')}{c^3(\mathbf{x}'')}dV(\mathbf{x}''), \tag{14}$$

represents a double-scattering mechanism: the source stimulates a scatterer at position $\mathbf{x}''$ which then stimulates a scatterer at position $\mathbf{x}'$ which is finally received at position $\mathbf{x}$. In this connection the second Born approximation is also called double-scattering approximation.

B. Green's function perturbations—Rytov approximation

An alternative representation of the perturbed Green's function was introduced by Rytov[24] in the form

$$G=G_0e^{\Delta\Psi}. \tag{15}$$

This representation emphasizes the phase perturbation $\Delta\Psi$. Taking into account that phase perturbations in the frequency domain reflect in wave form shifts in the time domain,[32] i.e., in arrival-time perturbations, the Rytov approximation is expected to be suitable for time-domain (arrival-time) calculations. Expanding the phase perturbation with respect to ϵ,

$$\Delta\Psi=\epsilon\Delta\Psi_1+\epsilon^2\Delta\Psi_2+O(\epsilon^3), \tag{16}$$

and using a Taylor expansion of Eq. (15) in the neighborhood of the unperturbed state ($\epsilon=0$) the perturbed Green's function can be written in the form

$$G = G_0\Big(1 + \epsilon\Delta\Psi_1 + \epsilon^2\Delta\Psi_2 + \frac{1}{2}(\epsilon\Delta\Psi_1 + \epsilon^2\Delta\Psi_2 + O(\epsilon^2))^2 + O(\epsilon^3)\Big) = G_0 + \epsilon G_0\Delta\Psi_1 + \epsilon^2 \times\Big(G_0\Delta\Psi_2 + G_0\frac{\Delta\Psi_1^2}{2}\Big) + O(\epsilon^3). \quad (17)$$

Equating the factors of corresponding orders in Eqs. (17) and (10) the following relations can be obtained between the terms of the Born and Rytov approximation:[33–35]

$$\Delta G_1 = G_0\Delta\Psi_1, \quad (18)$$

$$\Delta G_2 = G_0\Big(\Delta\Psi_2 + \frac{\Delta\Psi_1^2}{2}\Big). \quad (19)$$

This means that if the terms of the Rytov approximation are known the corresponding terms of the Born approximation can be calculated and vice versa:

$$\Delta\Psi_1 = \frac{\Delta G_1}{G_0}, \quad (20)$$

$$\Delta\Psi_2 = \frac{\Delta G_2}{G_0} - \frac{1}{2}\Big(\frac{\Delta G_1}{G_0}\Big)^2. \quad (21)$$

Thus from expressions (12) and (13) for the Born approximation, expressions for the corresponding terms of the Rytov approximation can be obtained though Eqs. (20) and (21). This is because the two approximations, though based on different types of expansions, have the same asymptotic ($\epsilon \to 0$) behavior. In this sense they are closely related to each other. Nevertheless, they are not equally efficient in describing travel-time variations, as will become clear from the numerical results in Sec. IV, due to their different form: The Born approximation focuses on variations of the Green's function itself, whereas the Rytov approximation focuses on variations of the phase.

III. NORMAL-MODE REPRESENTATION

Assuming the background ocean to be range independent, the background Green's function G_0 can be represented in terms of normal modes. Adopting a cylindrical coordinate system (r, z, θ) with origin at the sea surface and the source located on the vertical z axis (positive downwards) at depth $z = z_s$, G_0 at any location (r, z) in the water is expressed in the form[25,36]

$$G_0(r,z|z_s) = \frac{-j}{4\rho_w}\sum_{n=1}^{M}\phi_n(z_s)\phi_n(z)H_0^{(2)}(k_n r) + \int_0^{\omega/c_B}\mathcal{A}(k;r,z;z_s)dk + \int_0^{j\infty}\mathcal{B}(k;r,z;z_s)dk, \quad (22)$$

where ρ_w is the water density, $H_0^{(2)}$ is the Hankel function of the second kind and zeroth order, k_n and $\phi_n, n = 1, \ldots, M$, are the real eigenvalues and the corresponding eigenfunctions (propagating modes) of the vertical Sturm-Liouville problem:

$$\frac{d^2\phi_n(z)}{dz^2} + \frac{\omega^2}{c^2(z)}\phi_n(z) = k_n^2\phi_n(z), \quad (23)$$

supplemented by the conditions that $\phi_n = 0$ at the sea surface $(z = 0)$, ϕ_n and $\rho^{-1}d\phi_n/dz$ are continuous across the interfaces, and ϕ_n and $d\phi_n/dz$ are vanishing as $z \to \infty$.

The sum in Eq. (22) represents the contribution of the finite set of propagating modes with $\omega/c_B < k_n < \omega/c_{\min}$, $n = 1, 2, \ldots, M$, where $c_{\min}$ is the minimum sound speed in the water and c_B is the sound speed in the bottom half-space (c_B is assumed to be constant and also the highest sound speed in the propagation domain). The first integral in Eq. (22) represents the contribution of the half-space modes (high-order modes with grazing angle greater than critical entering the bottom half-space) whereas the second integral spans the evanescent spectrum (modes with imaginary k values and exponentially decaying contribution).[36] Both integrals are negligible in the water layer away from the source.

In the following we focus on the perturbation behavior of the low-order modes, with k_n close to $\omega/c_{\min}$, contributing to the late arrival pattern. Taking into account that the derivatives of modes (eigenvalues and eigenfunctions solving the vertical Sturm-Liouville problem) with respect to sound-speed perturbations can be expressed in terms of background eigenvalues and eigenfunctions, with the nearby modes (closest in terms of eigenvalues) playing the dominant role,[26] higher order half-space and evanescent modes will be omitted from the representation of the Green's function.

A. First Born approximation

Substituting the normal-mode expression for the background Green's function in the right-hand side of Eq. (12) and considering sound-speed perturbations in the water column of separable form

$$\Delta c(\mathbf{x}) = \Delta c_r(r)\Delta c_z(z)\Delta c_\theta(\theta), \quad (24)$$

where $\Delta c_r, \Delta c_z$, and Δc_θ are smooth, slowly varying functions of r, z, and θ, respectively, the first Born approximation can be written as

$$\Delta G_1(\mathbf{x}_r|\mathbf{x}_s) = \frac{\omega^2}{8\rho_w^2}\sum_{n=1}^{M}\sum_{m=1}^{M}\phi_n(z_s)\phi_m(z_r)\int_0^h\phi_n(z)\phi_m(z) \times\frac{\Delta c_z(z)}{c_0^3(z)}dz\int_0^{2\pi}\int_0^{\infty}\Delta c_\theta(\theta)\Delta c_r(r) \times H_0^{(2)}(k_n r)H_0^{(2)}(k_m\gamma)rdrd\theta, \quad (25)$$

where h is the water depth, $\gamma = \sqrt{r^2 + R^2 - 2Rr\cos\theta}$ is the horizontal distance from an arbitrary scattering location to the receiver, and R is the horizontal source-receiver distance. The use of the same axisymmetric Green's function to describe propagation from the source to the scattering point and from that point to the receiver is justified by the fact that the background environment is range-independent and thus axisymmetric with respect to any vertical axis including the source, the receiver, or the ar-

bitrary scattering point. In the following the double integral in line of Eq. (25), denoted I_{nm}, is evaluated.

Since the Hankel functions are singular for $r=0$ and $\gamma=0$, i.e., at the location of the source and the receiver, the integration domain is divided into three subdomains: two disks $\Gamma_{s,\beta}$ and $\Gamma_{r,\beta}$ of radius β centered at the source and receiver, respectively, over which the integral I_{nm} is evaluated analytically, and the remaining part S_β of the plane, where I_{nm} is evaluated by the stationary-phase method exploiting the oscillatory behavior of the kernel.

In the vicinity of the source we can assume that $\Delta c_\theta(\theta)$ and $\Delta c_r(r)$ are constants represented by Δc_θ and Δc_r. Using the addition theorem[37] the Hankel function $H_0^{(2)}(k_m\gamma)$ for $r<R$ can be expressed in terms of Bessel functions of the first kind and Hankel functions as follows:

$$H_0^{(2)}(k_m\gamma) = J_0(k_m r)H_0^{(2)}(k_m R) + 2\sum_{\ell=1}^{\infty} J_\ell(k_m r)H_\ell^{(2)}(k_m R) \times \cos(\ell\theta). \tag{26}$$

Thus the integral I_{nm} over a disk $\Gamma_{s,\beta}$ becomes

$$I_{nm}(\Gamma_{s,\beta}) = 2\pi\Delta c_\theta \Delta c_r H_0^{(2)}(k_m R)\int_0^\beta J_0(k_m r)H_0^{(2)}(k_n r) r dr. \tag{27}$$

The latter integral is first evaluated over an interval $[a,\beta]$, with $a\neq 0$,[38] and then the limit $a\to 0$ is taken. The final result reads

$$I_{nm}(\Gamma_{s,\beta}) = 2\Delta c_r \Delta c_\theta H_0^{(2)}(k_m R) \times \begin{cases} \dfrac{\beta}{k_n} & \text{for } n=m \\ \dfrac{2j}{k_m^2-k_n^2} + \text{o.t.} & \text{for } n\neq m \end{cases} \tag{28}$$

where o.t. stands for an oscillating term with respect to β averaging to zero. Δc_r and Δc_θ are taken at the source location. Similar expressions can be derived for $I_{nm}(\Gamma_r,\beta)$ over the disk centered at the receiver.

Assuming that the radius β is large enough the asymptotic expression for the Hankel functions can be used away from the source and receiver, such that the integral I_{nm} in the exterior domain S_β takes the form

$$I_{nm}(S_\beta) = \frac{2j}{\pi\sqrt{k_n k_m}}\int\int_{S_\beta} \Delta c_\theta(\theta)\Delta c_r(r) \times \frac{\exp(-j(k_m r + k_n\sqrt{r^2+R^2-2rR\cos\theta}))}{\sqrt{r\sqrt{r^2+R^2-2rR\cos\theta}}} r dr d\theta. \tag{29}$$

The exponential part in the kernel of this integral is a rapidly oscillating function of θ, see, e.g., Fig. 3 in Sec. IV. In this connection the method of stationary phase[39] can be applied for the evaluation of $I_{nm}(S_\beta)$. The phase is

$$\Phi_{nm} = k_m r + k_n\sqrt{r^2+R^2-2Rr\cos\theta} \tag{30}$$

and the stationary points are: $\theta=0$ and $\theta=\pi$. Using the stationary-phase formula for the θ-integral, $I_{nm}(S_\beta)$ can be approximated by

$$I_{nm}(S_\beta) = \frac{2\sqrt{2}}{k_n\sqrt{\pi k_m R}}\Bigg[\Delta c_\theta(0)e^{-j(k_n R-\pi/4)} \times \int_\beta^{R-\beta} e^{-j(k_m-k_n)r}\Delta c_r(r)dr + \Delta c_\theta(0)e^{j(k_n R+\pi/4)} \times \int_{R+\beta}^{\infty} e^{-j(k_m+k_n)r}\Delta c_r(r)dr - \Delta c_\theta(\pi)e^{-j(k_n R+\pi/4)} \times \int_\beta^{R-\beta} e^{-j(k_m+k_n)r}\Delta c_r(r)dr - \Delta c_\theta(\pi)e^{-j(k_n R+\pi/4)} \times \int_{R+\beta}^{\infty} e^{-j(k_m+k_n)r}\Delta c_r(r)dr\Bigg]. \tag{31}$$

This expression describes the range integration along the semiaxis $\theta=0$ from the source to the receiver (first integral) and beyond (second integral) as well as along the semiaxis $\theta=\pi$ (third and fourth integrals). The exponential kernels in Eq. (31) are all oscillatory except for the one in the first line for $m=n$. For $m\neq n$ this kernel oscillates with wavelength determined by the wave number difference k_m-k_n which becomes smallest for successive wave numbers, $n=m+1$, in which case the corresponding wavelengths coincide with the double-loop length of the corresponding rays,[27] see e.g. Fig. 4 in Sec. IV. In the following we assume range-dependent perturbations with horizontal scales large compared with the double-loop length of the corresponding rays (adiabatic range dependence), such that there is no contribution by the cross terms in the first integral—the integration result will be an oscillating term with respect to β averaging to zero. The remaining three integrals in Eq. (31) will have an even smaller contribution since the corresponding kernels oscillate at a higher rate governed by the wave number sums k_m+k_n. Thus, assuming weak (adiabatic) range dependence the term $I_{nm}(S_\beta)$ can be expressed as

$$I_{nm}(S_\beta) = \delta_{nm}2\Delta c_\theta(0)\frac{H_0^{(2)}(k_n R)}{k_n}\int_\beta^{R-\beta}\Delta c_r(r)dr + \text{o.t.}, \tag{32}$$

where δ_{nm} is the Kronecker delta, and the Hankel function has been restored from its asymptotic representation. Combining the expressions for $I_{nm}(\Gamma_s,\beta)$ and $I_{nm}(\Gamma_r,\beta)$, Eq. (28), with the above-noted expression for $I_{nm}(S_\beta)$ and omitting oscillating terms we finally obtain for I_{nm},

$$I_{nm} = \begin{cases} 2\Delta c_\theta(0)\dfrac{H_0^{(2)}(k_n R)}{k_n}\displaystyle\int_0^R \Delta c_r(r)dr & \text{for n = m} \\ 4j\Delta c_\theta(0)\left[\dfrac{H_0^{(2)}(k_m R)}{k_m^2-k_n^2}\Delta c_r(0) + \dfrac{H_0^{(2)}(k_n R)}{k_n^2-k_m^2}\Delta c_r(R)\right] & \text{for } n\neq m. \end{cases} \tag{33}$$

Substituting this expression into Eq. (25), the first-order Born approximation is finally written in the compact form

$$\Delta G_1(\mathbf{x}_r|\mathbf{x}_s) = -\frac{j\Delta c_\theta(0)}{4\rho_w}\sum_{n=1}^{m}\left\{\sum_{\substack{m=1\\ m\neq n}}^{M}\frac{Q_{nm}V_{nm}}{\Lambda_{nm}} + j\frac{Q_{nm}U_n}{k_n}\int_0^R \Delta c_r(r)dr\right\}H_0^{(2)}(k_nR), \quad (34)$$

where

$$\Lambda_{nm} = k_n^2 - k_m^2, \quad (35)$$

$$Q_{nm} = -\frac{2\omega^2}{\rho_w}\int_0^h \phi_n(z)\phi_m(z)\frac{\Delta c_z(z)}{c_0^3(z)}dz, \quad (36)$$

$$U_n = -\phi_n(z_s)\phi_n(z_r)/2, \quad (37)$$

$$V_{nm} = \phi_m(z_s)\phi_n(z_r)\Delta c_r(0) + \phi_n(z_s)\phi_m(z_r)\Delta c_r(R). \quad (38)$$

Expression (34) is a generalization of previous range-independent results.[26] The integral term in expression (34) is proportional to R, e.g., in the case of range-independent perturbations, but vanishes in the case of zero-mean range-dependent perturbations ($\int_0^R \Delta c_r(r)dr=0$). Equation (34) holds under the assumption of large-scale (adiabatic) range dependence, where the scale of Δc_r is large compared with the double-loop length of the corresponding rays.

The first term in expression (34), including the sum over m, is dominated by the orders m close to n for which the denominator Λ_{nm} becomes small. This explains the small contribution of the high-order half-space and evanescent modes to the perturbation behavior of the low-order modes (small n): the half-space modes (large m) are characterized by large Λ_{nm}, whereas for the evanescent modes (k_m imaginary) the differences Λ_{nm} become sums of the form $k_n^2 + |k_m|^2$ and thus even larger. As will become clear in Sec. III C the integral term in Eq. (34) is associated with travel-time variations, whereas the sum term is related with amplitude changes in the time domain—still, it gives rise to terms, associated with travel-time changes in the second order.

B. Second Born approximation

Substituting the normal-mode representation for the background Green's function and expression (34) for the first Born approximation into the first term in Eq. (13), denoted T_1, we obtain

$$T_1 = \int_0^h\int_0^{2\pi}\int_0^\infty \sum_{\ell=1}^{M}\phi_\ell(z)\phi_\ell(z_r) \times H_0^{(2)}(k_\ell\sqrt{r^2+R^2-2rR\cos\theta})\left(-\frac{j}{4\rho_w}\right)^2\Delta c_\theta(0) \times\left[-\frac{2\omega^2}{c_0^3(z)}\Delta c_z(z)\Delta c_\theta(\theta)\Delta c_r(r)\right] \times\sum_{n=1}^{M}\left\{\sum_{\substack{m=1\\ m\neq n}}^{M}\frac{Q_{nm}V_{nm}}{\Lambda_{nm}} + j\frac{Q_{nn}U_n}{k_n}\int_0^r \Delta c_r(r')dr'\right\} \times H_0^{(2)}(k_nr)rdrd\theta dz. \quad (39)$$

The integrals with respect to (r,θ) in the above-noted expression are of the general form

$$K_{n\ell} = \int_0^{2\pi}\int_0^\infty H_0^{(2)}(k_nr)H_0^{(2)}\left(k_\ell\sqrt{r^2+R^2-2rR\cos\theta}\right) \times\Delta c_\theta(\theta)F(r)rdrd\theta, \quad (40)$$

where F is a smooth slowly varying function of r. This integral is of the same type as the integral I_{nm} evaluated in the previous section. Applying the same method (analytical calculation close to the source/receiver, and stationary phase in the far field) the integral K_{nl} integral can be evaluated,

$$K_{n\ell} = \begin{cases} 2\Delta c_\theta(0)\dfrac{H_0^{(2)}(k_nR)}{k_n}\displaystyle\int_0^R F(r)dr & \text{for } k_n = k_\ell \\ 4j\Delta c_\theta(0)\left[\dfrac{H_0^2(k_\ell R)}{k_\ell^2-k_n^2}F(0) + \dfrac{H_0^2(k_nR)}{k_n^2-k_\ell^2}F(R)\right] & \text{for } k_n \neq k_\ell. \end{cases} \quad (41)$$

Using this result T_1 can be finally expressed in the form

$$T_1 = \frac{\Delta c_\theta^2(0)}{4\rho_w}\sum_{n=1}^{M}\sum_{\substack{m=1\\ m\neq n}}^{M}\frac{|Q_{nm}|^2U_n}{k_n(k_n^2-k_m^2)}H_0^{(2)}(k_nR)\int_0^R \Delta c_r^2(r)dr - \frac{j\Delta c_\theta^2(0)}{4\rho_w}\sum_{n=1}^{M}\frac{|Q_{nn}|^2U_n}{2k_n^2}H_0^{(2)}(k_nR) \times\int_0^R \Delta c_r(r)\int_0^r \Delta c_r(r')dr'dr - \frac{\Delta c_\theta^2(0)}{4\rho_w}\sum_{n=1}^{M}\sum_{\substack{m=1\\ m\neq n}}^{M}\frac{Q_{nn}Q_{nm}V_{nm}}{2k_n(k_n^2-k_m^2)}H_0^{(2)}(k_nR) \times\int_0^R \Delta c_r(r)dr. \quad (42)$$

In this expression only the terms proportional to R (double sums) and R^2 (single sum) are retained. The remaining (local) terms are of significance, with respect to travel-time variations, for the third order and higher, but not for the second order studied here.

Substituting the normal-mode representation for the background Green's function into the second term in Eq. (13), denoted T_2, and applying Eq. (41) to the integral with respect to (r,θ), T_2 finally becomes

$$T_2 = \frac{\Delta c_\theta^2(0)}{4\rho_w}\int_0^R \Delta c_r^2(r)dr\sum_{n=1}^{M}\frac{U_nQ_n'}{k_n}H_0^{(2)}(k_nR), \quad (43)$$

where

$$Q_n' = \frac{3\omega^2}{\rho_w}\int_0^h \phi_n^2(z)\frac{\Delta c_z^2(z)}{c_0^4(z)}dz, \quad (44)$$

where again the sum over the off-diagonal terms ($m\neq n$) has been omitted since it is not significant for the second order.

Combining relations (42) and (43) we obtain the following expression for the second-order term of the Born approximation.

$$\Delta G_2(\mathbf{x}_r|\mathbf{x}_s) = \frac{-j\Delta c_\theta^2(0)}{4\rho_w}\sum_{n=1}^{M}\left\{\frac{-jQ_{nn}}{2k_n}\sum_{\substack{m=1\\m\neq n}}^{M}\frac{Q_{nm}V_{nm}}{\Lambda_{nm}}\int_0^R \Delta c_r(r)dr + \frac{|Q_{nn}|^2U_n}{2k_n^2}\int_0^R \Delta c_r(r)\int_0^r \Delta c_r(r')dr'dr + \frac{jU_n}{k_n} \times\left(\sum_{\substack{m=1\\m\neq n}}^{M}\frac{|Q_{nm}|^2}{\Lambda_{nm}} + Q_n'\right)\int_0^R \Delta c_r^2(r)dr\right\}H_0^{(2)}(k_nR). \quad (45)$$

As in the case of the first Born approximation this is a generalization of previous results obtained for range-independent perturbations[26] and holds under the assumption of adiabatic range dependence. The above-presented expression contains the dominating terms, including the factors R and R^2, whereas the remaining terms (not essential for the second-order approximation) have been omitted.

C. Rytov-Keller approximation

By substituting expressions (34) and (45) for the first and second Born approximations into Eqs. (20) and (21), expressions for the first and second Rytov approximations can be obtained. To cope with the multimodal nature of ocean acoustic propagation, a variation of the standard Rytov method is applied, which was proposed by Keller.[40]

The standard Rytov approximation is sufficient for perturbations of single-component wave fields but fails in the case of multiple components such as in ocean acoustic propagation.[31] The reason is that each field component (mode) has its own phase, with different perturbation behavior, whereas the Rytov approximation assumes that the perturbed wave field can be described by a single phase perturbation. The variation proposed by Keller consists in applying the standard Rytov method to each wave (modal) component independently, rather than to the total wave field.[40] In this connection, the perturbed Green's function is written as follows:

$$G_{\mathrm{RK}} = \sum_{n=1}^{M} G_{0_n} e^{\Delta\Psi_n}, \quad (46)$$

where G_{0_n} is the contribution of the nth mode to the unperturbed Green's function. Based on Eqs. (20) and (34) the first-order phase variation for the nth mode according to the Rytov-Keller approximation is given from

$$\Delta\Psi_{1_n} = \frac{-j\Delta c_\theta(0)}{4\rho_w G_{0_n}}\left(\sum_{\substack{m=1\\m\neq n}}^{M}\frac{Q_{nn}V_{nm}}{\Lambda_{nm}} + j\frac{Q_{nn}U_n}{k_n}\int_0^R \Delta c_r(r)dr\right) \times H_0^{(2)}(k_nR), \quad (47)$$

where $Q_{nm}, V_{nm}, \Lambda_{nm}, U_n$ are the quantities defined in Eqs. (35)–(38). Using the normal-mode expression (22) for the background Green's function G_{0_n} the first-order phase perturbation of the nth mode can be finally written in the form

$$\Delta\Psi_{1_n} = \frac{-\Delta c_\theta(0)}{2U_n}\sum_{\substack{m=1\\m\neq n}}^{M}\frac{Q_{nm}V_{nm}}{\Lambda_{nm}} - j\frac{\Delta c_\theta(0)Q_{nn}}{2k_n}\int_0^R \Delta c_r(r)dr. \quad (48)$$

This expression holds under the assumption of adiabatic range dependence. The first term in Eq. (48) in this expression is real and represents attenuation effects. The imaginary integral term represents phase variations in the frequency domain which are associated with displacements in the time domain (travel-time variations). In the case of zero-mean range-dependent perturbations this term vanishes, such that there are no first-order effects of range dependence on travel times. This is in agreement with ray-theoretic results for environments with adiabatic range dependence.[21,22]

The second-order phase perturbation of the nth mode according to the Rytov-Keller approximation is obtained by substituting the nth component of the first- and second-order Born terms, Eqs. (34) and (45), into the Born-Rytov relations Eq. (21),

$$\Delta\Psi_{2_n} = \frac{-j\Delta c_\theta^2(0)}{4G_{0_n}\rho_w}\left\{\frac{-j}{2k_nQ_{nn}}\sum_{\substack{m=1\\m\neq n}}^{M}\frac{Q_{nm}V_{nm}}{\Lambda_{nm}}\int_0^R \Delta c_r(r)dr + \frac{|Q_{nn}|^2U_n}{2k_n^2}\int_0^R \Delta c_r\int_0^r \Delta c_r(r')dr'dr + \frac{jU_n}{k_n}\left(\sum_{\substack{m=1\\m\neq n}}^{M}\frac{|Q_{nm}|^2}{\Lambda_{nm}} + Q_n'\right)\int_0^R \Delta c_r^2(r)dr\right\} \times H_0^{(2)}(k_nR) - \frac{1}{2}(\Delta\Psi_{1_n})^2. \quad (49)$$

Substituting the modal representation for G_{0n} we finally obtain

$$\Delta\Psi_{2_n} = \frac{j\Delta c_\theta^2(0)}{4U_nk_nQ_{nn}}\sum_{\substack{m=1\\m\neq n}}^{M}\frac{Q_{nm}V_{nm}}{\Lambda_{nm}}\int_0^R \Delta c_r(r)dr - \frac{\Delta c_\theta^2(0)|Q_{nn}|^2}{4k_n^2}\int_0^R \Delta c_r(r)\int_0^r \Delta c_r(r')dr'dr - \frac{j\Delta c_\theta^2(0)}{2k_n}\left(\sum_{\substack{m=1\\m\neq n}}^{M}\frac{|Q_{nm}|^2}{\Lambda_{nm}} + Q_n'\right)\int_0^R \Delta c_r^2(r)dr - \frac{1}{2}(\Delta\Psi_{1_n})^2. \quad (50)$$

The dominating term in this expression in the case of a zero-mean range perturbation is the imaginary term in the third and last line. All other terms either vanish or they are real, which means that they are associated with attenuation effects (no effect on travel times).

The second-order Rytov-Keller approximation has strong similarities to the second-order adiabatic approximation of the Green's function (see the Appendix). This is a

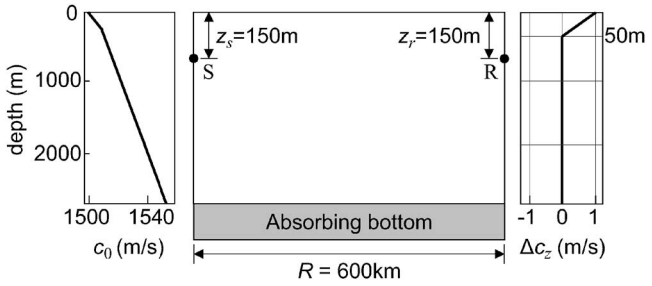


FIG. 1. Background sound-speed profile (left), problem geometry (middle), and depth mode $\Delta c_z(z)$ of the sound-speed perturbation (right).

consequence of the adiabatic assumptions made not only in the evaluation of the Born integrals (omission of cross terms) but also in the independent treatment of each modal component in the Keller approach.

IV. NUMERICAL RESULTS

This section presents some numerical examples for simple range-dependent ocean environments that are perturbations of a range-independent background state. The background sound-speed profile, shown in Fig. 1, is representative of winter conditions in the western Mediterranean sea. The water depth is taken 2500 m, the source and receiver depth 150 m, and the propagation range 600 km; these values are motivated from the Thetis-2 tomography experiment conducted from January to October 1994 in the Western Mediterranean.[41] The emitted signal is assumed to be a Gaussian pulse of central frequency 150 Hz and effective bandwidth 60 Hz. In the following calculations the Green's function (complex pressure), either exact or approximate, is evaluated at 501 frequencies from 100 to 200 Hz, with a step of 0.2 Hz, using a normal-mode code, and then fast Fourier transform is applied to obtain results in the time domain. An absorbing bottom is assumed filtering out the bottom-interacting part of the acoustic energy.[42]

Figure 2 shows the background arrival pattern corresponding to the reference profile of Fig. 1. Early arrivals have the form of distinct triplets corresponding to particular ray groups which propagate at steep grazing angles and thus sample the deep water layers. Late arrivals on the other hand correspond to acoustic energy propagating at low grazing angles, i.e., close to the surface, and thus they are affected most by near-surface range dependence. These arrivals are best described in terms of a few low-order modes.[4] In the following the calculations will concentrate on the late part of the arrival pattern shown in the bottom panel of Fig. 2.

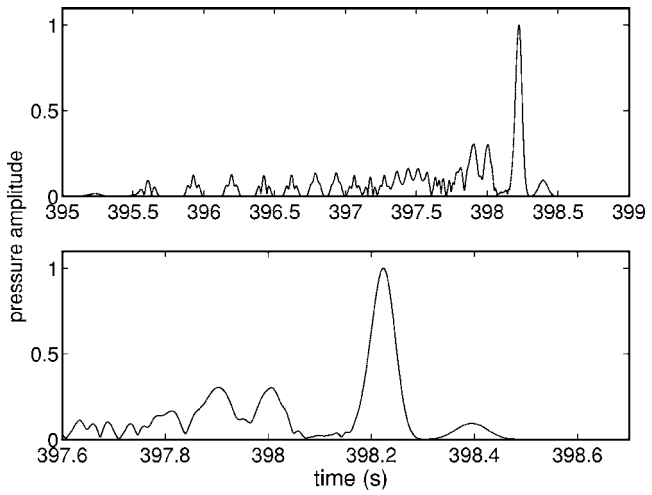


FIG. 2. Top panel: Background arrival pattern corresponding to the reference profile. Bottom panel: Late part of the background arrival pattern (detail).

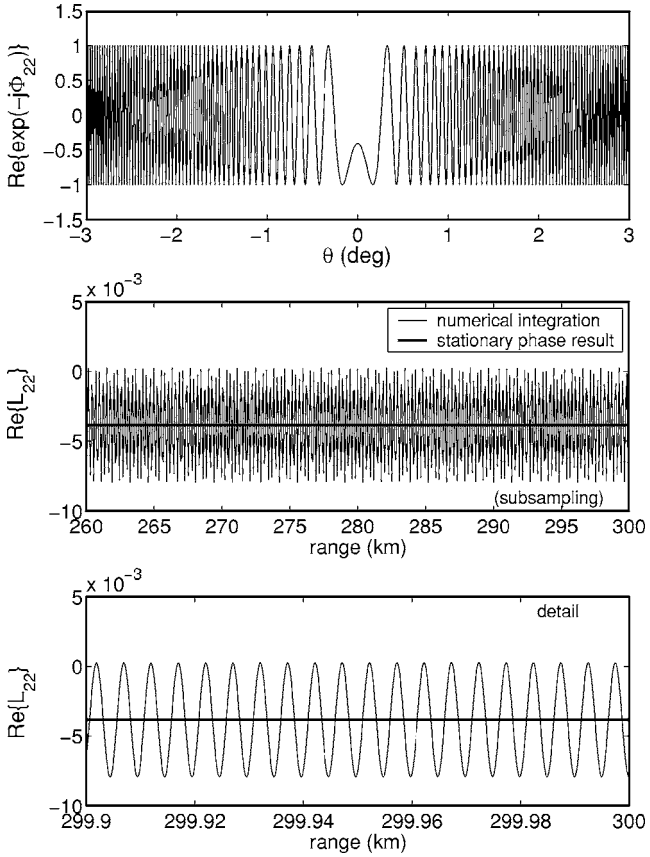


FIG. 3. Top panel: Real part of the exponential kernel of L_{22} vs angle in the vicinity of the stationary point for $r=300$ km. Middle panel: Numerical integration (real part) and stationary-phase evaluation (omitting oscillating terms) of L_{22} at various ranges from 260 to 300 km—the oscillation is subsampled. Bottom panel: Detailed view over a 100-m interval.

Figure 3 shows the evaluation of the integral

$$L_{nm} = \int_{-\pi}^{\pi} \frac{\exp(-j(k_m r + k_n \sqrt{r^2 + R^2 - 2rR\cos\theta}))}{\sqrt{r\sqrt{r^2 + R^2 - 2rR\cos\theta}}} r d\theta, \tag{51}$$

for the case $m=n=2$ corresponding to the highest peak in Fig. 2 (mode 2). L_{nm} represents the θ integral in Eq. (29) for $\Delta c_\theta = 1$. The upper panel in Fig. 3 shows the real part of the exponential term versus angle, over the interval $[-3°, 3°]$, for $r=300$ km. The stationarity of the phase at $\theta=0°$ is evident. The panel in the middle shows the real part of L_{22} evaluated through numerical integration and also through the stationary-phase approach (omitting the oscillating terms) for ranges from 260 to 300 km, whereas the bottom panel shows the same result in detail over a range interval

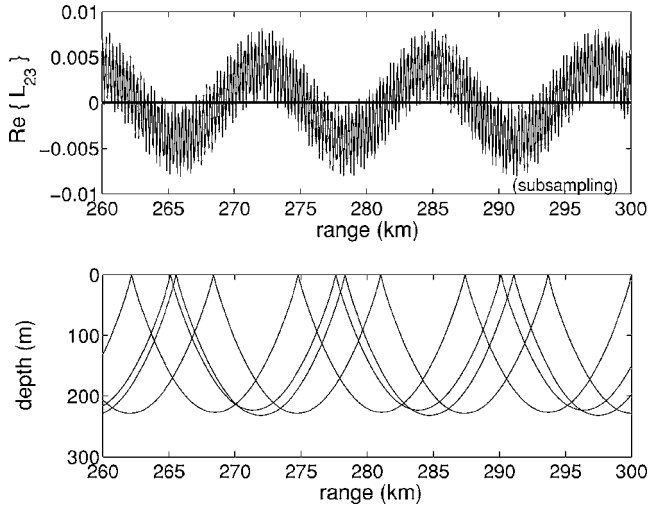


FIG. 4. Top panel: Numerical integration (real part) and stationary-phase evaluation (omitting oscillating terms) of L_{23} at various ranges from 260 to 300 km—the oscillation is subsampled. Bottom panel: Geometric rays corresponding to modes 2 and 3.

of 100 m length. The oscillation of L_{22} evaluated numerically is governed by twice the wave number k_2, cf. Eq. (31) [in this case $k_2=0.624\,541\ \mathrm{m}^{-1}$ resulting in a wavelength $2\pi/(2k_2)\approx 5$ m]. The oscillation in the middle panel is subsampled in range (using a range step of 79 m) in order to avoid a dark image, whereas in the lower panel the complete oscillation is shown. For large-scale perturbations Δc_r this oscillation is smoothed out by the integration in range carried out in Eq. (29), such that the remaining term coincides with the stationary-phase result.

Figure 4 (upper panel) shows the result of the numerical evaluation of L_{nm} for $n=2$ and $m=3$ for ranges from 260 to 300 km. In addition to the fast oscillation that we had in Fig. 3 (and which is again subsampled, as before) we also have a slow oscillation in this case, which is governed by the difference k_2-k_3, cf. Eq (31). In the particular case $k_2=0.624\,541\ \mathrm{m}^{-1}$ and $k_3=0.624\,046\ \mathrm{m}^{-1}$ such that the resulting wavelength is $2\pi/(k_2-k_3)=12\,693$ m. This wavelength corresponds to the double-loop length of the rays associated with modes 2 and 3 shown in the bottom panel of Fig. 4. From this figure it becomes clear that if the sound-speed perturbation has horizontal scale which is large compared with the double-loop length of the corresponding rays (adiabatic range dependence), the corresponding cross terms in Eq. (31) are smoothed out (cancelled) through the range integration, and thus they can be omitted.

Three cases of range-dependent perturbations are considered in the following. The first two are of large horizontal scale based on linear zero-mean and nonzero-mean range modes Δc_r combined with the depth mode Δc_z shown in the rightmost panel in Fig. 1. The particular depth mode is confined in the upper 50 m layer (range dependence is more pronounced close to the surface) attaining its maximum (1 m/s) at the surface, decreasing linearly to zero at 50 m depth and remaining zero thereafter. The third case concerns a sound-speed perturbation of small horizontal scale, for which the assumption of adiabatic range dependence does not hold.

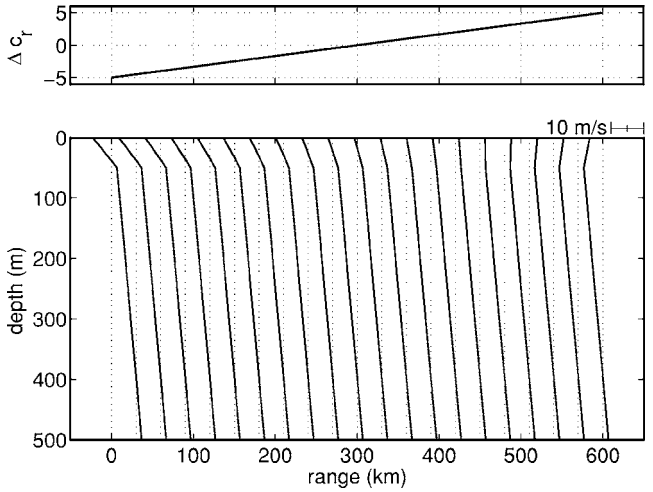


FIG. 5. Linear zero-mean range mode $\Delta c_r(r)$ of the sound-speed perturbation (top panel) and resulting sound-speed profiles at various ranges (bottom panel).

A. Large-scale, zero-mean range dependence

In the first numerical example the range mode Δc_r of the sound-speed perturbation is taken to be a linear function of range r, varying from −5 at the source to +5 at the receiver, Fig. 5 (top), and thus averaging to zero. The bottom panel of Fig. 5 shows the resulting sound-speed profiles at various ranges. The profiles close to the source are upward refracting with a strong surface duct whereas the ones close to the receiver form a weak channel at 50 m depth. The 10 m/s difference in the sound speed at the surface over the 600 km range corresponds to temperature gradients observed in the Western Mediterranean sea along the north-south axis.

Figure 6 shows the result of the exact adiabatic- and coupled-mode calculation of the perturbed arrival pattern corresponding to the range-dependent environment of Fig. 5; for these calculations the range-dependent environment was discretized into 21 range segments (piecewise constant discretization). The background arrival pattern is also shown in Fig. 6 for comparison. The difference between the adiabatic and coupled mode is very small indicating that propagation in this case can be considered as adiabatic. On the other hand the perturbed arrivals are significantly displaced with respect to their background positions: the late arrival (at 398.4 s) is delayed by as much as 70 ms whereas the earlier arrivals including the highest one (detail in upper right corner of the figure) are advanced by approximately 10 ms.

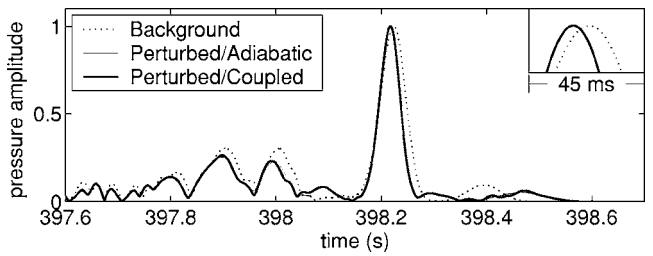


FIG. 6. Arrival pattern in the perturbed environment with linear zero-mean range dependence, predicted from exact adiabatic and coupled-mode calculation. Comparison with the background arrival pattern (dotted line). Upper right corner: Detail of the highest peak.

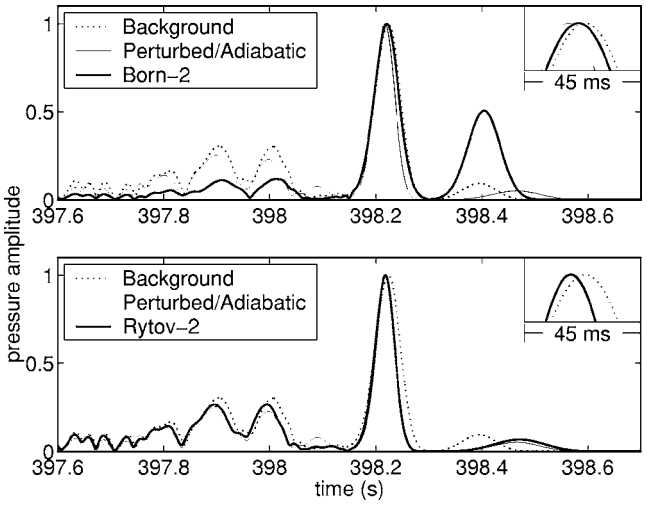


FIG. 7. Arrival pattern in the perturbed environment with linear zero-mean range dependence, predicted from the second Born approximation (top panel) and the second Rytov approximation (bottom panel). Comparison with exact adiabatic prediction (light solid line) and background arrival pattern (dotted line). Upper right corner: Detail of the highest peak.

Figure 7 shows the arrival pattern predicted from the second Born approximation (top) and the second Rytov approximation (bottom), together with the exact adiabatic prediction and the background arrival pattern. Since the range mode $\Delta c_r(r)$ of the sound-speed perturbation averages to zero, the first Born and first Rytov approximation results are practically the same as the background arrival pattern and in this connection they are not shown. The second-order Born approximation differs from the background arrival pattern in amplitude but hardly as far as the arrival times are concerned. Thus, the Born approximation fails to predict correct arrival times in the perturbed state. The second-order Rytov approximation, on the other hand, manages to describe efficiently the arrival shifts in nearly all cases. Thus, for the late arrival it reproduces the $\sim$70 ms delay, with respect to the background state, whereas for the earlier arrivals it reproduces the advancements. Further, the second-order Rytov approximation results in a prediction of the arrival amplitudes which is remarkably close to the exact prediction.

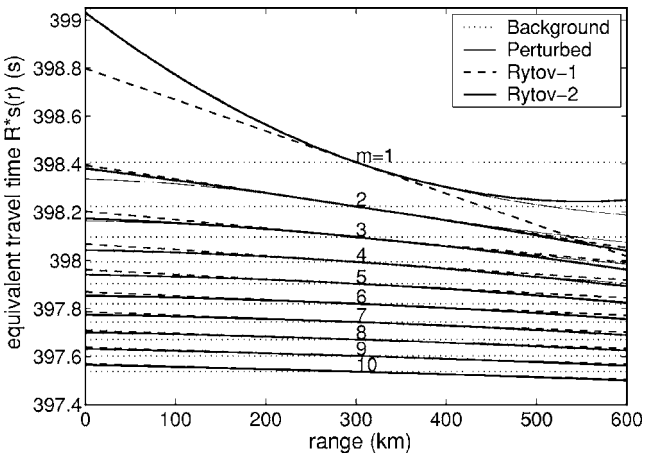


FIG. 8. Travel-time comparison for modes 1–10, at the central frequency (150 Hz).

Figure 8 presents a more detailed comparison of the travel times of the first 10 modes at the central frequency (150 Hz) predicted by the various approaches. The horizontal axis of this figure measures distance from the source whereas the vertical axis measures the equivalent travel time which is defined as the group slowness $s_{g,m}(r)$ of the mode m, which is different from range to range, multiplied by the source-receiver range R. The group slowness is obtained from the relation $s_{g,m}(r)=\partial k_m(r;\omega)/\partial\omega$ by applying numerical differentiation with respect to ω. The adiabatic arrival time of mode m at the receiver is given by

$$t_{g,m}=\int_0^R s_{g,m}(r)dr=\frac{1}{R}\int_0^R Rs_{g,m}(r)dr$$

and thus it is just the average of the equivalent travel time.

In Fig. 8 the equivalent travel times corresponding to the background and perturbed state are shown, as well as the first and second Rytov approximation. The background equivalent travel time for each mode is constant with respect to range and equals the corresponding group travel time. In the first Rytov approximation the phase has a linear dependency on the sound-speed perturbation and since the latter in this case is a linear function of range, the equivalent travel times are linear functions of range, as we see in Fig. 8, fully reflecting the zero-mean property of the range mode $\Delta c_r(r)$. In this connection the first Rytov approximation results in the same group travel times as in the background situation. In the second Rytov approximation the phase is a quadratic functional of the sound-speed perturbation and, since the latter varies linearly with range, the corresponding equivalent travel times are quadratic functions of range, and thus their average will be different than the background group travel times. In this sense the effect of range dependence on travel times is a second-order effect. We see from Fig. 8 that the second Rytov approximation lies close to the exact prediction as far as the equivalent travel times are concerned. This explains the good agreement between the Rytov approximation and the adiabatic prediction in Fig. 7.

Figure 9 shows the effect of the perturbation magnitude on the Green's function (real part) at the central frequency of 150 Hz, as well as on the travel time of the highest peak (mode 2). The predictions are based on the exact adiabatic and coupled-mode calculation as well as on the second-order Born and Rytov approximations. The horizontal axis in this figure measures the value of Δc_r at the receiver position, assuming linear, zero-mean perturbations in all cases. Thus, $\Delta c_r(R)=5$ corresponds to the perturbation shown in Fig. 5, whereas $\Delta c_r(R)=-5$ is the opposite perturbation (negative horizontal gradient). The values between, -5 and 5 in the horizontal axis of Fig. 9 correspond to linear range modes with smaller horizontal gradients, and finally the value 0 corresponds to the range independent background state (no perturbation). The adiabatic character of the propagation is verified once again by the agreement between the exact adiabatic and coupled-mode predictions. These predictions are approximated far better by the Rytov than by the Born approximation. The second-order character—in the neighborhood of the range independent background—is clear both in the fre-

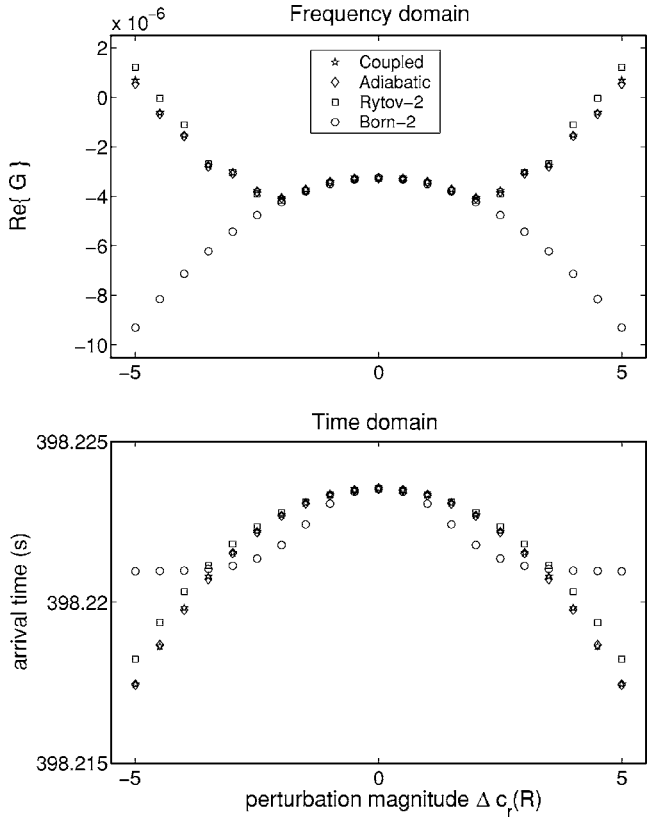


FIG. 9. The effect of the perturbation magnitude on the Greens function—real part—at 150 Hz (top panel), and on the travel time of the highest peak (bottom panel), as predicted by the exact adiabatic and coupled-mode calculation and the second-order Born and Rytov approximations.

quency and the time domain. Still, in the frequency domain this character is not preserved over the entire domain of variability, whereas in the time domain it is. This occurs because travel times are associated with the phase which has a second-order dependence on sound-speed changes, whereas the complex pressure depends on the exponential of the phase. This also explains why the Rytov approximation which focuses on the phase performs better than the Born approximation in Fig. 9. As seen in Fig. 9 the two approximations (Born and Rytov) are asymptotically ($\Delta c_r \rightarrow 0$) equivalent.

B. Large-scale, nonzero-mean range dependence

In the second case the range mode Δc_r is taken to be linear, but not zero mean, varying from 0 at the source to +5 at the receiver as shown in Fig. 10 (top panel). The bottom panel in Fig. 10 shows the resulting sound-speed profiles at various ranges, varying from upward refracting profiles close to the source to profiles with a channel of 50 m depth close to the receiver. Figure 11 shows the result of the exact adiabatic- and coupled-mode calculation of the late arrival pattern corresponding to the range-dependent ocean of Fig. 10, together with the background prediction. It is seen that the perturbed arrivals, especially the late ones, are advanced by more than 120 ms with respect to their background location. Further, the deviation between the adiabatic- and coupled-mode results is very small, indicating that propagation is adiabatic.

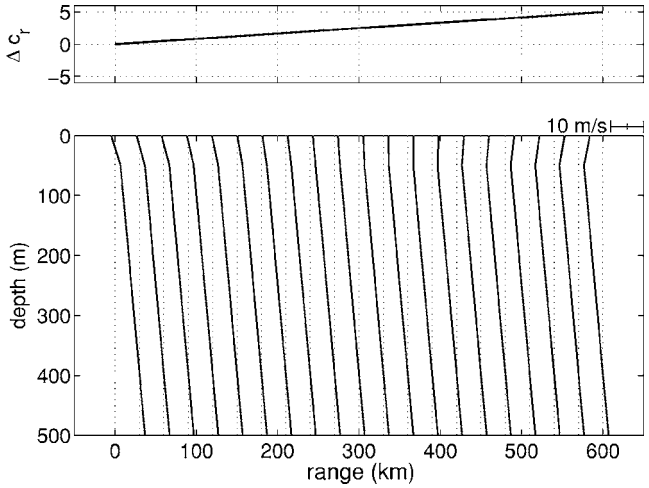


FIG. 10. Linear nonzero-mean range mode $\Delta c_r(r)$ of the sound-speed perturbation (top panel) and resulting sound-speed profiles at various ranges (bottom panel).

Figure 12 shows the late arrival pattern predicted from the first and second Born approximations (top panel) and first and second Rytov approximations (bottom panel). In the present case Δc_r is nonzero-mean, such that the first Born and the first Rytov approximations predict arrival patterns different from the background one. The background arrival pattern is also shown in Fig. 12 together with the adiabatic prediction for the perturbed state. The top panel of Fig. 12 shows that the two Born approximations differ from the background arrival pattern in amplitude but very little in the arrival times. From the comparison with the adiabatic prediction (target arrival pattern) it is seen that the approximation fails to predict the correct arrival times in the perturbed state, whereas there is a remarkable disagreement in the arrival amplitudes as well.

The bottom panel of Fig. 12 shows that the first Rytov approximation differs from the background arrival pattern mainly in amplitude but not very much in the arrival times. The second Rytov approximation offers a significant improvement in getting closer to the exact adiabatic prediction, still it does not describe efficiently the arrival shifts in all cases, especially in the case of the last two arrivals for which the discrepancies are as high as 20 ms.

Figure 13 shows a range-independent prediction of the perturbed arrival pattern by considering a range-independent perturbation equal to the average (+2.5) of the range mode. It is seen that the results of this approximation are very close,

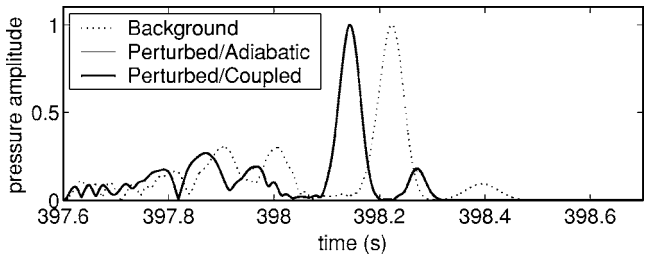


FIG. 11. Arrival pattern in the perturbed environment with linear nonzero-mean range dependence, predicted from exact adiabatic and coupled-mode calculation. Comparison with the background arrival pattern (dotted line).

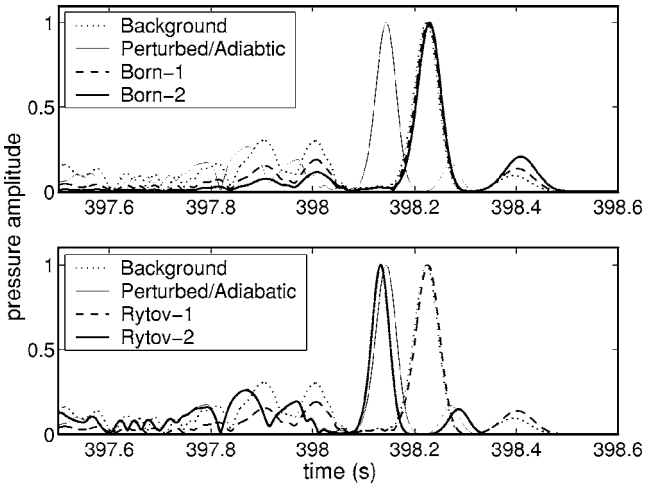


FIG. 12. Arrival pattern in the perturbed environment with linear nonzero-mean range dependence, predicted from the first and second Born approximation (top panel) and the first and second Rytov approximation (bottom panel). Comparison with exact adiabatic prediction (light solid line) and background arrival pattern (dotted line).

and in fact they are slightly better than the previous approximate prediction resulting from the range-dependent perturbation. This indicates that in case of perturbations whose range average is nonzero the dominant effect on travel times is that of the range average of the perturbation and not that of range dependence.

To focus on the effects of range dependence the perturbation analysis must use the range average as a background (reference) state. This is done in the following by taking a new background sound-speed profile which differs from the previous one by a term $+2.5c_z(z)$, thus coinciding with the range average of the perturbed range-dependent ocean environment. The range mode about this new background state varies from −2.5 to +2.5 and thus averages to zero. The late arrival pattern predicted from the second Rytov approximation in this case is shown in Fig. 14. It is seen that the prediction resulting from the perturbation of the new background coincides with the exact adiabatic calculation. This example points to the importance of the proper selection of the reference sound-speed profile for the application of the Rytov approximation to study the effects of range dependence.

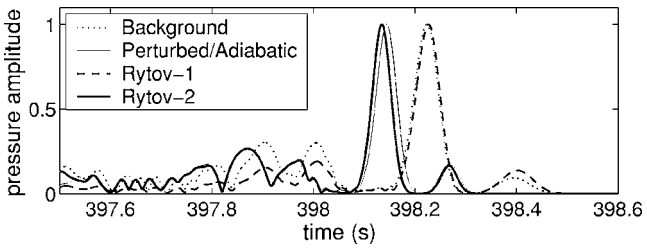


FIG. 13. Arrival pattern in the perturbed environment with constant range mode (equal to the average 2.5 of the range-dependent mode), predicted from the first and second Rytov approximation. Comparison with exact adiabatic prediction (light solid line) and background arrival pattern (dotted line).

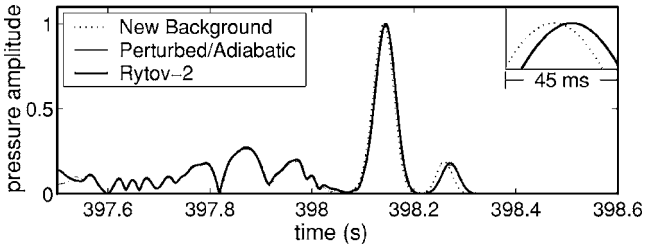


FIG. 14. Arrival pattern in the perturbed environment with linear range dependence, predicted from the second Rytov approximation about the new background state (range average of the distribution shown in Fig. 10). Comparison with exact adiabatic prediction (light solid line) and new background arrival pattern (dotted line). Upper right corner: Detail of the highest peak.

C. Small-scale, zero-mean range dependence

The approximations obtained in Sec. III are of adiabatic nature, see also the Appendix , and thus they are expected to be closer to the exact adiabatic prediction, than to the exact coupled-mode prediction, when these two differ. In order to check this, a perturbation case beyond the limits of adiabatic range dependence is considered. The range and depth mode defining this perturbation are shown in Fig. 15. The range mode is a single-loop sinusoid with scale smaller than the double-loop length of the rays corresponding to the main peak (highest peak in Fig. 2), cf. Fig. 4, whereas the depth mode is selected to concentrate around the turning depths of these rays, such as to maximize the perturbation influence.

Figure 16 shows the effect of the perturbation magnitude a on the Green's function (real part) at the central frequency of 150 Hz, and also on the travel time of the highest peak, as predicted by the exact adiabatic and coupled-mode calculation, by the second-order (adiabatic) Born and Rytov approximations, as well as by the corresponding first-order approximations including cross terms (ct) described in the following. The horizontal axis in this figure measures the deviation a of the first half-cycle of the range mode Δc_r. Positive a values correspond to the range mode as shown in Fig. 15, i.e., a positive half-cycle followed by a negative one.

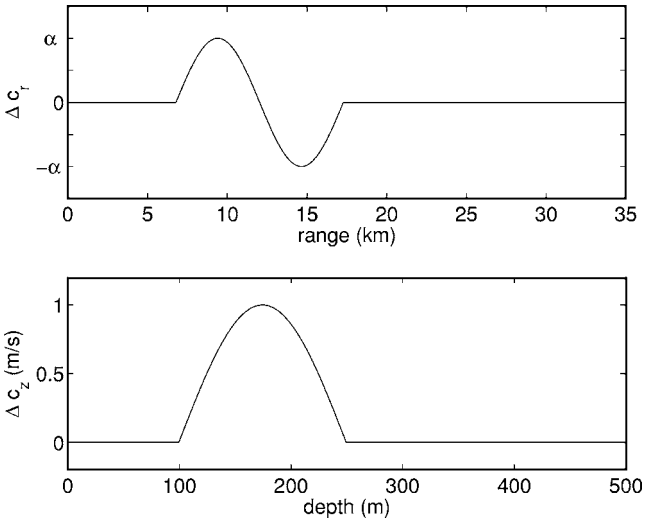


FIG. 15. Range mode (top panel) and depth mode (bottom panel) defining a small-scale sound-speed perturbation.

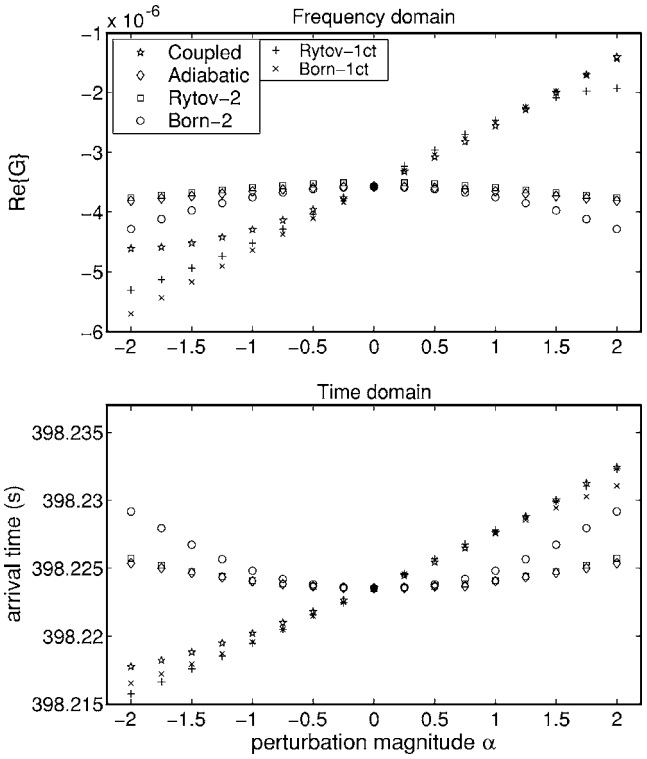


FIG. 16. The effect of the magnitude a of the small-scale perturbation on the Greens function—real part—at 150 Hz (top panel), and on the travel time of the highest peak (bottom panel), as predicted by the exact adiabatic and coupled-mode calculation, the second-order Born and Rytov approximations, as well as by the corresponding first-order approximations including cross terms (ct).

Negative a values correspond to the opposite function (negative half-cycle followed by a positive one). The background-state corresponds to $a=0$.

It is seen from Fig. 16 that while the exact adiabatic prediction is of second order, and the second-order Born and Rytov approximations behave similarly, the exact coupled-mode prediction exhibits a clear first-order component both in the frequency and the time domain. Thus, for small-scale range dependence the exact adiabatic prediction and the Born and Rytov approximations based on the assumption of adiabatic range dependence fail to describe the true (first-order) character of travel-time changes. However, by retaining the significant cross terms (first line) in Eq. (31) the Born and Rytov-Keller approximations can describe first-order effects in the case of small-scale perturbations.

The results denoted by (ct) in Fig. 16 are from the first-order Born and Rytov-Keller approximations including the above-mentioned cross terms. These results are in good agreement with the coupled-mode prediction both in the frequency and the time domain. It is remarkable that in this case the two approximations (Born and Rytov) describe the first-order character of the travel-time changes nearly equally well (the first-order Born approximation describes linear changes in the frequency domain, whereas the first-order Rytov approximation describes linear travel-time changes—this goes back to the first-order approximation of the phase in this case). Thus, by including cross terms, the Born and Rytov-Keller approximations can be enhanced to deal with small-scale perturbations giving rise to first-order effects.

V. DISCUSSION AND CONCLUSIONS

The objective of this work was to examine the feasibility of using Born and Rytov approximations of the Green's function for arrival-pattern calculations associated with long-range pulse propagation in weakly range-dependent ocean environments. Such calculations are important in the context of ocean acoustic travel-time tomography, particularly in connection with full-wave-form inversion.

The proposed approach is to use first- and second-order Born and Rytov approximations to calculate the perturbed frequency-domain Green's function associated with range-dependent sound-speed perturbations about a range-independent background state, repeat this for a large number of frequencies within the source bandwidth and then obtain the time-domain acoustic field through the inverse Fourier transform.

The use of the normal-mode representation of the background (unperturbed) Green's function, together with the assumption of adiabatic range dependence (sound-speed perturbations with horizontal scales large compared with the double-loop length of the rays corresponding to the arrivals under study), leads to closed-form expressions describing the perturbation behavior of the low-order modes. The performance of the various approximations was numerically studied for propagation over a range of 600 km in simple range-dependent ocean environments.

In the cases of large-scale (adiabatic) range dependence considered the Born approximation failed to describe the temporal displacement of arrivals caused by the sound-speed perturbations, unless the perturbations were very small, cf. Fig. 9. The reason is that the Born representation is based on a perturbation expansion of the Green's function itself in the frequency domain which can hardly be translated into shifts in the time domain (temporal displacements) through the inverse Fourier transform. The nature of the Born approximation makes it more suitable for backscattering problems focusing on changes in the intensity of the returning field than for transmission problems focusing on travel-time changes.[31]

The Rytov approximation, on the other hand, is based on a perturbation expansion of the phase in the frequency domain, which is directly associated with shifts in the time domain. In this connection the Rytov approximation is suitable for modeling perturbations in the time domain with emphasis on temporal displacements, as is the case in ocean acoustic travel-time tomography. However, since the acoustic field in the ocean is a multicomponent wave field, each component (mode) being characterized by a different phase and perturbation behavior the standard Rytov approximation cannot be applied, since it imposes a single phase perturbation to the whole field.

To address multimode propagation, an approach proposed by Keller (Rytov-Keller approximation) was adopted here. According to this approach the Rytov method is applied to each wave component (mode) independently, i.e., isolated from the other modes. In the case of adiabatic sound-speed perturbations the effect of range dependence on travel times is of second order, in agreement with previous ray-theoretic results.[21,22] This means that in the case of zero-mean range-

dependent perturbations of large horizontal scale first-order approximations predict no travel-time changes. In this case second-order approximations (or higher) are required. Further, for the study of the effects of range dependence on travel times it is essential to separate from the effects of range-independent perturbations. This can be done by using the range-average sound speed profile as a background (reference) state. Further, this will minimize range-dependent perturbations by turning them zero-mean.

The second-order character of travel times is a result based on the assumption of adiabaticity, i.e., of sound-speed perturbations with large horizontal scales compared with the double-loop lengths of the rays corresponding to the arrivals under study. In the case of small-scale perturbations (horizontal scales comparable to or smaller than the ray double-loop lengths) the adiabatic assumption and the second-order character of travel times no longer hold. In that case the cross terms in Eq. (31) become important and give rise to first-order contributions, even in the case of zero-mean range modes. This explains why in a recent work on travel-time sensitivity kernels[43] the effects of local sound-speed perturbations on travel times were described as first-order effects. By retaining the cross terms the Born and Rytov-Keller approximations can be enhanced to deal with small-scale perturbations and first-order effects. These effects become less and less important as the horizontal scale of the sound-speed perturbation increases, and finally at the adiabatic limit they disappear such that the second-order behavior prevails.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for helpful comments and suggestions.

APPENDIX

The axisymmetric acoustic field of a harmonic point source for a range-dependent environment is given in the adiabatic approximation by[25]

$$G'(R,z_r|z_s)=\frac{e^{-j\pi/4}}{\rho_w\sqrt{8\pi R}}\sum_{n=1}^{M}\frac{\phi_n(0,z_s)\phi_n(R,z_r)}{\sqrt{k_n(r)}}\times\exp\left(-j\int_0^R k_n(r)dr\right), \tag{A1}$$

where z_s is the depth of the source located on the z axis of the (r,z,θ) cylindrical coordinate system, R and z_r are the receiver range and depth. In the case of a range-dependent environment the quantities ϕ_n and k_n are functions of range as well, defined by the eigenvalue problem Eq. (23) with $c=c(r,z)$. The phase of each modal component is associated with the integral

$$\Psi_n=-j\int_0^R k_n(r)dr. \tag{A2}$$

Assuming a range-independent background environment, with sound speed profile $c_0(z)$, and an axisymmetric range-dependent sound-speed perturbation of the form $\Delta c_r(r)\Delta c_z(z)$, we can derive the perturbed phase using a second-order Taylor expansion of $k_n(r)$ with respect to the sound-speed about the background state[26]

$$k_n(r)=k_{n,0}+\frac{Q_{nn}}{2k_{n,0}}\Delta c_r(r)+\frac{1}{2k_{n,0}}\times\left[Q_n'+\sum_{\substack{m=1\\m\neq n}}^{M}\frac{|Q_{nm}|^2}{\Lambda_{nm}}-\frac{|Q_{nn}|^2}{4k_{n,0}^2}\right]\Delta c_r^2(r). \tag{A3}$$

Substituting this expression into the integral in Eq. (A2) we obtain

$$\Psi_n=-jk_{n,0}R-j\frac{Q_{nn}}{2k_{n,0}}\int_0^R\Delta c_r(r)dr-\frac{j}{2k_{n,0}}\times\left[Q_n'+\sum_{\substack{m=1\\m\neq n}}^{M}\frac{|Q_{nm}|^2}{\Lambda_{nm}}-\frac{|Q_{nn}|^2}{4k_{n,0}^2}\right]\int_0^R\Delta c_r^2(r)dr \tag{A4}$$

Comparing the right-hand side with the first and second Rytov-Keller approximation, Eqs. (48) and (50), we observe that the first-order adiabatic approximation is equivalent to the dominant part of the first Rytov-Keller approximation. Further, in case of zero-mean range dependent perturbations, the second-order adiabatic approximation has the same dominant part as the second Rytov-Keller approximation except for the last term in the brackets in Eq. (A4) which does not appear in the Rytov-Keller approximation (this term, however, is much smaller than the previous term including the sum over m). Thus, the Rytov-Keller approximation has a strong similarity with the second-order adiabatic approximation.

[1] W. H. Munk and C. Wunsch, "Ocean acoustic tomography: A scheme for large scale monitoring," Deep-Sea Res., Part A **26A**, 123–161 (1979).
[2] V. A. Del Grosso, "New equation for the speed of sound in natural waters (with comparisons to other equations)," J. Acoust. Soc. Am. **56**, 1084–1091 (1974).
[3] W. H. Munk, P. F. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, New York, 1995).
[4] W. H. Munk and C. Wunsch, "Ocean acoustic tomography: Rays and modes," Rev. Geophys. Space Phys. **21**, 777–793 (1983).
[5] E. C. Lo, J. X. Zhou, and E. C. Shang, "Normal mode filtering in shallow water," J. Acoust. Soc. Am. **74**, 1833–1836 (1983).
[6] P. J. Sutton, W. M. L. Morawitz, B. D. Cornuelle, G. Masters, and P. F. Worcester, "Incorporation of acoustic normal mode data into tomographic inversions in the Greenland Sea," J. Geophys. Res. **99**, 12487–12502 (1994).
[7] D. Y. Mikhin, O. A. Godin, S. V. Burenkov, Yu. A. Chepurin, V. V. Goncharov, V. M. Kurtepov, and V. G. Selivanov, "An experiment on acoustic tomography of the western Mediterranean from a moving ship," in *Proceedings of the Third European Conference on Underwater Acoustics*, edited by J. S. Papadakis (Crete University Press, Herakion, 1996), pp. 821–826.
[8] J.-P. Hermand and W. I. Roderick, "Acoustic model-based matched-filter processing for fading time-dispersive ocean channels: Theory and experiment," IEEE J. Oceanic Eng. **OE-18**, 447–465 (1993).
[9] J. -P. Hermand, "Model-based matched filter processing: A broad-band approach to shallow-water inversion," in *Full-Field Inversion Methods in Ocean and Seismo-acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht, 1995), pp. 189–194.
[10] K. J. McCann and F. Lee-McCann, "A narrow-band approximation to the acoustic pressure field," J. Acoust. Soc. Am. **89**, 2670–2676 (1991).
[11] E. K. Skarsoulis, "Second-order Fourier synthesis of broadband acoustic signals using normal modes," J. Comput. Acoust. **5**, 355–370 (1997).
[12] E. K. Skarsoulis, "Fast coupled-mode approximation for broadband pulse

propagation in a range-dependent ocean," IEEE J. Oceanic Eng. **24**, 172–182 (1999).
[13]G. N. Makrakis and E. K. Skarsoulis, "Asymptotic approximation of ocean-acoustic pulse propagation in the time domain," J. Comput. Acoust. **12**, 197–216 (2004).
[14]B. E. McDonald, M. D. Collins, W. A. Kuperman, and K. D. Heaney, "Comparison of data and model predictions for Heard island acoustic transmissions," J. Acoust. Soc. Am. **96**, 2357–2370 (1994).
[15]K. D. Heaney and W. A. Kuperman, "Frequency interpolation technique for broadband parabolic equation calculation," J. Comput. Acoust. **7**, 27–38 (1999).
[16]S. M. Flatte and R. Stoughton, "Predictions of internal-wave effect on the ocean acoustic coherence, travel-time variance, and intensity moments for very long-range propagation," J. Acoust. Soc. Am. **84**, 1414–1424 (1988).
[17]T. F. Duda, S. M. Flatte, J. A. Colosi, B. D. Cornuelle, J. A. Hildebrand, W. S. Hodgkiss, P. F. Worcester, B. M. Howe, J. A. Mercer, and R. C. Spindel, "Measured wave-front fluctuations in 1000-km pulse propagation in the Pacific Ocean," J. Acoust. Soc. Am. **92**, 939–955 (1992).
[18]J. A. Colosi, S. M. Flatte, and S. Bracher, "Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment," J. Acoust. Soc. Am. **96**, 452–468 (1994).
[19]J. A. Mercer and J. R. Booker, "Long-range propagation of sound through oceanic mesoscale structures," J. Geophys. Res. **88**, 689–699 (1983).
[20]J. L. Spiesberger, "Ocean acoustic tomography: Travel time biases," J. Acoust. Soc. Am. **77**, 83–100 (1985).
[21]W. H. Munk and C. Wunsch, "Biases and caustics in long-range acoustic tomography,"Deep-Sea Res., Part A **32**, 1317–1346 (1985).
[22]W. H. Munk and C. Wunsch, "Bias in acoustic travel time through an ocean with adiabatic range dependence," Geophys. Astrophys. Fluid Dyn. **39**, 1–24 (1987).
[23]M. Born, "Quantum mechanics of impact processes," Z. Phys. **38**, 803–827 (1926).
[24]S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radio-physics 4* (Springer, Berlin, 1989).
[25]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994).
[26]E. K. Skarsoulis, "Waveform perturbation of tomographic receptions due to sound-speed variations," Acta Acust. **89**, 789–798 (2003).
[27]L. Brekhovskikh and Yu. Lysanov, *Fundamentals of Ocean Acoustics* (Springer, Berlin, 1982).
[28]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw Hill, New York, 1953).
[29]The parameter ϵ controls the magnitude of the sound-speed perturbation and it is also used to discriminate the orders of the obtained approximations. The use of the same perturbation parameter ϵ in the Born and Rytov approximations has been justified in several model problems of propagation in layered media, see, e.g., W. J. Hadden and D. Mintzer, "Test of the Born and Rytov approximations using the Epstein problem," J. Acoust. Soc. Am. **63**, 1279–1286 (1978).
[30]M. Born and E. Wolf, *Principles of Optics–Electromagnetic Theory of Propagation, Interference, and Diffraction of Light* (Pergamon, Oxford, 1980).
[31]W. B. Beydoun and A. Tarantola, "First Born and Rytov approximations: Modeling and inversion conditions in a canonical example," J. Acoust. Soc. Am. **83**, 1045–1055 (1988).
[32]R. M. Bracewell, *The Fourier Transform and its Applications* (McGraw-Hill, Singapore, 1986).
[33]M. I. Saucer and A. D. Varvatsis, "A comparison of the Born and Rytov methods," Proc. IEEE **58**(1), 140–141 (1970).
[34]M. L. Oristaglio, "Accuracy of the Born and Rytov approximations for reflection and refraction at a plane interface," J. Opt. Soc. Am. A **2**, 2789–2798 (1985).
[35]H. T. Yura, C. C. Sung, S. F. Clifford, and R. J. Hill, "Second-order Rytov approximation," J. Opt. Soc. Am. **73**, 500–502 (1983).
[36]C. A. Boyles, *Acoustic Waveguides. Applications to Oceanic Science* (Wiley, New York, 1984).
[37]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1965).
[38]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1966).
[39]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (Mc-Graw Hill, New York, 1978).
[40]J. B. Keller, "Accuracy and validity of the Born and Rytov approximations," J. Opt. Soc. Am. A **59**, 1003–1004 (1969).
[41]S. Mauuary, Y. Desaubies, F. Gaillard, T. Terre, J. Papadakis, M. Taroudakis, E. Skarsoulis, C. Millot, U. Send, and G. Krahmann, "Acoustic observation of heat content across the Mediterranean Sea," Nature (London) **385**, 615–617 (1997).
[42]E. K. Skarsoulis and G. A. Athanassoulis, "Arrival-time perturbations of broad-band tomographic signals due to sound-speed disturbances. A wave theoretic approach," J. Acoust. Soc. Am. **97**, 3575–3588 (1995).
[43]E. K. Skarsoulis and B. D. Cornuelle, "Travel-time sensitivity kernels in ocean acoustic tomography," J. Acoust. Soc. Am. **116**, 227–238 (2004).

The viability of reflection loss measurement inversion to predict broadband acoustic behavior

Marcia J. Isakson[a)]
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

Tracianne B. Neilsen
Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602

(Received 18 March 2006; accepted 27 April 2006)

The viability of using reflection loss measurements for the determination of sediment parameters and the predictions of broadband acoustic behavior is studied as a function of frequency for a dispersive ocean bottom. For this study, a deterministic set of reflection loss values from an idealized dispersive sediment is calculated as a function of grazing angle over a large frequency band 100 Hz–1 MHz. In each of the four decades of frequency a simulated annealing optimization process is used to invert for the sediment parameters. In addition, a set of rotated coordinates is calculated that reveals the the relative sensitivities of the reflection data to each of the sediment parameters and therefore the ability of the inversion process to converge. The accuracy and precision of each estimate is analyzed. The predicted broadband acoustic behavior from the inverted parameters from each decade band was compared to the original model. Only the estimates obtained from data that include frequencies in the transition region can successfully predict broadband behavior. [DOI: 10.1121/1.2206515]

PACS number(s): 43.30.Pc, 43.20.Gp [WMC] Pages: 135–144

I. INTRODUCTION

The acoustic behavior of the ocean bottom can affect shallow-water transmission loss, reverberation estimation, ocean mapping, and acoustic communications. This behavior can be predicted by knowing the sediment parameters. Historically, ocean sediments have been characterized by five parameters: the compressional and shear sound speeds and attenuations and the density. However, recent data suggests that this parameterization is incomplete.[1–5] Current models such as the Biot-Stoll model,[6,7] the effective density fluid model,[8] and the Biot with contact squirt flow with shear drag model (BICSQS) (Ref. 9) require up to 13 parameters for complete characterization. Many of the parameters of these models are difficult if not impossible to measure. Additionally, current methods of sediment characterization are difficult, time consuming and invasive.[10]

Reflection loss measurements may provide a method of determining the parameter set for a poroelastic sediment which is both noninvasive and provides immediate results. In fact, there are several studies which attempt to use reflection loss measurements to predict sediment parameters in both a fluid and poroelastic framework.[11–15] However, many of these investigators use interrelations between the parameters or environmental measurements to estimate parameter values, rather than determining the parameters strictly from the data. For example, Holland[13] uses a form of the Kozeny-Carmen equation to estimate the value of the permeability based on the porosity. However, this method is only accurate for the case of well-sorted, unconsolidated sediments with roughly spherical grains.[16,17] In fact, for sandy sediments, the Kozeny-Carmen equation can overestimate permeability by an order of magnitude.[18]

In another study, Schock uses an inversion of normal reflection coefficient and attenuation measurements to obtain estimates of sediment parameters such as porosity, bulk density, permeability, and mean grain size.[15] However, Schock's method of determining the permeability requires the presence of a sub-bottom layer that is not always present. Also, the estimate for permeability assumes a frequency-independent value which has been challenged by Taylor-Smith, who demonstrated that permeability may vary up to an order of magnitude from the direct flow value.[19] Furthermore, Schock relies on interrelational expressions or measurements to determine the remaining poroelastic parameters. These expressions may not be valid. For example, his expression for the frame shear modulus is dependent on the total average stress and the porosity. However, as shown in the BICSQS model, the frame shear modulus may be dependent on the vicous drag of the interstitial fluid as well as the contact shear properties.[9]

It would be useful to devise a noninvasive method of determining the independent sediment parameters within a poroelastic framework from one measurement. The frequency- and angle-dependent reflection coefficient may provide a measurement set that can uniquely define all the parameters. The critical angle measurement is sensitive to the underlying sound speed, while the subcritical values are sensitive to shear speed and attenuation. Additionally, the normal reflection coefficient is sensitive to bulk density and porosity as shown by Schock.[15] However, there has been no detailed study to quantify the sensitivity of reflection coefficient loss measurements to poroelastic sediment parameters.

[a)]Electronic mail: misakson@arlut.utexas.edu

Also, since reflection loss has been shown to be frequency dependent, there have been no studies to quantify the ability of inversion results from one frequency band to predict the acoustic behavior over the broadband. The ability of one frequency band to predict the behavior of another band may be an important feature as high-frequency measurements are often easier to conduct.

The work presented herein contains a sensitivity study that strives to determine the viability of using estimates of seabed properties obtained from an inversion with angular- and frequency-dependent reflection coefficient data to predict broadband acoustic behavior. This study focuses on four features of the inversion. First, the relative sensitivity of each parameter is calculated. Second, the accuracy and precision of the inversion estimate is determined. Lastly, the parameter set from limited frequency ranges is used to predict broadband acoustic behavior.

It is understood that *in situ* measurements will be modified by interface roughness, volume inhomogeneity, range dependence, layering, and other experimental effects. However, the purpose of this work is to determine a baseline viability of the technique. In other words, given the complexity of the poroelastic model, are reflection coefficient measurements sensitive enough to provide estimates of sediment parameters?

The study was conducted using simulated data from 100 Hz to 1 MHz in four decade bands. The details of the simulated data calculation are given in Sec. II. The inversion methodology is described in Sec. III. Results are presented in Sec. IV and conclusion are made in Sec. V.

II. SIMULATED DATA

The simulated data are computed with the Biot-Stoll model. The data are computed for 50 frequencies on a logarithmic scale from 100 Hz to 1 MHz and for 180 angles from 1–90 deg. The Biot-Stoll formulation has been presented extensively in the literature[7] and is only summarized here. In Biot-Stoll theory, the equations are motion are governed by two coupled wave equations,

$$\nabla^2(He - C\zeta) = \frac{\partial^2}{\partial t^2}(\rho e - \rho_f \zeta), \tag{1}$$

$$\nabla^2(Ce - M\zeta) = \frac{\partial^2}{\partial t^2}(\rho_f e - m\zeta) - \frac{\eta}{\kappa}F(k)\frac{\partial \zeta}{\partial t}. \tag{2}$$

The fluid and frame move independently with displacements U and u, respectively. The volumetric strain is defined by $e=\mathrm{div}(u)$, and the local fluid content increase is $\zeta=\mathrm{div}(u-U)$. Equation (1) above describes the fluid and solid conservation, and Eq. (2) describes the fluid motion relative to the solid with two important first-order correction terms. The first term, $m\zeta$, describes the inertial coupling. This term is necessary at high frequencies since, according to Darcy's law, the fluid flow in the pores is not linearly related to the pressure gradient applied. In the Biot/Stoll formulation, m is given by $\tau\rho_f/\beta$, where τ is the tortuosity of the pores, ρ_f is the fluid density, and β is the porosity. The second correction term, $(\eta/\kappa)F(k)\ \partial\zeta/\partial t$, describes a coupling with viscosity η. Here, the fluid flow in the pores does not follow the Poiseuille law. Therefore, $F(k)$ is included to model the frequency dependence of viscous flow. The variable k is given by $a(\omega\rho_f/\eta)^{1/2}$, where a is the pore size parameter. The function, $F(k)$, approaches unity for low frequencies where the capillary flow can be modeled as parabolic. The moduli H, C, and M are derived by Stoll as[7]

$$H = \frac{(K_r - K_b)^2}{D - K_b} + K_b + \frac{4}{3}\mu, \tag{3}$$

$$C = \frac{K_r(K_r - K_b)}{D - K_b}, \tag{4}$$

$$M = \frac{K_r^2}{D - K_b}, \tag{5}$$

$$D = K_r\left[1 - \beta\left(\frac{K_r}{K_{f-1}}\right)\right]. \tag{6}$$

There are 13 independent parameters in the Biot-Stoll theory. The Biot parameters are generally grouped into three categories. The first category contains the bulk parameters: porosity β, fluid density ρ_f, fluid bulk modulus K_f, grain density ρ_g, and grain bulk modulus K_r. These parameters are generally well known and measurable. The fluid motion parameters form the next category: the viscosity η, the permeability κ, the pore size parameter a, and the tortuosity τ. These values are less well known and more difficult to measure. The last category is the frame response parameters: the frame shear modulus, $\tilde{\mu}=\mu+i\mu_i$, the frame bulk modulus, $\tilde{K}_b=K_b+iK_{b,i}$. These properties are not known and are practically impossible to measure directly.

A. Parameter bounds

It is necessary to determine realistic bounds for each of the Biot parameters that are explored during the inversion. By using a realistic parameter space, we are better able to predict the viability of the technique in a real scenario. The system is described by a half-space of water over fluid-saturated sediment. For the inversions that follow, the sound speed and density of the water column are considered known at $c=1530$ m/s and $\rho=1.03$ g/cm^3, respectively. The parameter values used to generate the simulated reflection data are based on nominal values for fluid-saturated sand and are listed in Table I in the column labeled "true." Twelve of the 13 parameters are varied in the inversion. However, in many applications, one or more of these parameters would be known. The limits of the search space for each of the parameters were chosen based on theoretical limits and possible experimental conditions as explained below.

1. Bulk properties

a. Porosity, β: (Limits: 0.26–0.47) Porosity is the ratio of volume of the pores to the total volume of the element. For spheres, a tetrahedral close packing yields a theoretical lower limit of 0.26; a cubic packing gives a value of 0.47,

TABLE I. Biot parameters used to generate the simulated reflection data and bounds on the parameters in the analysis. The pore size is not varied independently in the inversion but is calculated with the Kozeny-Carmen equation, Eq. (7).

Parameter	True	Min.	Max.
ρ_f-g/cm^3	1.03	1.02	1.04
K_f-GPa	2.3	2.0	2.5
η-kg/(m s)	1.0×10^{-3}	1.0×10^{-5}	1.5×10^{-3}
ρ_s-g/cm^3	2.69	2.6	2.7
K_r-GPa	36	32	49
β	0.38	0.26	0.47
τ-m^2	1.35	1.0	3.0
a_p-m	2.67×10^{-5}		
κ-m^2	2.5×10^{-11}	1×10^{-13}	1×10^{-9}
μ-Pa	3.0×10^{7}	1×10^{7}	2×10^{9}
μ_i-Pa	1.0×10^{6}	0	1×10^{8}
K_b-Pa	4.4×10^{7}	1×10^{7}	3.6×10^{8}
$K_{b,i}$-Pa	1.0×10^{6}	0	1×10^{8}

generally considered the upper limit for most sands.

b. Fluid density, ρ_f: (Limits: 1.020–1.040 g/cm^3) The fluid density can vary within small limits due to temperature or salinity. An uncertainty of ±0.020 g/cm^3 is reasonable for experimental conditions found in temperate ocean climates.

c. Fluid bulk modulus, K_f: (Limits: 2.0–2.5 GPa) The fluid bulk modulus is usually measured using the sound speed in the fluid, c, through $c^2=K_f/\rho_f$. An average value for the water sound speed in the ocean is between 1450 and 1530 m/s, giving a fluid bulk modulus between 2.2 and 2.5 GPa. However, it has been suggested that the fluid bulk modulus can also be dependent on the amount of absorbed gas or gas bubbles in the sediment produced by benthic activity.[20] This effect would lower the fluid bulk modulus in the sediment but not in the water column. Therefore, the lower limit of the fluid bulk modulus is set to accommodate up to 3 ppm of gas bubbles that may be produced by residual benthic activity.

d. Grain density, ρ_g: (Limits: 2.6 to 2.7 g/cm^3). The value of ρ_g has been measured for quartz sand at 2.650 g/cm^3 and should vary little when the grain material is uniform.

e. Grain bulk modulus, K_r: (Limits: 32 to 49 GPa) The bulk modulus of quartz is 36 GPa from standard references. In one study, the grain bulk modulus of sand in a laboratory tank was determined from a suspension of sand grains in a liquid with a density matching that of the sand grains using Wood's equation. This measurement produced a 95% confidence interval of 32 to 49 GPa.[21]

2. Fluid flow properties

a. Viscosity, η: [Limits: 1.0×10^{-5}–1.5×10^{-3} kg/(m s)]. The viscosity is found to vary little when measured.[3] The expected value of viscosity of water is $1.0E-3$ kg/(m s) from standard references.

b. Permeability, κ: (Limits: 1×10^{-13} to 1×10^{-9} m^2) Permeability measures quantitatively the ability of a porous medium to conduct fluid flow. By comparing a number of measurement techniques in order to quantify their relative inaccuracies. Taylor-Smith has found that permeability can vary as much as two orders of magnitude depending on the measurement.[19] Also, Taylor-Smith points out that in some cases the Biot model can predict permeabilities that are several orders of magnitude higher than what is measured, while in other cases, the Biot model predicts the same values that are found from direct flow methods. One of the possible explanations may be that permeability is frequency dependent. Although, much of the frequency dependence of the permeability is accounted for within the Biot model through the viscous loss term, some frequency dependence in the permeability can arise from grains which are not part of the frame and are floating. In a high-frequency regime, these grains remain suspended. While in the constant flow regime, these particles would tend to be pushed into pore openings, decreasing the permeability. Therefore, for this study, the permeability is given wide limits.

c. Pore size, a: (Limits: 2×10^{-6}–2×10^{-4} m) The pore size is the size of the pore space between the sand grains. The Kozeny-Carmen equation relates the values of permeability κ, tortuosity τ, porosity β, and pore size a[22]

$$\kappa=\frac{a^2\beta}{8\tau}. \tag{7}$$

For the purposes of this inversion, the pore size was calculated using the Kozeny-Carmen equation.

d. Tortuosity, τ: (Limits: 1–3 m^2) Tortuosity is the square of the ratio of the minimum path length of a contiguous path through the pore network to the straight-line path. For uniform cylindrical pores with axes parallel to the gradient, the tortuosity equals 1, while for a random system of uniform pores with all possible orientations, the theoretical maximum value is 3.

3. Frame properties

a. Frame shear modulus, $\mu+i\mu_i$: (Limits, real part: 1×10^7–2×10^9 Pa; imaginary part: 0–1×10^8 Pa). The frame shear modulus is the resistance of the frame to a tangential deformation, and generally, its value is unknown. However, in an elastic medium the shear wave speed c_s is related to the real part of the frame shear modulus μ by

$$c_s=\sqrt{\frac{\mu}{\rho}}. \tag{8}$$

Although this equation is not entirely valid for the poroelastic case because of the effects of the interstitial fluid, it is useful for determining the order of magnitude of the frame shear modulus. The shear wave speed has been measured between 70 and 147 m/s in sandy sediments at a frequency of 300 Hz to 10 kHZ.[23,24,10,25,26] There are no measurements of the shear wave speeds at the upper end of the frequency range because the high shear attenuation at these frequencies prohibits the measurement. Therefore, a shear wave speed of 1000 m/s will be taken as the upper limit. For a shear wave speed of 90 m/s, the frame shear modulus is 16 MPa. While for a shear wave speed of 1000 m/s, the frame shear modulus is 1.0 GPa. Therefore,

due to lack of knowledge of shear modulus, and the fact that the shear wave speed has not been measured for high frequencies in unconsolidated sediments, this parameter is allowed to vary widely.

Since frictional losses can occur with the deformation of the frame, an imaginary part of the frame shear modulus μ_i is included to account for the deformation and has a value that is related to the shear attenuation. It is difficult to obtain a value for the shear attenuation at high frequency, and data taken in the 1–20-kHz range suggest that the attenuation is not truly linear with frequency.[25] Therefore, large limits are set on μ_i to account for this uncertainty and frequency dependence.

b. Frame bulk modulus $K_b+iK_{b,i}$: (Limits, real part: $1\times10^7-3.6\times10^8$ Pa; imaginary part: $0-1\times10^8$ Pa) The frame bulk modulus is the resistance of the frame to a compressional deformation. The frame bulk modulus of water saturated sand is unknown and is allowed to vary widely.

III. INVERSION METHOD

The overall purpose of the inversion algorithm is to obtain estimates of the unknown parameters required to produce modeled values that best match the data. The elements of the inversion process are (1) the forward model; (2) the cost function used to quantify the match between model and data; and (3) the technique used to search the parameter space.

A. Forward model

For the present work, the OASES (Ocean Acoustics and Seismic Exploration Synthesis) Reflection Coefficient Module (OASR) from the OASES package version 2.1, is the forward model in the inversions. OASES is a collection of programs used to model acoustic propagation in layered fluids and sediments.[27] OASR is the submodule of the OASES analysis package that calculates reflection loss coefficients over a given range of sampling frequencies and grazing angles. In addition, OASR can accommodate both elastic and poroelastic models of the seafloor and its strata. A theoretical description of the reflection model is found in Ref. 28 and generally follows the derivation found in Appendix B of Ref. 7.

B. Optimization

The cost function to be optimized in the inversion quantifies the mismatch between the model and the data. When a perfect match is obtained, the value of the cost function is 1. The inversion attempts to maximize the cost function by adjusting the modeling parameters. The ideal cost function should have a varied landscape with significant gradients to optimize the chance of finding the global maximum.

The cost function considered for this work is a relative, least-squares cost function defined by

$$C(x)=1-\frac{1}{N_f N_\theta}\sum_{f,\theta}\left[\frac{R_d(f,\theta)-R_m(f,\theta,\mathbf{x})}{R_m(f,\theta,\mathbf{x})}\right]^2, \tag{9}$$

where $R_d(f,\theta)$ is the magnitude of the reflection coefficients measured at frequency f and grazing angle θ, and $R_m(f,\theta,\mathbf{x})$ are the corresponding modeled values obtained for the set of N parameters in $\mathbf{x}$. Because this is an analytic function, it is expected to have continuous and finite derivatives.

Inversions not only require a quantitative comparison of experimental and modeled data but also a strategy for sampling the N-dimensional parameter search space. To accomplish this task, a simulated annealing (SA) algorithm[29] is used to optimize the cost function. SA searches for the global maximum of the cost function using random perturbations and sequential variation of parameters with probabilistic criteria to accept and reject possible solutions. In this present work, OASR constructs a forward model based on the parameter variations from SA. The SA algorithm terminates when convergence is achieved.

C. Rotated coordinates

To obtain more information about how parameters are coupled to each other and the relative sensitivities of the reflection coefficient to changes in the parameters, rotated coordinates are computed that can be used to navigate the search space in the inversion.[30] The coordinate rotation of the parameter space is defined using the eigenvectors of K, the covariance matrix of derivatives of the cost function with respect to the N individual parameters. The components of K are defined as

$$K_{ij}=\int_\Omega \frac{\partial C}{\partial \hat{x}_i}\frac{\partial C}{\partial \hat{x}_j}d\Omega, \tag{10}$$

where $i,j=1,\ldots,N$, and C is the cost function. To effectively compare the parameters in $\mathbf{x}$, the dimensions are removed by dividing each parameter by the difference in the maximum and minimum values: $\hat{x}_i=x_i/(x_{\max i}-x_{\min i})$. Ω defines the bounds on the parameter space. The integration is carried out using Monte Carlo integration techniques.[31]

The N eigenvectors of K are the rotated coordinates and indicate how the parameters are coupled. The rotated coordinates $\mathbf{v}_i$ and the eigenvalues s_i provide information about the relative sensitivities of the cost function to changes in the individual parameters. For example, the rotated coordinate that corresponds to the largest eigenvalue of K defines the most efficient way to increase the cost function. Each rotated coordinate can be used in turn to vary the individual parameters in a fast simulated annealing algorithm[32] as in Ref. 30.

As an example, the rotated coordinates obtained for the Biot parameters, using the simulated reflection data from 0.1–1 kHz, the bounds listed in Table I, and 1200 samples in the Monte Carlo integration, are shown in Fig. 1. Each rotated coordinate, or eigenvector, is plotted as a row in Fig. 1, and the displacement of each circle from the dotted line indicates the amplitude of the element in the eigenvector corresponding to each parameter. The rotated coordinates have been sorted according to the eigenvalues, shown in Fig. 2. The eigenvalues are scaled by the largest eigenvalue and plotted on a logarithmic scale. In the first rotated coordinate, the top line in Fig. 1, the real part of the frame shear modulus, μ, and the real part of the frame bulk modulus have the largest amplitudes. Thus, over the bounds specified in Table I, the most effective way to increase the cost function for this

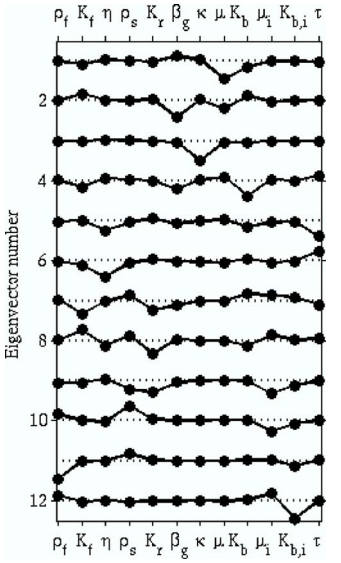


FIG. 1. Rotated coordinates for 12 of the Biot parameters calculated with the bounds shown in Table I for reflection data with frequencies 0.1–1 kHz.

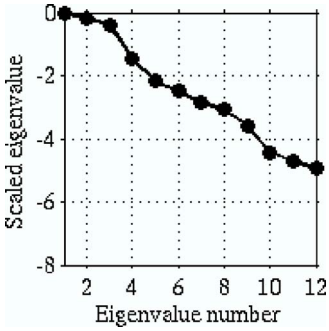


FIG. 2. Eigenvalues associated with the rotated coordinates in Fig. 1. The scaled eigenvalues are the ratios of each eigenvalue to the largest on a log scale.

frequency band is to change μ and K_b in the ratio indicated. The second rotated coordinate reveals coupling between the porosity, β, and the frame shear modulus. The third rotated coordinate has a large component from the permeability, κ. In the fourth rotated coordinate, the porosity, β, the real part of the frame bulk modulus and the fluid bulk modulus have contributions. The tortuosity has a large component in the fifth rotated coordinate. The sixth rotated coordinate is dominated by the viscosity, η, and the tortuosity, τ. The remaining rotated coordinates can be interpreted in the same manner. Every parameter is represented in at least one of the rotated coordinates. Parameters that only have large amplitudes in the higher order rotated coordinates, especially those which correspond to eigenvalues that are three orders of magnitude or more smaller than the largest eigenvalue, do not influence the cost function significantly as they are varied over the bounds Ω.

It is important to remember that the eigenvalues and rotated coordinates can be highly dependent on the bounds on the integration Ω in Eq. (10). For example, the limits on the well-known parameters, such as densities of the fluid and grains, are very small. Thus, varying these parameters within these limits has very little effect on the forward model, and the rotated coordinates. Therefore, the rotated coordinates with large amplitudes for those parameters are associated with much smaller eigenvalues.

IV. RESULTS

The simulated reflection data are analyzed in four ways. (1) The sensitivity of each parameter is determined by computing the rotated coordinates which indicate which parameters most influence the cost function over the specified bounds. (2) The accuracy of the parameter estimates is determined by comparison to the true values. (3) The precision of the inversion results is investigated by considering scatter plots of the accepted parameter values versus the the associated cost functions values. (4) The ability of the parameter estimates to predict broadband acoustic behavior is evaluated by comparing the predicted broadband compressional sound speed, attenuation, normal reflection coefficient, and shear speed.

A. Parameter sensitivity

Rotated coordinates are calculated separately for each decade of the frequency band 100 Hz to 1 MHz. The results of the sensitivity analysis are presented in Tables II–V. Parameters are ranked from most sensitive (MS) to least sensi-

TABLE II. Parameters estimates and sensitivities obtained from simulated annealing inversions for the 100-Hz to 1-kHz frequency band.

Parameter	Estimate	Sensitivity	Accuracy	Precision	Couplings
ρ_f-g/cm^3	1.031	LS	0.09%	NB	
K_f-GPa	2.23	MS (2,4)	3%	LB,UB	μ, β
η-kg/(m s)	0.788×10^{-3}	S(5,6)	21%	LB	
ρ_s-g/cm^3	2.67	LS	0.8%	UB, LB	
K_r-GPa	39.8	LS	10%	LB	
β	0.372	MS(2,4)	2%	LB, UB	μ, K_f
τ	1.32	S(5,6)	21%	LB, UB	η
κ-m^2	1.97×10^{-11}	MS(3)	21%	UB	
μ-Pa	3.0×10^{7}	MS(1,2)	0%	UB	K_b
μ_i-Pa	1.0×10^{6}	LS	0%	LB, UB	
K_b-Pa	4.33×10^{7}	MS(1,4)	2%	UB	μ
$K_{b,i}$-Pa	9.5×10^{6}	LS	850%	UB	

TABLE III. Parameters estimates and sensitivities obtained from simulated annealing inversions for the 1–10-KHz frequency band.

Parameter	Estimate	Sensitivity	Accuracy	Precision	Couplings
ρ_f–g/cm^3	1.020	LS	0.9%	NB	
K_f–GPa	2.05	MS(3,5,6)	11%	UB	
η–kg/(m s)	0.821×10^{-3}	LS	17%	LB	
ρ_s–g/cm^3	2.67	LS	3%	LB, UB	
K_r–GPa	43.4	LS	20%	LB	
β	0.342	MS(3,5,6)	10%	UB, LB	μ, μ_i
τ	1.25	S(4,5,6)	7%	LB, UB	
κ–m^2	2.01×10^{-11}	MS(2)	19%	UB	
μ–Pa	3.0×10^{7}	MS(1,4,5,6)	2%	UB, LB	β, μ_i
μ_i–Pa	1.0×10^{6}	S	0%	LB, UB	β, μ
K_b–Pa	4.48×10^{7}	S(6)	2%	UB, LB	
$K_{b,i}$–Pa	1.38×10^{6}	LS	38%	UB	

tive (LS) in the following manner. If the parameter has a significant magnitude in the first three eigenvectors, it is considered most sensitive (MS). If it is only represented in the fourth through sixth eignevector it is considered sensitive (S). If a parameter does not have a contribution in the first six eigenvectors it is considered least sensitive (LS). For the parameters from the first five eigenvectors, the eigenvector number is also annotated.

Also, some parameters are coupled as described in Sec. III C. The parameter couplings are provided in the column "couplings."

From Tables II–V, the inversion should obtain good estimates for the fluid bulk modulus K_f, porosity β, permeability κ, and frame shear and bulk moduli μ and K_b, respectively, for the lowest two frequency bands with less sensitivity for the frame bulk modulus as the frequency increases. Above 100 kHz, the inversion is less sensitive to the permeability while the dependence on the tortuosity increases. This is due to permeability being one of the major parameters affecting the transition region on the dispersion which is strongest in the 100 Hz–100-kHz region. (See Fig. 4.)

Although the inversion is sensitive to many parameters, estimating their value may be difficult due to coupling. For example, although there is a contribution of the fluid bulk modulus K_f, in many of the rotated coordinates, estimation of its values in this manner may prove difficult. This type of coupling may cause inaccuracies both in the estimation of K_f and β.

Lastly, as described in Sec. III C, the sensitivities calculated are dependent upon both the underlying model and the bounds chosen for the parameters. Therefore, these results should be used to determine of the viability of the entire technique rather than to gain insight of the coupling of parameters within the Biot model.

B. Accuracy of parameter estimates

The accuracy of the parameter estimates is determined by comparing the inversion results to the real values and computing the relative error as shown in Tables II–V. As seen in the tables, generally good estimates are obtained for the most sensitive parameters with a few exceptions. For example, in the highest frequency band, estimates for the porosity and tortuosity are less accurate due to the coupling between these parameters.

The inversion provides a good estimate of the porosity, fluid bulk modulus, and frame shear modulus over the entire band. Permeability estimates are good in the lower frequencies where the influence of the dispersion transition is appar-

TABLE IV. Parameters estimates and sensitivities obtained from simulated annealing inversions for the 10-100-kHz frequency band.

Parameter	Estimate	Sensitivity	Accuracy	Precision	Couplings
ρ_f-g/cm^3	1.0387	LS	0.8%	NB	
K_f-GPa	2.36	MS(2,6)	3%	LB, UB	β, K_b
η-kg/(m s)	3.6×10^{-4}	LS	63%	LB, UB	
ρ_s-g/cm^3	2.70	LS	0.3%	NB	
K_r-GPa	45.4	LS	26%	LB	
β	0.399	MS(2,3)	5%	LB, UB	K_f, τ, K_b
τ	1.45	MS(3,4)	7%	LB, UB	β, K_b
κ-m^2	0.759×10^{-11}	S	69%	UB	
μ-Pa	2.86×10^{7}	MS(1)	5%	LB, UB	
μ_i-Pa	1.08×10^{6}	LS	8%	LB, UB	
K_b-Pa	2.74×10^{7}	MS(2,3,4,6)	37%	UB	K_f, β, τ
$K_{b,i}$-Pa	2.04×10^{6}	LS	104%	UB	

TABLE V. Parameters estimates and sensitivities obtained from simulated annealing inversions for the 100-kHz–1-KHz frequency band.

Parameter	Estimate	Sensitivity	Accuracy	Precision	Couplings
ρ_f-g/cm^3	1.034	LS	0.3%	NB	
K_f-GPa	2.50	MS(2,6)	9%	LB	β, K_b
η-kg/(m s)	1.5×10^{-4}	LS	48%	LB	
ρ_s-g/cm^3	2.62	LS	3%	UB	
K_r-GPa	48	LS	33%	LB	
β	0.427	MS(2,3,4)	12%	LB, UB	K_f, τ, K_b
τ	2.05	MS(3,4)	51%	LB, UB	β
κ-m^2	0.923×10^{-11}	S(5)	63%	UB	
μ-Pa	2.29×10^{7}	MS(1)	23%	UB	
μ_i-Pa	0.581×10^{6}	LS	481%	LB, UB	
K_b-Pa	6.28×10^{7}	MS(2,3,4,6)	42%	LB, UB	K_f, β, τ
$K_{b,i}$-Pa	5.03×10^{6}	LS	403%	UB	

ent in the reflection coefficient. As the inversion becomes less sensitive to permeability in the upper frequencies, the estimate is also less accurate. Lastly, the estimate for tortuosity is accurate in the midfrequencies where it is sensitive but less accurate in the highest frequency band where it is strongly coupled with the porosity. This coupling also decreases the accuracy of the porosity estimate for the band.

C. Precision of parameter estimates

The precision of the parameter estimates can be determined by plotting each of the parameter estimates in the inversion model versus the cost function value. The estimates listed in Tables II–V occur at the highest cost function value. In order to reduce clutter on the scatter plots, an envelope is drawn to encompass all the highest values of the cost function as illustrated in Fig. 3. Additionally, a dashed vertical line is plotted at the correct value of the parameter.

Three parameters from the inversion, grain density ρ_s, porosity β, and permeability κ, are shown for the four frequency bands. The lowest frequency band is indicated by the lightest line and the highest by the darkest with increasing shades in between. A flat or wide envelope as with the grain density shown in Fig. 3(a) reveals that, even though the inversion returns an estimate for the parameter, the uncertainty associated with the estimate is essentially equivalent to the bounds on the parameter. For the data employed in the inversion, a parameter such as the one represented in Fig. 3(a) has no effect on the reflectivity when varied over these bounds. A tight envelope about the final inverted value indicates a well-determined parameter as in Fig. 3(b). As seen in the figure, the porosity estimates has both an upper and lower bound and is well defined. Note also that, although porosity is precisely determined for the highest band of frequencies, the estimate is not accurate. This is due to the coupling of porosity with other parameters such as tortuosity as described in the previous section. Lastly, a parameter may only have a lower or upper bound as illustrated with the permeability in Fig. 3(c).

Tables II–V summarize the precision of each parameter. "NB" indicates no bound as in Fig. 3(a). "LB,UB" indicates that there are both an upper and lower bound present. Lastly, "UB" or "LB" indicate an upper or lower bound as in Fig. 3(c).

Although this is a somewhat simplistic method of determining uncertainties, it provides a qualitative comparision of the reliability of the parameter estimates obtained by the inversion. More rigorous approaches to the problem of estimating uncertainty have been suggested.[33–35] However, these methods have not yet been applied to the Biot model.

There are four key points summarized in Tables II–V. (1) The most precise and accurate parameters as shown by the scatter plots are generally given by the parameters which have a large contribution to eigenvectors associated with the largest eigenvalues. (2) Parameters with very small bounds are generally not sensitive in the inversion. (3) A subset of parameters is sensitive across all frequency bands, namely porosity, fluid bulk modulus, and frame shear modulus. (4) There is a frequency dependence of the sensitivity of the Biot parameters. This is most evident in the permeability, which is sensitive and accurate in the lower frequencies and less sensitive and less accurate in the upper frequencies.

D. Implications for the predictive abilities of reflection coefficient inversion

The parameter estimates obtained from the inversion in each frequency range were used to calculate the broadband

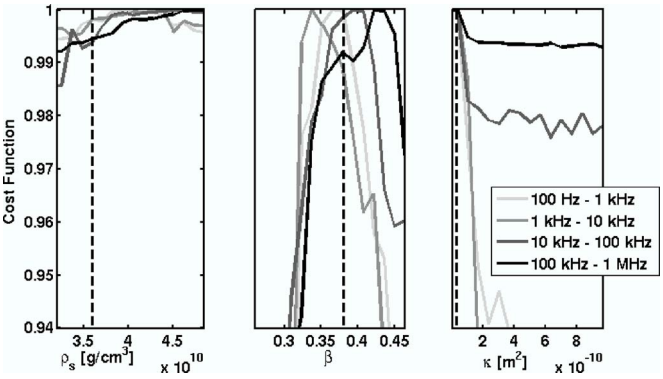


FIG. 3. Envelope plots for the grain density ρ_s, porosity β, and permeability κ for the four frequency bands.

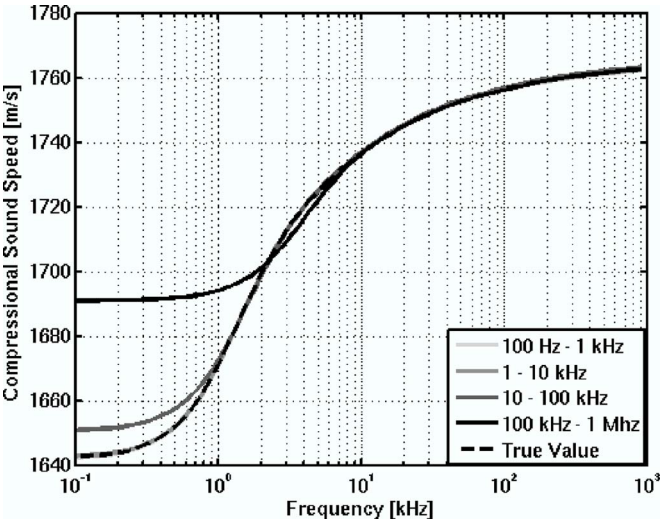


FIG. 4. The compressional sound speed associated with the parameters in Table I and the inversion results in Tables II–V.

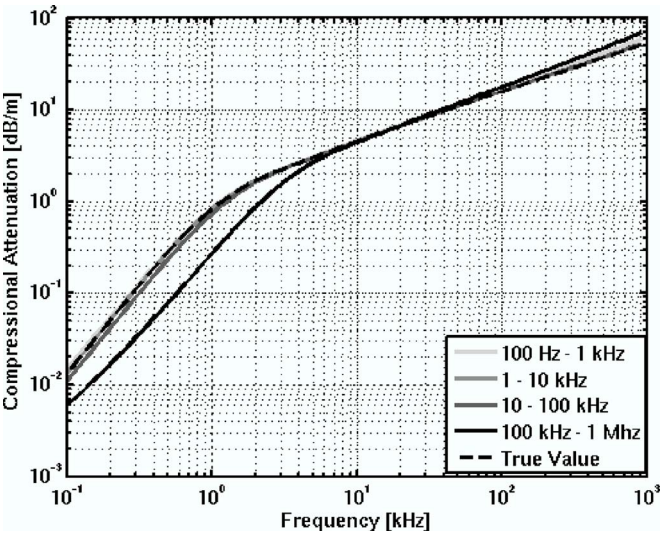


FIG. 5. The compressional attenuation associated with the parameters in Table I and the inversion results in Tables II–V.

compressional sound speed, attenuation, normal reflection coefficient, and shear sound speed. This analysis provides an estimation of the predictive abilities of inversions from particular frequency regimes and indicate which frequency ranges should be measured in order to predict the broadband behavior.

First, consider the dispersion curve in Fig. 4. Inversions from the two lowest frequency decades, 0.1–10 kHz, correctly predict the value and shape of the dispersion curve over the entire four-decade band, while the inversion estimates from the two highest decades do not. At the highest frequency, 100–1000 kHz, shown as the black line, the prediction obtained from the parameter estimates fails to match the values for the low-frequency sound speed and much of the transition region. This implies that inversions based on reflection coefficient measurements taken at high frequencies alone are not sufficient to describe broadband behavior. This is almost entirely due to the poor estimate of permeability. In contrast, inversions from the lower frequencies matched the entire curve very well. Therefore, if reflection coefficient inversions are to be used as a tool for sediment characterization, low-to midfrequency data covering the transition region of the dispersion curve should be used.

Similar results are evident in the attenuation (Fig. 5) and normal reflection coefficient (Fig. 6). It is clear from these figures that inversion based on low-or midfrequency reflection data are adequate to describe the broadband behavior while high-frequency data inversions are not. This is especially evident in the normal reflection coefficient where there is a large drop in the reflectivity over the transition region. The frequency of the nadir of the normal reflection coefficient is highly influenced by the value of the sediment permeability, while the depth of the decrease is dependent on the porosity. Therefore, for the range of frequencies over which the inversion is not sensitive to the permeability, the inversion estimates do not provide a good estimate of the frequency of the decrease, while if the porosity is poorly described the depth of the decrease is not accurate.

Lastly, the inversion estimates are used to predict the broadband shear sound speed. At the outset, good predictions for the shear wave speed were not expected because coupling between the water-borne wave and shear wave were expected to be weak due to the large sound-speed mismatch. However, the frame shear modulus was highly sensitive in the inversion and good estimates for this parameter were obtained. This is due to the influence of the shear wave on the reflection coefficient at low grazing angles. Figure 7 shows the influence of the frame shear modulus at a grazing angle of 3 deg. When the frame shear modulus is at the upper bound, there is significant coupling into the shear mode and the reflection coefficient is greatly affected. However, when the change in frame shear modulus is small, there is a much smaller change in the reflection coefficient that would be difficult to measure in practice. Because of the large effect at high values of the frame shear modulus, good estimates for the shear wave speed were obtained for every frequency band inversion with the lowest frequency bands

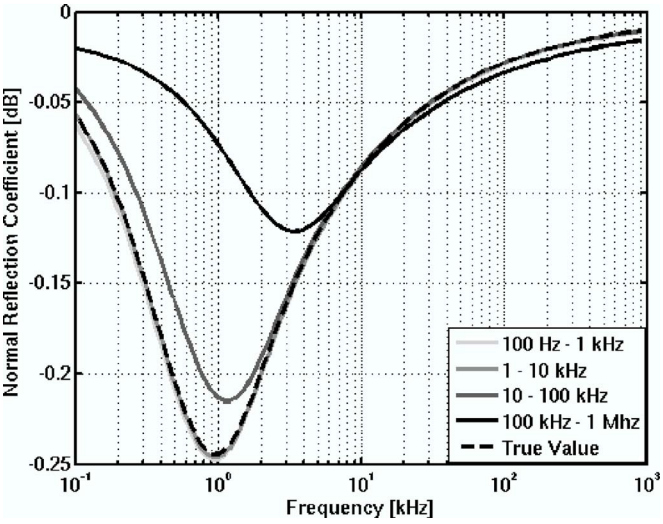


FIG. 6. The normal reflection coefficient associated with the parameters in Table I and the inversion results in Tables II–V.

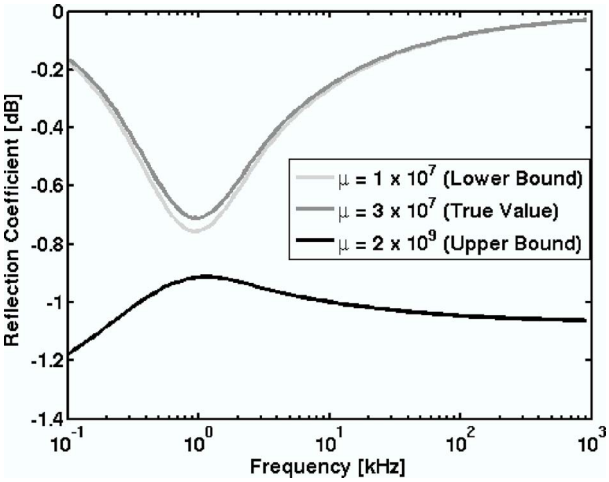


FIG. 7. The effect of the frame shear modulus μ on the value of the low grazing angle (3 deg) reflection coefficient.

the most accurate. (See Fig. 8.) Therefore, the shear wave speed may be predicted using reflection coefficient inversions with decreasing accuracy as frequency increases.

V. CONCLUSIONS

This study sought to determine the ability of parameter estimates obtained from inversions based on reflection data for a limited frequency range to predict broadband behavior. Simulated reflection coefficient data were produced over a large frequency range, 100 Hz to 1 MHz, using the Biot poroelastic model as formulated by Stoll. The data were analyzed in 4 decade regimes, 0.1–1, 1–10, 10–100, and 100–1000 kHz. The results were analyzed with respect to four criteria: (1) The sensitivity of the inversion to each parameter was determined by calculating the rotated coordinates. (2) The accuracy of the inversion was computed by comparing the parameter estimates to the true values. (3) The precision of the estimates was qualitatively determined using scatter plots of the accepted parameter values versus the cost function. (4) The estimates were considered for their ability to predict the broadband acoustic behavior over the broadband.

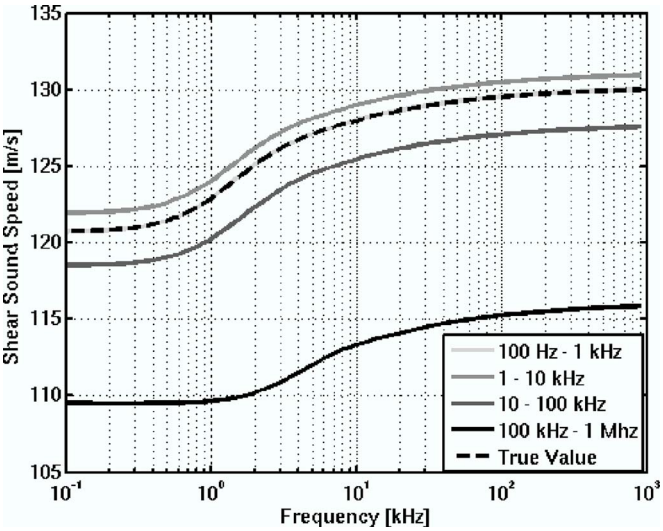


FIG. 8. The shear sound speed associated with the parameters in Table I and the inversion results in Tables II–V.

There are four primary results of this work. First, inversions based on reflection coefficient data give reasonable parameter estimates, especially for the most sensitive parameters. Three parameters were accurately determined for every frequency band: porosity, fluid bulk modulus, and frame shear modulus. Additionally, the lowest two frequency bands give good estimates for the permeability. The predicted accurate estimate of permeability and frame shear modulus demonstrates improvement over the normal reflection coefficient inversion which required additional measurements or interrelational expressions to estimate these quantities.[15] For the highest frequency band, the estimate for porosity is somewhat less accurate due to a coupling with the tortuosity. Since reflection coefficients can be measured noninvasively, there is potential that such inversions could provide a means of characterizing ocean seabeds.

Second, the sensitivity of the reflection coefficient data to the Biot parameters is frequency dependent. This is especially evident in the case of permeability. In the lowest two frequency regimes, in which the dispersion transition is apparent in the reflection coefficient, permeability is highly sensitive and a good estimate for permeability is determined. In contrast, for frequencies above the transition region, permeability is less sensitive and accurate estimates are not obtained.

Third, data taken over a limited high-frequency range may not be able to fully describe broadband acoustic behavior. Inversion results for single-decade frequency bands that included a portion of the low-frequency to high-frequency sound speed transition region successfully predict the dispersion, attenuation, normal reflection coefficient, and shear speed for the entire four-decade band. However, when only frequencies above the transition region were used in the inversion, the resulting parameter estimates were unable to accurately predict the broadband behavior.

Fourth, reasonable estimates of the shear speed were determined using reflection coefficient data. This result was unexpected due to the large speed mismatch of the waterborne sound and the sediment shear wave. However, the frame shear modulus was found to be sensitive due to the influence of the shear speed on low grazing angle reflection data.

Finally, it should be noted that these conclusions are based on inversions from data simulated by a particular realization of the Biot-Stoll model. Significant differences have been noted when comparing the Biot-Stoll model to experimental data,[9] implying a more complicated physical model. Also, experimental data may include effects of interface roughness, volume inhomogeneities, range dependence, layering, and other experimental effects. The influence of these effects on the inversion will be explored in subsequent publications.

ACKNOWLEDGMENTS

The authors would like to thank the Office of Naval Research and Robert Headrick for sponsoring this work.

Thanks also to Dr. Nicholas Chotiros for many helpful discussions. Lastly, thanks for Dr. Henrik Schmidt and Dr. Morris Stern, who developed the OASES code which was the basis of the forward model.

[1]N. Chotiros, "Inversion and sandy ocean sediment," in *Full Field Inversion Methods in Ocean and Seismic Acoustics*, Transport Processes in Porous Media edited by A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer Academic, Dordrecht, 1995).
[2]M. Isakson and T. Neilsen, "A comparison of elastic and poro-elastic models for inversions of reflection loss measurements of a smooth water/sand interface at high frequencies," in *Acoustic Inversion Methods and Experiments for the Assessment of the Shallow Water Environment* (Kluwer Academic, Dordrecht, in press).
[3]K. Williams, D. Jackson, E. Thorsos, D. Tang, and S. Schock, "Comparison of sound speed and attenuation measured in a sandy sediment to predictions based on the Biot theory of porous media," IEEE J. Ocean. Eng. **27**, 413–428 (2002).
[4]T. Yamamoto and A. Turgut, "Acoustic wave propagation through porous media with arbitrary pore size distributions," J. Acoust. Soc. Am. **83**, 1744–1751 (1988).
[5]J. Tatersall and D. Chizhik, "Application of biot theory to the study of acoustic reflection from sediments," in *Technical Report* NUWC-NLTR-10-115 (Naval Undersea Warfare Center, New London, 1992).
[6]M. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[7]R. Stoll, *Sediment Acoustics, Lecture Notes in Earth Science* (Springer, Berlin, 1983).
[8]K. Williams, "An effective density fluid model for acoustic propagation in sediments derived from Biot theory," J. Acoust. Soc. Am. **110**, 2276–2281 (2001).
[9]N. Chotiros and M. Isakson, "A broadband model of sandy ocean sediments: Biot-Stoll with contact squirt flow and shear drag," J. Acoust. Soc. Am. **116**, 2011–2022 (2004).
[10]M. Richardson, K. Briggs, L. Bibee, P. Jumars, W. Sawyer, D. Albert, R. Bennett, T. Berger, M. Buckingham, N. Chotiros, P. Dahl, N. Dewitt, P. Fleischer, R. Flood, C. Greenlaw, D. Holliday, M. Hulbert, M. Hutnak, P. Jackson, J. Jaffe, H. Johnson, D. Lavoie, A. Lyons, C. Martens, D. McGehee, K. Moore, T. Orsi, J. Piper, R. Ray, A. Reed, R. Self, J. Schmidt, S. Schock, F. Simonet, R. Stoll, D. Tang, D. Thistle, E. Thorsos, D. Walter, and R. Wheatcroft, "Overview of SAX99: Environmental considerations," IEEE J. Ocean. Eng. **26**, 26–51 (2001).
[11]N. Chotiros, "An inversion for biot parameters in water-saturated sand," J. Acoust. Soc. Am. **112**, 1853–1868 (2002).
[12]C. Harrison, "Noise modeling, noise experiment and noise inversion," J. Acoust. Soc. Am. **118**, 1844–1845 (2005).
[13]C. Holland and B. Brunson, "The Biot-Stoll sediment model: an experimental assessment," J. Acoust. Soc. Am. **84**, 1437–1531 (1998).
[14]C. Holland, J. Dettmer, and S. Dosso, "Remote sensing of sediment density and velocity gradients in the transition layer," J. Acoust. Soc. Am. **118**, 163–177 (2005).
[15]S. Schock, "A method for estimating the physical and acoustic properties of the sea bed using chirp sonar data," IEEE J. Ocean. Eng. **29**, 1200–1217 (2004).
[16]M. Clennell, "Tortuosity: A guide through the maze," in *Developments in Petrophysics, Geo. Soc. Special Pub. 122*, edited by M. Lovell and P. Harvey (Geo. Soc., London, U.K., 1997).
[17]F. Dullien, *Porous Media, Fluid Media and Pore Structure* (Academic, San Diego, CA, 1992).
[18]S. Forster, B. Bobertz, and B. Bohling, "Permeability of sands in the coastal areas of the Southern Baltic Sea: Mapping a grain-size related sediment property," Aquat. Geochem. **9**, 171–190 (2003).
[19]D. Taylor-Smith, "Geophysical-geotechnical predictions," in *Shear Waves in Marine Sediments*, edited by J. Hovem, M. Richardson, and R. Stoll (Kluwer Academic, Dordecht, 1993), pp. 725–734.
[20]R. Stoll, "Velocity dispersion in water-saturated granular sediment," J. Acoust. Soc. Am. **111**, 785–793 (2002).
[21]M. Richardson, K. Williams, K. Briggs, and E. Thorsos, "Dynamic measurement of sand grain compressibility at atmospheric pressure: Acoustic applications," IEEE J. Ocean. Eng. **27**, 593–601 (2002).
[22]P. Carmen, *Flow of Gases through Porous Media* (Academic, New York, 1956).
[23]B. Luke, *In situ measurement of stiffness profiles in the seafloor using the spectral-analysis-of-waves (SASW) method* (Technical Report under ARL:UT Independent Research and Development Program, Austin, TX) (1995).
[24]K. Briggs, "Comparison of measured compressional and shear wave velocity values with predictions from biot theory," in *Shear Waves in Marine Sediments*, edited by J. Hovem, M. Richardson, and R. Stoll (Kluwer Academic, Dordecht, 1991), pp. 121–130.
[25]B. Brunson, "Shear wave attenuation in unconsolidated laboratory sediments," in *Shear Waves in Marine Sediments*, edited by J. Hovem, M. Richardson, and R. Stoll (Kluwer Academic, Dordecht, 1991), pp. 141–148.
[26]D. W. Bell and D. J. Shirley, "Temperature variation of the acoustical properties of laboratory sediments," J. Acoust. Soc. Am. **68**, 227–231 (1980).
[27]H. Schmidt, *OASES Version 2.1 User Guide and Reference Manual* (Department of Ocean Engineering, Massachusetts Institute of Technology, Cambridge, MA, 1997).
[28]M. Stern, A. Bedford, and H. Millwater, "Wave reflection from a sediment layer with depth-dependent properties," J. Acoust. Soc. Am. **77**, 1,781–1,788 (1985).
[29]W. Goffe, G. Ferrier, and J. Rogers, "Global optimization of statistical functions with simulated annealing," J. Econometr. **60**, 65–99 (1994).
[30]M. Collins and L. Fishman, "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**, 1637–1644 (1995).
[31]W. W. H. Press, A. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (Cambridge University Press, Cambridge, U.K., 1992).
[32]H. Szu and R. Hartley, "Fast simulated annealing," Phys. Lett. A **122**, 157–162 (1987).
[33]S. Jaschke and N. Chapman, "Matched field inversion of broadband data using the freeze bath method," J. Acoust. Soc. Am. **106**, 1838–1851 (1999).
[34]S. Dosso and P. L. Neilsen, "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data," J. Acoust. Soc. Am. **111**, 143–159 (2002).
[35]S. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).

Background noise cancellation of manatee vocalizations using an adaptive line enhancer

Zheng Yan
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Christopher Niezrecki[a)]
Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854

Louis N. Cattafesta III
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Diedrich O. Beusse
College of Veterinary Medicine, University of Florida, P.O. Box 100126, Gainesville, Florida 32610-0126

(Received 29 November 2005; revised 10 April 2006; accepted 11 April 2006)

The West Indian manatee (*Trichechus manatus latirostris*) has become an endangered species partly because of an increase in the number of collisions with boats. A device to alert boaters of the presence of manatees is desired. Previous research has shown that background noise limits the manatee vocalization detection range (which is critical for practical implementation). By improving the signal-to-noise ratio of the measured manatee vocalization signal, it is possible to extend the detection range. The finite impulse response (FIR) structure of the adaptive line enhancer (ALE) can detect and track narrow-band signals buried in broadband noise. In this paper, a constrained infinite impulse response (IIR) ALE, called a feedback ALE (FALE), is implemented to reduce the background noise. In addition, a bandpass filter is used as a baseline for comparison. A library consisting of 100 manatee calls spanning ten different signal categories is used to evaluate the performance of the bandpass filter, FIR-ALE, and FALE. The results show that the FALE is capable of reducing background noise by about 6.0 and 21.4 dB better than that of the FIR-ALE and bandpass filter, respectively, when the signal-to-noise ratio (SNR) of the original manatee call is −5 dB. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2202885]

PACS number(s): 43.30.Sf, 43.60.Bf [EJS] Pages: 145–152

I. INTRODUCTION

According to the United States Coast Guard, the number of registered boats in Florida has grown to over 900 000 as of 2001 (United States Coast Guard, 2002). The population of the West Indian manatee (*Trichechus manatus latirostris*) has also increased slightly in recent years, reaching an estimated population of 3276. Between 1995 and 2002 the percentage of mortalities of the West Indian manatee due to watercraft strikes has risen from 22% to 31% (Florida Department of Environmental Protection, Division of Marine Resources, 1996; Florida Fish and Wildlife Conservation Commission, 2002). This has led to increased research into manatee avoidance technologies. A spatially fixed system to alert boaters of the presence of manatees is desired, as opposed to a device mounted on an individual boat that would be cost prohibitive to boaters and not likely adopted by the boating community. Once the presence of one or more manatees is detected within a navigable channel, their presence, and thus the need for boaters to slow to idle speed, could be signaled via a variety of methods, such as brilliant flashing strobe lights atop strategically placed marker pilings straddling that section of channel where manatees are present. Several methods to detect manatees have been proposed and include (1) a passive acoustic based detection system (Herbert *et al.*, 2002; Mann *et al.*, 2002; Niezrecki *et al.*, 2003), (2) an above water infrared detection system (Keith, 2002), and (3) an underwater active sonar based system (Bowles, 2002). A more detailed literature review of manatee vocalizations can be found in the paper by Niezrecki *et al.* (2003).

It is important to point out that in some situations manatees can exhibit long periods of silence when no vocalizations are made (Nowacek *et al.*, 2003). The implementation of a passive acoustic detection system will certainly have to account for periods when a manatee is not vocalizing. Assuming the detection ranges are sufficiently large for feasibility, the system could be implemented to slow boats down for a long period of time. If a manatee is detected, the warming system could be configured to indicate that a manatee is present in the waters for a period of hours or even a day.

[a)]Electronic mail: christopher_niezrecki@uml.edu

Therefore, calling rates are probably not as important to address as the manatee vocalization detection range and are not likely to limit the feasibility of a system.

Previous research has shown that background noise limits the manatee vocalization detection range of acoustic-based detection systems (Phillips *et al.*, 2005, Yan *et al.*, 2005). Primary examples include boat and snapping shrimp noise. Secondary noise sources are generated by wind, rain, water movement, and other biological species. Having a system with a large detection range is critical for practical implementation. By improving the signal-to-noise ratio (SNR) of the manatee vocalization signal, it is possible to extend the detectable range of manatee vocalizations while also reducing the false alarm rate and the number of missed calls.

An efficient method to improve the SNR of manatee calls using a finite impulse response (FIR) adaptive line enhancer (ALE) was previously proposed by Yan *et al.* (2005). However, the performance of the FIR-ALE is limited by several factors. First, the misadjustment, defined as the dimensionless ratio of the average excess mean square error to the minimum mean square error, is given by Widrow *et al.* (1976).

$$M = \mu L \phi_{xx}(0) \tag{1}$$

where μ represents the step size of the adaptive filter, L is the order (i.e., number of weighting coefficients) of the adaptive filter, and ϕ_{xx} is the autocorrelation function for the input $x(k)$ of the ALE,

$$\phi_{xx}(j) = E[x(k)x(k+j)], \tag{2}$$

where $E[]$ represents the expected value. Increasing L will narrow the filter pass band about the harmonics, thus improving the estimate of signal amplitude for a given SNR of the input. On the other hand, a small L is desired to minimize the misadjustment, as shown in Eq. (1), and reduce the required computation time. Furthermore, although an increase in M can be compensated for by decreasing the step size μ, a small step size will unfortunately decrease the convergence rate and the ability of the adaptive filter to track a nonstationary signal such as a manatee vocalization.

An alternative approach uses an infinite impulse response (IIR) structure. Although a FIR filter is easier to implement, its performance is generally inferior to an IIR filter with the same order. That is to say, the IIR structure ALE can provide the same performance as FIR-ALE with much lower order (Chang, 1993). The primary challenge associated with an IIR filter is maintaining its stability.

Within this paper, a constrained IIR adaptive line enhancer, proposed by Glover and Chang, which is called feedback ALE (FALE), is implemented to reduce the background noise (Glover and Chang, 1989). Additionally, a bandpass filter is used as a baseline to compare the performance of FIR-ALE and FALE. In the previous work by. Yan *et al.*., only four vocalizations were evaluated (Yan *et al.*, 2005). In this paper a library of 100 manatee calls was created, and each call was placed into one of ten categories. The performance of each method is evaluated via the 100 different calls. This allows a more comprehensive evaluation of the various algorithms.

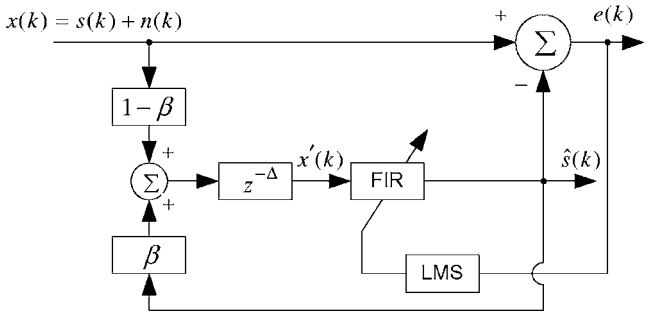


FIG. 1. Block diagram of the feedback adaptive line enhancer (FALE).

The paper is organized as follows. The theoretical development of the FALE algorithm is presented in Sec. II, and the library of the manatee calls is established in Sec. III. The simulation results are shown in Sec. IV, and the conclusions are discussed in Sec. V.

II. THEORETICAL DEVELOPMENT OF THE FEEDBACK ALE

Widrow *et al.* first proposed the adaptive line enhancer based on the Widrow-Hoff least-mean-square (LMS) adaptive algorithm (Widrow *et al.*, 1975). As stated earlier, the ALE can be classified into two main categories: FIR- (an all zero filter) based and IIR- (a pole-zero filter) based algorithms. The performance of the FIR-ALE implemented by Yan *et al.* on the background noise cancellation of manatee vocalizations is described in a prior paper (Yan *et al.*, 2005). A constrained IIR adaptive line enhancer called FALE was proposed by Glover *et al.* and is implemented to reduce the background noise of the manatee calls studied in this paper. The block diagram of the FALE is shown in Fig. 1. The primary input $x(k)$ is assumed to be of the form

$$x(k) = s(k) + n(k), \tag{3}$$

where, for the manatee problem, $s(k)$ represents the manatee vocalization, $n(k)$ represents the background noise, and $x(k)$ represents the observed manatee call corrupted by background noise, which is measured by a single hydrophone. The reference input $x'(k)$ is a delayed version of the signal that is the summation of the scaled version of the primary input $x(k)$ and the narrow-band output $\hat{s}(k)$:

$$x'(k) = \beta \hat{s}(k-\Delta) + (1-\beta)x(k-\Delta), \tag{4}$$

where β is the feedback constant and Δ represents the delay parameter that decorrelates the background noise. The enhanced narrow-band signal $\hat{s}(k)$ is added to the primary input (via the feedback path). The narrow-band output, when initially fed back, is matched in phase to that in the primary input but is smaller in amplitude and corrupted by residue noise (Glover and Chang, 1989). As the narrow-band output is refiltered, the noise component is progressively reduced. This process improves the correlation between the narrow-band signal in the reference and primary inputs.

The transfer function between the primary input and the narrow-band output for the algorithm of FALE, $H(z)$, is given by

$$H(z) = \frac{\hat{S}(z)}{X(z)} = \frac{(1-\beta)z^{-\Delta}F(z)}{1-\beta z^{-\Delta}F(z)}, \tag{5}$$

where $F(z)=\sum_{k=0}^{L-1} a_k z^{-k}$ is the z transform of the weights of the adaptive FIR filter. $z^{-\Delta}F(z)$ represents the transfer function of the FIR filter with the line delay included. $|F(z)|$ is less than 1 on the unit circle and increases to infinity at the origin because the poles of the FIR structure are located at the origin. Since $0<\beta<1$, and $1/\beta>1$, the roots of the denominator of $H(z)$ must lie within the unit circle. Therefore, FALE is stable as long as $|F(z)|$ is less than 1 on the unit circle (Chang, 1993).

The convergence rate of an adaptive filter is an important factor for tracking a nonstationary signal. However, the poles of the IIR filter decrease the convergence rate compared to an FIR filter. Therefore, a larger step size is needed for the FALE to track nonstationary signals. However, the feedback structure of the FALE moves the zeros and poles closer to the unit circle as $\beta \rightarrow 1$, which makes the bandwidth narrower. Therefore, a small disturbance in the weights may cause a large fluctuation in the frequency response of the FALE, and a smaller step size is required to maintain a stable adaptation process of the FALE. Hence, there exists a tradeoff between the stability and tracking ability of the FALE. Marshall suggested that if the FALE is able to track a nonstationary signal, increasing the amount of feedback β can improve the accuracy of its instantaneous frequency estimate (Marshall, 1994). However, the simulation results presented in Sec. IV show that if the FALE cannot track a nonstationary signal, its performance is worse than that of the FIR-ALE. Therefore, the step size should be large enough to track the nonstationary signal of interest yet small enough to maintain stability.

If $\beta=0$, then the FALE reduces to the FIR-ALE. There exists a range of β that presumably makes the FALE superior to the FIR-ALE. The simulations of Glover and Chang show that the "optimum" value of β varies from 0.4 for high SNR cases to somewhat less than 0.9 for lower SNR cases. For $\beta<0.4$, the impact of the feedback within the systems is not readily apparent, while for $\beta>0.9$, the estimation error is increased by the adaptation oscillations due to feedback (Glover and Chang, 1989). The feedback constant β is therefore set to 0.85 in this paper.

Since the fundamental frequency of a manatee calls typically lies between 2 and 5 kHz, a bandpass filter (tenth-order Butterworth IIR filter) is used as baseline system to compare and evaluate the performance of the FIR-ALE and FALE. The pass band of the filter is given by

$$f_1 < f < f_2, \tag{6}$$

where $f_1=1.2$ kHz and $f_2=20$ kHz. The bandpass filter is also used to preprocess the data before applying the adaptive filter. This greatly reduces the energy of noises at low frequencies that may degrade the performance of the ALE. In order to reduce the noise with low frequencies, the value of f_1 cannot be set too small. Since the highest frequency of the manatee calls is less than 20 kHz, the value of f_2 is set to 20 kHz. Experience has shown that a significant portion of the low-frequency noise can be reduced without significantly affecting a manatee call by preprocessing the signal with a bandpass filter.

It should be mentioned that the fundamental frequency of some manatee calls may be as low as 600 Hz and the manatee calls without harmonics are generally of higher frequency, around 4 or 5 kHz (Schevill and Watkins, 1965). The manatee calls with a fundamental frequency around 600 Hz generally have most of their energy in the higher harmonics and so they are likely to be passed by the high-pass filter.

III. ESTABLISHMENT OF THE LIBRARY OF THE MANATEE CALLS

O'Shea and the United States Geological Survey created an extensive library of manatee recordings between 1981 and 1984 (O'Shea, 1981–1984). Yan *et al.* used these recordings to quantify the performance of the FIR-ALE algorithm. However, only four manatee calls were used in their simulations which are not enough to fully evaluate the performance of the FIR-ALE versus the FALE (Yan *et al.*, 2005). In order to better test the performance of the different algorithms a library of the manatee calls is established. The library consists of ten different categories that include a total of 100

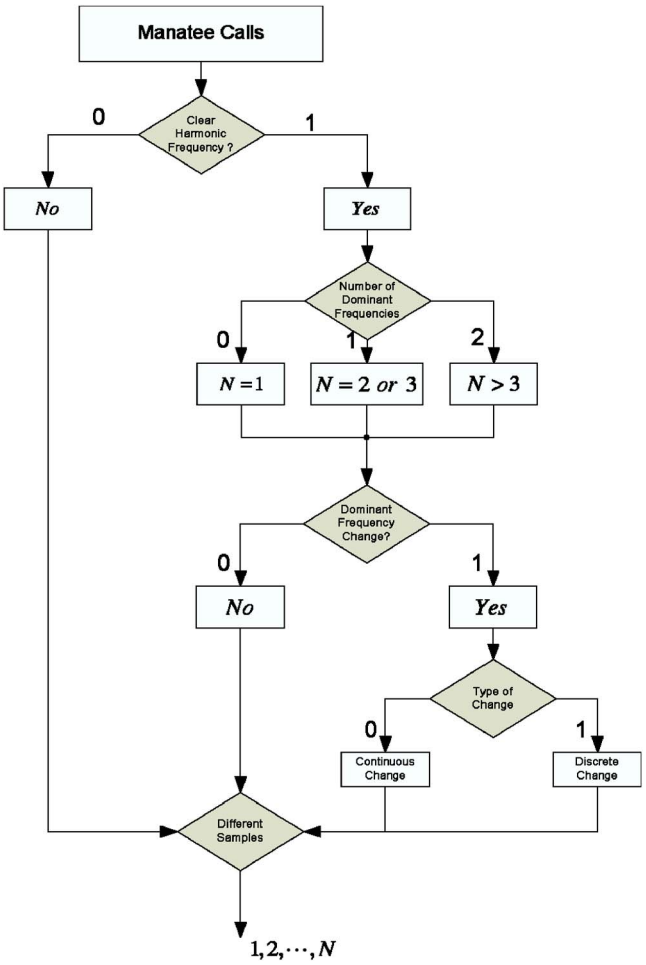


FIG. 2. The procedure used to categorize manatee calls.

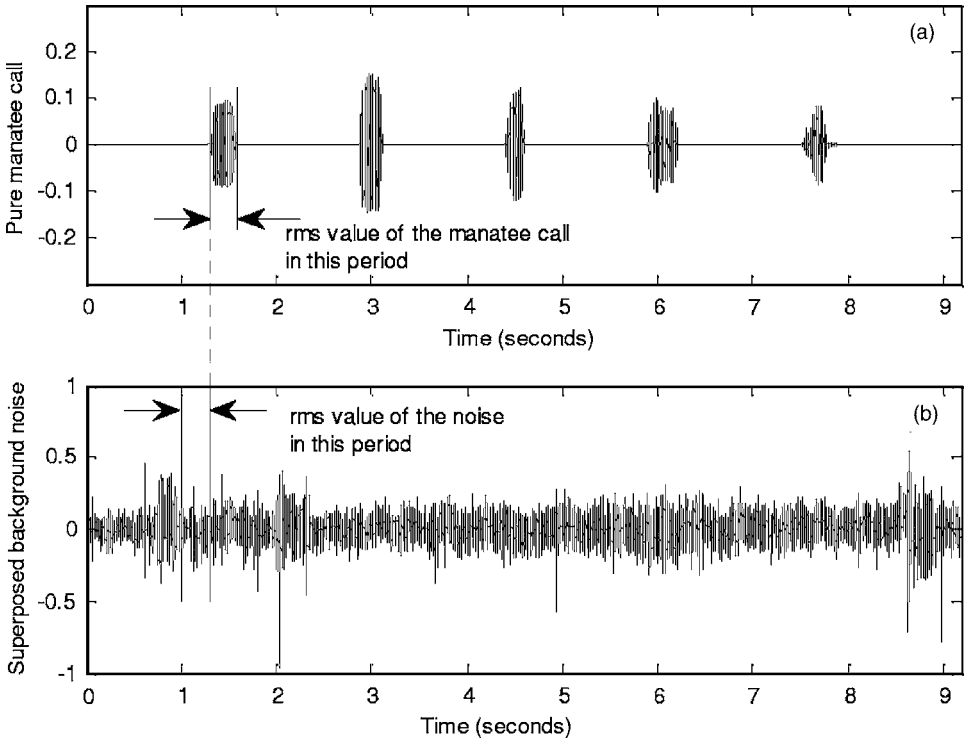


FIG. 3. The method used to compute the SNR of an original manatee call. (a) Pure manatee call and (b) background noise.

different manatee calls. Each category contains ten calls that were obtained from the extensive library of recordings created by O'Shea. Therefore, a total of 100 manatee calls are used to evaluate the performance of these three algorithms, i.e., bandpass filter, FIR-ALE, and FALE.

It is important to point out that categorization is performed purely from a signal detection perspective. No attempt is being made to infer what the significance of each category indicates in terms of manatee behavior. The categorization procedure is shown in Fig. 2. The first level of categorization discriminates a vocalization that either does or does not have some discernable harmonic content in which the dominant frequencies are at least 20 dB larger than the neighborhood frequencies. Likewise, if the powers of several frequencies of an individual manatee call are at least 20 dB larger than their neighborhood harmonic frequencies and the difference between them is less than 20 dB, all of these frequencies of the manatee call can be called dominant frequencies. Most manatee vocalizations do have harmonic structure. For the calls that do possess harmonics, a further subdivision is to identify calls that either have one, two (or three), or more than three dominant harmonics. The next level of categorization is to identify if the calls have a dominant frequency change. A frequency change is defined as a frequency shift of the strongest harmonic (for an individual manatee vocalization) in excess of 10%. These types of vocalizations can be used to test the frequency tracking ability of these two ALE algorithms. For those calls that do have a dominant frequency change, the last level of decomposition categorizes the frequency change as being continuous or discrete. The discrete frequency change is defined as a frequency shift in excess of 10% within a duration of 10 ms.

In order to discriminate between categories, a labeling system is adopted. The codes 0, 1, and 2 are used to represent different categories, and each manatee vocalization is categorized by a four digit number. For example, a manatee call with code 1111 indicates that it has clear harmonic frequency content, the number of dominant frequencies is two or three, and the dominant frequency changes discretely with time. The ten different possible categories in the flow chart are represented within the library by ten different manatee calls that all have the same characteristics. For a particular category, the following characteristics may be different from one call to the next: (1) the location of the dominant frequency, (2) the shape of the envelope of each manatee vocalizations in the time domain, (3) the power distribution of the harmonics, and (4) the overall power of the manatee call.

In some manatee calls, no distinct harmonic frequencies occur, but these calls can still be regarded as narrow-band

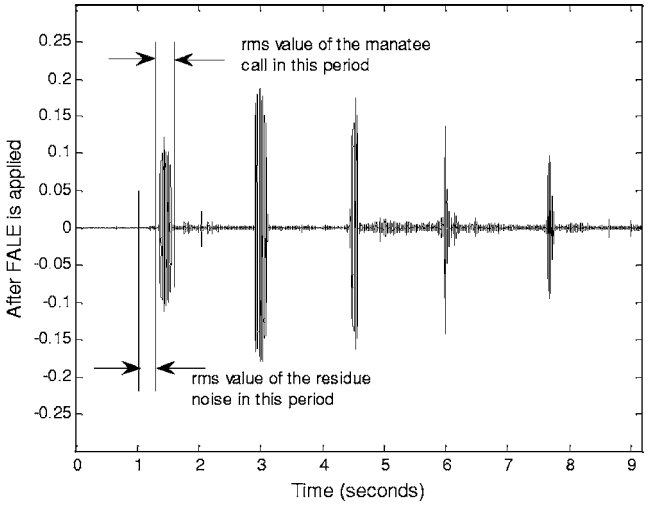


FIG. 4. The method used to compute the SNR of the manatee call after filtering (using the FALE in this case).

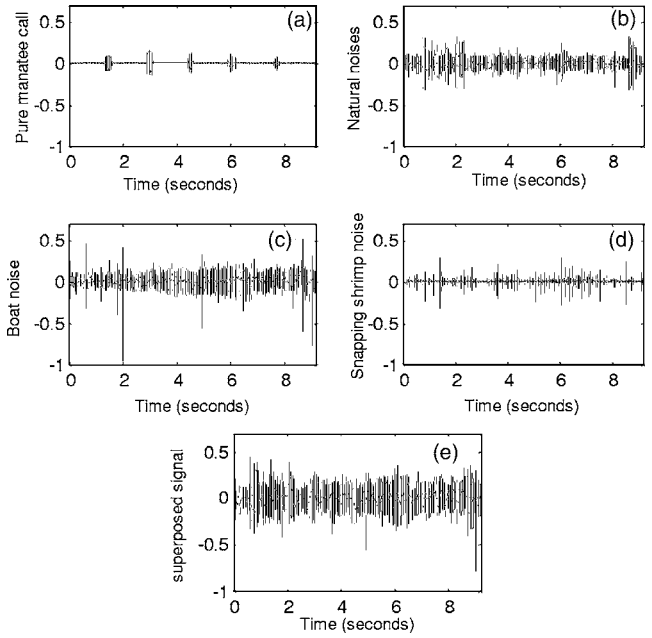


FIG. 5. (a) Pure manatee call, (b) natural noise, (c) boat-dominated noise, (d) snapping shrimp noise, and (e) superposition of manatee calls, natural noise, boat-dominated noise, and snapping shrimp noise.

signals when compared with the background noise. The narrower the bandwidth of the manatee calls and the wider the background noise, the better the performance of the ALE (Yan *et al.*, 2005). Therefore, for the manatee calls without distinct harmonic frequency, the performance of the ALE may degrade to some degree. Although the number of dominant frequencies of the manatee call may be one, two, three, or more, the energy of most manatee calls is dominated by one or two harmonics.

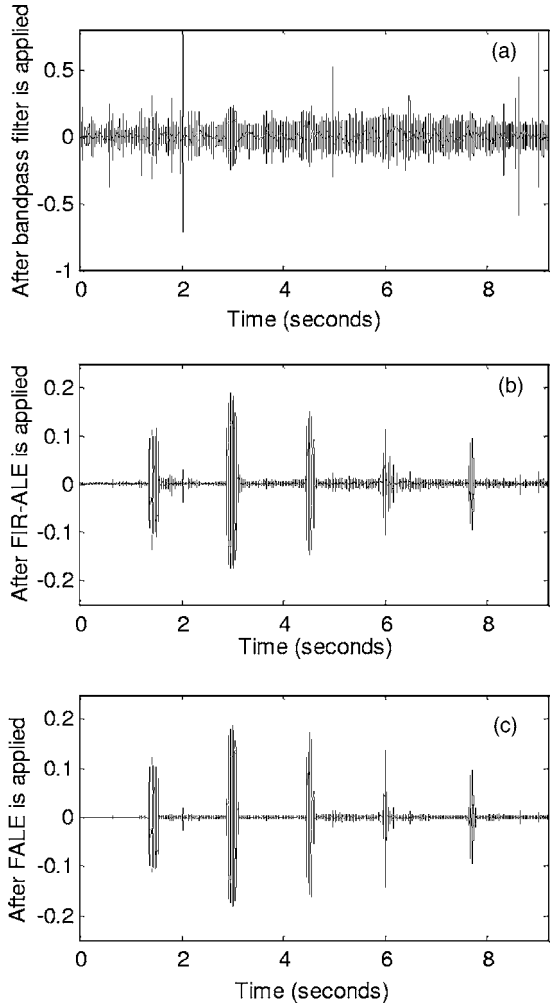


FIG. 6. (a) Superposed signal after the bandpass filter is applied, (b) superposed signal after FIR-ALE is applied, and (c) superposed signal after FALE is applied.

IV. SIMULATION RESULTS

Although there are many sources of underwater noise, the two primary sources of background noise that typically corrupt a manatee call are boat noise and snapping shrimp noise. In practice, a manatee call also may be corrupted by noise created by rain, wind, fish, marine mammals, human activity, wave motion, etc. Noise other than snapping shrimp noise and boat noise are classified as "natural noise" in these simulations. Therefore, three types of noise are considered: boat noise, snapping shrimp noise, and natural noise.

The performance of each algorithm is compared using the SNR. However, after filtering, the residue background noise cannot be separated from the manatee call; hence it is not possible to distinguish the noise and the manatee call during the time interval of the manatee call. Therefore a modified definition of SNR is used in this work. Five pure manatee calls and superposed background noise are shown in Figs. 3(a) and 3(b), respectively. $\mathrm{SNR}_{\mathrm{ori}}$ represents the estimated SNR of the original manatee call corrupted by noise. As shown in Fig. 3(a), $\mathrm{SNR}_{\mathrm{ori}}$ is computed by taking the root mean square (rms) value of the time domain signal in the region where the pure manatee call is present and dividing that value by the rms value over the same time interval just prior to the call where only the background noise is present. It is assumed that the background noise levels do not vary significantly over the duration of a manatee call. Hence, the same method is used to estimate the SNR of the manatee call after filtering (see Fig. 4). As a result, the residue noise during the interval of the manatee call is unavoidably added to the true signal power. Therefore, the SNR of the manatee call after filtering is a little biased.

In order to compare the performance of the FIR-ALE and FALE, the orders for these two algorithms are both set to 20 and the step sizes are set to 0.05 and 0.5, respectively. Typical time domain measurements of pure manatee calls (category 1000), natural noise, boat noise, snapping shrimp noise, and the superposition of these four signals are shown in Figs. 5(a)–5(e), respectively. The superposed signal after the bandpass filter, FTR-ALE, or FALE is applied are shown in Figs. 6(a)–6(c), respectively. A purely qualitative visual comparison of these results indicates that FALE is most effective at improving the SNR.

Next, the library of 100 manatee calls is used in the simulations to rigorously test these three algorithms. In simulations, the database of the manatee call is organized into 20 recordings. Each recording has five manatee calls. Therefore, each category has two recordings. In order to equally test

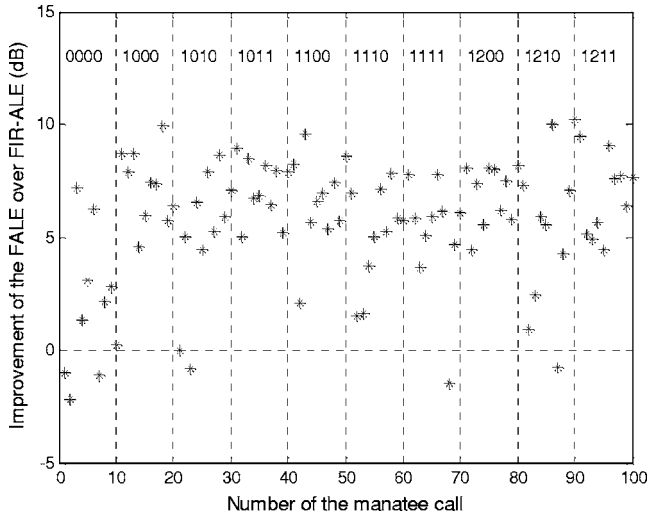


FIG. 7. SNR improvement of the FALE compared to FIR-ALE for each manatee call when SNR_{ori} is equal to −5 dB.

each category, the background noise used is the same for each recording. However, the background noise is different for five manatee calls within one recording. In order to equally compare performance, the SNR_{ori} is set to −5 dB for all manatee calls. The difference in the a SNR of the FALE versus the FIR-ALE for each manatee call is shown in Fig. 7. There are only six manatee calls (out of 100) for which the performance of the FALE is worse than that of the FIR-ALE. A qualitative inspection suggests that this is caused by the relatively large changes in the characteristic frequencies for three of the manatee calls and no clearly discernable harmonic frequencies for the other three manatee calls. The chosen settings of the FALE cannot adequately track the variation of the manatee call, especially when the SNR of the manatee call after bandpass filter is very low. As discussed earlier, the tracking ability of the FALE is worse than that of the FIR-ALE because of the feedback constant. From Eq. (4), the weight in front of the primary input is small when the feedback constant β is set to a large value. Thus, it is no surprise that each manatee call has a different optimal step size and feedback constant. Due to practical considerations, it is very important to select a fixed β with a corresponding proper step size for FALE, which not only facilitates tracking a nonstationary signal but also maintains stability.

Using the notation that the average SNR of each original manatee call category is represented by $\overline{SNR_{ori}}$, the performance of the bandpass filter, FIR-ALE, and FALE for the manatee calls in each category and the overall average are shown in Table I. The average SNR of the original manatee call for each category is given by

$$\overline{SNR_{ori}} = 10 \times \log_{10}\left[\left(\sum_{i=1}^{10} P_i^2\right) \Big/ \left(\sum_{j=1}^{10} Q_j^2\right)\right], \qquad (7)$$

where P_i represents the rms value of the ith manatee call for a particular category and Q_j represents the rms value of the jth noise component. Likewise, the method to compute the average SNR of the processed manatee calls after these three enhancement algorithms are applied is the same as that used to compute $\overline{SNR_{ori}}$. $G_{FIR\text{-}BPF}$ and $G_{FALE\text{-}BPF}$ represent the gain or SNR improvement of FIR-ALE and FALE over the bandpass filter, respectively. $G_{FALE-FIR}$ represents the SNR improvement of the FALE over FIR-ALE. These simulation results show that both the FIR-ALE and FALE are effective at reducing the background noise of a manatee call. The average performance of the FALE is about 21.4 and 6.0 dB better than that of the bandpass filter and FIR-ALE, respectively. From Table I, it is seen that the performance of these two algorithms for category 0000 (manatee calls with no clearly discernable harmonics) is a little worse than the others. The improvement of FALE over FIR-ALE for this category is also smaller than others categories of calls.

To further assess the performance of each filter for different levels of background noise, the bandpass filter, FIR-ALE, and FALE are compared when $\overline{SNR_{ori}}$ varies from −15 to 0 dB. The performance of each filter are shown in Figs. 8(a), 8(b), and 8(c), respectively. The results indicate that as the SNR of the original manatee call is reduced, the noise-reduction performance of all algorithms is also reduced. The best performance is again achieved by the FIR-ALE and the FALE for manatee calls with one dominant frequency, while the worst performance is achieved for manatee calls without distinct harmonic frequencies. As shown in Fig. 8(a), the

TABLE I. The average performance of bandpass filter, FIR-ALE, and FALE for the manatee calls corresponding to each category (dB).

Category	$\overline{SNR_{ori}}$	After bandpass	After FIR-ALE	After FALE	$G_{FIR-BPF}$	$G_{FALE-BPF}$	$G_{FALE-FIR}$
0000	−5.0	3.3	12.6	15.0	9.3	11.7	2.4
1000	−5.0	3.5	21.7	28.1	18.2	24.6	6.4
1010	−5.0	3.8	18.8	25.0	15.0	21.2	6.2
1011	−5.0	3.4	19.6	26.0	16.2	22.6	6.4
1100	−5.0	3.3	18.2	24.8	14.9	21.5	6.6
1110	−5.0	3.3	16.7	22.1	13.4	18.8	5.4
1111	−5.0	3.8	19.7	25.0	15.9	21.2	5.3
1200	−5.0	3.6	19.4	25.9	15.8	22.3	6.5
1210	−5.0	3.9	16.7	22.7	12.8	18.8	6.0
1211	−5.0	3.7	20.4	26.2	16.7	22.5	5.8
Average	−5.0	3.6	18.9	25.0	15.4	21.4	6.0

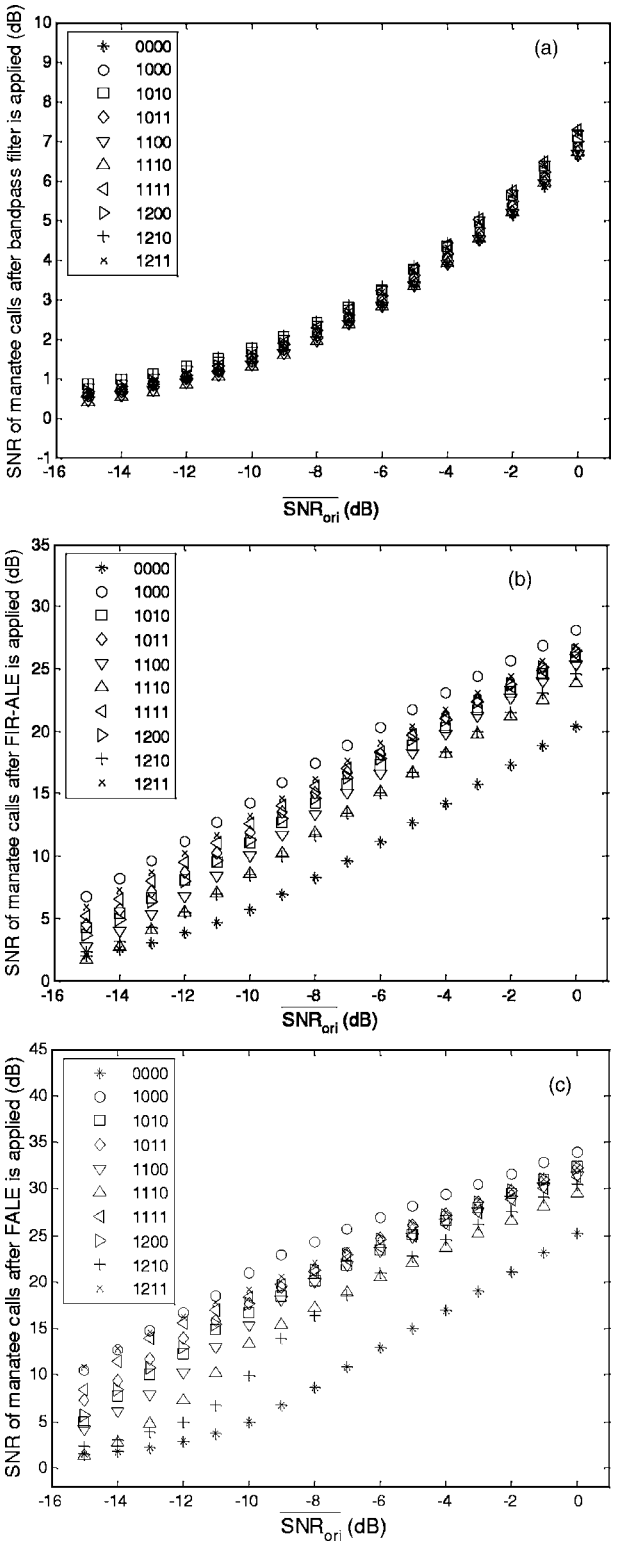


FIG. 8. (a) SNR of manatee calls after the bandpass filter is applied for each category and various SNR. (b) SNR of manatee calls after FIR-ALE is applied for each category and various SNRs. (c) SNR of manatee calls after FALE is applied for each category and various SNRs.

SNR improvement of the bandpass filter does not vary significantly from one category to the next as the background noise level is changed. However the FIR-ALE and FALE is dependent on the category of the manatee call selected as the background noise level is changed [see Figs. 8(b) and 8(c)].

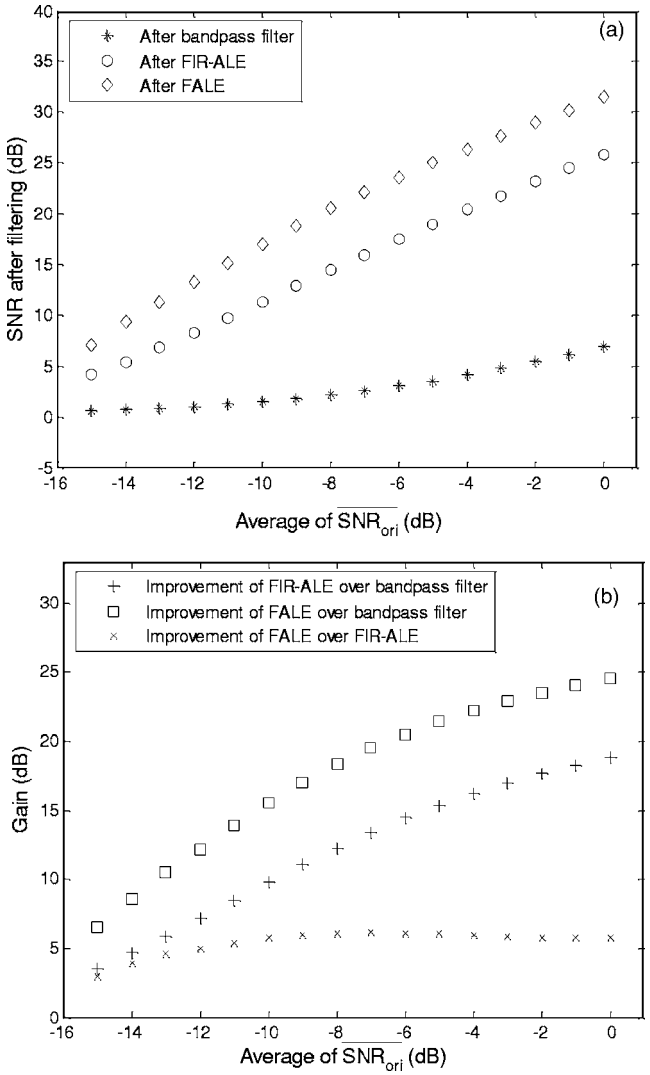


FIG. 9. (a) Overall performance comparison between the bandpass filter, FIR-ALE, and FALE as a function of the SNR. (b) Performance gains of the various algorithms using the same bandpass filter as a baseline as a function of the SNR.

Additional simulations are performed with the average SNR over ten categories to further evaluate the relative performance of the FIR-ALE and FALE algorithms. The overall average SNR of the original manatee calls, for all ten categories, is given by

$$\text{Average of } \overline{\text{SNR}_{\text{ori}}} = 10 \times \log_{10}\left[\left(\sum_{i=1}^{100} P_i^2\right) \Big/ \left(\sum_{j=1}^{100} Q_j^2\right)\right]. \tag{8}$$

Likewise, the method to compute the average SNR of the processed manatee calls after these three enhancement algorithms are applied is the same as that shown in Eq. (8). The performance comparison between the bandpass filter, FIR-ALE, and FALE are shown in Fig. 9(a) when the average of $\overline{\text{SNR}_{\text{ori}}}$ over ten categories varies from −15 to 0 dB. From this figure, the average SNR of the bandpass filter, FIR-ALE, and FALE are seen to be 0.6, 4.2, and 7.1 dB, respectively, when the average of $\overline{\text{SNR}_{\text{ori}}}$ is reduced to −15 dB.

A comparison of the performance achieved by the FIR-ALE and FALE compared to the baseline bandpass filter is shown in Fig. 9(b) when the average of $\overline{\mathrm{SNR}_{\mathrm{ori}}}$ varies from −15 to 0 dB. From this figure it is seen that the SNR improvement achieved by the FIR-ALE and FALE compared to the bandpass filter increases as the average of $\overline{\mathrm{SNR}_{\mathrm{ori}}}$ is increased. The improvement of the FALE over FIR-ALE becomes progressively larger when the average of $\overline{\mathrm{SNR}_{\mathrm{ori}}}$ is increased for low values of SNR up to −7 dB. This result is in agreement with the results by Chang (1993).

V. CONCLUSIONS AND FUTURE WORK

The practical implementation of an acoustic based manatee warning system is dependent on the minimum hydrophone spacing for the system. The required hydrophone spacing depends on the manatee vocalization strength, the decay of the acoustic signal strength with distance, and the background noise levels. In this paper, a feedback adaptive line enhancer algorithm (FALE) is used to reduce the underwater background noise in order to improve the SNR of a library of 100 manatee calls spanning ten different signal categories. Simulations over a range of SNR indicate that the FALE and FIR-ALE are much more effective than simply using a bandpass filter. The FALE is capable of improving the SNR by, on average, an additional 6 dB compared to the FIR-ALE when the original SNR exceeds −5 dB. Exceptions are found in instances when the FALE cannot track the rapid frequency changes of some manatee calls. The simulations also show that the feedback and step size are two important factors that affect the stability, tracking ability, and the performance of the FALE. The improved SNR can presumably be used to extend the detection range of manatee vocalizations and reduce the number of false alarms and the number of missed calls. The results of this work may ultimately be used to realize a practical system that can warn boaters of the presence of manatees.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Sea Grant, Florida Fish and Wildlife Conservation Commission, and University of Florida Marine Mammal Program in supporting this research.

Bowles, A. (**2002**). "Design for a manatee finder: sonar techniques to prevent manatee-vessel collisions," http://floridamarine.org/features/view_article.asp?id=14362

Chang, J. (**1993**). "The feedback adaptive line enhancer: a constrained IIR adaptive filter," Ph.D. dissertation, University of Houston.

Florida Department of Environmental Protection, Division of Marine Resources (**1996**). Save the Manatee Trust Fund, Fiscal Year 1995–1996, Annual Report, Florida Marine Research Institute, 100 Eighth Avenue S.E., St. Petersburg, FL 33701-5095.

Florida Fish and Wildlife Conservation Commission (**2002**). Save the Manatee Trust Fund, Fiscal Year 2001–2002, Annual Report, Florida Fish and Wildlife Conservation Commission, 620 South Meridian Street, OESBPS, Tallahassee, FL 32399-1600.

Glover, John R. Jr., and Chang, J. (**1989**). "The feedback adaptive line enhancer: a constrained IIR adaptive line enhancer," Twenty-third Asilomar Conference on Signals, Systems and Computers, 30 Oct.–1 Nov., Vol. **2**, pp. 568–570.

Herbert, T., Hitz, G., Mayo, C., Sermarini, C., Dobeck, G., Manning, B., Sandlin, M., Hansel, J., Bowden, T., and Artman, D. (**2002**). "Proof-of-concept for off the shelf technology to identify acoustic signature to detect presence of manatee(s)," http://floridamarine.org/features/view_article.asp?id=14362

Keith, E. O. (**2002**). "Boater manatee awareness system," http://floridamarine.org/features/view_article.asp?id=14362

Mann, D., Nowacek, D., and Reynolds, J. III. (**2002**). "Passive acoustic detection of manatee sounds to alert boaters," http://floridamarine.org/features/view_article.asp?id=14362

Marshall, B. H., Jr. (**1994**). "Analysis of the feedback adaptive line enhancer," Ph.D. dissertation, University of Houston.

Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am., **114**, 1640–1647.

Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). "Intraspecific and geographic variation of West Indian manatee," J. Acoust. Soc. Am. **114**(1), 66–69.

O'Shea, T. (**1981–1984**). "Manatee vocalization-catalog of sounds," produced by Coastal Systems Station, Naval Surface Warfare Center, Dahlgren Division, Panama City, FL.

Phillips, R., Niezrecki, C., and Beusse, D. O. (**2005**). "Detection ranges for acoustic based manatee avoidance technology," J. Acoust. Soc. Am. **117**, 2526.

Schevill, W. E., and Watkins, W. A. (**1965**). "Underwater calls of *Trichechus* (manatee)," Nature (London) **205**(4969), 373–374.

United States Coast Guard (**2002**). "Boating Statistics—2001," Commandant Publication P16754.15, 2100 Second Street SW, Washington, DC 20593-0001.

Widrow, B., McCool, J., Larimore, M., and Johnson, C., Jr. (**1976**). "Stationary and nonstationary learning characteristics of the LMS adaptive filter," Proc. IEEE, **64**, 1151–1162.

Widrow, B., Glover, J. R., Jr. McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E., Jr. and Goodlin, R. C. (**1975**). "Adaptive noise cancelling: principles and applications," Proc. IEEE, **63**(12), 1692–1716.

Yan, Z., Niezrecki, C., and Beusse, D. O. (**2005**). "Background noise cancellation for improved acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **117**, 3566–3573.

Theoretical detection ranges for acoustic based manatee avoidance technology

Richard Phillips
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250

Christopher Niezrecki[a)]
Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts, 01854

Diedrich O. Beusse
College of Veterinary Medicine, University of Florida, PO Box 100126, Gainesville, Florida 32610-0126

(Received 15 July 2005; revised 6 April 2006; accepted 18 April 2006)

The West Indian manatee (Trichechus manatus latirostris) has become endangered partly because of watercraft collisions in Florida's coastal waterways. To reduce the number of collisions, warning systems based upon detecting manatee vocalizations have been proposed. One aspect of the feasibility of an acoustically based warning system relies upon the distance at which a manatee vocalization is detectable. Assuming a mixed spreading model, this paper presents a theoretical analysis of the system detection capabilities operating within various background and watercraft noise conditions. This study combines measured source levels of manatee vocalizations with the modeled acoustic properties of manatee habitats to develop a method for determining the detection range and hydrophone spacing requirements for acoustic based manatee avoidance technologies. In quiet environments (background noise $\approx$70 dB) it was estimated that manatee vocalizations are detectable at approximately 250 m, with a 6 dB detection threshold. In louder environments (background noise $\approx$100dB) the detection range drops to 2.5 m. In a habitat with 90 dB of background noise, a passing boat with a maximum noise floor of 120 dB would be the limiting factor when it is within approximately 100 m of a hydrophone. The detection range was also found to be strongly dependent on the manatee vocalization source level.
 [DOI: 10.1121/1.2203597]

PACS number(s): 43.30.Sf, 43.30.Nb [WWA] Pages: 153–163

I. INTRODUCTION

According to Florida's Office of Boating and Waterways, the number of registered boats in Florida has grown to nearly one million as of 2003 (Florida Fish and Wildlife Conservation Commission, 2003b). During the last decade, the yearly percentage of mortalities of the West Indian manatee (Trichechus manatus latirostris) due to watercraft strikes has varied between 14 and 31 % (Florida Fish and Wildlife Conservation Commission, 2003a). This has lead to increased research into manatee avoidance technologies. One proposed method of manatee avoidance technology is based on detecting the presence of manatees by using hydrophones and listening for manatee vocalizations. The frequencies of manatee vocalizations, the source level of the vocalizations, the volume of the ambient background noise, and how that noise propagates in the relevant waterways, all affect the feasibility of an acoustic based detection system.

A comprehensive literature review on manatee vocalizations can be found in the work by Niezrecki *et al.* (2003). Nowacek *et al.* (2003) measured the mean received sound pressure levels of the peak frequency to be approximately 100 dB (re 1 μPa throughout the paper) and approximated the source levels to be within 6 to 15 dB of the measured level (Nowacek *et al.*, 2003). The received values were recorded with a hydrophone at approximately 20 m from a group of 50 manatees. Position estimation techniques have also been used with hydrophone arrays to approximate the location of the vocalizing manatee. By using the estimated position and the received sound pressure levels, the mean source level of the manatee was approximated to be 112 dB at 1 m (Phillips *et al.*, 2004).

Boat noise levels are also important to consider when detecting underwater acoustic signals. Boat noise is primarily generated underwater by both cavitating and noncavitating propellers. The formation and collapse of the bubbles created by cavitation create broadband sound. However, the majority of the acoustic energy generated by boating traffic is below 1 kHz. The Lloyd mirror effect and acoustic shadowing can both lead to significant levels of attenuation (Gerstein, 2002). Radiated noise from a 120 hp Mako 171 was measured at various speeds in the South Gandy Channel of Old Tampa Bay (Biddulph, 1993). The objective of their study was to obtain the tonal and broadband characteristics of the radiated noise at various speeds and aspects of a specific boat. Their study focused on the sound spectrum generated by the boat and not on the acoustic propagation of the sound. Another underwater noise study analyzed the relationship between ambient noise levels and manatee habitat use.

a)Electronic mail: Christopher_Niezrecki@uml.edu

Twenty-four habitats, including seagrass beds and dredged basins, were observed. The researchers showed that the high-use manatee, habitats had higher transmission losses and lower noise levels compared with the surrounding habitats (Miksis-Olds *et al.*, 2004, 2006).

The feasibility and practical implementation of an acoustic based manatee detection system is strongly dependent on its cost and the number of hydrophones required to monitor a channel or waterway. The required hydrophone spacing will depend on three important factors: (1) the intensity of the manatee sound signals; (2) the background noise levels; and (3) the decay of the signal's strength with distance. Without knowing all of this information it is difficult to quantify a detections system's useful range. The authors have previously determined the manatee vocalization source levels (Phillips *et al.*, 2004). Within this paper the other two missing components (2 and 3) are determined. This work compares several shallow water acoustic spreading models with data collected from the waterways where manatees frequent. The paper also provides a survey of the noise levels that are found in these waterways, which are attributed to boat noise as well as background noise. Within this paper knowledge of the manatee vocalization source levels is combined with the modeled acoustic properties of the habitats to develop a method for determining the minimum required hydrophone spacing for an acoustically based manatee detection system. This work contributes to the scientific knowledge of manatees and acoustically based manatee avoidance technology by providing information that is essential in determining the distance in which a manatee vocalization can be detected for a variety of background noise levels. The paper also addresses the feasibility of implementing an acoustic based manatee avoidance technology.

II. THEORETICAL DEVELOPMENT—ACOUSTIC SPREADING MODELS

In order to calculate the distance at which a manatee can be detected, the manner in which sound propagates in the Florida shallow water environments needs to be estimated. Acoustic modeling of shallow water environments can depend on numerous variables such as frequency, depth, substrate, interaction between upper and lower boundary layers, temperature, salinity, channel geometry, vegetation, and surface conditions. The locations relevant to this experiment are located in boating channels and rivers and consist of fairly flat or slightly sloping sea floors with depths less than 5 m. Because the distances and frequencies are relatively low, the sound absorption losses were neglected. Several different acoustic spreading models are now reviewed.

The two most basic acoustic spreading models are cylindrical spreading and spherical spreading. For spherical spreading, the sound pressure level difference between the acoustic source L_s and the received level L_r is shown in Eq. (1).

$$L_r = L_s - 20 \cdot \log(R) \quad R \geq 1. \tag{1}$$

For cylindrical spreading the pressure wave propagates between two parallel waveguides, such as the sea floor and surface. During cylindrical spreading the acoustic pressure decreases at approximately 3 dB with every doubling of the distance. The difference between the received and source levels during cylindrical spreading is shown by Eq. (2).

$$L_r = L_s - 10 \cdot \log(R) \quad R \geq 1. \tag{2}$$

An intermediate model between spherical and cylindrical spreading models is the mixed spreading model and is shown in Eq. (3) (Coates, 1989).

$$L_r = L_s - 15 \cdot \log(R) \quad R \geq 1. \tag{3}$$

A fourth model (Lurton, 2002) divides the spreading into a region of spherical spreading and a region of cylindrical spreading. Spherical spreading [Eq. (4)] is used at distances less than the transition range (r_0) and cylindrical spreading [Eq. (5)] is assumed after the transition range.

$$L_r = L_s - 20 \cdot \log(R) \quad 1 < R < r_0. \tag{4}$$

$$L_r = L_s - 20 \cdot \log(r_0) - 10 \cdot \log(R/r_0) \quad R \geq r_0. \tag{5}$$

The transition range is dependent upon the depth of the water (D), the sediment type, and the depth of the acoustic source. Considering that watercraft generate noise at or near the surface, the transition range r_0 is represented by Eq. (6)

$$r_0 = \frac{D}{\tan(\beta_0)}, \tag{6}$$

where β_0 is the critical angle of Snell's Law [Eq. (7)]. Beyond this angle all of the acoustic energy is reflected off of the sea floor and none is transmitted into the sediment. This angle is determined by the speed of sound in water (c_1) and the speed of sound in the sea floor (c_2). The physical and acoustic properties of sea floor sediments have been modeled by Hamilton and the sea floor at the experiment sites in this study are best described by Hamilton's definition of silt (Hamilton, 1980).

$$\beta_0 = \cos^{-1}(c_1/c_2). \tag{7}$$

Other more sophisticated models have also been created. Modified parabolic equation models have been used to determine transmission loss in shallow water environments (Collins and Chin-Bing, 1990; Jensen, 1984; Miksis-Olds and Miller, 2006; Smith 2001). An alternative model predicts transmission loss by using normal-mode and adiabatic normal-mode models with the sediment absorption calculated using the theory of Biot (Beebe *et al.*, 1982). Although these higher order models may be better at describing the detailed sound propagation characteristics within a particular channel having specific geometric, depth, and sedimentary characteristics, they are not necessarily more accurate for the analysis performed in this work. These higher order models typically predict a transmission loss that fluctuates with distance (or frequency) due to normal and refracted modes between the bottom and surface of the water. Therefore detailed transmission loss modeling at one location will not be valid at another, unless the depth, substrate, temperature, salinity, channel geometry, vegetation, and surface conditions are identical from place to place. This assumption is supported by the empirical data presented in the work by Miksis-Olds and Miller (2006) in which the Monterey-Miami parabolic

equation model generated results that were no more accurate than a simple cylindrical spreading model. The experimental results presented in the following section indicate that a mixed spreading. model is sufficient to estimate the transmission loss for this analysis.

III. EXPERIMENTAL MEASUREMENT OF ACOUSTIC SPREADING

In order to determine the most appropriate acoustic spreading model for the relevant shallow waters in Florida, the acoustic spreading was measured using two methods.

A. Chirp broadcast experiments

The first method used a chirp signal broadcast by a Clark Synthesis Aquacoustic AQ339 underwater speaker mounted approximately one meter from the bottom of the channel. A chirp signal is a sine wave whose frequency increases (or decreases) with time. For this experiment, the chirp signal consisted of a 1 kHz start frequency, a 10 kHz final frequency, and a 0.25 signal duration. Twenty averages were performed at each location to minimize the effects of any transient disturbances. The sound pressure level was then calculated from the rms pressure received at each hydrophone for the duration of the chirp signal. The sound pressure level was recorded on one hydrophone located 3.28 ft (1 m) from the speaker, mounted approximately one meter from the bottom of the channel. This hydrophone remained in place for all tests. A second roving hydrophone was used to record the signal at the following distances: 3.28, 15, 30, 45, 60, 75, and 100 ft (1, 4.6, 9.1, 13.7, 18.3, 22.9, and 30.5 m) for each chirp test. The signals were averaged 20 times to reduce the effects of any transient noise. This experiment was performed both parallel and perpendicular to the direction of the channel and at several locations.

Recordings were made at two locations at the edge of the channel in Crystal River, FL. Location "C.R. A" was at 28°53.745′N, 82°36.450′W. The water depth was approximately 10 to 12 ft in the area of the experiment. Location "C.R. B" was at 28°53.075′N, 82°36.794′W and the depth ranged from approximately 5 to 12 ft in the area surrounding the experiment. The substrate was predominantly muddy silt; however, certain areas inside the channel were rocky.

It should be noted that at location "C.R. A," two additional tests were conducted with the roving hydrophone for the distances previously specified. The roving hydrophone was placed 0.5 m below the surface and half way between the surface and the bottom. The results indicate that there was not a significant effect on the overall transmission loss measurements as a function of measurement depth for this location.

A second set of recordings were taken near Cedar Key, FL. Location "C.K. A" was at 29°07.641′N, 83°02.637′W and location "C.K. B" was at 29°06.590′N, 83°03.324′W. Both locations had a depth of approximately 6 to 7 ft and the bottom was sandy with a mixture of grassy areas.

A transfer function measurement between the roving hydrophone signal and the fixed hydrophone signal for a typical chirp test is shown in Fig. 1(a). The magnitude of the response is directly related to the ratio of the sound pressure at the roving hydrophone position compared to the sound pressure at the fixed hydrophone position. As expected the phase plot [Fig. 1(b)] shows a linear decrease as a function of frequency in which the slope of the line is directly related to the position of the roving hydrophone. Theoretically, if the hydrophones are located at the same position and distance from the source speaker, the transfer function measurement should have a flat response. As shown in the top curve of Fig. 1(a), there is variation in the transfer function measurement although it is generally within +/−5 dB throughout the spectrum. The deviation in the measurement from a flat response is attributed to differences in the sensitivities between the two hydrophones, the speaker having some directivity that changes with frequency, the fact that it is impossible to locate the hydrophones in the same geometrical position (25 m transverse separation, for the 1 m test), and the acoustical reverberation caused by the complex shallow water environment.

The results (calculated using the rms pressure received at each hydrophone for the duration of the chirp signal) from the various locations are displayed in Fig. 2 for the perpendicular ($\perp$) and parallel trials ($\parallel$). The test directions at Cedar Key were not based upon the channels; however trials 1 and 2 (C.K. B) were performed approximately perpendicular to each other. The four acoustic spreading models are also plotted for comparison.

The mixed spreading model best represents the data collected from the Crystal River locations, with a correlation coefficient (r^2) value of 0.9896. The correlation coefficient values from the four acoustic spreading models are displayed in Table I. At the Cedar Key locations the mean data is represented closely by either spherical spreading ($r^2=0.9441$) or the Lurton model ($r^2=0.9508$). The test located at Cedar Key B 1 has a transmission loss that is higher than predicted for spherical spreading. This can be attributed to the fact that the ocean bottom in that location and direction was covered with grass that may have attenuated the sound more rapidly than compared with a sandy bottom. For the mean of both locations, the Lurton model ($r^2=0.9308$) and mixed spreading model ($r^2=0.9307$) both closely represent the general features of the empirical data.

B. Boat pass experiments

The second method employed to measure the acoustic properties of the channel used a hydrophone in the channel and a boat passing over the hydrophone's position. The locations of this experiment are identical to the previous experiment. Additionally, the boat passes were performed at two locations in the Indian River, near Titusville, FL. Location "I.R. A" was 28°33.533′N, 80°46.031′W and location "I.R. B" was 28°35.802′N, 80°46.163′W. The depths at the Indian River locations were 7.5 and 5 ft respectively. The position of the experiments and the velocity of the boat were measured using a Garmin MAP-76 GPS receiver. The recordings were made using High Tech HTI-96-MIN series hydrophones with a TEAC RD-135T DAT data recorder

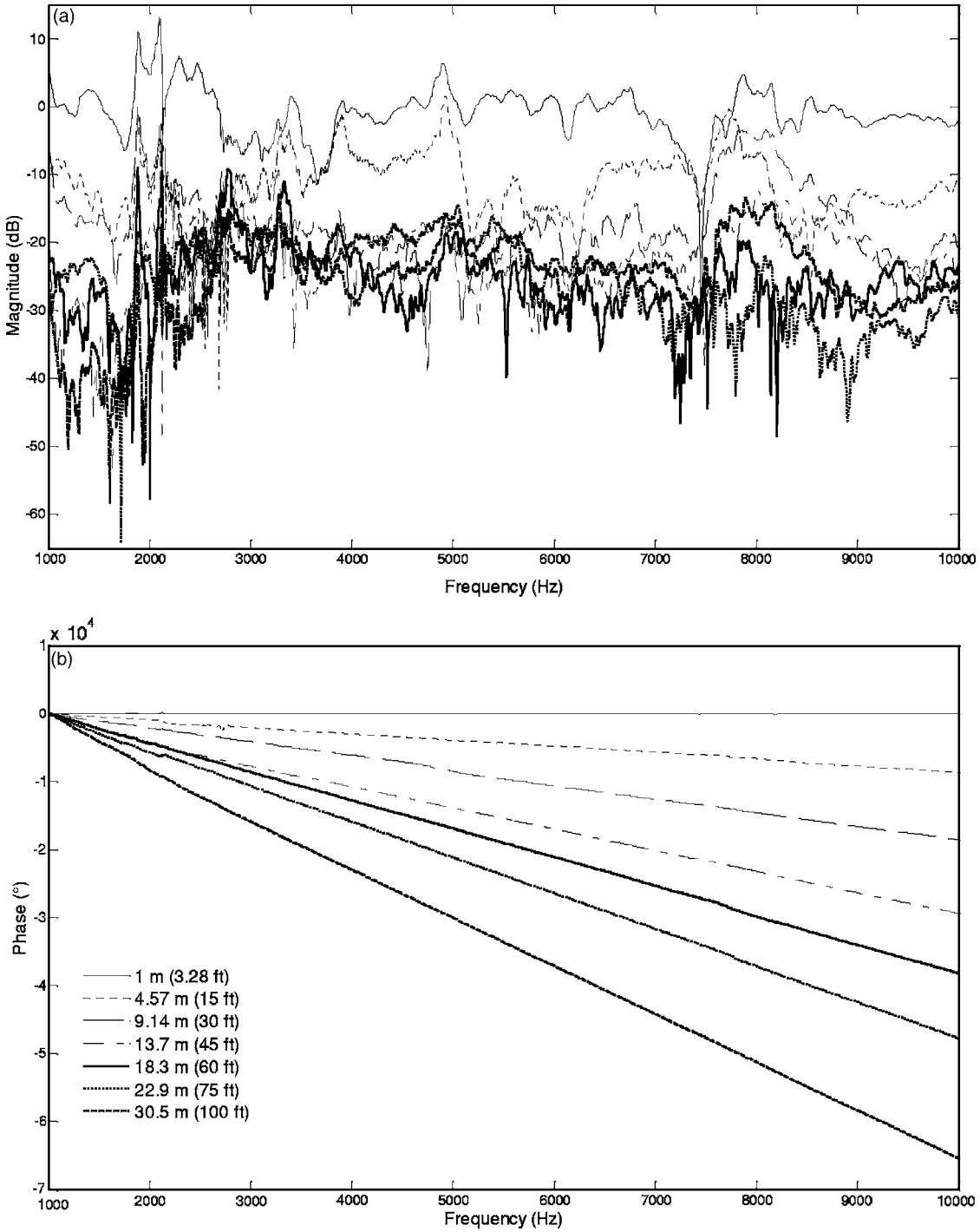


FIG. 1. Transfer function measurement between the roving hydrophone pressure signal and the fixed hydrophone pressure signal for a typical chirp test.

(sample rate of 48 kHz) and a Spectral Dynamics SigLab 2042 dynamic signal analyzer (sample rate of 51.2 kHz).

A 9.9 hp Jon boat was used during the boat passes at Crystal River and a 20 hp Jon boat at Cedar Key. The boat was run directly over the hydrophone location (~1 m underwater) at full throttle, approximately 13 mph, in both parallel and perpendicular to the direction of the channel at Crystal River. The actual speed for each run was measured using a Global Positioning System (GPS) receiver. At Cedar Key the boat was driven past the hydrophone three times, each having an orientation approximately 120° apart. The data was collected using the spectral mapping feature of SigLab. One frame of data was acquired for approximately every three feet traveled by the boat. For the Indian River locations, a 17 foot 70 hp boat was used and the average speed was approximately 32 mph. During the measurement of the data for all tests, the boats were operating at constant speed and throttle position. Therefore it is assumed that the sound emitted from the boat had the same magnitude from one frame to the next. A typical sample spectral map of the boat passing noise is shown in Fig. 3.

The transmission loss data was analyzed separately in two distinct regions (approaching and departing). The transmission loss for the boat departing from the hydrophone in-

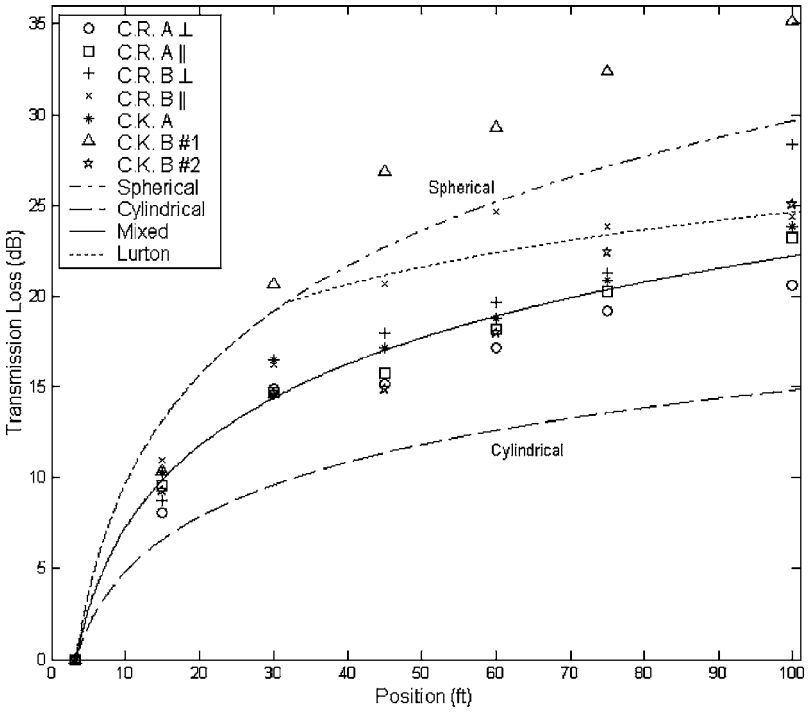


FIG. 2. Measured transmission loss (calculated using the rms pressure received at two hydrophones during the broadcast of 0.25 s chirp signal from 1–10 kHz) at Cedar Key (C.K.) and Crystal River (C.R.).

creased more rapidly compared to when the boat was approaching the hydrophone. The average data from the boat approaching and departing from the hydrophone is presented in Fig. 4, as well as the overall average transmission loss for the boat passes. To increase clarity of the plot, only every tenth data point is displayed.

For the boat passes, the average transmission loss most closely aligned with the spherical spreading model for both Crystal River (r^2=0.8652) and Cedar Key (r^2=0.8731). The correlation coefficient values for the models are given in Table II. At the Indian River locations the average transmission loss was also modeled closest by spherical spreading (r^2=0.9522). The overall mean for all locations is also beat represented by spherical spreading with an r^2 value of 0.9394.

The results shown in Fig. 4 indicate that for some situations, the sound spreading from the source has a transmission loss higher than would be expected for spherical spreading. For the departing test runs, an increase in the transmission loss can be attributed to the fact that the boat exhaust is located behind the boat and is underwater. The bubbles generated from the exhaust can scatter the sound from the propeller behind the boat. Therefore sound propagating behind the boat would be expected to have a transmission loss that is higher than that predicted for pure spherical spreading. Likewise for sound propagating in front of the boat, the mixed spreading model is the most appropriate. The transmission loss test results calculated using a single microphone and two microphones (using a chirp signal) are in agreement. For an approaching boat, a mixed spreading model should be used.

TABLE I. Correlation coefficient values for chirp broadcasts.

	Correlation coefficient (r^2)		
Spreading model	Crystal River	Cedar Key	Overall
Spherical	0.8229	0.9441	0.8827
Cylindrical	0.5825	0.5184	0.5431
Mixed	0.9896	0.8451	0.9307
Lurton	0.8941	0.9508	0.9308

IV. ENVIRONMENTAL FACTORS

In addition to the acoustic spreading properties in Florida's waterways, several other factors will affect the detection range of an acoustically based manatee warning system. Two such factors are the magnitude of boat noise and background noise and are now addressed.

A. Boat noise

A prominent sound source in Florida's coastal waterways is boating traffic. In order to obtain an **estimate** of the sound levels generated by a large number and variety of boats that typically travel in the waterways of interest, an experiment was conducted to survey the noise emitted from many boats in Florida. It should be noted that no attempt was made to identify the sound levels for one particular type of boat because there is considerable variability in boat engine size, propeller size, and operating speed. On July 7, 2003 and September 9, 2003 hydrophones were placed near the center of the channel in the mouth of Crystal River (28°55.537′N,82°41.837′W). On these dates, 83 boats were recorded while passing over the hydrophone. Essentially all the boats were traveling at speeds greater than 10 mph (16 kph). The channel was approximately 2 m deep and the hydrophones were positioned about 0.5 m from the sea floor. The observed engine sizes ranged from 25 to a 225 hp outboard motor. A commercial diesel fishing boat

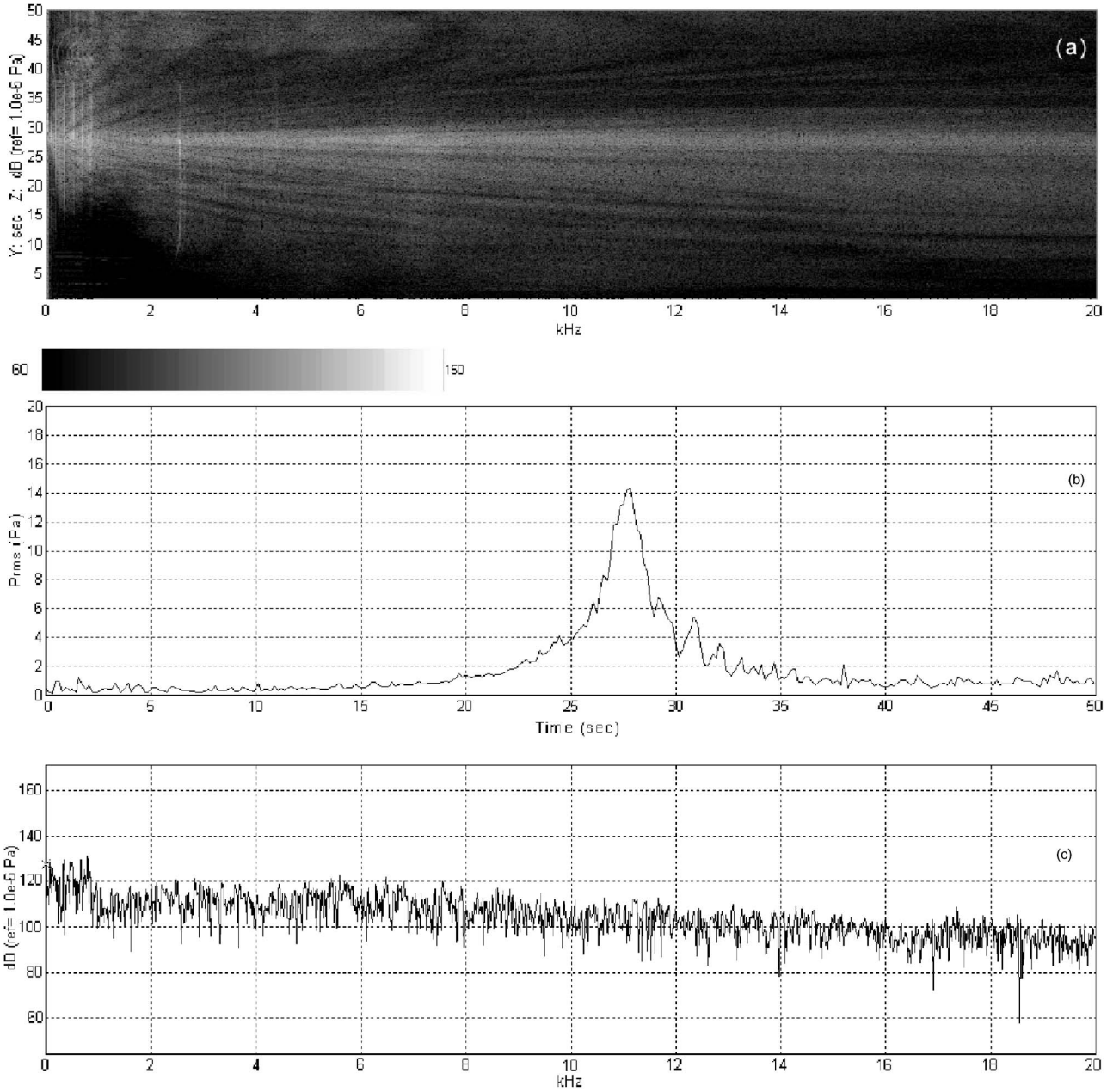


FIG. 3. Sample spectral map for a typical boat pass.

was also observed. The peak sound pressure level from each boat passes was then calculated. For the manatee detection problem, only sound above 2 kHz is significant, because the majority of energy in a manatee vocalization is contained in this range. The average sound pressure level of the passing boats was observed to be 140 dB (re1μ Pa; >2 kHz) with a standard deviation of 6.3 dB. The sound pressure level of the 83 passing boats ranged between 122 and 153 dB (see Fig. 5). When the frequencies below 2 kHz are also considered, the sound pressure levels of the boat traffic ranged from 129 to 169 dB at $\sim$1 m. Including the low frequencies increased the average sound pressure level to 146 dB with a standard deviation of 5.9 dB (see Table III).

Although the measured sound pressure levels of the boats ranged up to over 150 dB, a more appropriate metric for detection is the maximum noise floor of the boat noise in the frequency range of interest. Most of the energy in a manatee vocalization is typically located above 2 kHz. The frequency spectrum at the peak sound pressure level of the typical boat pass is displayed in the bottom graph shown in Fig. 3. This example has an overall sound pressure level (SPL) of 140 dB above 2 kHz, while the noise floor (>2 kHz) is approximately 120 dB. At frequencies above 2 kHz, 90% of the measured boat traffic has a maximum noise floor less than 120 dB. Therefore 120 dB is chosen as an evaluation metric in this study. It should be noted that this value is the best estimate based on the measured data.

B. Background noise

The overall ambient background noise was recorded at all of the test locations. Additional measurements were made

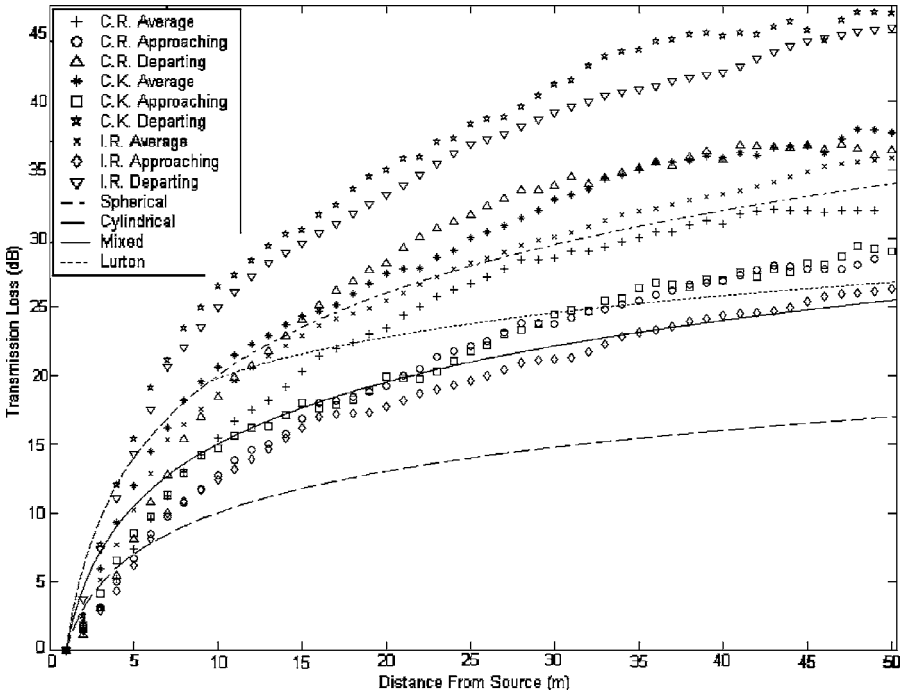


FIG. 4. Transmission loss of boat passing tests at Cedar Key (C.K.), Crystal River (C.R.), and Indian River (I.R.), calculated using the rms pressure received at a single hydrophone.

in Crystal River at the following locations: 28°55.537′N, 82°41.837′W; 28°55.103′N, 82°40.208′W; 28°54.371′N, 82°38.254′W; 28°53.627′N, 82°32.266′W; and 28°53.461′N, 82°35.913′W. The background noise in Crystal River decreased from 104 dB at the entrance to the Gulf of Mexico to 69 dB at locations farther from the mouth of the river. The average background noise near Cedar Key was measured to be 83 dB and in the Indian River, background noise was recorded between 102 and 105 dB.

V. THEORETICAL DEVELOPMENT—DETECTION RANGE

Manatee detection systems based upon acoustically detecting vocalizations are essentially passive sonar systems. The relevant equations are applicable to this problem and are detailed in the work by Caruthers (1977) and Ross (1976). These equations are now reviewed. The detection threshold (DT) of a passive sonar system is the minimum signal strength above the ambient noise level (NL) required to detect the signal. The sound pressure level (SPL) received by the hydrophone is the source level (SL), measured at 1 m, minus any transmission loss TL that may occur. The signal excess (SE) is the amount by which the detection threshold is exceeded [Eq. (8)], where DI is the directivity index of the system and AG is the array gain when multiple hydrophones are used. If the signal excess is greater than or equal to zero, the signal can be detected by the passive sonar system, and as the signal excess increases the probability of detection increases:

$$\mathrm{SE} = \mathrm{SL} - \mathrm{TL} - \mathrm{NL} + \mathrm{DI} + \mathrm{AG} - \mathrm{DT}. \tag{8}$$

When there is no signal excess (SE), and the SL and NL are known, Eq. (8) can be solved for the maximum allowable transmission loss [Eq. (9)]. The maximum allowable transmission loss is traditionally referred to as the figure of merit (FOM).

$$\mathrm{FOM} = \mathrm{SL} - \mathrm{NL} + \mathrm{DI} + \mathrm{AG} - \mathrm{DT}. \tag{9}$$

For the purpose of this study a single hydrophone (omnidirectional) system is considered and the manatee vocalizations are assumed nondirectional. These assumptions reduce the array gain and directivity index of equation 9 to zero [Eq. (10)].

$$\mathrm{FOM} = \mathrm{SL} - \mathrm{NL} - \mathrm{DT}. \tag{10}$$

In addition to ambient background noise, this study also considers the effects of boat noise [Eq. (11)]. To compensate for noise from boats the ambient noise level is considered to be the maximum of the ambient noise level (BL) and the received boat noise (RN).

$$\mathrm{ROM} = \mathrm{SL} - \max(\mathrm{BLRN}) - \mathrm{DT}. \tag{11}$$

The received boat noise is the source level of the boat (SL_B) with any transmission losses (TI_B) subtracted. The maximum allowable transmission loss for a manatee vocalization is shown in Eq. (12), where SL_M is the source level of the manatee vocalization.

$$\mathrm{FOM} = \mathrm{SL}_M - \max(\mathrm{BL}, \mathrm{SL}_B - \mathrm{TL}_B) - \mathrm{DT}. \tag{12}$$

TABLE II. Correlation coefficient values for boat passes.

Spreading model	Correlation coefficient (r^2)				
	Crystal River	Cedar Key	Indian River	Overall	Approaching
Spherical	0.8652	0.8731	0.9522	0.9394	0.6941
Cylindrical	0.4725	0.4802	0.4822	0.4789	0.5010
Mixed	0.6109	0.5383	0.5723	0.5684	0.8944
Lurton	0.5797	0.5220	0.5634	0.5549	0.6953

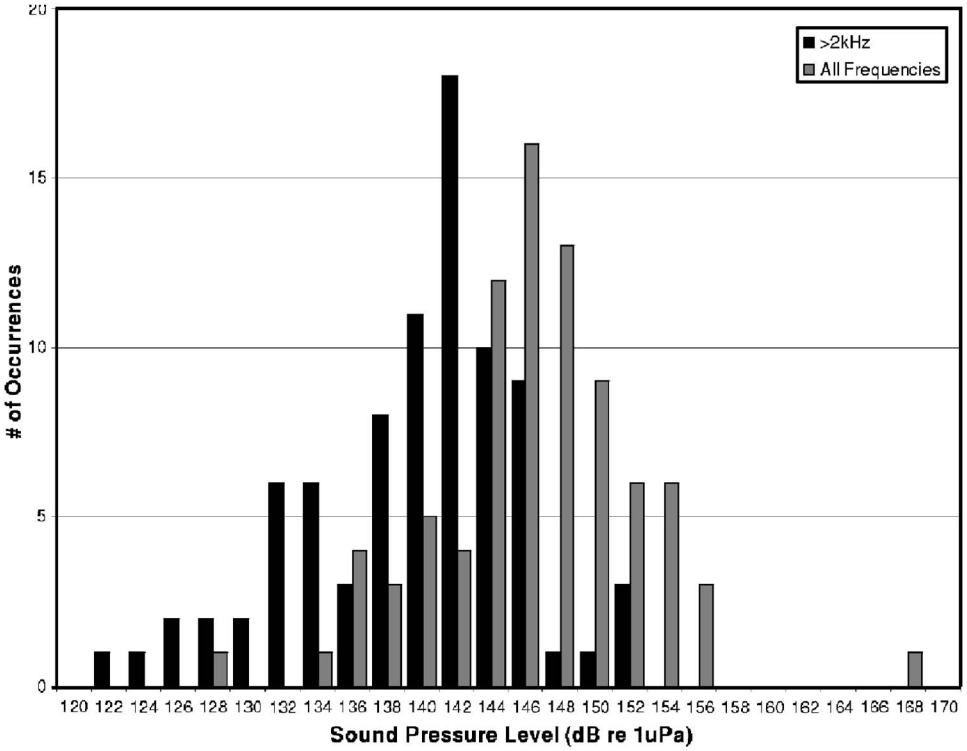


FIG. 5. Sound pressure level histogram of passing boats.

The transmission losses due to acoustic spreading of the manatee vocalizations and boat noise are best approximated by a mixed spreading model [see Sec. III and Eq. (3)].

$$\mathrm{TL} = 15 \cdot \log(\mathrm{R}) \quad R \geq 1. \tag{13}$$

To obtain the sonar equation in terms of the manatee detection distance (R_M) and the boat distance (R_B), Eq. (13) is substituted into Eq. (12) for the transmission loss of the boat noise and the figure of merit.

$$15 \cdot \log(R_M) = \mathrm{SL}_M - \max[\mathrm{BL}, \mathrm{SL}_B - 15 \cdot \log(R_B)] - \mathrm{DT}. \tag{14}$$

Rearranging Eq. (14), the sonar equation yields the maximum distance a manatee is detectable and is given by

$$R_M = 10\text{^}\left(\frac{\mathrm{SL}_M - \max[\mathrm{BL}, \mathrm{SL}_B - 15 \cdot \log(R_B)] - \mathrm{DT}}{15}\right). \tag{15}$$

The maximum distance in which a manatee is detectable (R_M) is dependent on the source level of the manatee vocalization (SL_M), the ambient background noise level (BL), the source level of the boat (SL_B), the distance of the boat from the hydrophone (R_B), the type of acoustic spreading model chosen (mixed spreading), and the system detection threshold (DT).

VI. RESULTS—DETECTION RANGE

The average source level of a manatee vocalization has been previously estimated to be 112 dB at 1 m, ref 1 μPa by using a hydrophone array and source localization. This average value was calculated with the assumption that the acoustic wave propagated in a cylindrical manner (Phillips *et al.*, 2004). If the source levels are recalculated assuming a mixed spreading model for underwater sound transmission instead of a cylindrical spreading model, the average source level of a manatee vocalization is estimated to be 118 dB at 1 m, ref 1 μPa. Likewise, it has been previously shown that the manatee source level is variable (Nowacek *et al.* (2003); Phillips *et al.*, 2004). To account for variations in the source levels, a range of detection distances are calculated based on source levels varying from 109 to 118 dB. It is important to note that a Lombard vocal response exists for many mammals but no prior research studies have been reported that indicate the presence or absence of such a response in manatees. If a Lombard vocal response is present in manatee vocalizations, then the assumed manatee source levels would be low and the results presented conservative.

Ambient background noise has also been measured at several locations likely to be inhabited by manatees. Although the dominant background noise is caused by boats and snapping shrimp, other noise may also be generated by wind, rain, water movement, and wildlife. The ambient background noise measurements typically vary between 70 and 105 dB. It is important to note that in some situations (during storms or high wind) the ambient background noise may be higher than this range.

Knowing the source level of the manatee vocalization and the ambient background noise levels, the maximum de-

TABLE III. Estimated sound pressure level (SPL, re 1 μPa) data for boating traffic.

	Broadband	>2 kHz
Minimum SPL	128.5	122.0
Maximum SPL	168.5	152.9
Mean SPL	146.3	139.6
Std. deviation	5.9	6.3

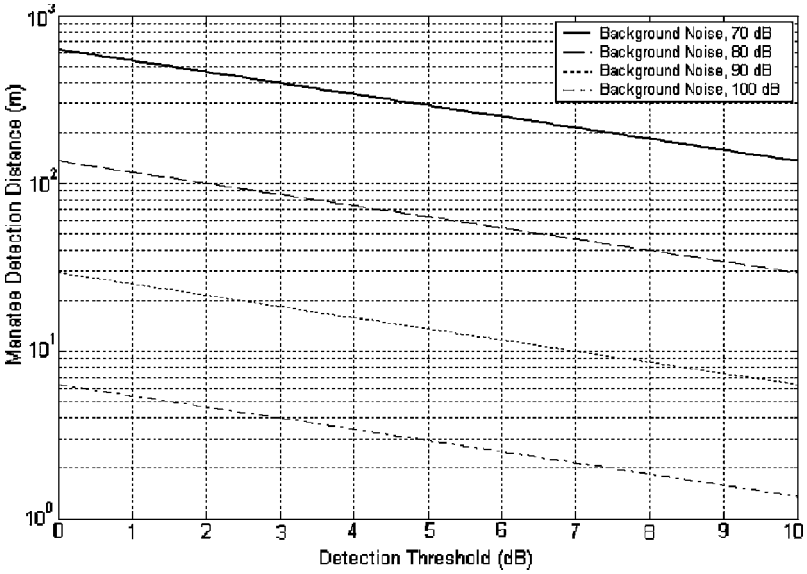


FIG. 6. Maximum manatee detection ranges at various background levels assuming a manatee source level of 112 dB at 1 m, ref 1 μ Pa (with no boat noise present.)

tection range can be estimated with respect to the detection threshold by using Eq. (15). The results shown in Fig. 6 do not consider boat noise. The maximum detection range at selected detection thresholds is also shown in Table IV for a variety of manatee source levels and background noise levels. The best scenario would be a detection system with a low detection threshold placed in an environment with minimal ambient noise. A detection system with a 3 dB detection threshold surrounded by 70 dB of ambient noise would allow for theoretical detection ranges up to almost 400 m (SL_M =112 dB). However, in habitats with an abundance of snapping shrimp and other background noise the maximum detection range will be significantly smaller. If a detection algorithm requires a 9 dB detection threshold and is in an environment with 100 dB of ambient noise, the maximum theoretical detection range is approximately 1.5 m. Additionally, as the manatee source level varies (+/−3 dB) the detection range changes significantly, especially if the background noise levels are low.

TABLE IV. Manatee detection ranges (m) for various ambient background levels, manatee source levels (SL_M), and detection thresholds.

Detection range (m)		Detection threshold (dB)		
SL_M (dB)	Background level (dB)	3	6	9
109	70	251.2	158.5	100.0
109	80	54.12	34.15	21.54
109	90	11.66	7.356	4.642
109	100	2.512	1.585	1.000
112	70	398.1	251.2	158.5
112	80	85.88	54.12	34.15
112	90	18.48	11.66	7.356
112	100	3.981	2.512	1.585
115	70	631.0	398.1	251.2
115	80	135.9	85.77	54.12
115	90	29.29	18.48	11.66
115	100	6.310	3.981	2.512
118	70	1000	631.0	398.1
118	80	215.4	135.9	85.77
118	90	46.42	29.29	18.48
118	100	10.00	6.310	3.981

Using a manatee vocalization source level of 112 dB, estimated mean ambient background noise levels, and including a boat generating a 120 dB noise floor (above 2 kHz) the maximum theoretical detection range can be calculated. Cases with ambient background noise levels at 83 and 90 dB, and boat distances from 0 to 300 m are presented in Figs. 7 and 8, respectively. Each of the three graphs in Figs. 7 and 8, contain the same information presented in a slightly different manner. For example if the reader peruses the lower right hand graph of Fig. 7, for a detection threshold of 3 dB, it is evident that the detection range is ~55 m when the boat is located 300 m away from the hydrophone. If the boat moves closer and is located 100 m from the hydrophone, the manatee detection range drops to ~18.5 m. If the boat moves even closer and is located less than 5 m from the hydrophone, the boat noise levels have exceeded the detection threshold (3 dB) required to detect the manatee.

As the ambient background noise levels increase from 70 dB, the results presented in Fig. 7 will not change until ~83 dB for a boat operating within 300 m of the hydrophone and a manatee source level of 112 dB. With ambient background noise levels less than 83 dB, the limiting factor in the manatee detection distance is the boat noise, for a boat cruising within 300 m of the manatee. If the ambient background noise levels are increased to 90 dB (see Fig. 8), the detection range is governed by the background noise until the boat is located closer than ~100 m from the hydrophone. Within the range <100 m, the boat noise limits the detection range, while if the boat is further than 100 m, the detection range is limited by the background noise. If the ambient background levels are increased to 100 dB, the background noise limits detection range until the boat is within ~25 m. Beyond this range the manatee detection distance is limited by the boat noise. For a manatee source level of 112 dB, in

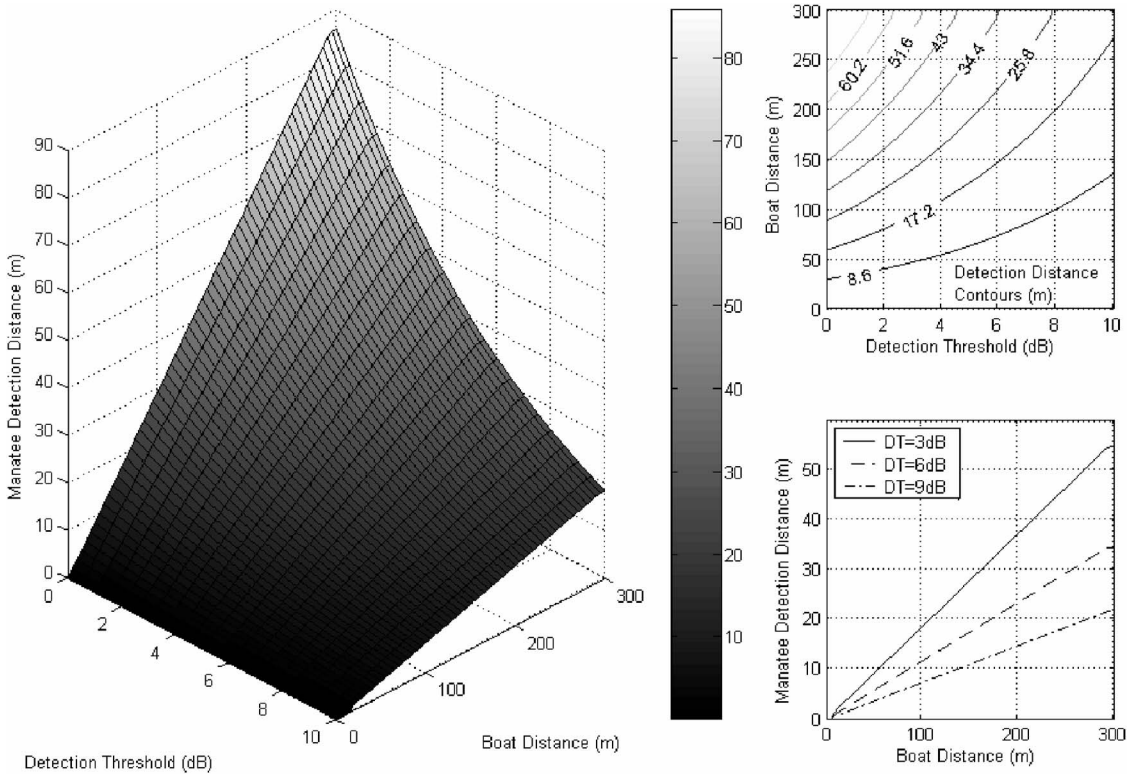


FIG. 7. Maximum manatee detection ranges at BL=83 dB, SL_M=112 dB, and SL_B=120 dB.

all of the cases the hydrophone is saturated by engine and propeller noise when the boat is closer than 5 to 10 m, depending on the detection threshold.

If the manatee source level is increased from 112 to 118 dB for an ambient background noise of 90 dB, the detection range graphs will have a similar appearance to the ones shown in Fig. 8, however the values will be higher. For example, the maximum detection range increases from 18.48 to 46.42 m for a 90 dB ambient background noise environment (DT=3 dB) as shown in Table IV. This clearly indicates that the detection range estimation is sensitive to variability in the manatee vocalization source levels.

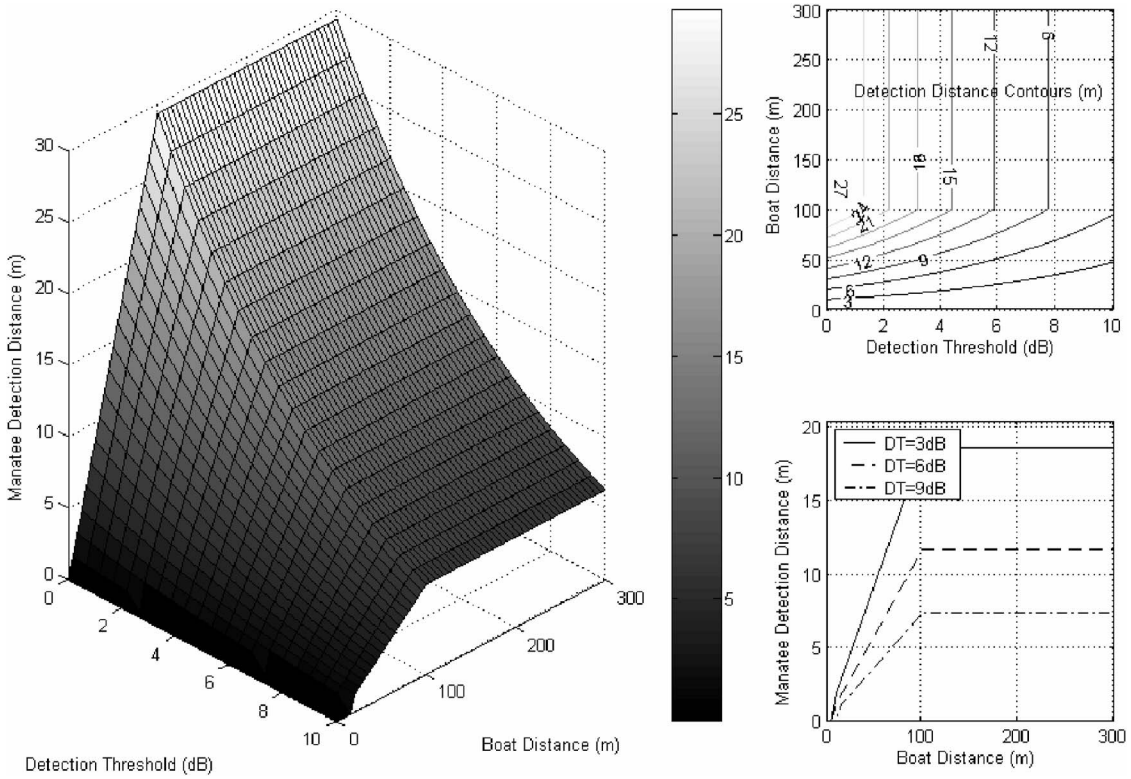


FIG. 8. Maximum manatee detection ranges at BL=90 dB, SL_M=112 dB, and SL_B=120 dB.

VII. DISCUSSION

When operating in low ambient noise environments with sparse marine traffic, manatee detection systems relying on vocalizations could be feasible with hydrophones spaced several hundred meters apart. However, 53% of the watercraft related manatee fatalities occurred in six Florida counties and approximately 240 000 watercraft are registered in these counties (Florida Fish and Wildlife Conservation Commission, 2003a; Florida Fish and Wildlife Conservation Commission, 2003b). The dense concentration of boating traffic combined with potentially high ambient background noise levels could lead to the required hydrophone spacing every tens of meters. Manatee protection and avoidance systems are needed in these habitats; however, these environments are also the least efficient areas for hydrophone spacing with the current technology.

By combining advances in other acoustic and signal processing fields, hydrophone spacing can be increased and acoustic manatee detection systems will become more economically feasible. Multiple independent hydrophone signals can be combined into arrays and noise reduction algorithms can be used to reduce the effects of ambient background and boating noise (Yan *et al.*, 2005). Using techniques such as these, the detection ranges of manatee avoidance systems relying upon vocalizations could possibly be realized, at a reasonable cost. At present, it is the authors' conclusion that an acoustic based manatee avoidance technology is not realizable unless background noise cancellation technology is also implemented in parallel with the detection system.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Fish and Wildlife Conservation Commission, Florida Sea Grant, and the University of Florida College of Veterinary Medicine, Marine Mammal Program in supporting this research.

Beebe, J. H., McDaniel, S. T., and Rubano, L. A. (**1982**). "Shallow water transmission loss prediction using the Biot sediment model," J. Acoust. Soc. Am. **71**(6), 1417–1426.

Biddulph, T. W. (**1993**). "Radiated Noise Report: Florida Marine Research Institute's 17 FT. MAKO 171," Code 3813 Range Analysis Branch, Naval Undersea Warfare Center, AUTEC Detachment West Palm Beach, AUTOVON 483-7469, Commercial (407)832-8566 EXT 7469.

Caruthers, J. W. (**1977**). *Fundamentals of Marine Acoustics (Elsevier Oceanography Series; 18)* (Elsevier Scientific Publishing Co., Amsterdam, NL), pp. 63–66.

Collins, M. D., and Chin-Bing, S. A. (**1990**). "A three-dimensional parabolic equation model that includes the effects of rough boundaries," J. Acoust. Soc. Am. **87**, 1104–1109.

Coates, R. F. W. (**1989**). *Underwater Acoustic Systems* (Halsted Press, a division of John Wiley & Sons, Inc., New York), pp. 18–19.

Florida Fish and Wildlife Conservation Commission (**2003a**). Marine Mammal Pathobiology Laboratory, Manatee Mortality Yearly Summaries, Florida Fish and Wildlife Conservation Commission, 620 South Meridian Street, OESBPS, Tallahassee, FL 32399-1600.

Florida Fish and Wildlife Conservation Commission (**2003b**). 2003 Boating Accident Statistical Report, Florida Fish and Wildlife Conservation Commission, Division of Law Enforcement -Office of Boating and Waterways, 620 South Meridian Street, Tallahassee, Florida 32399-1600.

Gerstein, E. R. (**2002**). "Manatees, Bioacoustics and Boats," Am. Sci. **90**(2), 154–163.

Hamilton, E. L. (**1980**). "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**(5), 1313–1340.

aJensen, F. B. (**1984**). "Numerical models in underwater acoustics," *Hybrid Formulation of Wave Propagation and Scatering*, Martinus Nijhoff, Dordrecht, Netherlands, pp. 295–335.

Lurton, X. (**2002**). *An Introduction to Underwater Acoustics Principles and Applications* (Praxis Publishing Ltd, Chichester, UK), pp. 31–32.

Miksis-Olds, J. L., Miller, J. H., and Tyack, P. L. (**2004**). "The acoustic environment of the Florida manatee: correlation with level of habitat use," J. Acoust. Soc. Am. **115**(5), 2558.

Miksis-Olds, J. L., and Miller, J. H. (**2006**). "Florida manatee habitat usage and transmission loss: Quieter is better," J. Acoust. Soc. Am., submitted.

Niezrecki, C., Phillips, R., Meyer, M., and Beusse, D. O. (**2003**). "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**(3), 1640–1647.

Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A. (**2003**). "Intraspecific and geographic variation of West Indian manatee," J. Acoust. Soc. Am. **114**(1), 66–69.

Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**(1), 422–428.

Ross, D. (**1976**). *Mechanics of Underwater Noise* (Pergamon Press Inc., Elmsford, New York), pp. 9–11.

Smith, K. B. (**2001**). "Convergence, stability, and variability of shallow water acoustic predictions using a split-step Fourier parabolic equation model," J. Comput. Acoust. **9**, 243–285.

United States Coast Guard. (**2002**). Boating Statistics – 2001, Commandant Publication P16754.15, 2100 Second Street SW, Washington, DC 20593–0001.

Yan, Z., Niezrecki, C., and Beusse, D. O. (**2005**). "Background Noise Cancellation for Improved Acoustic Detection of Manatee Vocalizations," J. Acoust. Soc. Am. **117**(6), 3566–3573.

A scanning laser Doppler vibrometer acoustic array

Benjamin A. Cray,[a] Stephen E. Forsythe, Andrew J. Hull, and Lee E. Estes
Naval Undersea Warfare Center Division, 1176 Howell Street, Newport, Rhode Island 02841

(Received 6 September 2005; revised 3 April 2006; accepted 2 May 2006)

Experiments confirm that a laser Doppler vibrometer can be used to detect acoustic particle velocity on a fluid-loaded acoustically compliant, optically reflective surface. In these experiments, which were completed at the Acoustic Test Facility of the Naval Undersea Warfare Center, Scotchgard™ reflective tape was affixed to the interior surface of a standard acoustic window. The polyurethane array window had a thickness of 0.9525 cm (0.375 in.) and a material density of 1000 kg/m^3. The surface velocity measured, using a commercial scanning laser vibrometer system (SLVS), was beamformed conventionally and flawlessly detected and localized acoustic signals. However, the laser Doppler vibrometer used in the experiments had relatively poor acoustic sensitivity, presumably due to high electronic noise in the photodetector, speckle noise, standoff distance, and drifting laser focus. An improved laser Doppler vibrometer, the simplified Michelson interferometer laser vibrometer sensor (SMIV), is described in brief. The SMIV achieves sensitivity of 61 $dB/\mu Pa$ in a 1-Hz band at 11.2 kHz, which represents a 39-dB acoustic sensitivity improvement.
 [DOI: 10.1121/1.2207569]

PACS number(s): 43.30.Wi, 43.60.Fg, 43.20.Ye [LPF] Pages: 164–170

I. INTRODUCTION

In principle, a passive high-bandwidth sonar system can be realized using a scanning laser vibrometer—an optical instrument that directly measures the surface velocity of a structure. A commercial scanning laser vibrometer system (SLVS), modified for underwater scanning by the Acoustic Test Facility (ATF) at the Naval Undersea Warfare Center (NUWC) Division, Newport, RI,[1] has been used in a variety of applications to determine the vibrational response of fluid-loaded structures.[2] The SLVS, a type of laser Doppler vibrometer system, can sample a grid of 512×512 points, with each grid point having a spot size of 10 μm (0.0004 in.). This sampling could be used to create an essentially continuous acoustic aperture; the upper cutoff frequency for such a finely sampled array would be greater than 10 MHz. Acoustic grating lobes would be eliminated at all frequencies of practical interest. Hence, rather than measuring acoustic pressure, as with a conventional array of hydrophones, the array window's surface velocity would be measured. Theoretical and experimental research in developing a laser-based hydrophone and sonar system has been documented by Antonelli.[3–6] For example, Antonelli devised a means for remote, aerial detection of underwater sound. In this application, a narrow laser beam is directed onto the water surface to measure the velocity of the surface vibrations that occur as the underwater acoustic signal reaches the water surface. This system enables the sensing of underwater sound from a remote and potentially safer position in the air without requiring underwater hydrophone equipment.

Rather than a single-point measurement on a variable surface, the system described here allows for conventional (weight, delay, and sum) beamforming of acoustic particle velocity over a uniform surface—as would be done with any array of hydrophones. The focus here is on measurements that examine fundamental sonar performance parameters, such as the noise floor limit of the SLVS (or the minimum detectable signal level), amplitude and phase tolerance (the ability to accurately measure acoustic particle velocity over a broad range of frequencies and incidence angles), and point-directivity measurements, that is, the angular beam pattern of individual sample *points* (not elements) on the surface of the acoustic window. These points replace the hydrophone sensors of a conventional sonar.

Section II presents a review of the completed investigative measurements, including a description of the commercial laser Doppler vibrometer system. The experimental results are presented in Sec. III, along with a subsection on a prototype of a simplified Michelson interferometer vibrometer (SMIV). Theoretical characteristics of the SMIV system are described, and it is shown that the design maximizes acoustic sensitivity.

II. SUMMARY OF INVESTIGATIVE MEASUREMENTS

In its simplest form, Euler's equation for a propagating harmonic plane wave reduces to

$$p(x,t) = \rho c v(x,t), \tag{1}$$

where $p(x,t)$ is acoustic pressure and $\rho c v(x,t)$ is the product of the medium's characteristic impedance (ρc) and acoustic particle velocity $v(x,t)$. Thus, measuring acoustic pressure with an array of conventional hydrophones is equivalent to measuring acoustic particle velocity with an array of velocity sensors. This equivalence has been the basis for sonobuoy designs for decades and, more recently, for innovative submarine sonar systems.

Consider a thin plate (membrane) of thickness h, insonified by an acoustic plane wave of amplitude (P_i) as seen in Fig. 1, with reflected and transmitted amplitudes, P_r and P_t,

[a]Author to whom correspondence should be addressed. Electronic mail: crayba@npt.nuwc.navy.mil

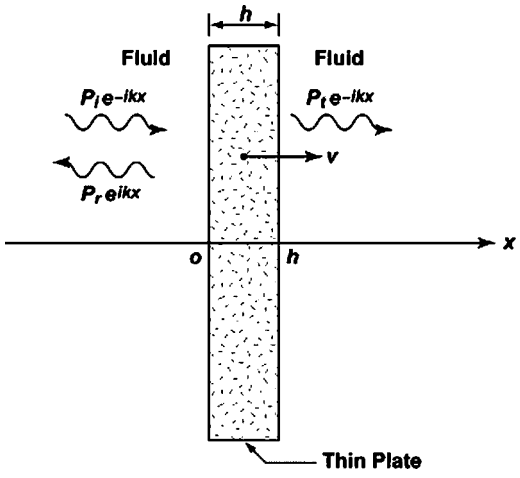


FIG. 1. Response of a thin plate to incident pressure.

respectively. The plate is thin relative to the incident acoustic wavelength, that is, $kh \ll 1$, where k is the acoustic wave number, and the harmonic time dependency is ignored.

Given this, the plate's velocity can be assumed constant throughout the plate, $v(0,t)=v(h,t)$. Applying Euler's equation to both surfaces of the plate gives

$$\frac{p_i - p_r}{\rho c} = v = \frac{p_t}{\rho c}. \tag{2}$$

The amplitude of the plate's velocity is $V=|v|$. A force balance between each plate boundary yields

$$(P_i + P_r)A = F(0), \tag{3}$$

$$P_t A = F(h), \tag{4}$$

and

$$F(0) - F(h) = i\omega\rho_1 hAV = (P_i + P_r)A - P_t A, \tag{5}$$

where A is a unit surface area and ρ_1 is the plate's density. Combining Eqs. (2) and (5) yields the incident pressure to window velocity transfer function,

$$\frac{P_i}{V} = \rho c + i\omega\rho_1(h/2), \tag{6}$$

where ω is the angular frequency of harmonic excitation. Note that as $h \to 0$, Eq. (6) reduces to Euler's equation, $P_i = \rho c V$, and the plate's velocity is directly proportional to the incident acoustic pressure. This is, of course, a simplification.

Two test configurations were used to examine the laser-based array concept. The laser-array configuration used in the first set of measurements, conducted in May 2002 at NUWC Division Newport's ATF, consisted of a water-backed acoustic window and a laser beam that propagated through water. In the second and preferred (improved signal gain) configuration, the laser system was completely sealed within the forward bulkhead of a cylinder. Thus, the array acoustic window was air-backed and the laser beam propagated through air.

The initial water-backed measurements completed in 2002 were limited. For example, at oblique angles, the incident sound field insonified both the acoustic window and the optical lens of the underwater housing containing the laser. Furthermore, water particles in the path of the laser were insonified. This distorted the time-series measurements made on the surface of the acoustic window. (No distortion was seen for broadside, or normal, incidence.) Placing the laser Doppler vibrometer within a large air-filled cylinder eliminated this source of interference, thus permitting insonification of the acoustic window at all angles of incidence. Wave-vector frequency analysis could then be used to examine the structural-acoustic response of the window and to compare that response to theoretical predictions.[7]

A laser Doppler system utilizes interference[8] of two coherent light beams to measure either surface displacement (by counting interference fringes) or surface velocity (by detecting the Doppler shift due to the motion of the surface). It is a heterodyne system, that is, an additional and known oscillator signal, at frequency f_{rf}, is added to the signal received by the photodetector. In this manner, the polarity or velocity direction can be determined. The system has a photodetector that measures the time-dependent intensity of the sum of the reference and measurement (or signal) beams.

Deflection mirrors automatically steer the helium-neon (He-Ne) laser beam (at a wavelength of 633 nm) within a $40° \times 40°$ (horizontal by vertical) field of view on to the vibrating surface. A simple geometry calculation determines the required standoff distance (d) for a field of view (θ) and a given aperture size L; $d=L/[2\tan(\theta/2)]$. Thus, a 40° scan would require a standoff distance of $d=1.4L$. The scan resolution of the commercial SLVS is stated to be very precise at 0.01° (the corresponding point-to-point positional resolution would be determined from the surface-to-photodetector standoff distance). The normal component of velocity is always measured. For oblique angles of laser beam incidence, the system automatically compensates via a cosine correction. A video camera monitors the scan surface and the scanning beam.

III. EXPERIMENTAL RESULTS

A. Minimum detectable acoustic signals

The SLVS vendor claims a velocity resolution down to 0.25 μm/s semipeak in a 1-Hz bandwidth, independent of frequency (1 Hz to 1.5 MHz). However, for fast Fourier transform (FFT) temporal processing, the maximum sampling rate of the data acquisition system is 400 kHz (with a 2048-point FFT), which limits the upper frequency for two-channel measurements to approximately 200 kHz. (Users may bypass the data processing hardware and access the voltage time-series data directly for processing independently with other data analysis hardware.)

Measurements were made to determine the minimum signal detection capabilities of the SLVS (to estimate the lowest level acoustic signal that could be detected by the laser-window array configuration). The measurements focused on two possible limiting sources of noise: noise generated by the laser's photodetector, or shot noise,[9,10] and environmental noise due to various acoustic and vibration

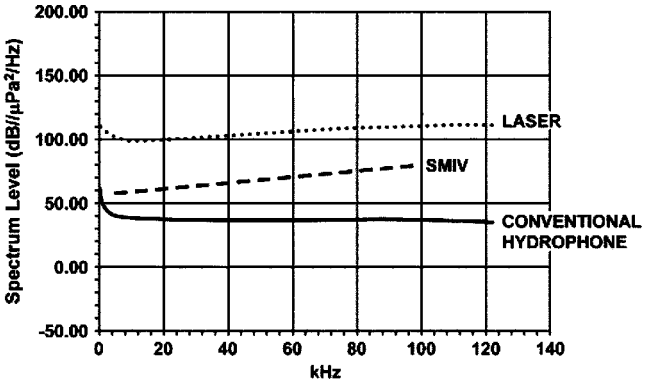


FIG. 2. Noise floor comparison of monitor hydrophone versus the SLVS. (The dashed line represents the predicted performance of the simplified Michelson interferometer vibrometer described in Sec. III B when turbulence and laser phase noise are eliminated.)

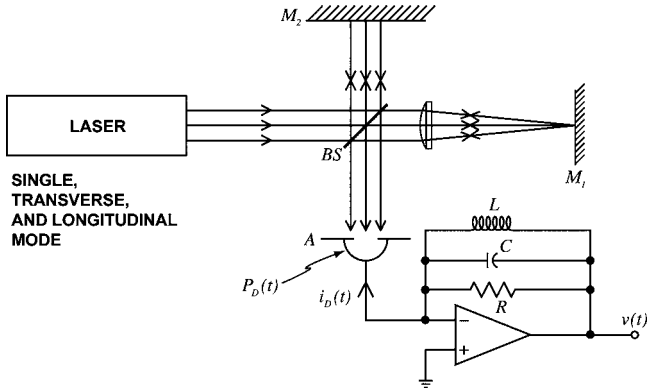


FIG. 3. The simplified Michelson interferometer laser vibrometer. Light reflected by the reference mirror M_2 is combined by the beamsplitter BS with light reflected from the signal mirror M_1 to produce interfering beams with optical power $P_D(t)$ that are detected by the sensor with aperture A to produce the photocurrent $i_D(t)$ and the amplifier output voltage $v(t)$.

sources impinging on the surface of the window. One would want the limiting noise to be environmental acoustic noise.

Figure 2 compares the measured noise floor of the SLVS to that of a standard H-52 monitor hydrophone. The laser noise (dotted line) is clearly much greater than that of the monitor hydrophone, and increases with increasing frequency to an equivalent acoustic plane-wave level of approximately 120 dB//1 μPa^2/Hz. For reference, from Euler's equation, the sound-pressure level (SPL) equivalent to an acoustic particle velocity of 0.25 μm/s is 111.7 dB//1 μPa in a 1-Hz band. The monitor hydrophone (solid line) measured the ambient noise within the acoustic tank. The electronic, or background, noise within the laser's photodetector is more than 50 dB greater than the ambient noise. Both in-air and in-water measurements were made, all with the laser directly centered on the window, without scanning. Single-shot (not shown) minimum SPLs as well as spectral averaged (16-average) SPLs were examined. These differences resulted in relatively minor changes in laser system noise. Also shown in Fig. 2 is the predicted noise floor (dashed line) for a simplified Michelson interferometer laser vibrometer (SMIV).

B. Simplified Michelson interferometer vibrometer (SMIV)

To focus on fundamentals, a noise model was developed for a simplified interferometer rather than the usual heterodyne or dual-track homodyne configurations. A brief summary of the main results will be presented here; additional details of this work are available in Refs. 11 and 12.

The SMIV analyzed is depicted in Fig. 3. In the analysis the suboptical wavelength sinusoidal position variations of mirror M_1 represent the signal to be measured. Optimum measurement sensitivity is achieved by positioning the reference mirror M_2 so that the reference beam and the mean measurement beam are in phase quadrature. The detector circuit consists of a photodetector and a transimpedance amplifier with a resonant feedback circuit that is tuned to the signal frequency. The noise sources modeled[11] were detection shot noise, laser light amplitude noise, atmospheric turbulence noise, laser light phase noise, mirror and beamsplitter thermal vibration noise (including waterborne thermal noise that would impinge on a membrane in the ocean), and circuit feedback resistor thermal noise. Further, the analysis assumes a single transverse and longitudinal mode He–Ne laser operating at a 633 nm wavelength, with 1 mW power and with typical phase and amplitude noise.

Table I presents the noise predicted by the model when the environment and feedback resistor temperature was taken to be $T=300$ K; the sensor quantum efficiency was 0.6; the measurement and reference arm lengths were 1 and 0.15 m, respectively; and the index of refraction structure constant

TABLE I. Modeled SMIV noise voltage variances and surface movement sensitivity at 10 kHz over a 0.32-Hz bandwidth.

Definition	Value
Atmospheric turbulence noise voltage variance	$\sigma^2_{Vn}=1.99\times10^{-7}$ V^2
Laser light phase noise voltage variance	$\sigma^2_{VLP}=7.19\times10^{-8}$ V^2
Laser light amplitude noise voltage variance	$\sigma^2_{VLN}=1.17\times10^{-9}$ V^2
Shot noise voltage variance	$\sigma^2_{VLN}=9.78\times10^{-11}$ V^2
Amplifier resistor thermal voltage noise variance	$\sigma^2_{VR}=8.28\times10^{-15}$ V^2
Target mirror+beamsplitter+reference mirror thermal vibration noise voltage variance	$\sigma^2_{VT}=6.14\times10^{-16}$ V^2
Total surface displacement (rms)	$a_{S\ \mathrm{rms}}=1.72\times10^{-13}$ m
Total surface velocity (rms)	$V_{S\ \mathrm{rms}}=1.08\times10^{-8}$ m/s

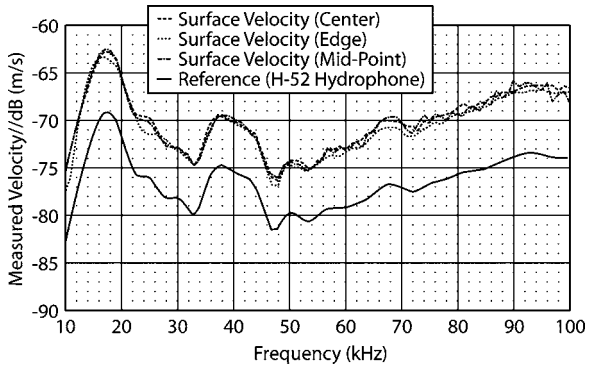


FIG. 4. Surface velocities measured on the air-backed window compared to the equivalent acoustic particle velocity measured by the reference hydrophone (solid curve).

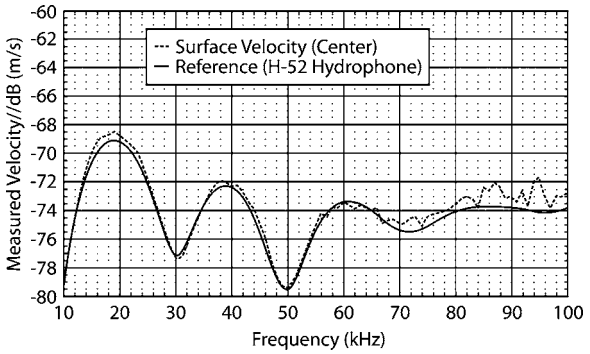


FIG. 5. Surface velocity measured on the water-backed window compared to the equivalent acoustic particle velocity measured by the reference hydrophone (solid curve).

was $C_n^2 = 10^{-17}$ m$^{-2/3}$. The feedback resistance was 10^6 ohms, and the detector resonant frequency and bandwidth were 10 kHz and 0.32 Hz, respectively.

Examining Table I, the dominant noise in this SMIV is due to turbulence, followed by laser phase noise. If the mean lengths of the interferometer arms are set equal and if measures such as using a nearly common propagation path for the two beams are employed, the laser phase noise and atmospheric turbulence noise can be approximately eliminated. Without laser phase and atmospheric turbulence noise, the model predicts the noise performance shown by the dashed curve in Fig. 2.

To test the SMIV model, an SMIV with parameters and features similar to those used in the above example was designed, built, and tested at NUWC Division Newport. Care was taken to balance the interferometer properly and to eliminate the effects of turbulence by using a nearly common optical path for both interferometer arms. The detector was implemented by using a hybrid photodiode followed by two stages of bandpass amplification and a low-pass filter. Measurements made at 11.2 kHz produced a sensitivity of 61 dB//μPa in 1 Hz. This is in close agreement with the model predictions in Fig. 2 and represents a 39-dB improvement over the results obtained with the commercial vibrometer.

C. Measured surface velocity and directivity

The surface velocity (dB//1 m/s) measured by the laser array system, with an air-backed window, is shown in Fig. 4. The figure compares the normal components of the surface velocity measured at three points—the window center (dashed line), midpoint (dash-dot line), and edge (dotted line)—to the equivalent velocity measured by the reference type 52 hydrophone (solid line). The sound-pressure level measured by the reference hydrophone was converted to an equivalent velocity level via Euler's equation, with $c_o = 1467$ m/s and $\rho_o = 1000$ kg/m^3. Differences in gains and spreading losses between the reference hydrophone and the measured surface velocity were accounted for. Thus, Figs. 4 and 5 compare the equivalent free-field acoustic particle velocity at the surface of the window to that measured on the window. The near-constant 6-dB difference between the window surface velocities and the reference hydrophone is due to air-backing, which creates an observed pressure-release boundary; the large fluctuations seen in both figures (Fig. 4 and Fig. 5) are due to the source (BQR-7 projector). The BQR-7 projector is essentially omnidirectional in the horizontal plane within the frequency range shown.

The surface velocity measured on the water-backed configuration at the window center point is shown in Fig. 5. Here, the water-backing eliminates the pressure release boundary condition. In Fig. 5, the difference between the theoretical (solid line) and measured (dashed line) particle velocity was less than 1.5 dB, or no more than approximately 18%, over the frequency band from 10 to 80 kHz.

Directivity measurements were also made of the angular response of the window to acoustic insonification. These measurements are comparable to beam patterns of conventional hydrophones. Here, the change in the normal component of velocity on the window with incidence angle was measured. Surface velocity on an idealized pressure-release boundary varies sinusoidally (as a cosine if broadside is defined to be normal incidence) with incidence angle.

Figure 6 compares the measured (air-backed) directivity (solid line) at frequencies of 10, 20, and 40 kHz to an ideal cosine dependence (dashed line). As shown at 20 kHz, a full 360° beam measurement was made. The front-to-back plane ratio is greater than 25 dB.

D. Wave-vector frequency measurements—Methodology

The data recorded in NUWC Division Newport's ATF were analyzed to give k_x-k_y and k-ω spectra for the synthesized array. An FM sweep (10 kHz to 100 kHz, 7-ms duration) was used to obtain fine granularity in the frequency domain. A grid of 36×36 points, with 5-mm (0.2-in.) spacing, was acquired by the SLVS. Hence, the vertical and horizontal array aperture was approximately 18×18 cm (7 in.^{2}). The time series for the SLVS responses to the FM sweep at each point on the grid were recorded and matched-filtered offline using standard correlation processing techniques. The resulting impulse response was time gated using the same gate for all channels to maintain phase relations at all frequencies. The time-gated signals were then subjected to spectral processing to derive spatial and temporal frequency representations of the response of the window. A set of inci-

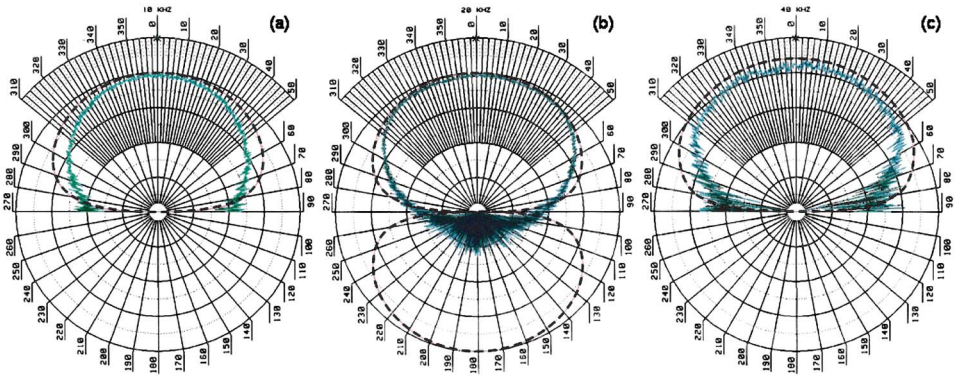


FIG. 6. (Color online) Directivity at the geometrical center of the window at (a) 10 kHz; (b) 20 kHz; and (c) 40 kHz. Dashed line represents theoretical dipole response.

dence angles was examined, including 0°, 30°, 45°, and 60° in azimuth (horizontal) and 0° in elevation (vertical).

Due to ambient static pressure, the acoustic window, or membrane, on the cylinder nose deformed into a concave shape—the z displacement of the center was about 2 cm inward from the extreme edge of the array. The shape of the cavity was determined to be accurate to about 1 mm (2σ fit) by fitting the z coordinate (derived from the time displacement of the impulse responses of the individual sensor positions) to the following parabolic surface:

$$z = ax^2 + by^2 + cxy + dx + ey + f. \tag{7}$$

This allows for adjustments to length, width, and orientation of the cavity shape relative to the (x,y) scanning grid. Of course, this fitting can only be done based on pulse arrival times from a source at normal incidence in the far field. Once the cavity shape was determined, the conventional k_x-k_y processing was modified to account for the window's curvature. Thus, the wave number spectrum, which directly relates to the array's beamforming ability, can be assessed without the distortion that the conventional 2D FFT would produce.

E. Wave-vector frequency measurement—Results

The wave-vector $(k_x$-$k_y)$ spectra at a frequency of 100 kHz are shown with and without the curvature, or z-coordinate correction, in Figs. 7 and 8, respectively. The magnitude of the incoming energy, as a function of wave vectors k_x and k_y, is normalized to the maximum response at broadside $(\theta=0°)$. Note the major differences between the corrected (Fig. 7) and uncorrected (Fig. 8) spectra at a high frequency, where the acoustic wavelength of 1.5 cm is less than the maximum cavity depth of about 2 cm. At lower frequencies, the effects of, and need for, the corrections become less obvious.

Figure 9 shows the wave-vector spectra for azimuthal angle of incidence of 30° at 60 kHz, with the elevation angle held constant at 0° incidence (all data are corrected for concavity).

Figure 10 shows the results of k-ω processing with the spatial coordinate chosen along the horizontal (x) direction. The vertical dependence is ignored. To reduce sidelobe levels generated by the finite-sampling aperture, a Hamming window was imposed on the raw time-series data. In addition, at each frequency, all of the known variations in sensitivity due to source level and other gains were removed.

The most interesting feature here is the uniformity of the beams at all angles of incidence and over nearly a decade of frequency bandwidth. The measured responses are near ideal and demonstrate a very capable system. Note that there is little indication of significant aliasing within the acoustic region—this was true even at 60° azimuth.

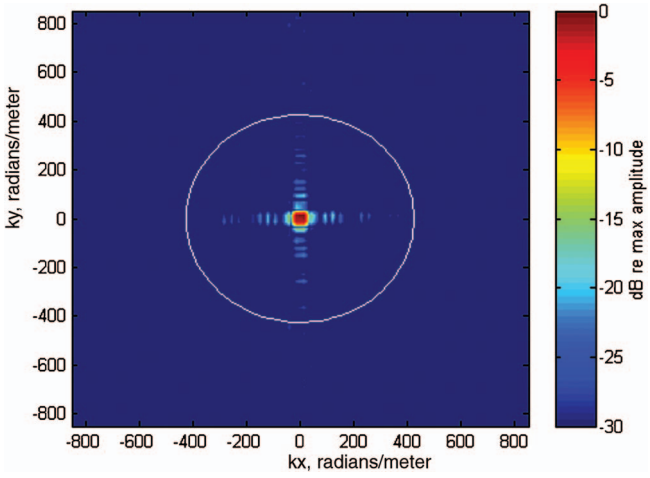


FIG. 7. Wave-vector spectrum for 100 kHz with correction for concavity (white circle represents highest acoustic wave number).

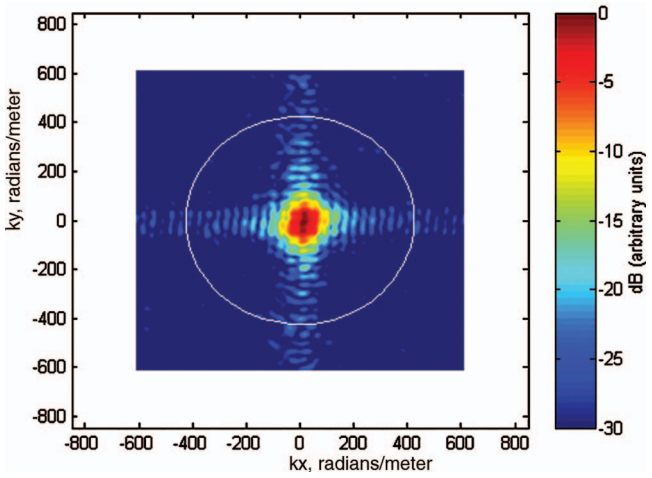


FIG. 8. Wave-vector spectrum for 100 kHz without correction for concavity (white circle represents highest acoustic wave number).

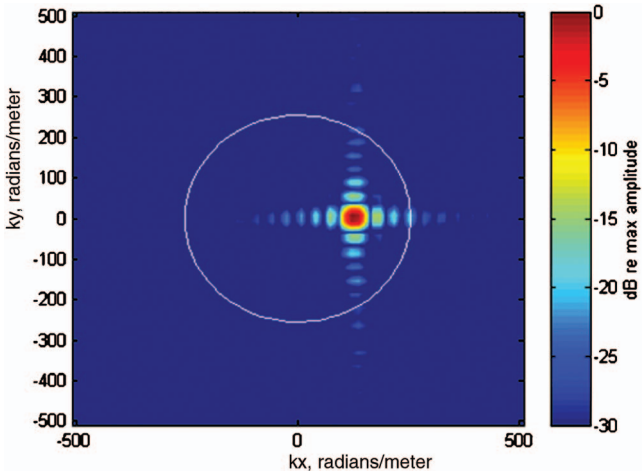


FIG. 9. Wave-vector spectrum for (30°, 0°) incidence angle at 60 kHz (white circle represents highest acoustic wave number).

IV. SUMMARY AND CONCLUSIONS

The research presented here has demonstrated the capabilities of a high-frequency laser sonar system. The results reveal a capable receiving array having exceptional frequency bandwidth and angular coverage. For the chosen configuration of the array window and material properties, there was no indication of significant aliasing within the acoustic region, even at an upper frequency of 100 kHz and array steering angle of 60° azimuth. These measurements verified a frequency bandwidth of a decade, from 10 to 100 kHz, and the ability to steer, without grating lobes, to an angle of 60°. However, the high-frequency laser sonar concept was validated only for high signal-to-noise ratios and under ideal (quiet ambient) test tank conditions.

Some of the questions regarding feasibility of the SLVS array system have been examined. A theoretical foundation was formulated using both a simple membrane model for the vibrational response of an acoustic window[12] and a detailed viscoelastic continuous plate model.[7] In theory, the normal velocity of the internal surface of a well-designed acoustic window will be directly proportional to the acoustic particle velocity of an incidence plane wave. For the high-frequency sonar application described here, viscoelastic neoprene or polyurethane windows with a thickness of 1.27 cm (0.5 in.) were desirable.

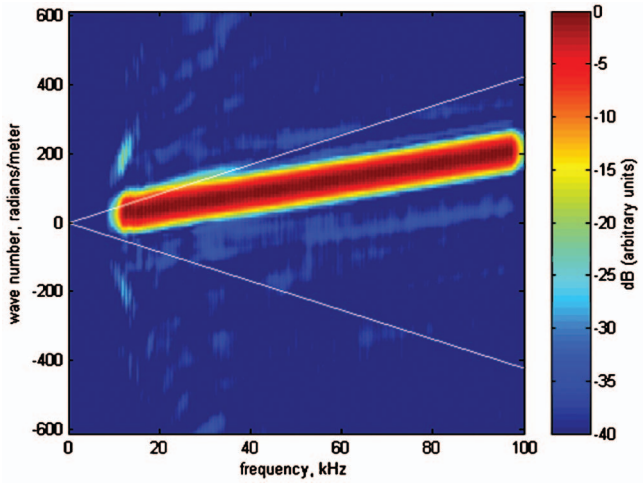


FIG. 10. Wave-number spectrum for (30°, 0°) incidence angle versus frequency with normalization and Hamming shading (white line represents highest acoustic wave number).

The poor acoustic sensitivity of the commercial SLVS remains a primary feasibility issue. Recall that Fig. 2 compared the measured noise floor of the commercial laser to that of the H-52 monitor hydrophone. Within the quiet test tank, the laser noise is at least 50 dB greater than the ambient noise. A passive receive array would require better sensitivity; the sensitivity of the commercial system (approximately 120 dB‖1 μPa) would limit the sonar severely. Even though the wavelength of a coherent He-Ne laser beam is quite small at 633 nm, which registers very small Doppler shifts, the velocities associated with easily detectable acoustic plane waves are even smaller. For example, the acoustic particle velocity at an SPL of 20 dB‖1 μPa is approximately 0.08 Å/s. As is, the commercial system would be useful for only active reception of a strong transmitted signal. However, analytical modeling and measurements made on a proposed simplified Michelson interferometer laser vibrometer produced a 39-dB sensitivity improvement at 11.2 kHz.

There are other feasibility concerns, such as required scanning rate,[13] vibration isolation, and demodulation hardware. However, the scanning laser Doppler vibrometer sonar can adapt to varying signal and noise fields using reconfigurable array geometry, element sampling, and element staving, which may be useful for sonar self-noise mitigation and noise cancellation.

ACKNOWLEDGMENTS

The authors would like to acknowledge the measurement support received from Walter Boober, Rene LaFleur, and Hugo F. Mendoza of the Acoustic Test Facility (Code 8211) at the Naval Undersea Warfare Center Division, Newport, RI. The authors would like to thank the Office of Naval Research, Program Officers Dr. David Drumheller (ONR 332) and Dr. Michael Traweek (ONR 321MS), for financial support of this research. The authors appreciate, as well, the insightful and thorough remarks and observations provided by the reviewers of this manuscript.

[1]"Acoustic Test Facility, Naval Undersea Warfare Center Division Newport," http://www.npt.nuwc.navy.mil/atf8211

[2]W. Boober, D. Morton, and C. Gedney, "System for wave number-frequency analysis of underwater structures," 138th Meeting of ASA, Session 5aSA, Columbus, OH (1999).

[3]L. Antonelli, K. Walsh, and A. Alberg, "Laser interrogation of the air-water interface for in-water sound detection: Initial feasibility tests," 138th Meeting of the Acoustical Society of America (ASA), Columbus, OH, J. Acoust. Soc. Am. **106**, (1999).

[4]L. Antonelli and I. Kirsteins, "Empirical acousto-optic sonar performance versus water surface conditions," Proceedings of Marine Technology Soc./IEEE International Conference on Engineering in the Ocean Environment, Oceans 2001 **3**, 1546–1552 (2001).

[5]L. Antonelli and F. A. Blackmon, "Experimental investigation of optical, remote, aerial sonar," Proceedings of Marine Technology Soc./IEEE International Conference on Engineering in the Ocean Environment, Oceans 2002 **4**, 1949–1955 (2002).

[6]L. Antonelli and F. Blackmon, "Experimental demonstration of remote, passive acousto-optic sensing," J. Acoust. Soc. Am. **116**(6), 3393–3403 (2004).

[7]A. J. Hull and B. A. Cray, "Wavevector-frequency analysis for the structural acoustic response of three fluid-loaded plates," NUWC Division, Newport, RI, Tech Report NUWC-NPT TR 11,430, 16 May 2003.
[8]F. W. Sears, M. W. Zemansky, and H. D. Young, *University Physics*, 5th ed. (Addison-Wesley, Reading, MA, 1979), pp. 710–712.
[9]S. Hanish, *A Treatise on Acoustic Radiation, Volume II—Acoustic Transducers*, 3rd ed. (Naval Research Laboratory, Washington, DC, 1989), pp. 377–397.
[10]J. F. Vignola, Y. H. Berthelot, and J. Jarzynski, "Laser detection of sound," J. Acoust. Soc. Am. **90**(3), 1275–1286 (1991).
[11]L. E. Estes, "Noise analysis of a simplified Michelson interferometer vibrometer," NUWC Division, Newport, RI, Tech Report 11,465, 30 September 2004.
[12]B. A. Cray and S. E. Forsythe, "Adaptive high-frequency laser sonar: A feasibility study," NUWC Division, Newport, RI, Tech Report 11,427, RI, 14 April 2003.
[13]L. E. Estes and R. L. Murray, "Scanning and data reduction techniques for a laser vibrometer sonar array system," NUWC Division, Newport, RI, Tech Report 11,696, 1 August 2005.

Vector sensors and vector sensor line arrays: Comments on optimal array gain and detection

Gerald L. D'Spain
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 93940-0701

James C. Luby
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

Gary R. Wilson and Richard A. Gramann
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713

(Received 28 July 2005; revised 18 April 2006; accepted 2 May 2006)

This paper examines array gain and detection performance of single vector sensors and vector sensor line arrays, with focus on the impact of nonacoustic self noise and finite spatial coherence of the noise between the vector sensor components. Analytical results based on maximizing the directivity index show that the particle motion channels should always be included in the processing for optimal detection, regardless of self noise level, as long as the self noise levels are taken into account. The vector properties of acoustic intensity can be used to estimate the levels of nonacoustic noise in ocean measurements. Application of conventional, minimum variance distortionless response, and white-noise-constrained adaptive beamforming methods with ocean acoustic data collected by a single vector sensor illustrate an increase in spatial resolution but a corresponding decrease in beamformer output with increasing beamformer adaptivity. Expressions for the spatial coherence of all pairs of vector sensor components in homogeneous, isotropic noise show that significant coherence exists at half-wavelength spacing between particle motion components. For angular intervals about broadside, an equal spacing of about one wavelength for all components provides maximum directivity index, whereas each of the component spacings should be different to optimize the directivity index for angular intervals about endfire. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2207573]

PACS number(s): 43.30.Yj, 43.30.Wi, 43.60.Bf, 43.60.Cg [RAS] Pages: 171–185

I. INTRODUCTION

Acoustic vector sensors have been used in directional sonobuoys deployed by the U.S. Navy community for the past half-century (Wolf, 1998). "DIFAR" (DIrectional low Frequency Analysis and Recording) sonobuoys simultaneously measure the two horizontal components of acoustic particle motion (or acoustic pressure gradient) and acoustic pressure. These sonobuoys have not often been used in scientific studies [for an exception, see Wilson, *et al.* (1985)], but this situation has been changing over the past half-decade, especially in the field of large baleen whale vocalization studies (e.g., Thode *et al.* 2000; McDonald, 2004; Greene *et al.* 2004).

Recent advances in piezocrystal materials have enabled the design of shear mode single crystal sensors that allow all three acoustic particle motion components to be measured simultaneously by a compact package that can be oriented in any direction with respect to the earth's gravitational field (Traweek, 2004; Wilcoxon, 2004). These new-design vector sensors are being incorporated into line arrays for towed array applications because of their ability (among others) to break the left/right ambiguity associated with line arrays of omnidirectional elements. Tow tests have shown that with proper sensor mounting, flow noise at tow speeds of 2.6 m/s does not severely affect beamforming performance at frequencies above 300 Hz (Richards, 2005).

Motion-induced self noise on the particle motion channels has always been the greatest challenge in using vector sensors for ocean acoustics applications. In addition, electronic self noise on these channels can be a problem, e.g., the self noise on the accelerometer channels of the present-design vector sensors exceeds sea state zero ocean ambient noise levels at lower frequencies. Therefore, algorithms designed for use with these types of sensors should account for these sources of contamination. A focus of this paper is examining the impact of this self noise and accounting for it in an optimal way.

The basic processing of vector sensor data is well established (this point is reflected by the fact that most of the important literature references in this paper were published about 50 years ago), but published beamforming results have been limited almost exclusively to conventional "cardioid" beamforming. All the various types of beamforming techniques developed for spatially distributed hydrophones also can be applied to vector sensor data once the vector sensor processing approach is placed in a general matched field processing framework. The theoretical basis for processing the data from these types of sensors is provided by the fundamental conservation laws of physics, e.g., the conservation of linear momentum and the conservation of mass, and Taylor's theorem. The approach in this paper is to use a simple analytical formulation from the optimization of the directivity index to provide insight into the optimal processing of

vector sensor data, particularly for autonomous systems, and the design of vector sensor arrays. The results extend the work of Cray and Nuttall (2001) to cases where self noise on the particle motion channels is present, the interelement spacings differ from half-wavelength spacing, and the individual vector sensor components have different spacings. Maximizing the directivity index is equivalent to minimizing the output variance of the beamformer under the constraint of unity gain in the look direction, i.e., "minimum variance distortionless response" (MVDR) processing, for the case of a single plane wave in a homogeneous, isotropic noise field. The solutions to these optimization problems are a set of channel weightings that, in effect, spatially whiten the data. The formulation suggests a simple beamforming approach that provides optimal detection if certain assumptions are satisfied. [A formal analysis of optimal beamforming for detection with vector sensor arrays is provided in Traweek (2003).] Section II provides the background material for vector sensor processing followed by a discussion of some of the issues with optimal array gain and detection with single vector sensors in Sec. III. Results from applying conventional, MVDR, and white-noise-constrained adaptive plane wave beamforming methods with single vector sensor data collected at sea also are presented in this section. Section IV is devoted to an examination of a line array of vector sensor components. The focus in this section is on the spatial coherence properties of the various vector sensor components in noise fields. Conclusions from this work are summarized in the final section.

II. BACKGROUND FOR VECTOR SENSOR PROCESSING

The standard approach in ocean acoustics for measuring the directional properties of the ocean sound field is to deploy a spatially distributed set of acoustic pressure sensors. An alternative approach can be obtained from the Taylor series expansion of the acoustic pressure field, $p(\underline{x},t)$, about a single measurement point in space, $\underline{x}_0$ (e.g., D'Spain *et al.*, 1991). Assuming that no acoustic sources exist in the region about this point, then

$$p(\underline{x},t) = p(\underline{x}_o,t) + \nabla p(\underline{x}_o,t)\cdot\Delta\underline{x} + \frac{1}{2}\Delta\underline{x}^T\begin{bmatrix}\text{matrix of}\\ \text{2nd derivatives}\end{bmatrix}_{x_o}\Delta\underline{x} + \cdots, \quad (1)$$

where $\Delta\underline{x}\equiv\underline{x}-\underline{x}_o$. This equation says that the measurement of acoustic pressure and its higher-order spatial derivatives at a single point in space is equivalent to the measurement of acoustic pressure in a volume about the measurement point. It provides the theoretical basis for array processing with measurements at a single point in space.

The phenomenon of superdirectivity is directly related to the Taylor series expansion of the pressure field. That is, superdirectivity arises for a spatially distributed hydrophone array when the directivity index is maximized as a function of the element weights, and the interelement spacing becomes smaller than half the acoustic wavelength (Pritchard, 1954). One can show that, as the ratio of the interelement spacing to the acoustic wavelength approaches zero, the weights for a "linear point array" approach the finite difference approximations to the spatial derivatives of pressure given in Eq. (1). The instability that results when the weights become large and of opposite sign can be avoided by the use of alternative transduction methods suggested by the physical interpretation of the spatial derivatives of pressure, as discussed in the next paragraph.

The acoustic wave equation is derived from two conservation laws, the conservation of linear momentum and the conservation of mass. The conservation of linear momentum under the linear acoustic approximation is

$$\rho_0(\underline{x})\frac{\partial\underline{v}(\underline{x},t)}{\partial t} + \nabla p(\underline{x},t) = 0. \quad (2)$$

The ambient density is $\rho_0(\underline{x})$ and the acoustic field variables, $\underline{v}(\underline{x},t)$ and $p(\underline{x},t)$, are the particle velocity and pressure, respectively. The linearized conservation of mass with the equation of state included is

$$\frac{\partial p(\underline{x},t)}{\partial t} + \kappa_s(\underline{x})\,\nabla\cdot\underline{v}(\underline{x},t) = 0. \quad (3)$$

The fluid properties, i.e., the adiabatic incompressibility, $\kappa_s(\underline{x})$, and the density, are assumed to be fixed in time over the time scales of acoustic signal propagation. Deviations from an ideal fluid (e.g., viscous dissipation) are assumed negligible. These two equations contain all the physics needed for the understanding of basic sound propagation. They hold in regions where no acoustic sources exist.

Since Eq. (2) shows that acoustic particle velocity at a given frequency is proportional to the first-order spatial gradient of pressure, then the simultaneous measurement of acoustic pressure and acoustic particle velocity at a single point in space is equivalent to the measurement of acoustic pressure in a small volume about the measurement point. In fact, an acoustic vector sensor is analogous to a four-element superdirective pressure sensor array where the two sensor types may use different transduction principles. Both sensor types require four data telemetry channels to measure all components. An important implication of Eq. (2) is that if the acoustic pressure field is measured (or otherwise known) everywhere within a given region of space, the corresponding acoustic particle velocity field can be derived everywhere within this region. That is, measurement of acoustic particle velocity adds no new information beyond that available from spatially distributed measurements of pressure.

III. OPTIMAL ARRAY GAIN AND DETECTION WITH A SINGLE VECTOR SENSOR

A. Detection

The mathematical theory of detection is based on statistical hypothesis testing (Van Trees, 1968). The situation of interest is that of low signal-to-noise ratio, SNR $\ll 1$. Application of results from information theory (Brown and Rowlands, 1959) shows that no better method exists to increase the SNR of signals at low SNR than to add the outputs of a hydrophone array. Therefore, since the simultaneous measurement of acoustic pressure and acoustic particle motion at a single point is equivalent to a volumetric measurement of

acoustic pressure about the measurement point, then the optimal approach to detection with vector sensors is to apply additive beamforming techniques, i.e., to sum the pressure and particle motion signals together. In vector intensity processing (e.g., D'Spain *et al.*, 1991), the pressure and particle motion signals are multiplied or correlated together rather than summed so that this approach is suboptimal for detection. Many other multiplicative processing approaches can be defined (e.g., D'Spain, 1990), but the Brown and Rowlands results indicate that none can exceed additive beamforming in detection performance. The degradation in detection performance over additive beamforming has been derived for the specific case of combining pressure and particle velocity time series multiplicatively (Cox and Baggeroer, 2003).

The conclusion that additive beamforming techniques, rather than multiplicative processing or other approaches, are optimal for detection of weak signals in noise also is consistent with standard detection theory. That is, in the case of signal completely unknown, the optimal processing approach is a linear filter followed by a square-law ("energy") detector (Van Trees, 1968). In narrow-band additive beamforming, the filtering process is composed of temporal filtering with a fast Fourier transform followed by the spatial filtering by the beamformer. Square-law detection then is performed on the beamformer output. Therefore, optimal beamforming techniques are directly related to optimal detection in that they provide optimal spatial filtering and their outputs squared comprise the detection statistic. The focus of this paper is on this optimal spatial filtering step.

In contrast to detection, the process of localization and signal parameter estimation typically involves signals with significantly higher SNR. When excess signal energy is available, multiplicative processing such as intensity processing can be quite advantageous, particularly when narrow beams and reduction in size and/or the number of elements is required. For example, Brown and Rowlands (1959) show that a multiplicative array of only two elements can be designed to have the same beam pattern as a linear array of an arbitrary number of equally spaced elements. An added benefit of vector intensity processing is that it provides a physics-based way of studying the properties of the sound field.

B. Vector sensor beamforming

The expression for the linear beamformer output of an acoustic vector sensor (three orthogonal components of acoustic particle velocity and acoustic pressure) for look direction, $\hat{l}$, is

$$D(\hat{l}) = a_0 p(t) + \sum_{j=1}^{3} a_j Z_j [v_j(t) + n_j(t)] \cos(\beta_j). \tag{4}$$

The components of the look direction are the direction cosines, $\hat{l} \equiv (\cos(\beta_1), \cos(\beta_2), \cos(\beta_3))$, given in terms of azimuth, θ, and elevation angle, ϕ, as $\cos(\beta_1) = \cos(\theta)\cos(\phi)$, $\cos(\beta_2) = \sin(\theta)\cos(\phi)$, and $\cos(\beta_3) = \sin(\phi)$. The quantities a_0 through a_3 (also expressed in this paper as a_0, a_x, a_y, and a_z) are weightings of the individual component time series determined by some criterion [e.g., to maximize the directivity index (DI)], $n_j(t)$ is the self noise time series on the jth velocity component (the self noise on the hydrophone channel is assumed negligible) and the quantities Z_j's are factors used to convert acoustic particle velocity into pressure. In vector notation, Eq. (4) is (e.g., D'Spain *et al.*, 1992)

$$\begin{aligned} D(\hat{l}) &= \underline{e}^H \cdot \underline{d}(t) \\ &= [a_0^*, a_x^* \cos(\beta_x), a_y^* \cos(\beta_y), a_z^* \cos(\beta_z)] \\ &\quad \times [p(t), Z_x(v_x(t) + n_x(t)), Z_y(v_y(t) + n_y(t)), Z_z(v_z(t) \\ &\quad + n_z(t))]^H, \end{aligned} \tag{5}$$

where the superscript H represents the complex conjugate transpose operation. The set of channel weights, a_0, a_x, a_y, and a_z is set so that the output amplitude for a plane wave arrival in the look direction at a given frequency is preserved; $|\underline{e}| = 1/\sqrt{2}$. The output variance (or mean squared amplitude) of the vector sensor beamformer at circular frequency ω then can be written as

$$\begin{aligned} &\underline{e}^H [S(\omega)] \underline{e} \\ &= \underline{e}^H \begin{bmatrix} S_p(\omega) & Z_x S_{px}(\omega) & Z_y S_{py}(\omega) & Z_z S_{pz}(\omega) \\ \cdots & Z_x Z_x^* S_x(\omega) & Z_x Z_y^* S_{xy}(\omega) & Z_x Z_z^* S_{xz}(\omega) \\ \cdots & \cdots & Z_y Z_y^* S_y(\omega) & Z_y Z_z^* S_{yz}(\omega) \\ \cdots & \cdots & \cdots & Z_z Z_z^* S_z(\omega) \end{bmatrix} \underline{e}, \end{aligned} \tag{6}$$

where the symbol "*" indicates complex conjugation and $[S(\omega)] = E\{\underline{d}(\omega)\underline{d}^H(\omega)\}$ is the data cross spectral density matrix, typically estimated from averaging statistically independent snapshots under the ergodic assumption (Bendat and Piersol, 1986). The formulation in Eqs. (5) and (6) is in the same form as plane wave beamforming with spatially distributed hydrophone arrays, and, more generally, with matched field processing formulations. Therefore, all of the methods developed for use with hydrophone arrays are directly applicable to vector sensor data (see Sec. III E for some examples). The physics contained in the cross spectral density matrix is describable in terms of the energetics of acoustic fields (D'Spain *et al.*, 1991). That is, the sum of the four autospectra (the trace of the matrix) is proportional to the total acoustic energy density, where the acoustic pressure spectrum (in the first row and first column) is proportional to the acoustic potential energy density and the sum of the three individual particle velocity autospectra is proportional to the acoustic kinetic energy density. The cross spectra between the pressure and the particle velocity (the off-diagonal elements in the first row and column) are proportional to the components of vector acoustic intensity. The real part of these cross spectra, the "active" intensity, is a measure of the net flux of acoustic energy and the imaginary part (signifying its time average is zero) is called the "reactive" intensity and is the energy flow required to support any spatial heterogeneity in the sound field. The properties of the 3×3 particle velocity submatrix are determined by the polarization of the

particle motion (D'Spain 1999). Sound propagation in the ocean is not necessarily rectilinear; it can take on various degrees of ellipticity, including purely circular motion.

In plane wave beamforming, the conversion factors from particle velocity to pressure are equal to the characteristic impedance of the medium, i.e., $Z_j = \rho_0 c$. In this case, the output of the vector sensor beamformer can be expanded as

$$\begin{aligned} E\{DD^*\} &= a_0^2 E\{pp^*\} + (\rho_0 c)^2 \sum_{j=1}^{3} a_j^2 [E\{v_j v_j^*\} + \sigma_j^2] \cos^2(\beta_j) \\ &+ 2(\rho_0 c) a_0 \sum_{j=1}^{3} a_j \, \mathrm{Re}\{E\{p v_j^*\}\} \cos(\beta_j) \\ &+ 2(\rho_0 c)^2 [a_1 a_2 \, \mathrm{Re}\{E\{v_1 v_2^*\} \\ &+ E\{n_1 n_2^*\}\} \cos(\beta_1)\cos(\beta_2) + a_1 a_3 \, \mathrm{Re}\{E\{v_1 v_3^*\} \\ &+ E\{n_1 n_3^*\}\} \cos(\beta_1)\cos(\beta_3) + a_2 a_3 \, \mathrm{Re}\{E\{v_2 v_3^*\} \\ &+ E\{n_2 n_3^*\}\} \cos(\beta_2)\cos(\beta_3)], \end{aligned} \tag{7}$$

where $E\{\cdots\}$ is the expectation operator, $\mathrm{Re}\{\cdots\}$ is the real part of a complex quantity, and the asterisk superscript signifies the complex conjugate operation. The self noise on the jth particle velocity channel, whose variance is σ_j^2, is assumed independent of the acoustic signal and acoustic noise on that component, and of the hydrophone component time series. However, correlation of the self noise between the various particle motion channels is allowed to be nonzero in order to take account of coupling from vibration, strum, etc.

C. Directivity index for a single vector sensor

The gain of an array (AG) is defined as ten times the logarithm, base 10, of the output signal-to-noise ratio of the array ($\mathrm{SNR}_{\mathrm{array}}$) normalized by the SNR of a single omnidirectional element ($\mathrm{SNR}_{\mathrm{omni}}$) (Urick, 1986):

$$AG \equiv 10 \log_{10}[\mathrm{SNR}_{\mathrm{array}}/\mathrm{SNR}_{\mathrm{omni}}]. \tag{8}$$

The directivity index (DI) is defined as the array gain for the case of a single, perfectly spatially coherent plane wave in the presence of homogeneous, isotropic ocean acoustic noise. Assume that the mean squared amplitude of the plane wave is $\bar{A}^2/2$, that the variance of the ocean acoustic pressure noise field is σ^2, and that the vector sensor is beamformed in the direction of arrival of the plane wave. Then, the beamformer output mean squared amplitude for signal only is $\bar{A}^2/2$ and the $\mathrm{SNR}_{\mathrm{omni}}$ equals $\bar{A}^2/(2\sigma^2)$. Therefore, the only quantity that remains to be determined is the denominator of $\mathrm{SNR}_{\mathrm{array}}$, i.e., the vector sensor beamformer output due to noise alone.

In a homogeneous, isotropic noise field, no net flux of acoustic energy occurs. Therefore, the active acoustic intensity, i.e., the real part of $E\{pv_j^*\}$, is zero in every direction. In addition, the imaginary part, the reactive intensity, is zero because the field is homogeneous (the acoustic pressure spectrum is everywhere the same). Similarly, the average polarization of particle motion in this ocean acoustic noise field is zero so that $E\{v_j v_k^*\}=0$ for $j \neq k$. Note that $E\{n_j n_k^*\}$ is not assumed to be zero for $j \neq k$ in order to account for any particle motion channel coupling of motion-induced self noise. Since the vector sensor is steered in the direction of the plane wave arrival, only the velocity component in that direction, v_l, must be considered in calculating DI. The kinetic energy density equals the potential energy density (proportional to σ^2) in a homogeneous, isotropic noise field and each particle velocity component contributes equally to the kinetic energy density so that $(\rho_0 c)^2 E\{v_l v_l^*\} = \sigma^2/3$. Substituting these results into Eq. (7) gives

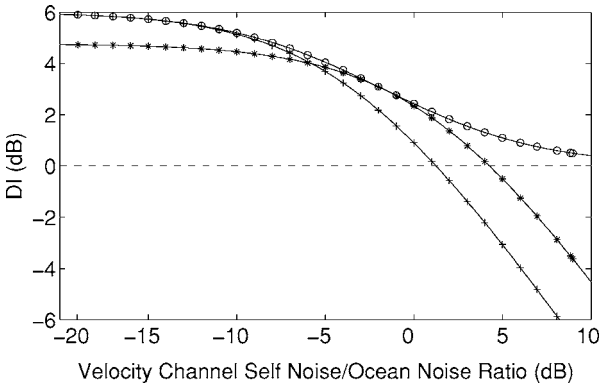


FIG. 1. Plots of the directivity index for a single vector sensor as a function of the ratio of the self noise variance on the particle motion channels to the ocean acoustic pressure noise variance. The curve of connected asterisks ("*") is the result for equal channel weighting, the curve of connected plus signs ("+") is for optimal channel weighting assuming no self noise, and the curve of connected circles ("○") pertains to optimal weighting that accounts for the level of self noise.

$$E\{DD^*\}/\sigma^2(\text{noise only}) = a_0^2 + a_v^2[1/3 + \eta]. \tag{9}$$

The quantity a_v is the weighting of the velocity component in the direction of the plane wave arrival, σ_v^2 is its self noise variance, and $\eta \equiv (\rho_0 c)^2 \sigma_v^2/\sigma^2$ is the ratio of the velocity channel self noise variance to ocean acoustic pressure noise variance. The DI then follows as

$$\mathrm{DI}(\text{vector sensor}) = -10 \log_{10}[a_0^2 + a_v^2(1/3 + \eta)]. \tag{10}$$

The conventional (Bartlett) approach to beamforming with vector sensors uses equal weighting of the pressure and particle velocity components. The resulting beam pattern is a cardioid and the directivity index is, from Eq. (10),

$$\begin{aligned} \mathrm{DI}(\text{cardioid}) &= 6.0 \text{ dB} - 10 \log_{10}[4/3 + \eta] \\ &= -10 \log_{10}[1/3 + \eta/4]. \end{aligned} \tag{11}$$

A plot of the directivity index for this conventional "cardioid" beamformer as a function of the velocity channel self noise to ocean noise ratio, η, is shown as the curve with asterisks in Fig. 1. At a self noise to ocean noise ratio of about 4.3 dB, the DI is zero, equal to that of a single omnidirectional hydrophone. As the self noise increases relative to the ocean noise beyond this value, the DI becomes negative and the velocity channels should be discarded when using equal weighting for the detection of weak signals. (For localization, the velocity channels still may provide useful information for large SNR signals, e.g., by reducing ambiguity).

Another beamforming approach is to determine the component weighting based on some specified optimization criterion. This optimization criterion can be defined in order to maximize the spatial filtering performance of the beamformer prior to the square-law detection. A useful criterion is to maximize the directivity index. As mentioned in Sec. III E, maximizing the directivity index is equivalent to minimizing the beamformer output variance under the constraint of unity gain in the look direction [i.e., minimum variance distortionless response (MVDR) beamforming] for the case of a single plane wave in homogeneous, isotropic noise. Minimizing the argument of the logarithm in Eq. (10) under the constraint that $a_0+a_v=1$ (equivalent to the constraint of unity gain in the look direction for this case) gives the solution for the optimal weighting as $a_v=(\frac{4}{3}+\eta)^{-1}$ and $a_0=(1+3\eta)/(4+3\eta)$. The ratio of the two weightings, $a_0/a_v=\frac{1}{3}+\eta$, is reasonable in that the ocean noise variance on the acoustic pressure channel is three times that on the particle velocity channel (because of the equivalence of acoustic potential and kinetic energy densities in a homogeneous field) and so is weighted a third as much. In effect, optimal weighting whitens the noise across channels. The resulting maximum DI is

$$\mathrm{DI}(\max) = -10\log_{10}\left[1-1/\left(\tfrac{4}{3}+\eta\right)\right] = -10\log_{10}[(1+3\eta)/(4+3\eta)]. \quad (12)$$

This result also is plotted in Fig. 1, as a curve of connected circles. When the self noise on the velocity channels is negligible, the maximum DI is 6.0 dB in comparison to the 4.77 dB attainable with equal weighting. At a value of the self noise to ocean noise ratio equal to −1.76 dB $[10\log_{10}(\frac{2}{3})]$, the maximum DI equals the DI for equal channel weighting. At noise ratios smaller than −1.76 dB, the velocity channel is weighted more heavily than the pressure channel to achieve maximum DI, whereas it is weighted less heavily at larger noise ratios. A benefit of using this optimal channel weighting is that the directivity index never decreases below zero no matter how large the self noise to ocean noise ratio becomes. This result also indicates that to optimize detection, the particle motion channels should never be discarded except in the limit of infinite self noise, given that the level of self noise is accounted for in determining the channel weighting.

The corresponding plane wave response patterns with channel weightings that maximize the directivity index for a selection of ratios of velocity channel self noise to ocean noise are plotted in Fig. 2. As the ratio increases starting with very small values, the main lobe becomes broader and the side lobe in the 180° direction becomes smaller. The cardioid plane wave response with a null in the 180° direction, corresponding to equal channel weighting, is obtained at a noise ratio of −1.76 dB. For increasing noise ratios beyond this value, the plane wave response becomes progressively more circular in nature.

It is tempting to neglect the self noise on the particle motion channels because the channel weighting then is independent of frequency. However, as the curve for equal channel weighting in Fig. 1 (connected asterisks) illustrates, detection performance can be severely degraded by this self noise. The decrease in directivity index with increasing self noise is even more rapid if the channel weighting is set to maximize DI for the case of no self noise, as shown by the curve of connected plus signs in Fig. 1. Therefore, significant improvements in detection performance can be achieved by taking channel self noise into account even though the problem then becomes frequency dependent (assuming the self noise is frequency dependent).

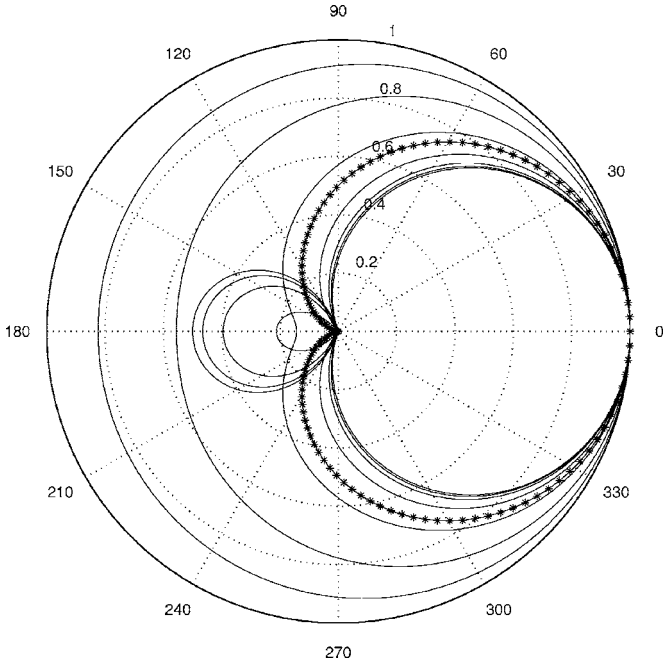


FIG. 2. Plots of the optimal beam pattern in 2D for a single vector sensor for various levels of self noise on the particle motion channels. The cardioid beam pattern for the case of a self noise to ocean noise ratio of −1.76 dB (equal channel weighting) is marked by asterisks. The curves for progressively smaller values of self noise, at −5, −10, −15, and −∞ dB, show increasingly narrower main lobes and increasingly larger side lobes in the 180° direction. Conversely, for progressively larger particle motion self noise levels, at 0, 5, 10, and ∞ dB, no nulls exist and the beam pattern becomes progressively more omnidirectional.

These results suggest a simple algorithm for optimal detection using single vector sensor data when the self noise on the particle velocity channels is electronic in nature and the background noise field is approximately homogeneous and isotropic (e.g., dominated by wind-generated ocean surface wave activity). It is based on a periodic estimate of the variance of the ocean acoustic pressure background noise field and *a priori* knowledge of the frequency dependence of the electronic self noise levels on the velocity channels. This information is used to derive a time-and frequency-dependent self noise to ocean noise ratio, $\eta(f,t)$. The self noise variance on each of the three channels can be obtained from laboratory measurements and included as a look-up table or as a curve expressed as a function of frequency. The value of $\eta(f,t)$ then determines the channel weighting using the formulas just above Eq. (12). This approach eliminates the need to estimate all terms in, and invert, data cross spectral density matrices as required by the adaptive beam forming methods discussed in Sec. III E.

The levels of nonacoustic self noise on the particle mo-

tion channels (motion-induced noise from vibration, strumming, etc., as well as electronic) can be estimated from the data as long as all four vector sensor components (acoustic pressure and the three components of acoustic particle velocity) are measured simultaneously, as discussed in the next part of this section.

D. Distinguishing acoustic and nonacoustic noise

Evaluation of the contribution of electronic noise, motion-induced contamination, and other forms of nonacoustic self noise in ocean acoustics measurements provides essential information for improving sensor performance and array design. The ability to distinguish this nonacoustic self noise from ocean acoustic signal and ocean acoustic noise on the particle motion channels depends upon the fact that this contamination does not satisfy the basic conservation laws of acoustic fields. The vector properties of acoustic intensity are useful in this regard in that they provide relationships between the various elements in the vector sensor data cross spectral density matrix. They are (Mann *et al.*, 1987)

$$\nabla \cdot \underline{C}_{p\underline{v}}(\omega) = 0, \tag{13a}$$

$$\nabla \times \underline{C}_{p\underline{v}}(\omega) = \frac{\omega}{c^2}\left\{\underline{C}_{p\underline{v}}(\omega) \times \underline{Q}_{p\underline{v}}(\omega)\right\} \Big/ \left\{\frac{1}{2}\frac{1}{\kappa_s} S_p(\omega)\right\}, \tag{13b}$$

$$\nabla \times \underline{Q}_{p\underline{v}}(\omega) = 0, \tag{13c}$$

$$\nabla \cdot \underline{Q}_{p\underline{v}}(\omega) = -\omega\left[\rho_o \sum_{j=1}^{3} S_j(\omega) - \frac{1}{\kappa_s} S_p(\omega)\right]. \tag{13d}$$

Equations (13a) and (13d) are valid outside regions containing acoustic sources and Eqs. (13b) and (13c) assume the spatial gradient of the fluid ambient density is negligible ($\nabla\rho_o \approx 0$). Since the divergence of the active intensity is zero, this form of energy flux can be modeled as an incompressible fluid flow using streamlines and stream functions for 2D flow (Waterhouse and Feit, 1986). Similarly, the curl-free nature of the reactive intensity indicates that it is analogous to potential fluid flows. Since (D'Spain *et al.* 1991)

$$\nabla S_p(\omega) = -2\omega\rho_o \underline{Q}_{p\underline{v}}(\omega), \tag{14}$$

then the scalar potential for reactive intensity is proportional to the acoustic pressure spectrum.

Equation (13d) is applicable to the estimate of overall level of self noise on the particle motion channels. For example, assume the ocean noise field is approximately homogeneous, i.e., the levels are locally independent of position. This assumption can be verified by examining the levels of the reactive intensity spectrum using Eq. (14). Given that the reactive intensity spectrum is approximately zero, then Eq. (13d) indicates that the scaled sum (scaled by the square of the medium's characteristic impedance) of the individual particle velocity autospectra [i.e., the trace of the 3×3 particle velocity submatrix in Eq. (6)] equals the acoustic pressure autospectrum. Therefore, any excess levels in the scaled sum of the particle velocity channels are a measure of the contribution from a nonacoustic component.

Similarly, correlation between particle motion channels caused by vibration or other forms of self noise will not be reflected in a corresponding correlation between pressure and particle velocity channels. One approach to estimating the degree of self noise that causes correlation between particle motion channels is to examine the imaginary parts of the cross spectra of the off-diagonal elements in the 3×3 particle velocity submatrix. The following relationship can be derived between the z component of the curl of the active acoustic intensity and the imaginary part of the xy particle motion cross spectrum (D'Spain, 1990):

$$\nabla \times \underline{C}_{p\underline{v}}(\omega)\big|_z = 2\omega\rho_o Q_{xy}(\omega) \tag{15}$$

and similarly for the other two components. From Eq. (13b), then

$$[Q_{yz}(\omega), Q_{zx}(\omega), Q_{xy}(\omega)] = \frac{\underline{C}_{p\underline{v}}(\omega) \times \underline{Q}_{p\underline{v}}(\omega)}{S_p(\omega)}. \tag{16}$$

Equation (16) shows that if the sound field is locally spatially homogeneous (i.e., $\underline{Q}_{p\underline{v}} \approx 0$), then the particle velocity cross spectral density matrix must be purely real, assuming (1) the spatial gradient of the fluid ambient density is negligible and (2) the field is quasi-monotone. Therefore, the degree to which the particle velocity cross spectral density matrix is not purely real in a homogeneous sound field is a measure of the self noise contamination that causes coherence between particle motion channels.

Since both acoustic signals and acoustic noise pass these tests, the tests are not used as part of the detection process. Rather, they are useful as one way of evaluating the quality of the vector sensor data.

E. Robust beamforming in a nonisotropic noise field

As mentioned above, maximizing the directivity index is equivalent to minimizing the beamformer output variance under a constraint equivalent to unity gain in the look direction for the case of a single plane wave in homogeneous, isotropic noise. This beamforming approach therefore provides an optimal spatial filtering for square-law (beamformer output) detection. The first (apparently) report of the use of adaptive beamforming techniques with acoustic vector sensor data (D'Spain *et al.*, 1992) presented only results from the use of the minimum variance/distortionless response beamformer (Capon, 1969). [Note that the use of adaptive beamforming methods with vector sensor data in fields outside underwater acoustics was reported much earlier, e.g., Oltman-Shay and Guza (1984).] However, more robust beamformer methods can be used. For example, white-noise-constrained (WNC) adaptive beamforming methods are designed to provide some of the higher spatial resolution and interferer cancellation capabilities of adaptive beamforming while maintaining some of the robustness of conventional beamforming (Cox *et al.*, 1987; Gramann, 1992). The constraint value is a free parameter that allows the beamformer to be "tuned" to the properties of a given signal and noise

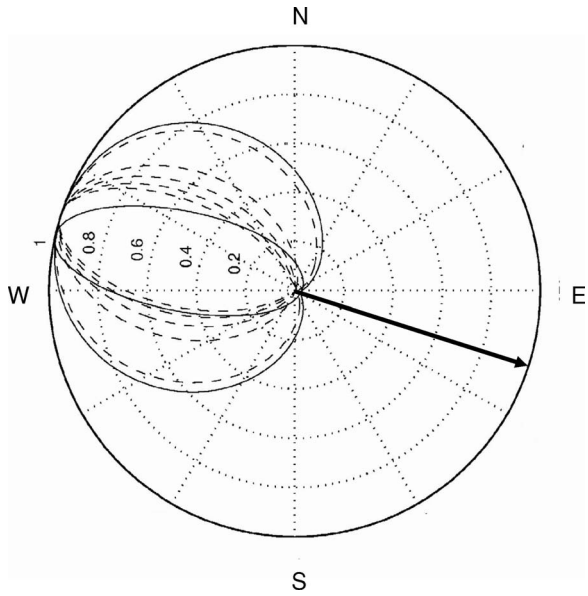


FIG. 3. The results of using various types of beamformers with ocean acoustic data collected by a single vector sensor. The sensor was deployed at 1385-m depth during an experiment in the deep northeast Pacific Ocean in July 1989. The processing was performed in the 14-Hz frequency band, corresponding to the frequency of a tone transmitted by an underwater projector 1700 km to the west/northwest of the receiver location. The outer solid curve with the cardioid shape is the conventional (equal weighting) beamformer output and the inner, oblong-shaped, solid curve is for the MVDR beamformer. The dashed curves represent the white-noise-constrained adaptive beamformer output with constraint values from 6 dB (i.e., 0.02 dB down from conventional) in the outermost dashed curve to 2 dB (4.02 dB down from conventional) in the innermost curve. Also plotted as a thick black line terminated with an arrow is the horizontal projection of the active acoustic intensity vector for this 14-Hz frequency bin.

structure given deviations from the underlying assumptions made in the processing (e.g., that the sensor components are accurately calibrated with respect to each other, that the signal arrival structure is adequately modeled, e.g., as a plane wave, etc.). It, in effect, places a constraint on the length of the channel weighting vector applied to the individual component data prior to summation.

Figure 3 shows the results of applying white-noise-constrained (WNC) adaptive beamforming as well as conventional and minimum variance distortionless response processing to data collected by a single vector sensor. Also plotted is the horizontal projection of the active acoustic intensity vector which measures the net horizontal flux of acoustic energy away from the source. These data were collected by a freely drifting, neutrally buoyant Swallow float during the 1989 Downslope Conversion experiment (D'Spain *et al.*, 1991). The sensor was deployed to a depth of 1385 m and recorded the transmissions of a 14-Hz tone from a ship-deployed acoustic source at a range of 1700 km to the west-northwest. White-noise-constraint values of 0.02, 1.02, 2.02, 3.02, and 4.02 dB down from $10\log_{10}(M)$ were used, where $M=4$ is the number of data channels. Setting the constraint to $10\log_{10}(M)$ is equivalent to conventional beamforming with the beamformer becoming increasingly more adaptive with decreasing constraint value (the 4.02-dB down constraint represents the most adaptive WNC beamformer in Fig. 3). The results indicate that the spatial resolution steadily increases as the beamformer becomes more adaptive, with the greatest resolution provided by the pure minimum variance distortionless response (MVDR) processor. However, the MVDR processor shows a bias of 5° compared to the other processors (283° versus 288°). In addition, not shown in Fig. 3, because all the beamformer outputs are normalized to the same value, is the fact that the maximum output of the MVDR processor at 83 dB *re* $1\mu\text{Pa}^2/\text{Hz}$ is more than 9 dB lower than the conventional beamformer output. About 0.5 dB of beamformer output is lost with each 1-dB increase in adaptability of the WNC beamformer for the results in Fig. 3. The bias and decrease in beamformer output are due to the well-known sensitivity of MVDR to mismatch (Cox 1973a). The two main sources of mismatch with these ocean acoustic data are calibration errors and acoustic propagation that differs from that assumed in using plane wave steering vectors. Calibration errors include not only those in amplitude of the various channels but also those in relative phase between the hydrophone and particle motion channels and between two particle motion channels. An error in relative phase calibration results in a quadrature component that otherwise would not exist. Propagation effects also can result in a quadrature component between pairs of vector sensor channels. In either case, a mismatch results between the data and the assumed plane wave steering vector in which the pressure and all particle velocity components are in phase and all quadrature components are zero.

For adaptive beamforming in azimuth with a vector sensor, two nulls are provided by the horizontal particle velocity components. The beamformer has some flexibility to steer these nulls in order to minimize the output variance. For example, simulations show that for the case of uncorrelated plane wave arrivals in a homogeneous, isotropic background noise field, the MVDR processor can resolve up to two sources spaced in azimuth by 60° to 65° when both arrivals have a signal-to-background-noise (SNR) of 12 dB, 75° to 80° when both arrivals have 6 dB SNR, and 65° to 70° when one source has an SNR of 15 dB and the other has an SNR of 9 dB.

Nonisotropic ambient noise results in a net flux of acoustic energy on average (i.e., the active acoustic intensity is nonzero). The very broad main lobe of a conventional vector sensor beamformer will allow much of the noise energy from high-noise-level directions to contaminate most look directions (except those at and near 180° from the high-noise directions), thereby decreasing SNR and therefore decreasing detection capability. An adaptive beamformer works to steer nulls in the directions of high levels of noise. Its ability to significantly reduce this noise depends both on the angular spread of the region of high noise levels and its proximity to the look direction of interest, as well as the number of high-noise directions. If the ambient noise field is inhomogeneous in addition to being anisotropic, imaginary components in the off-diagonal terms in the data cross spectral matrix appear. These imaginary components can have an impact equivalent to that of correlated multipath and can cause highly adaptive beamformers to perform very poorly unless accounted for in the steering vectors, as mentioned above.

IV. A LINE ARRAY OF VECTOR SENSORS

Consider a line array of vector sensor components, not necessarily equally spaced. Assume without loss of generality that the line array is oriented in the $\hat{z}$ direction and that the v_3 component is parallel to this direction (i.e., $v_3 \equiv v_z$). Also assume without loss of generality that the two velocity components perpendicular to the line of the array have been rotated so that the $\hat{x}$ direction corresponds to the azimuth of the single plane wave arrival). (The azimuth, θ, is the angle of the arrival in the $\hat{x}$-$\hat{y}$ plane measured clockwise from the $\hat{y}$ axis and the elevation angle, ϕ, is the angle between the arrival and the $\hat{x}$-$\hat{y}$ plane, so that $\phi=0$ corresponds to broadside to the line array.) Then, only three components need to be considered in evaluating the array's DI, p, v_x, and v_z. These components are obtained from plane wave beamforming the M vector sensors in the elevation angle direction of arrival of the single plane wave (ϕ):

$$
\begin{aligned}
p(\phi) &= \sum_{m=1}^{M} b_m p_m(\omega) \exp[ikz_m^p \sin(\phi)], \\
v_x(\phi) + n_x &= \sum_{m=1}^{M} c_m [v_x^m(\omega) + n_x^{(m)}(\omega)] \exp[ikz_m^x \sin(\phi)], \\
v_z(\phi) + n_z &= \sum_{m=1}^{M} d_m [v_z^m(\omega) + n_z^m(\omega)] \exp[ikz_m^z \sin(\phi)].
\end{aligned}
\tag{17}
$$

The spatial weightings of the individual components are specified by the b_m 's, c_m 's, and d_m 's. The quantities $z_m^{p,x,z}$ are the distances of the mth vector sensor components from the array phase center, where this phase center is the same for all components. The individual component outputs now are a function of frequency since the relative shift between sensors ($kz_m^{p,x,z} \sin(\phi)$) is a function of frequency through the wave number $k=\omega/c$. (Since all components in a single vector sensor are located at the same point in space, no such frequency-dependent phase shift occurs.) In effect, the vector sensor array can be viewed as a single vector sensor located at the phase center of the array and whose components [Eq. (17)] are directional in elevation angle. (The Appendix contains a discussion of the order in which this beamforming process occurs.)

As before for a single vector sensor, the only quantity in the expression for DI that must be evaluated is the beamformer output due to noise alone given that the beamformer points in the direction of arrival of the single plane wave. The output variance of the beamformer in Eq. (4), with the expressions for the various components given by Eq. (17), can be obtained directly from Eq. (7). The terms associated with the ocean acoustic noise field that must be evaluated in the resulting expression contain samples from the spatial cross spectra between each of the vector sensor components. The properties of these cross spectra now will be discussed.

A. Spatial cross spectra between vector sensor components

The spatial cross spectra between the various components of a vector sensor can be derived from the spatial cross spectrum of the acoustic pressure field (Eckart, 1953), $S_{pp}(\underline{x}_1, \underline{x}_2, \omega)$. This latter quantity is defined as

$$
S_{pp}(\underline{x}_1, \underline{x}_2, \omega) \equiv \int E\{p(\underline{x}_1, t) p(\underline{x}_2, t-\tau)\} \exp[-i\omega\tau] d\tau,
\tag{18}
$$

where the term involving the expectation is the pressure field's spatial correlation function. This expression is valid for temporally stationary fields. Taking the spatial gradient with respect to each of the positions, $\underline{x}_1$ and $\underline{x}_2$, and using Eq. (2) from Sec. II, then

$$
\nabla_{\underline{x}_1} S_{pp}(\underline{x}_1, \underline{x}_2, \omega) = i\omega\rho_o \underline{S}_{\underline{v}p}(\underline{x}_1, \underline{x}_2, \omega),
\tag{19a}
$$

$$
\nabla_{\underline{x}_2} S_{pp}(\underline{x}_1, \underline{x}_2, \omega) = i\omega\rho_o \underline{S}_{p\underline{v}}(\underline{x}_1, \underline{x}_2, \omega),
\tag{19b}
$$

$$
\begin{aligned}
\nabla_{\underline{x}_1} \nabla_{\underline{x}_2} S_{pp}(\underline{x}_1, \underline{x}_2, \omega) &= \nabla_{\underline{x}_2} \nabla_{\underline{x}_1} S_{pp}(\underline{x}_1, \underline{x}_2, \omega) \\
&= -(\omega\rho_o)^2 [S]_{\underline{v}\underline{v}}(\underline{x}_1, \underline{x}_2, \omega),
\end{aligned}
\tag{19c}
$$

where, for example, $\underline{S}_{\underline{v}p}(\underline{x}_1, \underline{x}_2, \omega)$ is a vector whose three components are the spatial cross spectra between acoustic pressure at $\underline{x}_1$ and each of the three components of acoustic particle velocity at $\underline{x}_2$. In tensor analysis, a scalar quantity (e.g., pressure or the spatial cross spectrum between spatially separated measurements of pressure) is a zeroth rank tensor and a vector [such as the quantities in Eqs. (19a) and (19b)] is a first rank tensor. The gradient operator increases the rank of a tensor by one. Therefore, the quantity in Eq. (19c), the spatial coherence between the various particle velocity components, is a second rank tensor, indicated by the brackets.

In inhomogeneous, time-stationary fields, the quantity $E\{p(x_1, t) p(x_2, t-\tau)\}$ is not symmetric about $\tau=0$, but rather

$$
E\{p(\underline{x}_1, t) p(\underline{x}_2, t+\tau)\} = E\{p(\underline{x}_2, t) p(\underline{x}_1, t-\tau)\}.
\tag{20}
$$

Therefore, $S_{pp}(\underline{x}_1, \underline{x}_2, \omega)$ is complex in general. However, in homogeneous, isotropic fields, the spatial cross spectra do not depend on the actual measurement locations, $\underline{x}_1$ and $\underline{x}_2$, but only on their separation $|\underline{x}_1 - \underline{x}_2|$, defined here as s. [In range-independent ocean acoustic waveguides, the noise field is homogeneous in the horizontal direction but not in the vertical direction (Kuperman and Ingenito, 1980)]. In this case, the pressure field's spatial correlation function is symmetric about $\tau=0$ and the corresponding spatial cross spectrum becomes purely real. The expression for the pressure field's spatial cross spectrum under these conditions is a scaled version of the sinc function (e.g., Eckart, 1953), which is equal to the spherical Bessel function of zeroth order, $j_0(ks)$ (Abramowitz and Stegun, 1964):

$$S_{pp}(s,\omega) = S_p(\omega)\mathrm{sinc}(ks) = S_p(\omega)\frac{\sin(ks)}{ks} = S_p(\omega)[j_0(ks)]. \tag{21}$$

The quantity $S_p(\omega)$ is the acoustic pressure autospectrum. It is specified explicitly as a function of frequency (versus the previous use of σ^2) because of the frequency dependence of the spatial coherence functions through $k=\omega/c$. Now

$$\nabla S_{pp} = \frac{\partial S_{pp}}{\partial s}\hat{s} = \frac{\partial S_{pp}}{\partial s}\left[\frac{\Delta x}{s}\hat{x} + \frac{\Delta y}{s}\hat{y} + \frac{\Delta z}{s}\hat{z}\right] = -kS_p(\omega)j_1(ks)\hat{s}, \tag{22}$$

where $(\Delta x)^2+(\Delta y)^2+(\Delta z)^2=s^2$ and the function $j_1(ks)$ is the spherical Bessel function of first order (Abramowitz and Stegun, 1964). Applying Eq. (19a) and (19b) gives

$$\underline{S}_{p\underline{v}}(s,\omega) = \frac{i}{\rho_0 c}S_p(\omega)j_1(ks)\left[\frac{\Delta x}{s}\hat{x} + \frac{\Delta y}{s}\hat{y} + \frac{\Delta z}{s}\hat{z}\right]. \tag{23}$$

Similarly,

$$\nabla\nabla S_{pp} = \nabla\left[\frac{\partial S_{pp}}{\partial s}\hat{s}\right] = \frac{\partial^2 S_{pp}}{\partial s^2}\hat{s}\hat{s} + \frac{\partial S_{pp}}{\partial s}\nabla\hat{s}. \tag{24}$$

The two second rank tensors in this equation are

$$\hat{s}\hat{s} = \begin{bmatrix} \frac{(\Delta x)^2}{s^2} & \frac{\Delta x\Delta y}{s^2} & \frac{\Delta x\Delta z}{s^2} \\ \frac{\Delta x\Delta y}{s^2} & \frac{(\Delta y)^2}{s^2} & \frac{\Delta y\Delta z}{s^2} \\ \frac{\Delta x\Delta z}{s^2} & \frac{\Delta y\Delta z}{s^2} & \frac{(\Delta z)^2}{s^2} \end{bmatrix} \tag{25}$$

and

$$\nabla\hat{s} = \frac{1}{s}\begin{bmatrix} 1-\frac{(\Delta x)^2}{s^2} & -\frac{\Delta x\Delta y}{s^2} & -\frac{\Delta x\Delta z}{s^2} \\ -\frac{\Delta x\Delta y}{s^2} & 1-\frac{(\Delta y)^2}{s^2} & -\frac{\Delta y\Delta z}{s^2} \\ -\frac{\Delta x\Delta z}{s^2} & -\frac{\Delta y\Delta z}{s^2} & 1-\frac{(\Delta z)^2}{s^2} \end{bmatrix} = \frac{1}{s}([I]-\hat{s}\hat{s}), \tag{26}$$

so that

$$\nabla\nabla S_{pp} = \left[\frac{\partial^2 S_{pp}}{\partial s^2} - \frac{1}{s}\frac{\partial S_{pp}}{\partial s}\right]\hat{s}\hat{s} + \frac{1}{s}\frac{\partial S_{pp}}{\partial s}[I] = k^2 S_p(\omega)j_2(ks)\hat{s}\hat{s} - \frac{k}{s}S_p(\omega)j_1(ks)[I]. \tag{27}$$

The quantity $j_2(ks)$ is the spherical Bessel function of second order and $[I]$ is the tensor with 1's along the diagonal and 0's in the off-diagonal positions. Applying the expression in Eq. (19c) yields

$$[S]_{\underline{v}\underline{v}}(s,\omega) = \frac{1}{(\rho_0 c)^2}S_p(\omega)\left[-j_2(ks)\hat{s}\hat{s} + \frac{1}{ks}j_1(ks)[I]\right]. \tag{28}$$

Equations (23) and (28) hold for any orientation of the particle velocity components with respect to the direction of their separation. For the line array oriented in the $\hat{z}$ direction as presented at the beginning of this section, $\Delta x=\Delta y=0$ and $s=\Delta z$, and these equations simplify to

$$S_{px}(\Delta z,\omega) = 0, \tag{29a}$$

$$S_{pz}(\Delta z,\omega) = \frac{i}{\rho_0 c}S_p(\omega)[j_1(k\Delta z)], \tag{29b}$$

$$S_{xx}(\Delta z,\omega) = \frac{1}{(\rho_0 c)^2}S_p(\omega)\left[\frac{1}{k\Delta z}j_1(k\Delta z)\right], \tag{29c}$$

$$S_{xz}(\Delta z,\omega) = 0, \tag{29d}$$

$$S_{zz}(\Delta z,\omega) = \frac{1}{(\rho_0 c)^2}S_p(\omega)\left[\frac{1}{k\Delta z}j_1(k\Delta z) - j_2(k\Delta z)\right]. \tag{29e}$$

These equations along with Eq. (21) provide all the terms needed to evaluate the vector sensor line array beamformer output variance in a homogeneous, isotropic noise field. Note that the following limits hold as the component spacing decreases to zero (Abramowitz and Stegun, 1964)

$$\lim_{s\to 0} j_0(ks) = 1,$$

$$\lim_{s\to 0} j_1(ks) = 0,$$

$$\lim_{s\to 0} \frac{j_1(ks)}{ks} = \frac{1}{3},$$

$$\lim_{s\to 0} j_2(ks) = 0. \tag{30}$$

Therefore, $(\rho_0 c)^2[S_{xx}(0,\omega)+S_{yy}(0,\omega)+S_{zz}(0,\omega)]=S_p(\omega)$, a reflection of the fact that the kinetic energy density equals the potential energy density in a homogeneous acoustic field. Also note that S_{pp}, S_{xx}, and S_{zz} are purely real and symmetric about $k\Delta z=0$, whereas S_{pz} is purely imaginary and antisymmetric about 0.

The spatial coherences for various pairings of channel types and elevation angles, ϕ, are plotted in Fig. 4. Specifically, the following quantities are plotted using the indicated symbols:

$$\Psi_{pp}(x) = j_0(x)\cos(x\sin\phi) \text{ at } 0^\circ\text{:square},$$

$$\Psi_{xx}(x) = 3\frac{j_1(x)}{x}\cos(x\sin\phi) \text{ at } 0^\circ\text{:circle},$$

$$\Psi_{zz}(x) = 3\left[\frac{j_1(x)}{x} - j_2(x)\right]\cos(x\sin\phi) \text{ at } 0^\circ\text{:plus},$$

$$\Psi_{pz}(x) = \sqrt{3}j_1(x)\sin(x\sin\phi) \text{ at } 90^\circ\text{:asterisk}.$$

A factor of 3 is used to scale the particle velocity component spatial coherences so that they have a value of unity at zero separation rather than one-third. Analogously, a factor of $\sqrt{3}$ is applied to the pressure/vertical particle velocity coher-

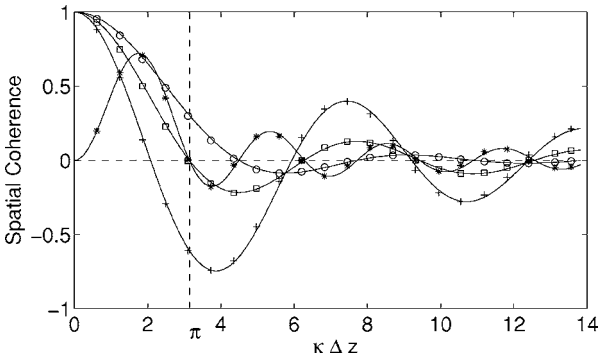


FIG. 4. Plots of the spatial coherence in homogeneous, isotropic noise as a function of the product of acoustic wave number and separation distance for various combinations of vector sensor components: two pressure sensors (connected squares), two particle motion sensors perpendicular to the direction of separation (connected circles), two particle motion sensors parallel to the direction of separation (connected plus signs), and a pressure sensor with a particle motion sensor parallel to the direction of separation (connected asterisks). The analytical expressions for the curves are presented in the text.

ence curve. This figure shows that significant spatial coherence exists at half-wavelength spacing (at a value of π marked by the vertical dashed line) between two vertical particle velocity components, $\Psi_{zz}(x)$, and between two horizontal particle velocity components, $\Psi_{xx}(x)$. It also shows that the locations of the zeros in the functions $\Psi_{xx}(x)$ and $\Psi_{zz}(x)$ are not periodic in $\kappa\Delta z$. Therefore, the typical approach of designing a hydrophone line array for effective detection at a specific frequency by placing the elements at the same interelement interval corresponding to the distance of the first zero in the spatial coherence function (half wavelength) does not apply with particle motion sensors. Figure 4 also suggests that different intercomponent spacings should be considered for each of the vector sensor component types. Whereas it may be desirable to space the hydrophones at integral multiples of $\frac{1}{2}$ wavelength, the first zero in the spatial coherence function for the two particle motion components perpendicular to the direction of the line array (in the $\hat{x}$ and $\hat{y}$ directions) is approximately at 0.72 wavelengths, about halfway between $\frac{2}{3}$ and $\frac{3}{4}$ wavelength (the first zero in the curve of connected circles in Fig. 4 is at 1.43π). In contrast, the first zero for the particle velocity component aligned with the direction of the spacing between components (the $\hat{z}$ direction) occurs at about $\frac{1}{3}$ wavelength (the first zero in the curve of connected "+" 's in Fig. 4 is at 0.66π). For this particle motion component, consideration also must be given to the spatial coherence between it and the acoustic pressure components as given by the $\Psi_{pz}(x)$ function.

Optimal spacings for the three vector sensor components is determined numerically at the end of this section. However, before presenting these results, note that the plots in Fig. 4 indicate that the noise spatial coherence for the various component pairs is minimized at component spacings close to larger integral multiples of π (e.g., 4π represents a spacing of 2 wavelengths). An array of M four component vector sensors in which all components are mutually incoherent in the acoustic noise field has a maximum directivity index of $10\log_{10}(M)+10\log_{10}(4)$. The first term is the well-known result for a hydrophone array (Urick, 1983) and the second term is the maximum directivity index of a single four-component vector sensor, as shown in Sec. III C. This result is equal to ten times the logarithm of the total number of data channels, exactly the same as for hydrophone arrays. It provides a benchmark with which to compare the directivity indices of vector sensor arrays where the intercomponent spatial coherence in the noise field is not negligible. [Although the product theorem states that the beam pattern of a spatially distributed array of identical directional sensors is equal to the product of the beam pattern of the equivalent spatial array of omnidirectional sensors and the beam pattern of the single directional sensor (Urick, 1983), the array gain only satisfies this property for all arrival angles when all components of the spatially distributed sensors are spatially uncorrelated.]

One problem with spacing the vector sensor array components at these larger values is that compact array design provided by the vector sensor/Taylor series approach is sacrificed. Also, the assumption of perfect spatial coherence for the signal of interest may no longer be valid at these larger apertures. In addition, as the self noise increases on the particle motion channels and their data are weighted less heavily, grating lobes begin to adversely impact source localization. [Grating lobes do not exist for a line array of 4–component vector sensors where self noise is not significant]. Therefore, in the frequency bands where self noise on the particle motion channels is of concern, array element spacing should be determined by criteria other than just the directivity index.

The foregoing discussion on the spacings of the various components for which the spatial coherence is zero holds for all elevation angles. An additional zero is contributed by the (co-)sinusoidal term involving the elevation angle. For example, the $\Psi_{zz}(x)$ function in Fig. 4 pertains to an elevation angle of 0° (of course, the output of particle motion components oriented in the $\hat{z}$ direction to an arrival at this elevation angle is zero). As the elevation angle increases, the first zero from the cosine term decreases monotonically, starting from infinity, to smaller values and reaches a value of 0.5π at 90°. At an elevation angle of about 49.3°, the zero from this additional term coincides with the first zero in the expression inside the brackets in Eq. (29e). Knowledge of the behavior of these elevation-angle-dependent zeros can be used in array design for detection of weak signals from known directions as in low-level underwater communication systems, acoustic barrier systems, etc.

B. Directivity index for a line array of vector sensors

Using the results in Sec. IV A, the output of a line array of vector sensor components in the presence of homogeneous, isotropic noise, normalized by the pressure autospectrum, is

$$E\{D(\phi,\omega)D^*(\phi,\omega)\}/S_p(\omega) = a_o^2\left[\sum_{m=1}^{M} b_m^2 + 2\sum_{m=2}^{M}\sum_{n=1}^{m-1} b_m b_n j_0(k\Delta z_{mn}^p)\cos(\Phi_{mn}^p)\right]$$
$$+ a_x^2\left[\frac{1}{3}\sum_{m=1}^{M} c_m^2 + 2\sum_{m=2}^{M}\sum_{n=1}^{m-1} c_m c_n \frac{j_1(k\Delta z_{mn}^x)}{k\Delta z_{mn}^x}\cos(\Phi_{mn}^x) + \eta_x\right]\cos^2(\phi)$$
$$+ a_z^2\left[\frac{1}{3}\sum_{m=1}^{M} d_m^2 + 2\sum_{m=2}^{M}\sum_{n=1}^{m-1} d_m d_n\left[\frac{j_1(k\Delta z_{mn}^z)}{k\Delta z_{mn}^z} - j_2(k\Delta z_{mn}^z)\right]\cos(\Phi_{mn}^z) + \eta_z\right]\sin^2(\phi)$$
$$+ a_0 a_z\left[\sum_{m=2}^{M}\sum_{n=1}^{m-1}[b_m d_n j_1(k\Delta z_{mn}^{pz})\sin(\Phi_{mn}^{pz}) + b_n d_m j_1(k\Delta z_{nm}^{pz})\sin(\Phi_{nm}^{pz})] + \sum_{m=1}^{M} b_m d_m j_1(k\Delta z_{mm}^{pz})\sin(\Phi_{mm}^{pz})\right]\sin(\phi)$$
$$+ 2a_x a_z \eta_{xz}\cos(\phi)\sin(\phi) \tag{31}$$

with $\Delta z_{mn}^p \equiv z_m^p - z_n^p$, $\Delta z_{mn}^{pz} \equiv z_m^p - z_n^z$, and similarly for the other Δz terms. Also, $\Phi_{mn}^{p,x,z}$ is defined as $\Phi_{mn}^{p,x,z} \equiv k\Delta z_{mn}^{p,x,z}\sin(\phi)$.

The full solution for the maximum directivity index can be calculated from the inverse of the $3M \times 3M$ data cross spectral density matrix whose elements are determined by the terms inside the summation signs in Eq. (31) setting the b_m's, c_m's, and d_m's to unity. Figure 5(a) presents the result of calculating the maximum DI in this manner for a family of vector sensor arrays. Each of the elements of these arrays are four-component vector sensors that are equally spaced, with the spacing varying from 0.5 to 5.0 wavelengths in steps of 0.1 wavelengths. The level of self noise on all particle motion channels is zero. The plot indicates that at the larger interelement spacings, the maximum DI shows less variation with changes in elevation angle and spacing because of decreasing spatial coherence of the acoustic noise. The value approaches 6.02 dB above $10\log_{10}(M)$, the maximum DI in the case of mutual incoherence between sensors. However, at the smaller spacings, the maximum DI is significantly greater than the mutually incoherent value for some elevation angles and interelement spacings. The arcs along which the DI is greatest start at broadside (0° elevation angle) at approximately integral multiples of a wavelength in spacing. The reason for this periodicity is that the Ψ_{pp} and Ψ_{xx} functions attain negative values simultaneously at this same periodicity, as seen in Fig. 4. [A value of π along the abscissa in Fig. 4 corresponds to 0.5 wavelengths in Fig. 5(a).] Negative values of these spatial cross spectra partially offset the contribution to the beamformer output variance from the positive-valued autospectral terms (the terms along the diagonal in the data cross spectral density matrix), thereby increasing DI. Knowledge of the behavior of these arcs of increased DI is valuable in the design of vector sensor arrays for detection of narrow-band signals arriving within a given angular interval, providing more than 1-dB improvement in certain cases. In addition to this one-wavelength periodicity, a periodicity at 0.5 wavelength is apparent in Fig. 5(a), associated with the approximate π periodicity in the functions in Fig. 4.

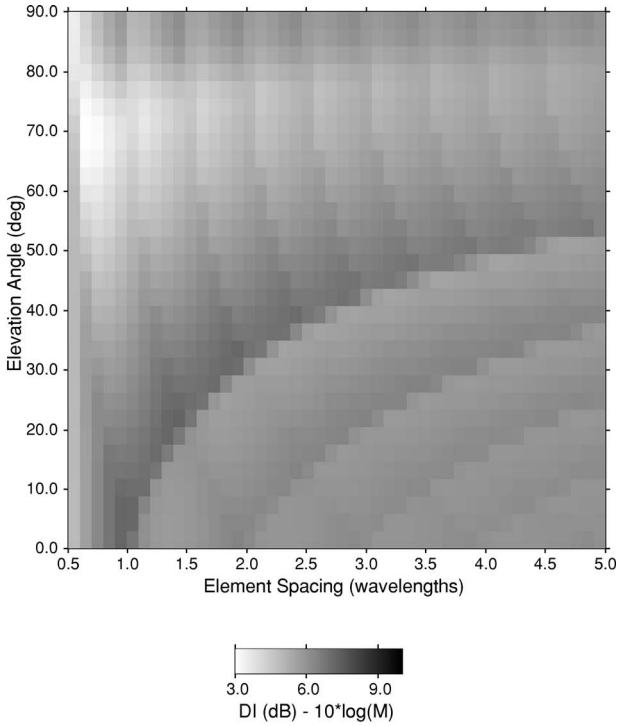
The actual values for the optimal $3M$ weights [the product of a_o with the M b_m weights, a_x with the M c_m's, and a_z with the M d_m weights in Eq. (31)] also can be calculated from the inverse of the data cross spectral density matrix. Results show that these optimal weights for one or more of the vector sensor components often oscillate between two widely separated values, sometimes between positive and negative values, as a function of position along the array length. An alternative, approximate processing approach is to first beamform in elevation angle, as in Eq. (17), and to set the weights of all components of a given type to equal values, i.e., $b_m = c_m = d_m = 1/M$. This approach assumes that the self noise levels on all components of the same type are approximately equal. After this first step, the impact of self noise on the particle motion channels and the effects of nonzero spatial coherence of various component pairs in the noise field on detection performance can be evaluated in a way identical to that used with a single vector sensor. In this case, only three weights, a_o, a_x, and a_z, need to be determined in the optimization. Another benefit of this approach is that it can be applied in situations where MVDR beamforming cannot be used in elevation angle because of problems with correlated multipath, e.g., when the vector sensor array is deployed in the vertical direction in the ocean (D'Spain *et al.*, 1992). Figure 5(b) shows the results of this approximate approach under exactly the same conditions as in Fig. 5(a). The relationship between these two processing approaches is discussed in the Appendix.

Following this approach, the equal $1/M$ weighting can be factored out of Eq. (31) and the equation can be rewritten in matrix form with separation of the contribution from component spatial coherence as

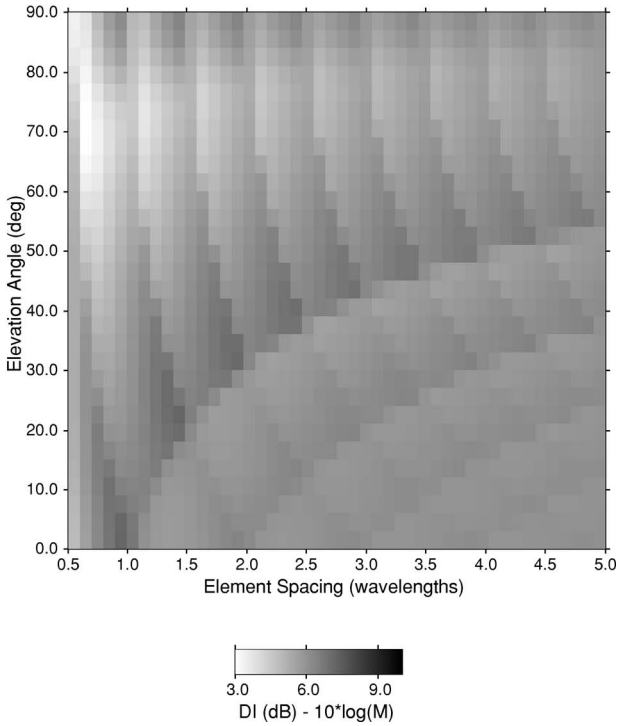
$$M \cdot E\{D(\phi,\omega)D^*(\phi,\omega)\}/S_p(\omega) = \underline{e}^H[[S]_{\text{single}} + [\Delta]]\underline{e}. \tag{32}$$

The replica (steering) vector $\underline{e}$ is defined as in Sec. III and

$$[S] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 + \eta_x & \eta_{xz} \\ 0 & \eta_{xz} & 1/3 + \eta_z \end{bmatrix} \tag{33}$$



(a)



(b)

FIG. 5. (a) The maximum directivity index minus $10\log_{10}(M)$ for a vector sensor line array as a function of angle with respect to broadside ("elevation angle") and element separation. The result is obtained by inverting the complete $3M\times 3M$ data cross spectral density matrix. All vector sensor components have the same intercomponent spacing. The value of M, the number of vector sensor elements, is equal to 11. (b) The maximum directivity index minus $10\log_{10}(M)$ for a vector sensor line array as in Fig. 5(a) but now with equal weighting of the components of the same type when first beamforming in elevation angle, i.e., $b_m=c_m=d_m=1/M$ in Eq. (17) in the text.

is the cross spectral density matrix in the form identical to that for a single vector sensor. The $[\Delta]$ matrix of spatial coherence contributions is

$$[\Delta]=\begin{bmatrix}\gamma_p-1 & 0 & \gamma_{pz}\\ 0 & (\gamma_x-1)/3 & 0\\ \gamma_{pz} & 0 & (\gamma_z-1)/3\end{bmatrix}. \tag{34}$$

The expressions for the γ terms are obtained from those involving the summations inside the brackets in Eq. (31), after multiplication by M. This form of the normalized beamformer output illustrates that under certain conditions, the $[\Delta]$ matrix can be considered as an additional source of self noise on the equivalent vector sensor channels that reduces the DI of the array when the gamma terms along the diagonal are greater than unity. However, for values of the gamma terms less than one, the directivity index actually can be greater than the result for uncorrelated sensors, as discussed above. In addition to its effect on the particle motion components, spatial correlation adds "noise" to the hydrophone channel and can introduce coherence between the pressure and $\hat{z}$ components.

The effects of channel self noise and spatial coherence between vector sensor components for an equal half-wavelength-spaced vector sensor array using this approximate approach are illustrated in Figs. 6 and 7. As in Figs. 5(a) and 5(b), these figures show plots of the directivity index after subtraction of $10\log_{10}(M)$ ($M=11$ in this case). Results of changes in the ratio of channel self noise to ocean noise for both conventional (equal-channel) weighting (Fig. 6) and for the channel weighting that maximizes DI (Fig. 7) are presented. For comparison, the corresponding single vector sensor results in Fig. 1 are replotted as the curves of connected asterisks in both figures. The adverse effect of spatial coherence at half-wavelength spacing for a plane wave arrival at broadside (the curves of connected circles in both figures) is less than that for an arrival at endfire (curves of connected plus signs). The degradation from the single vector sensor value is greatest in Fig. 7 with optimal weighting, increasing from slightly more than 1-dB degradation of maximum DI at broadside to a nearly 3-dB degradation at endfire for negligible self noise levels. Increases in channel self noise decrease the effects of spatial coherence.

Finally, the full $3M\times 3M$ matrix inversion approach was used to calculate the maximum directivity index as a function of angle with respect to broadside for various combinations of component spacings. The components of a given type (p, v_x, or v_z) were equally spaced, but the spacing for each of the three components was allowed to be different. The self noise levels were set to zero. The quasi-periodic nature of the optimal DI results in Figs. 5(a) and 5(b) suggests that the calculations do not need to be performed over an interval of intercomponent spacings greater than a wavelength or so, allowing the number of calculations to be kept to a manageable level. In addition, since the maximum DI approaches the fixed value of $10\log_{10}(M)+10\log_{10}(4)$ independent of elevation angle and spacing with increasing intercomponent spacing, only smaller intercomponent spacings need to be considered. Therefore, each component type was allowed to assume one of 12 equal-intercomponent spacings from 0.4 to 1.5 wavelengths in 0.1-wavelength steps. Calculations were performed for all $12^3=1728$ possible combina-

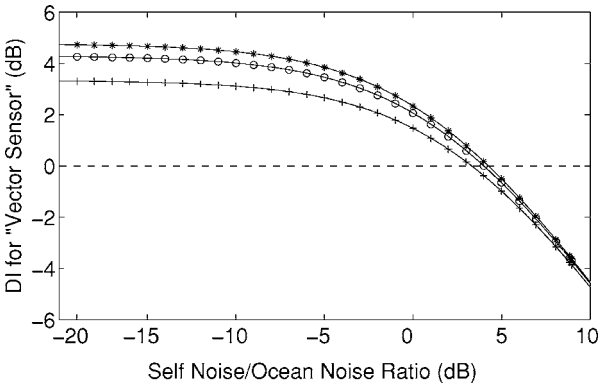


FIG. 6. The directivity index as a function of the particle motion self noise to ocean noise ratio for three arrays using equal channel weighting. The curve of connected circles is for a vector sensor line array having half-wavelength spacing for all components and a plane wave arrival at broadside. The curve of connected plus signs also is for an equal half-wavelength-spaced line array but now for an arrival at endfire. The total directivity index in these two cases is the sum of $10\log_{10}(M)$, where $M \equiv 11$, and the value given on the plot. For comparison, the curve of connected asterisks is for equal channel weighting with a single vector sensor and is the same as the curve of connected asterisks in Fig. 1.

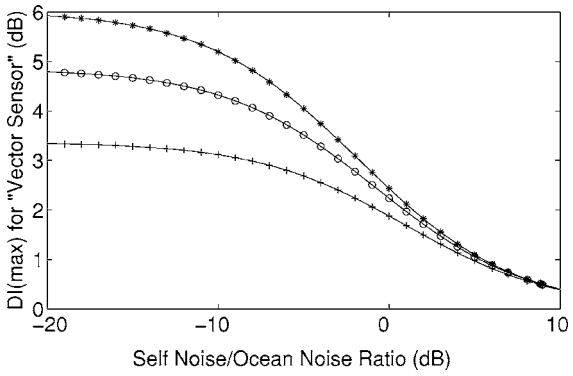


FIG. 7. The directivity index as a function of the particle motion self noise to ocean noise ratio for the same three cases as in Fig. 6, but now using the channel weighting that optimizes the directivity index. The curve of connected asterisks is the same as the curve of connected circles in Fig. 1.

tions and the minimum beamformer output variance for each combination was averaged over a specified elevation angle interval. Somewhat surprisingly, for an angular interval about broadside up to ±21°, the averaged output variance is minimized by making all intercomponent spacings equal to 0.9 wavelengths. In fact, up to angle intervals as wide as ±51° about broadside, the optimal spacings for each of the components are within ±0.2 wavelengths of 1 wavelength. Therefore, if detection of arrivals near broadside is of greatest interest, the components should all have the same spacing of about 1 wavelength. This optimal spacing takes advantage of the spatial coherence between components to improve detection; e.g., for an angular interval of ±12° about broadside, the equivalent directivity index exceeds the value of $10\log_{10}(M)+10\log_{10}(4)$ by more than 1 dB. In all cases up to angular intervals about broadside of nearly ±80°, the optimal intercomponent spacing for the pressure is equal to that of the $\hat{z}$ component of particle velocity.

For angular intervals about endfire, the result is quite different. That is, for angular intervals of almost all widths about endfire, the optimal spacings are the same as when averaging over the complete ±90° angular interval: 0.7 wavelengths for pressure, 0.5 wavelengths for v_x, and 1.4 wavelengths for the $\hat{z}$ component of particle velocity.

V. CONCLUSIONS

An alternative approach to the use of spatially distributed hydrophone arrays for determining the directional properties of a sound field is to measure acoustic pressure and its higher order spatial derivatives at a single point in space, as given by a Taylor series expansion of the field. Since acoustic particle velocity is proportional to the gradient of acoustic pressure at a given frequency, then an acoustic vector sensor is equivalent to a volumetric acoustic pressure array. Its equivalent physical aperture "adapts" to the temporal frequency of the field, resulting in a beam pattern that is independent of frequency. Therefore, all the various types of beamforming techniques developed for spatially distributed hydrophone arrays also can be applied to vector sensor data. The formulation can be placed in the more general context of matched field processing.

Given the equivalence between vector sensor measurements and those of a volumetric hydrophone array, vector acoustic intensity processing is equivalent to multiplicative processing with spatially distributed hydrophone array data. It has been known for nearly a half-century that additive beamforming techniques, rather than multiplicative or other nonlinear processing methods, are optimal for detection of weak signals in noise. However, multiplicative array processing methods can be very useful in parameter estimation problems such as source localization when excess signal-to-noise ratio is available to the processor.

The great advantage of the vector sensor approach is that the directionality of a sound field can be determined with a sensor that is a small fraction of an acoustic wavelength in size. This advantage allows, for example, the left/right ambiguity of a line array of omnidirectional sensors to be broken. The major disadvantage is the sensitivity of these types of sensors to contamination from motion-induced and other forms of nonacoustic self noise on the particle motion channels. This nonacoustic self noise can have a significant adverse impact on detection performance of a vector sensor or an array of vector sensor components. A simple analytical approach to maximizing the directivity index (equivalent to minimum variance beamforming with unity gain in the look direction in a field composed of a single plane wave in homogeneous, isotropic noise) of a single vector sensor shows that detection performance is optimized by always including the particle motion channels in the processing regardless of the level of self noise (except in the limit of infinite self noise). This performance is achieved by incorporating the level of channel self noise with respect to the ocean acoustic noise into the determination of the channel weighting used in beamforming.

Improvements in vector sensors and vector sensor array design depend upon evaluation of the degree of motion-induced contamination and other types of nonacoustic self noise in at-sea measurements. The fact that this noise does

not obey the same laws of physics as acoustic fields can be used to distinguish it from ocean acoustic noise contributions. In particular, the vector properties of vector acoustic intensity, both the active and reactive components, describe relationships between quantities in the first row (column) and the 3×3 particle motion submatrix in the single vector sensor data cross spectral density matrix. Deviations of actual ocean measurements made by well-calibrated sensors from the predictions made by these relationships can be associated with nonacoustic contributions.

Another consequence of the equivalence between vector sensor measurements and those of a volumetric hydrophone array is that all the various types of array processing methods developed for hydrophone arrays also are applicable to vector sensor data. Ocean acoustic data collected by a freely drifting, neutrally buoyant, vector sensor in the northeast Pacific Ocean are used to illustrate the results of applying conventional beamforming, minimum variance distortionless response processing, and white-noise-constrained adaptive beamforming techniques to the estimation of the azimuth of arrival from a 14-Hz controlled source at 1700-km range. An increase in beamformer adaptivity provides an increase in spatial resolution, but a decrease in beamformer output and possible introduction of estimation bias due to the effects of mismatch.

Analytical expressions for the spatial coherence between the various vector sensor components in a homogeneous, isotropic noise field show that significant spatial coherence exists between pairs of particle motion components at half-wavelength spacing. In addition to the zeros of these spatial coherence functions not being located at half-wavelength spacing as they are for a pair of acoustic pressure sensors, they are not spaced periodically. The maximum directivity index for a four-component vector sensor array of M elements approaches the value of $10\log_{10}(M)+10\log_{10}(4)$ for intercomponent spacings greater than a few wavelengths because of decreasing spatial coherence between components. This value in the case of mutual incoherence between all components equals ten times the logarithm of the number of data channels, exactly the same result as with an array of hydrophone elements. At smaller intercomponent spacings, the spatial coherence can increase the directivity index of a line array of vector sensor components to values greater than $10\log_{10}(M)+10\log_{10}(4)$ over large angular intervals (approximately ±60° about broadside and ±75° about endfire). For most angular intervals about broadside, the directivity index is maximized by spacing each of the vector sensor components at equal intervals about 1 wavelength apart. In contrast, maximizing the directivity index over angular intervals about endfire requires spacing each of the vector sensor components by a different amount: 0.7 wavelength spacing for the pressure component, 0.5 wavelengths for the particle velocity component perpendicular to the direction of separation between components, and 1.4 wavelengths for the particle motion component parallel to the direction of separation.

ACKNOWLEDGMENTS

We have had fruitful discussions with Harry Cox, Mike Traweek, Roger Richards, Ben Cray, Bruce Abraham, Clay Shipps, John Polcari, and Vladimir Shchurov. Helpful suggestions were provided by John P. Ianniello of SAIC after reading a first draft of this paper. The single vector sensor discussed in Sec. III E was designed and built at the Marine Physical Lab primarily by Greg Edmonds, Victor Anderson, and Fred Spiess, with assistance from Richard L. Culver and Marvin Darling. This work was supported by the Office of Naval Research (ONR) and the Office of Naval Technology. Recent support from ONR Code 321(OA) is gratefully acknowledged.

APPENDIX: ORDER OF BEAMFORMING

In Sec. IV, beamforming with the line array of vector sensors is first done over the angle with respect to broadside (elevation angle, ϕ) for each type of vector sensor component separately [Eq. (17)]. The array element weights are set equal to $1/M$ to be consistent with maximizing DI for a line array of identical, equally spaced components. The resulting beamformed components for a given elevation angle, $p(\phi)$, $v_x(\phi)$, and $v_z(\phi)$, then are used to determine the optimum equivalent vector sensor weights with an approach identical to single vector sensor processing. An alternative method is to process the whole set of $3M$ data channels in one step. The purpose of this Appendix is to demonstrate the relationship between these two approaches.

The first step of beamforming in elevation angle each vector sensor component separately and then collecting the results into a single-vector-sensor equivalent data vector can be written as

$$\begin{aligned}\underline{d}^H(\phi)_{1\times3} &\equiv [\underline{r}_p^H(\phi,z)\underline{p}(z),\underline{r}_x^H(\phi,z)\underline{v}_x(z),\underline{r}_z^H(\phi,z)\underline{v}_z(z)]\\ &= [p(\phi),v_x(\phi),v_z(\phi)].\end{aligned}\tag{A1}$$

The quantities p, v_x, and v_z are the $M\times1$ complex data vectors at a given frequency for pressure and the two components of particle velocity, respectively. The replica vector for each component is allowed to be different. The 3×3 cross spectral density matrix for a given elevation angle is formed by the outer product, dd^H, and can be written as

$$\begin{aligned}[S]_{3\times3}(\phi_i) &= \underline{d}(\phi_i)\underline{d}^H(\phi_i)\\ &= \begin{bmatrix}\underline{r}_p^H(\phi_i) & \underline{0} & \underline{0}\\ \underline{0} & \underline{r}_x^H(\phi_i) & \underline{0}\\ \underline{0} & \underline{0} & \underline{r}_z^H(\phi_i)\end{bmatrix}\\ &\quad\times\begin{bmatrix}\underline{p}\underline{p}^H & \underline{p}\underline{v}_x^H & \underline{p}\underline{v}_z^H\\ \underline{v}_x\underline{p}^H & \underline{v}_x\underline{v}_x^H & \underline{v}_x\underline{v}_z^H\\ \underline{v}_z\underline{p}^H & \underline{v}_z\underline{v}_x^H & \underline{v}_z\underline{v}_z^H\end{bmatrix}\\ &\quad\times\begin{bmatrix}\underline{r}_p(\phi_i) & \underline{0} & \underline{0}\\ \underline{0} & \underline{r}_x(\phi_i) & \underline{0}\\ \underline{0} & \underline{0} & \underline{r}_z^H(\phi_i)\end{bmatrix}\\ &\equiv [R]_{3\times3M}^H[C]_{3M\times3M}[R]_{3M\times3}.\end{aligned}\tag{A2}$$

Vector sensor processing then is performed as specified in Eq. (6), i.e.,

$$\underline{e}^H[S]\underline{e} = \underline{e}^H[[R]^H[C][R]]\underline{e} = [[R]\underline{e}]^H[C][[R]\underline{e}] \qquad \text{(A3)}$$

with $\underline{e}$ as defined in Sec. III B. Equation (A3) is precisely the same equation as for the alternative method of processing the whole set of $3M$ data channels in one step, where $[C]$ is the full array $3M \times 3M$ data cross spectral density matrix and the $3M \times 1$ replica vector is $[[R]\underline{e}]$. Therefore, these two approaches are identical when conventional processing is used. For minimum variance beamforming, the element weights are determined by inverting the matrix $[S]$ in the two-step method, whereas they are determined by the inverse of $[C]$ in the one-step full-array approach.

Abramowitz, M., and Stegun, I. A. (**1964**). *Handbook of Mathematical Functions*, (Dover, New York).

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data: Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York).

Brown, J. L., and Rowlands, R. O. (**1959**). "Design of directional arrays," J. Acoust. Soc. Am. **31**, 1638–1643.

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

Cox, H. (**1973a**). "Resolving power and sensitivity to mismatch of optimum array processors," J. Acoust. Soc. Am. **54**, 771–785.

Cox, H., and Baggeroer, A. B. (**2003**). "Performance of vector sensors in noise," J. Acoust. Soc. Am. **114**, pt. 2 2426.

Cox, H., Zeskind, R. M., and Owen, M. M. (**1987**). "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**(10), 1365–1376.

Cray, B. A., and Nuttall, A. H. (**2001**). "Directivity factors for linear arrays of velocity sensors," J. Acoust. Soc. Am. **110**, 324–331.

D'Spain, G. L. (**1990**). "Energetics of the Ocean's Infrasonic Sound Field," Ph.D. thesis, University of California, San Diego.

D'Spain, G. L. (**1999**). "Polarization of acoustic particle motion in the ocean and its relation to vector acoustic intensity," Proc. of the 2nd International Workshop on Acoustical Engineering and Technology, Harbin, China, pp. 149–164.

D'Spain, G. L., Hodgkiss, W. S., and Edmonds, G. L. (**1991**). "Energetics of the deep ocean's infrasonic sound field," J. Acoust. Soc. Am. **89**, 1134–1158.

D'Spain, G. L., Hodgkiss, W. S., Edmonds, G. L., Nickles, J. C., Fisher, F. H., and Harriss, R. A. (**1992**). "Initial analysis of the data from the vertical DIFAR array," Oceans '92, Newport, RI, pp. 346–351.

Eckart, C. (**1953**). "The theory of noise in continuous media," J. Acoust. Soc. Am. **25**, 195–199.

Gramann, R. A. (**1992**). "ABF algorithms implemented at ARL-UT," ARL Tech. Letter ARL-TL-EV-92-31, Applied Research Laboratories, The University of Texas at Austin, Austin, TX.

Greene, C. R., McLennan, M. W., Norman, R. G., McDonald, T. L., Jakubczak, R. S., and Richardson, W. J. (**2004**). "Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration," J. Acoust. Soc. Am. **116**, 799–813.

Kuperman, W. A., and Ingenito, F. (**1980**). "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996.

Mann, J. A., Tichy, J., and Romano, A. J. (**1987**). "Instantaneous and time-averaged energy transfer in acoustic fields," J. Acoust. Soc. Am. **82**, 17–30.

McDonald, M. A. (**2004**). "DIFAR hydrophone usage in whale research," Can. Acoust. **32**(2), 155–160.

Oltman-Shay, J., and Guza, R. T. (**1984**). "A data-adaptive ocean wave directional spectrum estimator for pitch and roll type measurements," J. Phys. Oceanogr. **14**, 1800–1810.

Pritchard, R. L. (**1954**). "Maximum directivity index of a linear point array," J. Acoust. Soc. Am. **26**, 1034–1039.

Richards, R. (**2005**). "Acoustic vector sensor line arrays," ONR Joint Review of Unmanned Systems Technology Development, Panama City Beach, Fla., 31 January–4 February.

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**, 1286–1300.

Traweek, C. M. (**2003**). "Optimal Spatial Filtering for Design of a Conformal Velocity Sonar Array," Ph.D. thesis, The Pennsylvania State University.

Traweek, C. M. (**2004**). "A collaborative roadmap for vector sensor towed arrays enabled by piezocrystal materials," ONR powerpoint presentation, Office of Naval Research.

Urick, R. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).

Van Trees, H. L. (**1968**). *Detection, Estimation, and Modulation Theory, Part 1*, (Wiley, New York).

Waterhouse, R. V., and Feit, D. (**1986**). "Equal-energy streamlines," J. Acoust. Soc. Am. **80**, 681–684.

Wilson, O. B., Wolf, S. N., and Ingenito, F. (**1985**). "Measurements of acoustic ambient noise in shallow water due to breaking surf," J. Acoust. Soc. Am. **78**, 190–195.

Wilcoxon Research (**2004**). "The vector sensor," specification sheet, www-.wilcoxon.com.

Wolf, G. W. (**1998**). "U.S. Navy sonobuoys—Key to antisubmarine warfare," Sea Technol. **39**, 41–44.

Design and performance of a microprobe attachment for a $\frac{1}{2}$-in. microphone[a)]

Gilles A. Daigle[b)] and Michael R. Stinson
Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

(Received 18 July 2005; revised 28 April 2006; accepted 7 May 2006)

It is often necessary to measure sound fields in confined spaces where minimum disturbance of the sound field is important. In applications where the confinement of the space is extreme such as very small cavities, existing probe microphones are too large. In this paper the probe section of a Brüel and Kjær probe microphone type 4170 is redesigned using smaller diameter probe tubes. New designs (microprobes) are first considered through simulations in the case of probe tubes having inside diameters ranging from 1.25 down to 0.1 mm. Several microprobes were constructed and their performance measured and compared to the simulated results. Good agreement was found between the measured and simulated frequency response and sensitivity. Measurements show that a sensitivity of about 0.02 mV/Pa could be obtained from a probe tube with 0.2 mm o.d. and 0.1 mm i.d. if the length is less than 1 cm. The input impedance of the probe orifice is estimated to be greater than 3×10^{10} kg s^{-1} m^{-4}. Viscous resistance and thermal conduction due to the small diameter provides a well-behaved frequency response that is flat within 5 dB between 200 and 6000 Hz and to within 12 dB up to 10 000 Hz.
[DOI: 10.1121/1.2208454]

PACS number(s): 43.38.Kb, 43.58.Vb [AJZ] Pages: 186–191

I. INTRODUCTION

Probe microphones have been available for many years and they are typically used to measure sound pressure levels in spaces in which the size of a standard microphone would interfere with the sound field. For example, Shaw[1] designed a probe microphone with horn coupling in order to measure the sound field within external ears.[2] The Brüel and Kjær (B&K) 4170 horn-coupled probe microphone[3] employs the principles developed by Shaw. In the B&K 4170 probe, the protective cap of a $\frac{1}{2}$-in. condenser microphone is removed and the microphone is mounted into a specially designed housing. The probe tube is coupled to a cavity in front of the microphone diaphragm via an exponential horn. In order to avoid standing wave formation in the high frequency range, the probe orifice is damped by a fine wire mesh. Matching at the larger end of the horn is ensured by a resistance which is vented into a cavity, coupled in turn to a larger cavity in order to equalize the frequency response. A frequency response of 30 Hz–8 kHz within 4 dB is thereby obtained. The probe tube of the horn-coupled microphone causes the sensitivity to be about 20 dB lower than the sensitivity of the $\frac{1}{2}$-in. condenser microphone (typically about 1.5 mV/Pa). Further, for measurements in small cavities the probe orifice impedance must be high enough to not noticeably modify the measured sound field (the acoustical impedance of the probe orifice is greater than 10^9 kg s^{-1} m^{-4}). More recently B&K introduced the 4182 probe microphone.[4] The B&K 4182 probe uses a $\frac{1}{4}$-in. condenser microphone and a different impedance matching tube to provide the microphone with its characteristic smooth response. The sensitivity of the B&K 4182 probe is typically about 3.16 mV/Pa and the probe orifice impedance is approximately 8×10^8 kg s^{-1} m^{-4}.

The probe end of both the B&K 4170 probe and 4182 probe has a diameter of about 1.25 mm. Applications can arise that require sound pressure level measurements in very restricted physical space, such as small cavities a few millimeters in size, where even a 1.25 mm probe diameter would interfere with the sound field. An example would be the rapidly varying sound field within an ear canal.[5] In particular, our application involves fitting model ear canals with hearing aids and measuring the sound field as close as possible along the surface of the hearing aid and down the ear canal. In order to measure the detailed variations of the sound field, especially around the receiver and at the higher frequencies, the existing B&K probe must be redesigned to use the smallest possible diameter probe tubes.

There have been relatively few papers published on alternative probe-tube designs. Copeland and Hill[6] designed an adapter for a 1-in. microphone. Franzoni and Elliott[7] discuss designs to build probe-tubes with a smooth transfer function without introducing damping material. However, as far as we are aware, there has been no work to reduce significantly the size of the probe-tube entrance. In this paper, the horn attachment of the B&K 4170 is redesigned using very small diameter (<1 mm) stainless steel tubing (microprobes). The goal is to design a probe with the smallest possible diameter while retaining a reasonably flat frequency response and a usable signal-to-noise ratio in order to make accurate measurements within model ear canals.

[a)]Portions of this work were presented at the 148th meeting of the Acoustical Society of America in San Diego, CA.
[b)]Corresponding author; electronic mail: gilles.daigle@nrc-cnrc.gc.ca

TABLE I. The inside diameters (i.d.) and outside diameters (o.d.) of the Stainless steel tubing.

i.d. (mm)	o.d. (mm)
0.38	0.51
0.15	0.25
0.10	0.20

II. DESCRIPTION OF THE NEW DESIGNS

Stainless steel tubing is commercially available in a variety of very small inside and outside diameters. The dimensions for tubing considered in this paper are shown in Table I. The sketches in Fig. 1 show how the small tubing was used to design the three microprobes that were studied in this paper. In Fig. 1(a), the damping material is removed from the end of the existing B&K probe and replaced by a small-diameter tubing of length ℓ_1. Because of its small inside diameter, the microprobe will possess some inherent damping due to air viscosity and thermal conductivity. The two other options replace the B&K probe entirely by using a series of stainless steel tubing of decreasing diameter, as illustrated in Figs. 1(b) and 1(c). The larger tubing at the extreme right has the proper diameter to couple the probe to the cavity in front of the microphone diaphragm. The two-stage design is the simplest, but it was anticipated that the three-stage design might result in better acoustical coupling.

III. THEORETICAL MODELING

If S_0 is the sensitivity of the $\frac{1}{2}$-in. condenser microphone, then with the probe tube in place there is an overall sensitivity,

$$S = S_0 p_L / p \tag{1}$$

where p is the pressure at the tip of the probe and p_L is the pressure at the condenser microphone. It is thus necessary to model the propagation of the sound from the tip of the probe to the front of the condenser microphone.

A. Existing B&K probe

The propagation along the existing B&K probe is modeled using the Webster horn equation[8] in the case of a horn

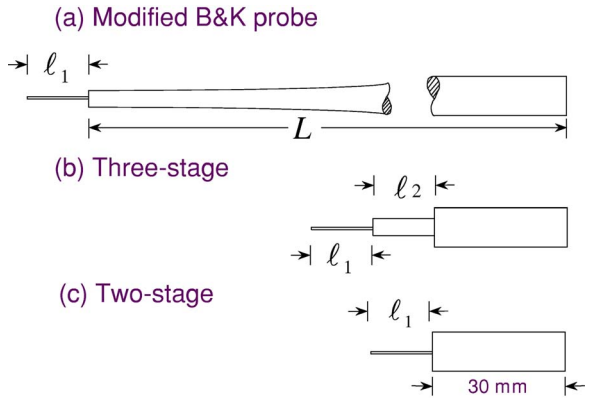


FIG. 1. (Color online) Sketch showing the three new designs considered in this paper.

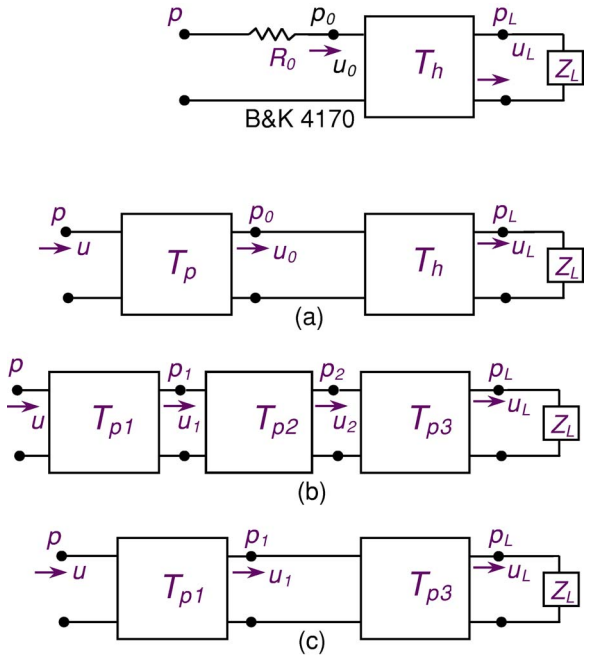


FIG. 2. (Color online) Sketch showing the schematic representation of the probe designs, existing B&K 4170, (a) modified B&K probe, (b) three-stage design, and (c) two-stage design.

with constant flare.[9,10] If d_0 is the diameter of the B&K probe tip, L the length of the exponential horn, and d_t the diameter of the throat at the condenser microphone, then the flare constant m is

$$m = \frac{\ln(d_t/d_0)}{L}. \tag{2}$$

The schematic at the top of Fig. 2 (labeled B&K 4170) represents the B&K probe where u_0 is the volume velocity at the tip of the probe, R_0 is the resistance of the wire mesh at the tip, T_h is the transfer matrix representing the propagation along the probe, u_L is the volume velocity at the condenser microphone, and Z_L is the load resistance. The transfer matrix T_h is

$$\begin{pmatrix} p_0 \\ u_0 \end{pmatrix} = \begin{pmatrix} T_{h11} & T_{h12} \\ T_{h21} & T_{h22} \end{pmatrix} \begin{pmatrix} p_L \\ u_L \end{pmatrix}, \tag{3}$$

where

$$T_{h11} = e^{mL}\left(\cos \beta L - \frac{m}{\beta} \sin \beta L\right), \tag{4a}$$

$$T_{h12} = \frac{i\rho\omega e^{-mL}}{\beta A_0} \sin \beta L, \tag{4b}$$

$$T_{h21} = \frac{iA_0 k^2 e^{mL}}{\rho\omega\beta} \sin \beta L, \tag{4c}$$

$$T_{h22} = e^{-mL}\left(\cos \beta L + \frac{m}{\beta} \sin \beta L\right). \tag{4d}$$

In Eq. (4), $\beta = \sqrt{k^2 - m^2}$ and $A_0 = \pi(d_0/2)^2$ where k is the wave number ω/c, where ω is the angular frequency and c is the speed of sound in the tube.

Actual measurements on several B&K probes suggest that they are better modeled as a compound exponential horn with two sections. The first section has length $L_a=7.5$ cm and the intermediate diameter is $d_a=0.154$ cm. If T_a is the transfer matrix of the first section and T_b the transfer matrix of the remaining second section, then $T_h=T_aT_b$. Finally, relating the pressure p at the tip of the probe,

$$p = p_0 + u_0 R_0 \tag{5}$$

and noting that $u_L=p_L/Z_L$, the ratio p_L/p required in Eq. (1) is obtained from Eqs. (3) and (5):

$$p_L/p = \frac{1}{(T_{h11} + T_{h12}/Z_L) + R_0(T_{h21} + T_{h22}/Z_L)}. \tag{6}$$

The B&K instruction manual[3] for the probe microphone gives a detailed cross-sectional view of the 4170. Thus it is possible to construct a simple equivalent circuit representation of the 4170 in order to estimate Z_L. The circuit is then optimized to find the load resistance Z_L that produces frequency response curves that match published curves[3] for the 4170.

B. Modified B&K probe

The sketch in Fig. 2(a) is a schematic representing the probe design shown in Fig. 1(a). The fine wire mesh is removed from the B&K probe and a length of small circular tubing (microprobe) is attached. The propagation along the microprobe is represented by the transfer matrix T_p. Thus $R_0=0$, but because of the small inside diameter of the microprobe the transfer matrix T_p includes some inherent damping due to viscosity and thermal conduction.

The propagation of sound in a uniform, circular tube is a fundamental problem that has been well studied in acoustics. In particular, Zwikker and Kosten[11] have introduced an approximate treatment where the effects of viscosity and thermal conductivity are treated separately.[12] Their formulation is used to model the propagation along the microprobes using the transmission line approach. Thus, if u is the volume velocity at the tip of the microprobe, then the transfer matrix T_p is

$$\begin{pmatrix} p \\ u \end{pmatrix} = \begin{pmatrix} T_{p11} & T_{p12} \\ T_{p21} & T_{p22} \end{pmatrix} \begin{pmatrix} p_0 \\ u_0 \end{pmatrix}, \tag{7}$$

where

$$T_{p11} = \cosh(\Gamma \ell), \tag{8a}$$

$$T_{p12} = Z \sinh(\Gamma \ell), \tag{8b}$$

$$T_{p21} = \sinh(\Gamma \ell)/Z, \tag{8c}$$

$$T_{p22} = \cosh(\Gamma \ell). \tag{8d}$$

In Eq. (8), $\ell=\ell_1$ is the length of the microtubing (see Fig. 1) and

$$\Gamma = \frac{i\omega}{c}\left[\frac{F_\alpha}{F_\beta}\right]^{1/2}, \tag{9}$$

$$Z = \frac{\rho c}{\pi a^2}\frac{1}{[F_\alpha F_\beta]^{1/2}}, \tag{10}$$

where $a=a_1$ is the inside diameter of the microtubing and

$$F_\alpha = 1 + \frac{2(\gamma-1)J_1(\alpha a)}{\alpha a J_0(\alpha a)}, \tag{11a}$$

$$F_\beta = 1 - \frac{2J_1(\beta a)}{\beta a J_0(\beta a)}. \tag{11b}$$

In Eq. (11)

$$\alpha = \left(-\frac{i\omega\rho N_{\mathrm{Pr}}}{\mu}\right)^{1/2}, \tag{12}$$

$$\beta = \left(-\frac{i\omega\rho}{\mu}\right), \tag{13}$$

where ρ is the density, ω is the angular frequency, μ is the coefficient of viscosity, and N_{Pr} is the Prandtl number. Finally, the ratio p_L/p is easily obtained from Eqs. (3) and (7).

C. Two- and three-stage designs

The schematic representing the three-stage design sketched in Fig. 1(b) is shown in Fig. 2(b). The transfer matrices T_{p1}, T_{p2}, and T_{p3} represent the propagation along the three sections of tubing, respectively. Thus,

$$\begin{pmatrix} p \\ u \end{pmatrix} = T_{p1}\begin{pmatrix} p_1 \\ u_1 \end{pmatrix}, \tag{14}$$

where the elements of the matrix T_{p1} are calculated using Eqs. (8)–(13) with $\ell=\ell_1$ and $a=a_1$.

The transfer matrix representing the propagation through the middle section is

$$\begin{pmatrix} p_1 \\ u_1 \end{pmatrix} = T_{p2}\begin{pmatrix} p_2 \\ u_2 \end{pmatrix}, \tag{15}$$

where the elements of the matrix T_{p2} are also calculated using Eqs. (8)–(13) but with $\ell=\ell_2$ and $a=a_2$ where a_2 is the inside diameter of the middle section of tubing. Finally, the transfer matrix representing the propagation along the final section of tubing is

$$\begin{pmatrix} p_2 \\ u_2 \end{pmatrix} = T_{p3}\begin{pmatrix} p_L \\ u_L \end{pmatrix}, \tag{16}$$

with $\ell=30$ mm in Eq. (8) and $a=3.27$ mm in Eqs. (10) and (11). Finally, the ratio p_L/p is obtained from Eqs. (14) to (16).

In the case of the two-stage design in Figs. 1(c) and 2(c), the transfer matrix in Eq. (14) and the following matrix,

$$\begin{pmatrix} p_1 \\ u_1 \end{pmatrix} = T_{p3}\begin{pmatrix} p_L \\ u_L \end{pmatrix}, \tag{17}$$

are used to obtain the ratio p_L/p.

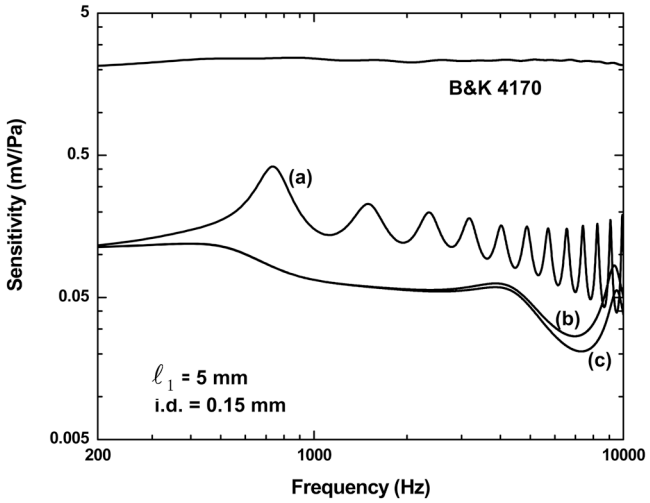


FIG. 3. Comparison of the predicted sensitivities of the three new designs and with that of the existing B&K 4170; (a) modified B&K probe, (b) three-stage design, and (c) two-stage design.

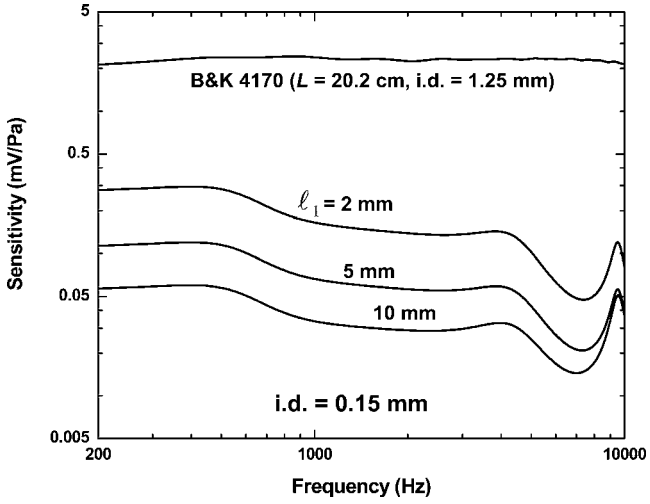


FIG. 4. The predicted sensitivity of the microprobe as a function of microprobe length.

IV. CALCULATED SENSITIVITY

In Fig. 3, curves (a)–(c) are the sensitivities calculated in the case of the new designs shown in Figs. 1(a) and 1(c), respectively. The calculated sensitivity of the unmodified B&K probe is also shown for comparison. In all three cases, the microprobe is assumed to have a length ℓ_1=5 mm and an i.d.=0.15 mm. In the case of curve (b), ℓ_2=5 mm and i.d.=1.25 mm. It was initially believed that maintaining the use of the exponential B&K horn would provide the best sensitivity. Although this did turn out to be case, it was not possible to dampen the observed standing waves. In the original B&K probe, the damping material is optimized to dampen the standing waves in the frequency range of interest. Tuning the length and i.d. of the microprobe in order to achieve the same optimization led to geometries that were not practical. It was also believed that the three-stage design might provide better sensitivity than the two-stage design. The prediction shows that there is only a marginal improvement at the higher frequencies. Further, the smaller i.d. of the microprobe is expected to increase the orifice impedance. Calculations show that the probe orifice impedance of the two-stage design of curve (c) is comparable to the B&K 4170. Therefore the two-stage design shown in Fig. 1(c) was retained for further analysis.

In order to measure correctly, a probe microphone must be more sensitive to sound pressures at the tip of the probe than it is to unwanted sound transmission through the probe housing and tube walls. This means that the sensitivity with the probe tube open must be greater than the acoustical signal measured when the tube is blocked. Published curves[3] show that the sensitivity of the 4170 with a closed probe orifice is 30–40 dB below the sensitivity when the orifice is open. The sensitivity of curve (c) in Fig. 3 is 25–35 dB below the sensitivity of the 4170 between 200 and 6000 Hz (and 40 dB around 8000 Hz). Thus, *if* the unwanted transmission of sound through our microprobe walls and housing is comparable to that of the 4170 probe then our reduced sensitivity leaves only about 5 dB between the sound from the orifice and that from the housing and microtube walls. This would be marginal for most applications. In our intended application, however, measurements will be made within a sealed cavity by inserting the probe into the cavity through the appropriate size holes, so the main transmission of unwanted sound would be expected through the microprobe walls and not from transmission through the housing. This issue is addressed experimentally in the next section.

The curves in Fig. 4 show the predicted changes in sensitivity as a function of the microprobe length ℓ_1 in the case of an i.d.=0.15 mm. The sensitivity of the 10 mm microprobe is 6 dB lower than the 5 mm probe at the lower frequencies and nearly the same at 10 kHz.

The curves in Fig. 5 show the changes in sensitivity as a function of the i.d. for a microprobe length of 5 mm. The acoustical input impedance for these probes is given in Table II along with the values for the two existing B&K probes.[3,4] The sensitivity of the microprobe with an i.d. of 0.1 mm is predicted to be about 40 dB below the sensitivity of the B&K 4170 in the midfrequency range. However, even

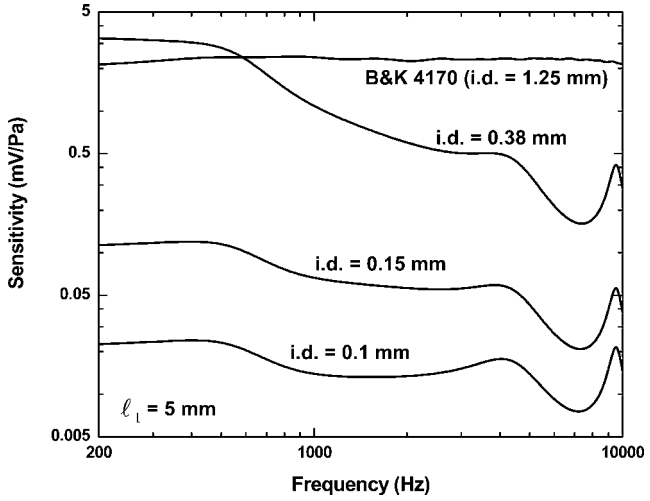


FIG. 5. The predicted sensitivity of the micrcprobe as a function of microprobe i.d.

TABLE II. The acoustical input impedance of the two-stage microprobes and the existing B&K probes.

Probe	Impedance (kg s^{-1} m^{-4})
i.d. 0.38 mm	$>2\times10^{8}$
i.d. 0.15 mm	$>7\times10^{9}$
i.d. 0.10 mm	$>3\times10^{10}$
B&K 4170	$>1\times10^{9}$
B&K 4182	8×10^{8} (approx.)

though insufficient signal-to-noise is expected from the smallest microprobe, the smaller diameter would allow measurements with a finer resolution within a model ear canal and the smallest microprobe will be included in the experimental investigation. Further, the smallest microprobe is predicted to have an acoustical input impedance that is an order of magnitude greater than the B&K 4170.

V. MEASURED SENSITIVITY

In order to determine the sensitivity limit experimentally, three microprobes were built from stock stainless steel tubing with dimensions specified in Table I. In all three cases $\ell_1=5$ mm. The detail of the construction is shown in Fig. 6(a). The microprobe was offset to provide the greatest flexibility when using the microprobe for measurements in model ear canals. The details on the right-hand side allow the proper coupling to the B&K 4170 housing. Figure 6(b) shows the microprobe mounted in the B&K housing. The smallest assembled microprobe is shown in the photograph in Fig. 7 mounted in the B&K 4170 housing (bottom right). Also shown for comparison are the probe end of the original B&K 4170 with its protection tube (top) and a conventional $\frac{1}{2}$-in. condenser microphone (bottom left).

The measured frequency responses of the three microprobes are shown in Fig. 8. The measurements were carried out in a small cavity driven with a miniature receiver. The vertical axis is arbitrary and only relative differences are relevant. For comparison, a measurement was also made using the existing B&K 4170. Since the frequency response of the B&K 4170 is flat, the curves include the frequency response

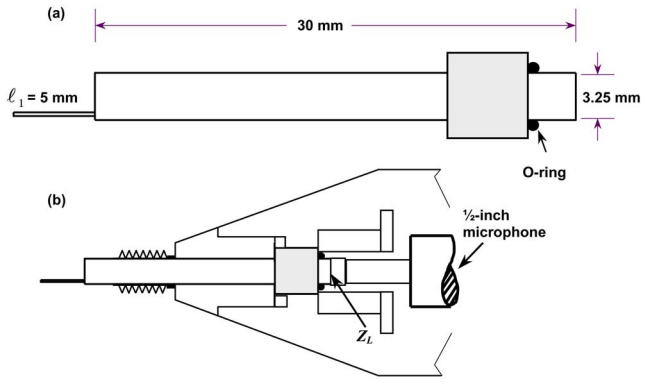


FIG. 6. (Color online) Sketch showing the details of the constructed microprobe. (a) Implementation details of the two-stage design. The relative dimensions are drawn to scale, (b) The two-stage microprobe shown mounted in the front end of the B&K 4170 housing. The quantity Z_L is the load resistance at the condenser microphone.

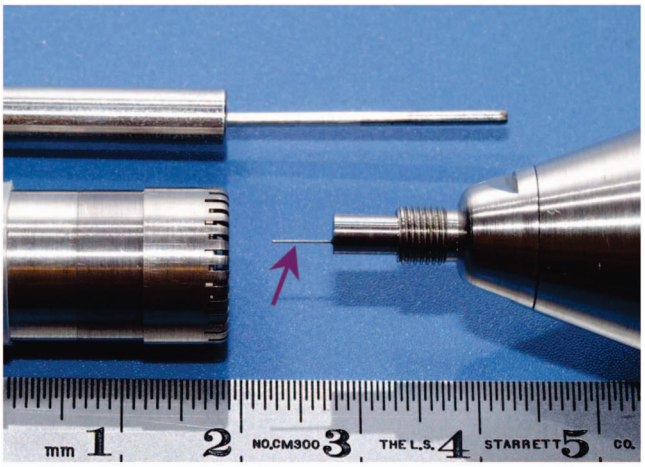


FIG. 7. The new microprobe mounted in the B&K housing (right); the probe end of the existing B&K horn attachment with its protection tube (top); and a $\frac{1}{2}$-in. condenser microphone (left).

of the receiver and cavity used for the measurements. Note that the measured responses of the two smallest microprobes show traces of noise at the lower and higher frequencies. The rms voltage applied to the miniature receiver was 35.5 mV. When the applied voltage is increased, the noise disappears.

In order to provide a direct comparison with the calculated values in Fig. 5, the measured sensitivity of the three microprobes is shown in Fig. 9. The measured sensitivity is obtained by removing the response of the receiver and cavity from the measured responses shown in Fig. 8 using the known response of the B&K 4170. There is reasonable agreement between the calculated and measured sensitivities. The measured sensitivity of the smallest microprobe (i.d. =0.10 mm) is greater than expected. The most likely explanation is that the diameter of the stock tubing used to fabricate the smallest microprobe was slightly larger than specified. Calculations indicate that the probe orifice impedance of the smallest microprobe is expected to be greater than 10^{10} kg s^{-1} m^{-4}.

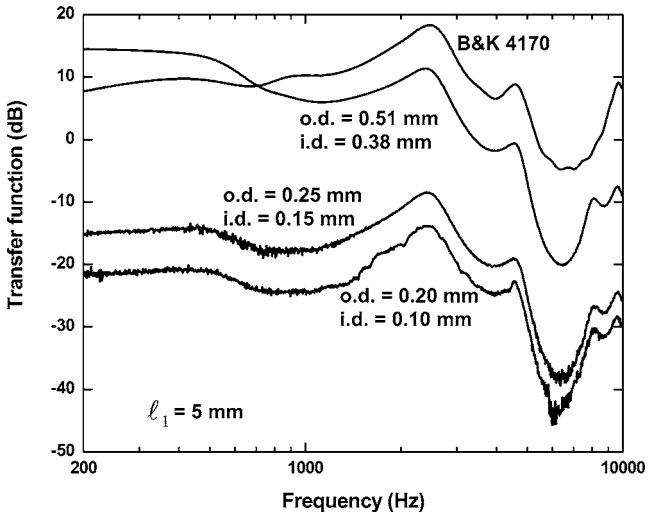


FIG. 8. Measured frequency responses of three microprobes of different sizes. Also shown is the measured frequency response of the B&K 4170.

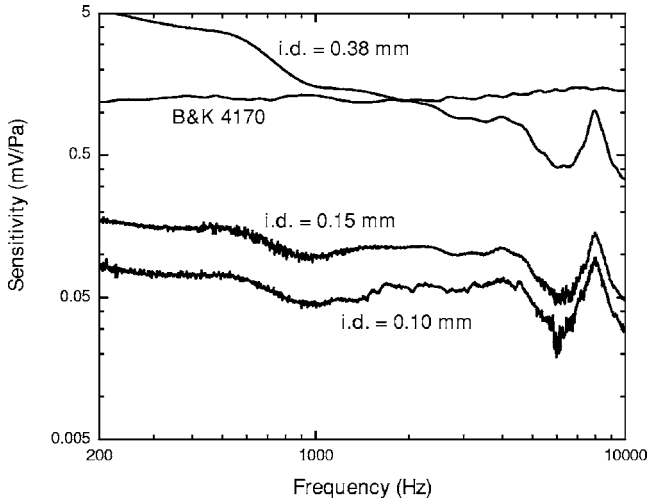


FIG. 9. Measured sensitivity of the three microprobes. The sensitivity is obtained by removing the response of the receiver and cavity from the measured frequency responses shown in Fig. 8 using the known response of the B&K 4170.

In total, six of the smallest microprobes were constructed and one was intentionally blocked. The top curves in Fig. 10 show the measured responses of the five working microprobes. The applied voltage was 140 mV and generated a sound pressure level of 80 dB at 1 kHz within the cavity at the probe tip (note the cleaner response and that here, the cavity has an open vent and the low frequency response was modified). The measured responses of the five working microprobes are indistinguishable from each other. The lower darker curve labeled "blocked" in Fig. 10 is the response of the blocked probe, giving an estimate of the inherent acoustic noise floor of the system. Since doubling the length of the microprobe is expected to lower the sensitivity by 6 dB (see Fig. 4), it can be inferred that the length of the smallest microprobe could be increased up to 10 mm. From Fig. 9, it can be inferred that the longer microprobe would have a measured sensitivity of about 0.02 mV/Pa in the midfrequency range. When the entire probe is acoustically sealed (the probe is insulated and inserted into a brass pipe; the pipe is wrapped with heavy felt) the lower lighter curve labeled "sealed" is measured, giving an estimate of the inherent electrical noise floor of the system. The lower lighter curve is in agreement with the published noise data for the equivalent SPL (55 dB re 2×10^{-5} Pa)[3] of the B&K 4170.

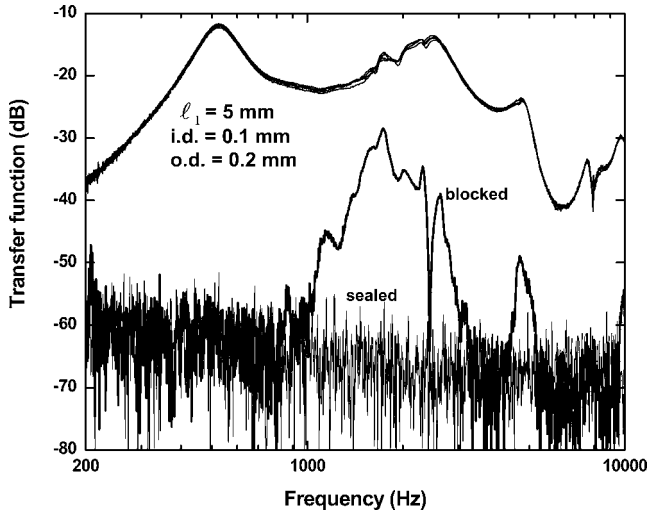


FIG. 10. Measured frequency responses of five identical microprobe (top curves). Acoustical noise floor of the system with the probe blocked (darker bottom curve). Inherent electrical noise when the entire probe is acoustically sealed (lighter bottom curve).

VI. CONCLUSION

The horn attachment of the B&K 4170 has been redesigned using very small diameter stainless steel tubing. Three different designs were first considered theoretically and one of the new designs was chosen for further modeling. A number of different microprobes were built using the theoretical designs. The sensitivity and frequency response of each microprobe were measured and compared well with the theoretical predictions. An i.d. as small as 0.1 mm is possible if the length is less than 1 cm, with a sensitivity of about 20 μV/Pa. The microprobes have a frequency response that is flat within 5 dB in the midfrequency range due to viscous and thermal damping of the small tubes and the probe-to-probe variability is negligible.

[1]E. A. G. Shaw, "Prove microphone with horn coupling," J. Acoust. Soc. Am. **33**, 1679(A) (1961) and Cdn Patent No. CA 750013 (3 January 1967).
[2]E. A. G. Shaw, "Earcanal pressure generated by a free sound field," J. Acoust. Soc. Am. **39**, 465–470 (1966).
[3]Brüel and Kjær Instruments, Inc., *Horn-coupled Probe Microphone Type 4170: Instruction Manual* (Nærum Offset, Nærum, 1982).
[4]Brüel and Kjær Instruments, Inc., *Probe Microphone Type 4182: Product Data* (B&K Headquarter, Nærum, 2005).
[5]M. R. Stinson and G. A. Daigle, "Comparison of an analytic horn equation approach and a boundary element method for the calculation of sound files in the human ear canal," J. Acoust. Soc. Am. **118**, 2405–2411 (2005).
[6]A. B. Copeland and D. Hill, "Design of a probe-tube adapter for use with a 1-inch condenser microphone," J. Acoust. Soc. Am. **48**, 1036–1039 (1970).
[7]L. P. Franzoni and C. M. Elliott, "An innovative design of a probe-tube attachment for a 1/2-in. microphone," J. Acoust. Soc. Am. **104**, 2903–2910 (1998).
[8]A. G. Webster, "Acoustical impedance and the theory of horns and of the phonograph," Proc. Natl. Acad. Sci. U.S.A. **5**, 275–282 (1919).
[9]V. Salmon, "Generalized plane wave horn theory," J. Acoust. Soc. Am. **17**, 199–211 (1946); "A New Family of Horns," **17**, 212–218 (1946).
[10]A summary of Webster horn equation and the case of a horn with constant flare can also be found in A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 7.
[11]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949), Chap. 2.
[12]The main theory can also be found in D. H. Keefe, "Acoustical wave propagation in cylindrical ducts," J. Acoust. Soc. Am. **44**, 616–623 (1968) or M. R. Stinson, "The propagation of plane sound waves in narrow and wide circular tubes and generalization to uniform tubes of arbitrary cross-sectional shapes," *ibid.* **89**, 550–558 (1991).

Active control of drag noise from a small axial flow fan

Jian Wang and Lixi Huang[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

(Received 10 November 2005; revised 12 April 2006; accepted 20 April 2006)

Noise sources in an axial flow fan can be divided into fluctuating axial thrust forces and circumferential drag forces. For the popular design of a seven-blade rotor driven by a motor supported by four struts, drag noise dominates. This study aims to suppress the drag noise globally by active control schemes. Drag noise features a rotating dipole and it has to be cancelled by a secondary source of the same nature. This is achieved experimentally by a pair of loudspeakers positioned at right angles to each other on the fan rotational plane. An adaptive LMS feedforward scheme is used to produce the control signal for one loudspeaker and the time derivative of this signal is used to drive the other loudspeaker. The antisounds radiated by the two loudspeakers have a fixed phase relation of 90° forming a rotating dipole. An open-loop control scheme is also implemented for the purpose of comparison and easier implementation in real-life applications. The results show that the globally integrated sound power is reduced by about 13 dB for both closed- and open-loop schemes. A possible limiting factor for the cancellation performance is found to be the presence of higher order modes of drag noise. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2204443]

PACS number(s): 43.50.Ki [SFW] Pages: 192–203

I. INTRODUCTION

Small axial flow fans make noise much like large compressors and turbines, but the small number of rotor blades renders easier physical interpretation of the source characteristics (Huang, 2003). A brief review of the noise sources and noise abatement techniques specific to small axial flow fans is given in a previous study (Wang *et al.*, 2005). Unsteady forces acting on the blades are the origin of most fan noise. The forces acting on a blade can be divided into a thrust component along the rotating axis, and a drag component in the circumferential direction of the rotor. The final radiated noise is a result of complex acoustic interference between these two force components on each blade, and of sounds from all blades. The previous study deals with the active control of noise radiated by the thrust forces arising from the rotor-strut interactions. This study extends the control to drag noise, which is a far more complex component. In what follows, a brief summary is given for the existing active fan noise control with emphasis on the issue of directivity pattern, followed by the description of the configurations used in the current study.

Active fan noise control is further divided into active minimization of the source strength by interfering with the aerodynamics (e.g., Neuhaus *et al.*, 2003; Rao *et al.*, 2001; Simonich *et al.*, 1993), and the cancellation of the radiated sound by secondary sources (e.g., Gerhold, 1997; Thomas *et al.*, 1993, 1994; Quinlan, 1992; Lauchle *et al.*, 1997; Gee and Sommerfeldt, 2004). One common feature in most reported works is that the rotational plane of the fan is placed in a baffle in order to simplify the acoustic field before the global control is contemplated. The effect of such a baffle on the acoustic directivity has not been quantified. However, it may be speculated that the effects on different components of the noise sources are different. As a result, the acoustic interference between these components would also change, leading to a changed directivity pattern. Recently, Gerard *et al.* (2005) tried to use a single loudspeaker to cancel the tonal noise of an axial flow fan without using a baffle. They assume that the fan noise source can be represented by a single dipole source, but their measured acoustic directivity shows a tilted pattern and, not surprisingly, only part of the sound field can be cancelled. In fact, the tilted directivity pattern is a result of acoustic interference between waves radiated by the axial thrust component and the circumferential drag component. A detailed analysis of such interference has been given by Huang and Wang (2005) based on experimental data.

The acoustic radiation efficiency depends strongly on the difference between the spatial index of spinning pressure modes and the frequency index of the radiated sound (Tyler and Sofrin, 1962; Lowson, 1970), $\nu = mB - kS$, where B and S are the numbers of rotor blades and stationary struts, respectively, m is the harmonic index and k is any integer. Careless designs of a cooling fan often create a rotor-strut interaction in which the effective number of strut is $S=1$, for which the thrust noise radiates at the leading mode of $\nu=0$ for $k=mB$. The leading mode thrust noise has a simple directivity pattern and a previous study has demonstrated the effectiveness of active control by using a single loudspeaker (Wang *et al.*, 2005). A more careful design of a cooling fan, however, features more drag noise than thrust noise (Huang and Wang, 2005). Drag noise is radiated by rotating dipoles and its leading order mode features $\nu = \pm 1$. It is far more complex acoustically than the thrust noise and, to the best of the authors' knowledge, there has not been a three-dimensional measurement of a fan noise radiation in which such a leading mode

[a)]Author to whom correspondence should be addressed. Electronic mail: mmlhuang@polyu.edu.hk

 0001-4966/2006/120(1)/192/12/$22.50

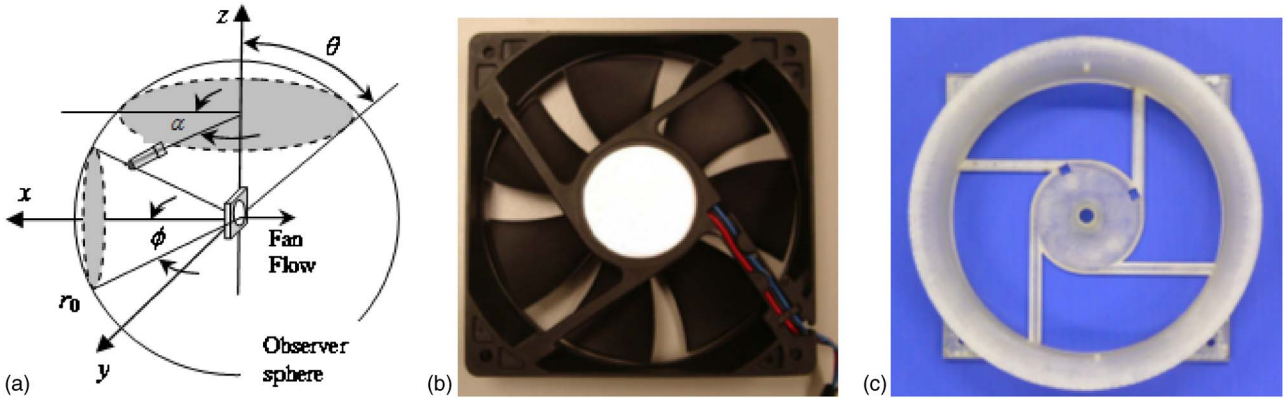


FIG. 1. (Color online) (a) Coordinate system used in theory and experiments. (b) The back view of the sample fan. (c) The front view of the modified fan casing.

dominates. As is shown in later sections, the leading mode drag noise can be approximated by two stationary dipoles operating in a fixed phase relation. This being true, two loudspeakers can be used to construct the antisound for the drag noise. This study aims to find out, experimentally, whether such antisound works and to what extent it may globally suppress the drag noise. The specific objectives of this study are the following: (*i*) to measure the percentage of sound power from a well-designed fan that can be attributed to the leading mode drag noise; (*ii*) to study, by numerical simulation, the extent to which the sound radiated by a pair of loudspeakers can globally cancel the drag noise from the interaction of seven rotor blades and four struts; (*iii*) using the filtered-X least-mean-square algorithm, conduct the active control for the drag noise with a set of optimal parameters found by numerical simulation; (*iv*) compare the performance of the closed-loop control with an open-loop control; and (*v*) analyze the results and determine the crucial factor that controls the overall performance.

In what follows in Sec. II, the characteristics of a real sample fan and the design improvement are described briefly to demonstrate how a well-designed fan features mainly the drag noise. In Sec. III, numerical simulation is conducted to predict the parametric influence of various factors present in a real control rig, such as the location of the error microphone and the distance between the fan center and the secondary sources. Section IV describes the experimental results of both closed-loop and open-loop controls. Analysis of the residual noise is also described, leading to conclusions in Sec. V.

II. ACOUSTICS OF THE SAMPLE FAN

The coordinate system is defined in Fig. 1(a). The axial-flow fan is shown standing vertically up along the $+z$ axis, and the sound radiated by the fan is surveyed by a microphone over a sphere of radius r_0 from the fan center. When viewed from upstream, the fan rotates counter-clockwise. The observer sphere is described by a latitudinal angle $\theta \in [0,\pi]$ and a longitudinal angle $\alpha \in [0,2\pi]$. An alternative latitudinal angle is ϕ measured from the $+x$ axis. Note that ϕ overlaps with α when ϕ is measured on the central horizontal plane of $\theta=90°$, or $z=0$, but they are, strictly speaking, different. Note that ϕ is used in theoretical derivations while α is used for directivity measurements and discussions on the central horizontal plane of $z=0$. To measure the effectiveness of the global control, three directivity measurement planes are used: $\theta=30°, 60°, 90°$, in addition to the top point at $\theta=0°$.

A. Theory

In order to analyze the noise made by the sample fan, the basic theory of the rotor-strut interaction acoustics is summarized below. The tonal sound radiated by the unsteady force on the rotor blades arising from the interaction of B blades with S struts of equal size and uniform spacing is given by (Lowson 1965, Lowson 1970).

$$c_{mB}^{(\mathrm{rotor})} = \frac{im\omega B^2 S}{2\pi c_0 r_0} \sum_{k=-\infty}^{+\infty} i^{-\nu}\left(T_{kS}^{(\mathrm{rotor})}\cos\phi - \frac{\nu}{mB\mathrm{M}} D_{kS}^{(\mathrm{rotor})}\right) \times J_\nu(mB\mathrm{M}\sin\phi), \quad \nu = mB - kS, \tag{1}$$

where $c_{mB}^{(\mathrm{rotor})}$ is the complex pressure amplitude at the frequency of $mB\times$rps, rps is the rotations per second for the rotor, $\omega=2\pi(\mathrm{rps})$, kS is the frequency index in the spectrum of the unsteady force components of thrust T and drag D on the rotor blades, c_0 is the speed of sound, M is the Mach number defined as $\omega r_s/c_0$, r_s is the radius at which the interaction occurs, and k is any integer. The frequency index differential, $\nu=mB-kS$, or the index of the spinning pressure mode (Tyler and Sofrin, 1962), is the most important parameter. The noise radiated by the corresponding interaction forces on the struts is found when the source terms of $T_{kS}^{(\mathrm{rotor})}$, $D_{kS}^{(\mathrm{rotor})}$ are replaced by $T_{mB}^{(\mathrm{strut})}$, $D_{mB}^{(\mathrm{strut})}$ in which the frequency index kS is replaced by mB as each strut experiences the interaction events with B rotor blades per rotational cycle.

The two most effective modes of sound radiation are explained physically by Huang (2003). When $mB=kS$, the thrust forces exerted by all blades occur simultaneously and simply add up, and the noise radiated is a simple dipole whose axis is along the rotational axis. However, the situation for the drag force is different. Drag force changes direction once per cycle, so the frequency perceived by a stationary observer is $kS\pm1$ and no noise is radiated at $mB=kS$, as

can be seen by the numerator ν, which vanishes, in the drag noise term in Eq. (1). For this reason the leading drag noise radiation mode has $\nu=\pm 1$. Higher order modes radiate sound by way of the Doppler effect, which is small for typical computer cooling fan operating at a low Mach number below 0.1. However, their presence alters the appearance of the acoustic directivity quite dramatically (Huang and Wang, 2005). The directivity patterns of the leading and the next higher order modes are summarized below:

$$p_{T0} \propto \cos\phi, \quad p_{T1} \propto \sin\phi\cos\phi,$$
$$p_{D1} \propto \sin\phi, \quad p_{D2} \propto \sin^2\phi, \tag{2}$$

where p is the radiated sound pressure, subscripts T and D indicate the sources of thrust and drag forces, respectively, and the numerical subscripts 0, 1, and 2 indicate the value of $|\nu|$. The distinct directivity patterns make it possible to separate the four mechanisms by simultaneously measuring sound at four symmetrical angular positions of $\alpha_1, \alpha_2=\pi-\alpha_1, \alpha_3=\pi+\alpha_1, \alpha_4=2\pi-\alpha_1$, where $\alpha_1 \in [0, \pi/2]$ is the position of the first microphone on a horizontal measurement plane. The four noise components given in Eq. (2) may be extracted as follows (Huang and Wang, 2005):

$$p_{T0}=\frac{p_1-p_2-p_3+p_4}{4}, \quad p_{T1}=\frac{p_1-p_2+p_3-p_4}{4},$$
$$p_{D1}=\frac{p_1+p_2-p_3-p_4}{4}, \quad p_{D2}=\frac{p_1+p_2+p_3+p_4}{4}. \tag{3}$$

B. Noise from the original sample fan

The sample fan used in this study of active control is the improved version of a computer cooling fan taken from the market (Delta AFB1212SH series), which is shown in Fig. 1(b). The original fan has a casing diameter of 120 mm, $B=7$ rotor blades, and $S=4$ downstream struts. The design speed is 3000 rpm. Two aspects of this sample fan contribute to loud noise. The first is that the circular inlet flow passage is intercepted by the square frame distorting the inlet flow to become one with a four-lobe pattern. The effect is similar to a set of four inlet vanes. Vortices are generated by the four edges and they are not entirely coordinated with a phase locked to the rotation. As such, they should contribute to both discrete and broadband parts of the noise spectrum. The second feature is that the strut carrying the electrical wires is larger than the other three. The effect of the extra size of the large strut can be considered to be that of a single strut, $S=1$, which is a very efficient noise source for both thrust and drag noise components when it interacts with a rotor of any blade number. Details of the noise mechanism of these two features are studied by Huang and Wang (2005) for a smaller cooling fan 90 mm in diameter. Correction of these two features led to a power reduction of about 10 dB in tonal noise for that small fan. The same procedure is followed here, and the new casing design is shown in Fig. 1(c). A full circular inlet bellmouth is installed and the four struts are made equal in size. It is emphasized that the active control technique is applied on the acoustically improved design shown in Fig. 1(c) instead of the original fan.

The acoustic directivity patterns for the original and improved fan are shown in Fig. 2. Figures 2(a)–2(c) are for the original fan shown in Fig. 1(b), while Figs. 2(d) and 2(e) are for the improved fan whose casing is shown in Fig. 1(c). Figure 2(a) shows the overall noise (outer dashed line), random noise (dash-dot line with a pattern almost parallel with the horizontal axis), rotary noise (thin solid line), and the BPF component of the rotary noise (thick solid line). Here, random noise is defined as the difference between the overall noise and the rotary noise, and the rotary noise is obtained by the synchronous average with the help of the tachometer signal. The air flow is drawn from the left, where $\alpha=0$ is labeled, to the right.

As shown in Fig. 2(a), the major axis of the overall noise pattern is tilted along the direction of $\alpha=30°$, 210°. This oblique distribution is a result of the interference between the drag noise, which spans out on the rotational plane, and the thrust noise, which beams along the rotational axis. The two components are separated in Fig. 2(b). The integrated power of the BPF component of the drag noise (thin line) is $SWL_{\mathrm{BPF}}^{(\mathrm{drag})}=50.4$ dB, which is 4.8 dB higher than the thrust noise (thick line). Such a prominent contribution from the thrust noise is caused by the extra size of the strut carrying the wires, which acts like a single strut, $S=1$. For the spectral component of $k=B$, it gives the leading mode thrust noise radiation with $\nu=mB-kS=0$. Meanwhile, the modes of $k=B-1, B+1$ also give the leading mode drag noise radiation with $\nu=\pm 1$, which is superimposed on the radiation of the drag noise by

$$\nu=mB-kS=1\times 7-2\times 4=-1$$

caused by the interaction between the seven rotor blades with the four struts with $k=2$. The extracted drag noise pattern shown in Fig. 2(b) also features asymmetry with respect to the rotational axis. More noise is radiated towards $\alpha=270°$ than $\alpha=90°$. This is caused by the interference between the leading mode drag noise, $p_{D1} \propto \sin\phi$, and the higher order drag noise, $p_{D2} \propto \sin^2\phi$, the latter being, presumably, originated from the factor of $S=1$. Using the method of noise source and modal decomposition of Eq. (3), the leading mode drag noise has the sound power of $SWL_{\mathrm{BPF}}^{(D1)}=50.8$ dB while the higher order drag noise has $SWL_{\mathrm{BPF}}^{(D2)}=33.2$ dB. The typical sound pressure level (SPL *re* 20 μPa) spectrum measured at the angle of $\alpha=230°$ is shown in Fig. 2(c), which shows high peaks for the first few BPF harmonics. Note that the peaks on the harmonics of the BPF may be a little higher than the result from the raw data, and this is caused by the time-base correction for the rotational speed changes in the synchronous averaging procedure described by Huang and Wang (2005).

C. Noise from the improved fan

When the square inlet frame is replaced by the full-circle bellmouth and the large strut is trimmed down to form a set of four equal struts, as shown in Fig. 1(c), the only leading

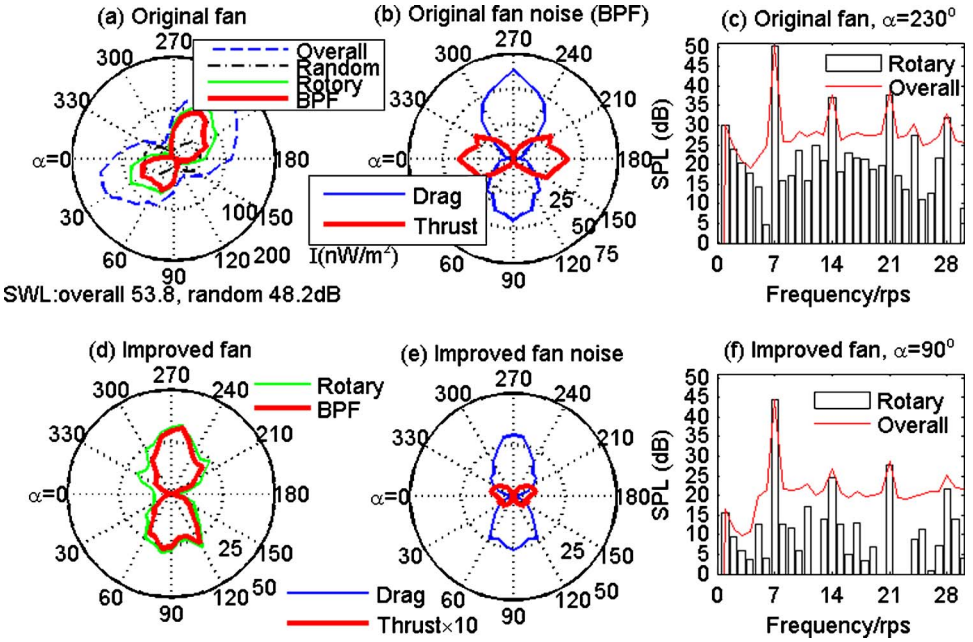


FIG. 2. (Color online) Comparison of sound intensity directivity and spectra from the original and the improved fans. (a) is the directivity of the original fan, (b) is the separation into the drag and thrust noise components, and (c) is the typical spectra at α=230°. (d), (e), and (f) are, respectively, the directivity, noise component separation, and spectra for the improved fan.

mode noise radiation comes from the drag component. The directivity of the measured sound intensity for the BPF is shown in Figs. 2(d) and 2(e). Since the random noise does not change much by the modifications, only the rotary noise is shown. Compared with the original fan, the total rotary noise is reduced from 52.4 to 48.5, or by 3.9 dB. The thrust noise power at the BPF, $SWL_{\mathrm{BPF}}^{(\mathrm{thrust})}$, is reduced from 45.6 dB for the original fan to 31.4 dB for the improved fan, which is a more significant reduction than that of the overall rotary noise. The rotary sound intensity shown in Fig. 2(d) is no longer tilted away from the rotational plane, and there is improved symmetry between the sound measured at 270° and that at 90°. Using the method of noise source and modal decomposition of Eq. (3), the leading mode drag noise has $SWL_{\mathrm{BPF}}^{(D1)}$=48.1 dB, while the higher order drag noise has $SWL_{\mathrm{BPF}}^{(D2)}$=31.2 dB. The improved fan has a total BPF drag noise of $SWL_{\mathrm{BPF}}^{(\mathrm{drag})}$=47.7 dB, while that of the thrust noise is 16.3 dB below this level. As a result, the thrust noise can only be shown in Fig. 2(e) after being amplified by ten times. A typical SPL spectrum for α=90° is shown in Fig. 2(f) for the improved fan. The first BPF peak is still about 20 dB above the broadband floor but, compared with the original fan, the improved fan is already very quiet and is chosen to be the starting point for the proposed active noise control scheme.

III. NUMERICAL SIMULATION OF ANC

The fan is represented by four unsteady drag forces distributed uniformly around the circumference with a rotating phase relation. The action of the active control is simulated by choosing a proper linear superposition of the two sound fields such that the pressure oscillation at the position of the error microphone is forced to be zero. All simulations are conducted for the fundamental blade passing frequency.

A. Secondary source model

The timing of the interaction is determined by the relative position between a rotor blade and the stationary struts. In this sense, the unsteady lift force occurs mainly when a rotor blade passes by a strut. As a result, the location is fixed relative to the stationary struts. However, as one blade comes to interact with a set of struts in a fixed sequence, the unsteady lift repeats from the position of one strut to the next with a fixed time delay, forming a pattern which can also be considered to be an oscillating force rotating continuously in space. Whether the source is better described by such a rotating force or fixed force with a rotating phase relation is purely a mathematical choice. Physically, the latter description is easier to model. The drag component of each interaction site, usually near the tip of a blade span, can be further decomposed into two parts, one in the horizontal direction, F_y, and another in the vertical direction, F_z, cf. Fig. 1(a). A normal fan has a set of evenly spaced rotor blades and struts; the summation of F_y from all interaction sites can be simulated by a concentrated F_y applied at the fan center. Since one oscillating force radiates an acoustic dipole, the difference between the distributed interaction forces and the total force located at the fan center represents two tightly coupled dipoles, or a quadrupole, which has a much smaller sound power and can be ignored in the current study. The same applies to F_z, and the result of all drag forces in all interaction sites can be represented by a pair of two forces at the fan center. The two component forces are

$$F_y = Ae^{i(\omega t+\pi/2)}, \quad F_z = Ae^{i\omega t}, \tag{4}$$

in which the y component leads the z component by 90° as the fan rotates from the $+y$ axis towards the $+z$ axis, and A is the amplitude. The sound radiated by each point force,

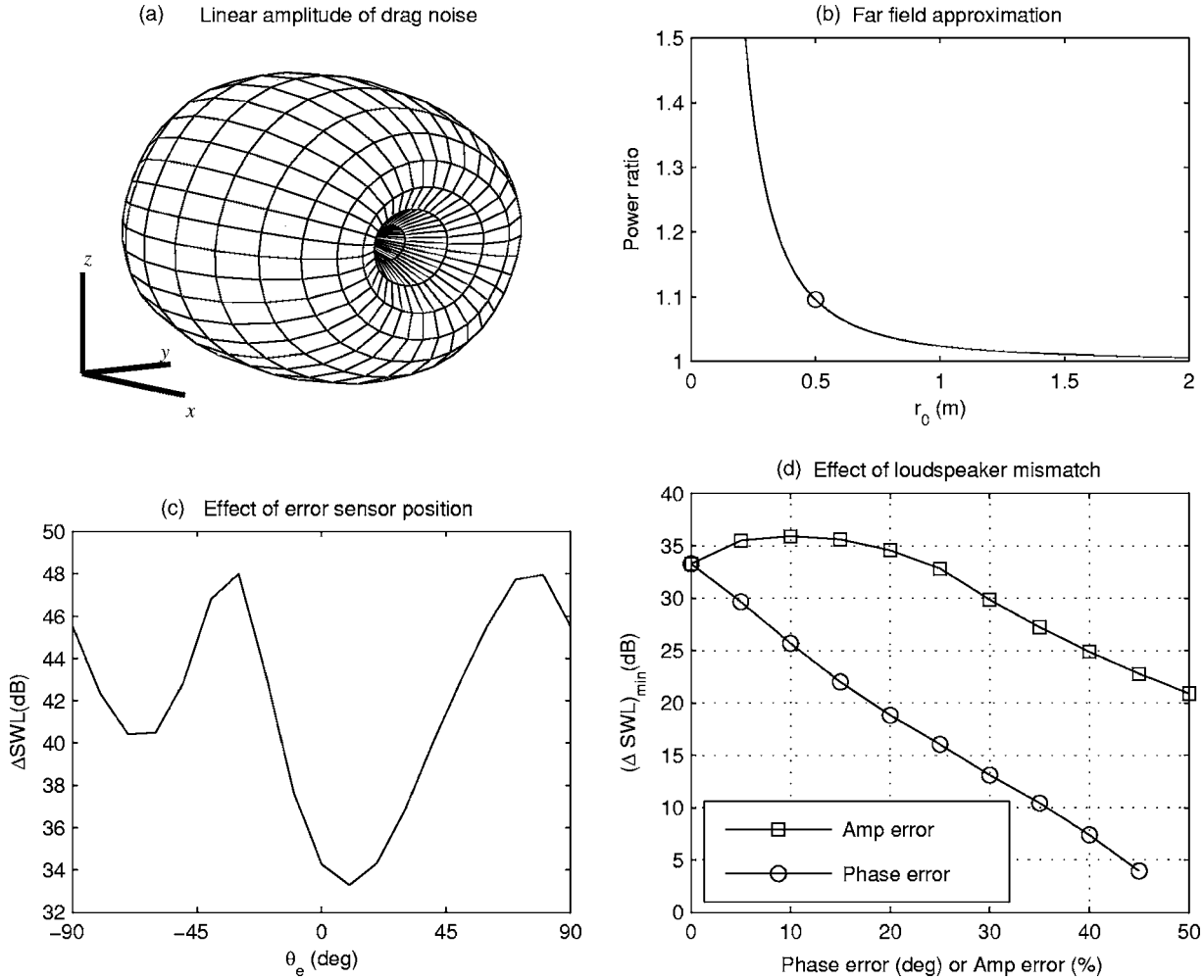


FIG. 3. Dipole simulations. (a) Fan noise simulated by four-point dipoles. (b) Ratio of the approximate to the accurate sound powers as a function of the measurement sphere radius. (c) The effect of the error microphone location on the control performance. (d) Effect of the loudspeaker phase and amplitude mismatch.

say $F_z e^{i\omega t}$, can be simulated by the following formulas (Dowling, 1998),

$$p|_{F_z} = \frac{i\omega \cos\theta}{4\pi r_0 c_0}\left(1 + \frac{c_0}{i\omega r_0}\right) F_z e^{i\omega(t-r_0/c_0)}, \quad \cos\theta = \frac{z}{r_0},$$

$$u_r = \frac{i\omega \cos\theta}{4\pi\rho_0 c_0^2 r_0}\left(1 + \frac{2c_0}{i\omega r_0} - \frac{2c_0^2}{\omega^2 r_0^2}\right) F_z e^{i\omega(t-r_0/c_0)}, \tag{5}$$

$$u_\theta = \frac{\sin\theta}{4\pi\rho_0 c_0 r_0^2}\left(1 + \frac{c_0}{i\omega r_0}\right) F_z e^{i\omega(t-r_0/c_0)}, \quad u_\alpha = 0,$$

where p is the total oscillating pressure inclusive of propagating sound and the near field, u_r, u_θ, u_α are the particle velocity in the radial, latitudinal, and longitudinal directions, respectively, and ρ_0 is the undisturbed density of fluid. The particle velocity is useful for the calculation of sound intensity and is not elaborated further. The sound generated by F_y is obtained similarly. The only difference is that z/r_0, which derives from $\cos\theta$ in the first expression of Eq. (5), should be changed to y/r_0 for $p|_{F_y}$. Together, the radiations by F_z and F_z form a rotating dipole whose pressure is given below,

$$p = p|_{F_z} + p|_{F_y} = \frac{i\omega(z+iy)}{4\pi r_0^2 c_0}\left(1 + \frac{c_0}{i\omega r_0}\right) A e^{i\omega(t-r_0/c_0)},$$

$$|z+iy| = r_0 \sin\phi,$$

$$|p| = \frac{A\omega\sqrt{1+(\omega r_0/c_0)^{-2}}}{4\pi r_0 c_0} \sin\phi. \tag{6}$$

The rotating dipole has the directivity of $|p| \propto \sin\phi$, which has the shape of a ring around the x axis, which is given in Fig. 3(a).

The above simulation results are obtained by placing the two forces at the same point which is meant to be the fan center. Experimentally, each component dipole is realized by a loudspeaker with a finite size. In fact, it is impossible to place the two loudspeakers at the fan center position without seriously blocking the flow. Under such structural constraints, one pair of loudspeakers is put on the top edge of the fan frame and another pair under the bottom edge of the fan frame. This arrangement is illustrated in Fig. 4(b) and the photo is shown in Fig. 4(c). The two pairs approximate a rotating dipole placed at the center of the fan. Numerical simulation is easily conducted to see the difference between

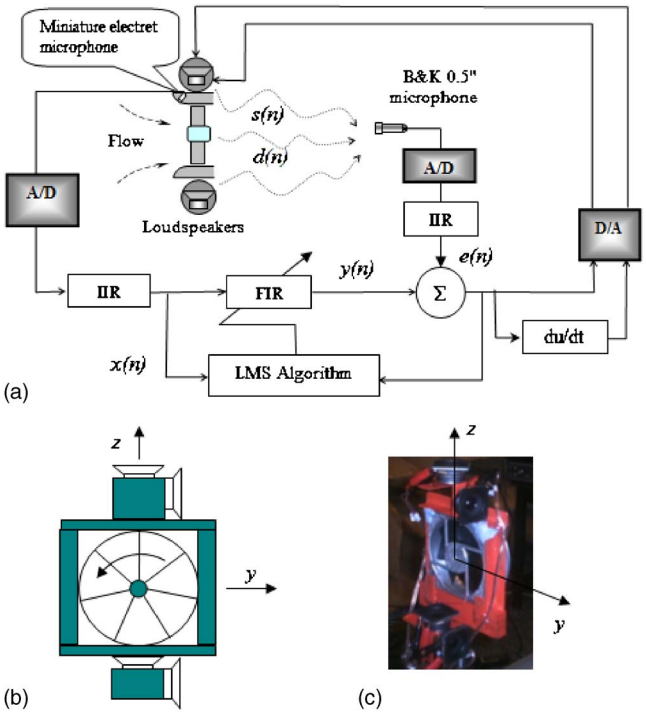


FIG. 4. (Color online) Experimental setup. (a) Overall view of the control system. (b) Schematic of the secondary source arrangement. (c) The back view (photo) of the two pairs of loudspeakers, two horizontal and two vertical.

the sounds from the two pairs of loudspeakers and a single rotating dipole with perfect source collocation with the fan center. The difference is found to be negligible for the parameters relevant to the current experiment.

The dipole sound has both far-field and near-field components, and the exact sound intensity should be calculated as $I=\frac{1}{2}\,\mathrm{Re}(pu_r^*)$, where u_r^* is the conjugate of the radial component of the acoustic particle velocity. In the experiment, however, a simple measurement of the free field takes only the local pressure. The far-field approximation, $I=p_{\mathrm{rms}}^2/(\rho_0 c_0)$, is used for sound intensity calculation, where p_{rms} is the local root-mean-square value of the measured pressure oscillation. Ideally, the measurement microphone is placed as far away as possible from the source, but a close proximity is necessary if the source is weak and a good signal-to-(electronic) noise ratio is desired. The percentage error caused by the far-field approximation for the BPF of 350 Hz is simulated by the dipole shown in Fig. 3(a), and the result is shown in Fig. 3(b). The compromise reached in the current experiment is to place the error microphone at $r_0=0.5$ m from the fan center, and the expected deviation in sound power estimation is 9.6% or $10\log_{10}(1.096)=0.4$ dB, which is marked by an open circle in Fig. 3(b).

B. Optimization of the error microphone position

The secondary source is driven by a signal produced in such a way that the sound radiated cancels exactly the primary noise at the position of the error microphone. The sound radiated by a single loudspeaker follows Eq. (5), while that by a rotating dipole follows Eq. (6). The signal obtained at the error microphone is used to adjust the signal fed to the secondary source such that the combination of the fan noise, say p_{fan}, and the antisound, Cp_{anti}, cancel at this point, $p_{\mathrm{fan}}+Cp_{\mathrm{anti}}=0$, where p_{anti} is the antisound generated by the two pairs of tightly coupled loudspeakers at the position of the error microphone by a unit voltage amplitude, and C is the complex amplitude to be determined either manually in an open loop control, or automatically in a closed-loop control. The position of the error microphone affects the global control results, and many strategies exist to place the error microphone or microphones so that a certain global measure of residual noise is minimized. In the current problem, the primary noise to be controlled has been shown to have the directivity of $p_{D1}\propto\sin\phi$, so a single error microphone position can be used. The primary noise is simulated by four circumferential point forces at an equal angular interval of 90° and the source radius is $r_s=5$ cm, which is about 83% of the fan radius of 6 cm. The first source is placed on the $+y$ axis at $x=z=0, y=r_s$. The four antisound loudspeakers are simulated by two rotating dipoles, one above the fan at $z=10$ cm, and one below the fan at $z=-10$ cm, both with $x=y=0$. For obvious reasons, the error microphone is chosen to be located on the rotational plane of $x=0$ or $\phi=90°$. The exact location is further specified by the angular value denoted by θ_e, for which $\theta_e=0,+90°,-90°$ represents the intersection points of the observer sphere with the $+z,+y,-y$ axes, respectively. The total sound power reduction, ΔSWL, is found as a function of θ_e, as shown in Fig. 3(c). The variation curve is asymmetrical but the asymmetry vanishes when the radial position of the error microphone, r_0, is increased towards infinity. In other words, the asymmetry is caused by the near-field effect. For $r_0=0.5$ m, the optimal angular position for the error microphone is found to be close to the horizontal plane position of $\theta_e\approx 90°$, but in fact the angular position is relative to the first source force position, which is not actually known in experiment. However, the lowest value of sound power reduction, $\Delta SWL=33.3$ dB, is expected to be realized in experiment.

C. Effect of loudspeaker mismatch

Two loudspeakers are used to construct one rotating dipole as antisound. In order to simplify the control rig, only one output signal is used. The signal is used to drive one loudspeaker, say the one facing the horizontal direction. The time derivative of this signal is used to drive the other loudspeaker in order to make sure that the two form a 90° phase difference. Such configuration is based on the assumption that the two loudspeakers behave identically. The fact that there is bound to be some difference in their response to input signals calls for amplitude and phase correction. A fixed amount of correction can be embedded in the control circuit without difficulty. However, such correction cannot account for variations in loudspeaker performance. Simulation is thus conducted to see the sensitivity of the loudspeaker mismatch and the global noise suppression performance. The minimal sound power reduction obtained for the worst error microphone position relative to the source location is used. In other words, the trough of the curve in Fig. 3(c) is used for a given configuration. The configuration is

then changed to one in which the two loudspeakers facing the vertical directions are given an amplitude or phase angle mismatch with the response of the two loudspeakers facing the horizontal direction. The variations in amplitude and phase are tested separately, and the resulting minimal sound power reductions are plotted together in Fig. 3(d). Both curves begin with the value of 33.3 dB for the reference configuration in which the two loudspeakers are identical in responses. For the performance of ΔSWL to deteriorate to 20 dB, the amplitude mismatch should be roughly 50% (end of the upper curve), while the phase mismatch is close to 20°. It may be said that the performance is more sensitive to phase mismatch. During the actual experiment lasting for one hour or so, the loudspeaker is found to vary in amplitude within a band of around 30%, while the phase variation is around 15°. The results of this simulation seem to indicate that both are tolerable as far as a target of around 20-dB sound power reduction is concerned.

In order to obtain the best result using the close-loop control, multi-channels may be necessary. In this case, the loudspeakers facing the horizontal direction control the horizontal component of the rotating dipole, and their input signals should be adjusted by signals from an error microphone located on the central horizontal plane. The second channel for the two vertical-facing loudspeakers should be adjusted by an error microphone placed at the top of the fan. Both error microphones should be placed on the rotational plane. The use of multi-channel control is beyond the scope of the current study which is entirely motivated by the pursuit of simplicity and practicality of a possible active control method.

IV. ACTIVE CONTROL STUDIES

As shown by Huang and Wang (2005), the configuration of $B=7$ blades with $S=4$ struts features rotating dipoles for the fundamental BPF tone and the third harmonic, while the second harmonic is in the higher spinning pressure mode with $\nu=2$. The latter can be seen as a set of tightly coupled dipoles or an approximate quadrupole. The sound power levels for the three harmonics are experimentally found to be 47.8, 32.7, and 35.4 dB, respectively. The second harmonic is indeed low and is not much of a concern here, while the third harmonic is higher than the second due to the leading mode radiation. However, its absolute level is also not very high and is left out of the control scheme.

A. Adaptive and open-loop feed-forward control

The experimental setup for the closed-loop control is shown in Fig. 4(a) schematically. The reference signal is provided by a miniature electret microphone (151 series supplied by Tibbet industry) located on the bellmouth of the fan. It senses the near-field aerodynamic pressure on the bellmouth surface, and the signal is a saw-tooth-like waveform, which has a richer BPF content than the narrow pulses provided by a photoelectric tachometer, the difference being around 20 dB. The electret microphone has a flat frequency response of 0.018 V/Pa from 300 Hz to 5 kHz. This sensitivity is very low compared with that of the condenser-type

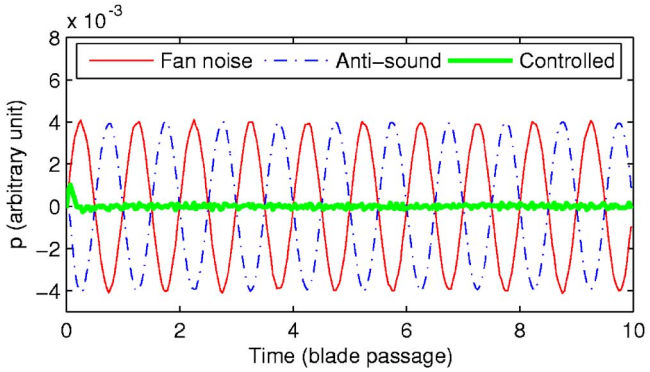


FIG. 5. (Color online) The computer simulation of the LMS control algorithm.

microphone used for the directivity measurement. In other words, the aerodynamic pressure oscillation sensed is much higher than the radiated sound, eliminating a possible feedback path in the control rig. The reference signal is bandpass filtered to keep the components of the BPF tone. A six-order Chebyshev infinite-impulse-response (IIR) filter is constructed by using the least p-norm optimal IIR filter design in the SIMULINK of MATLAB® (Wang *et al.*, 2005). The error signal is taken by a half-inch B&K microphone located at $x=z=0$, $y=0.5$ m, which is level with the fan center on the rotational plane. A time-domain adaptive filtered-X LMS feedforward controller (described below) constructs the antisound signal to drive the secondary sound sources, which are made by the 2-in. loudspeakers. The output signal from the LMS circuit is divided into two paths, one direct path, which drives the horizontally oriented loudspeaker, and another with a time derivative, which gives a $\pi/2$ phase delay to drive the vertically oriented loudspeaker. Together, they form a dipole rotating counter-clockwise, simulating the drag noise from the fan rotating in the same direction. Physically, these two loudspeakers cannot be located at the center of the cooling fan. They can only be located outside the casing of the fan, and the result would be two not-so-tightly-coupled dipoles with their centers away from the fan center. In order to align the antisound with the fan noise source, two pairs are used, one above and one below the center of the fan. The back view of the fan equipped with four loudspeakers is shown in Fig. 4(c) and the schematic for this arrangement is shown in Fig. 4(b).

A 16-order normalized LMS adaptive filter is adopted. The FXLMS algorithm is tested in the SIMULINK and the result is shown in Fig. 5. The disturbance is a harmonic signal of 357 Hz plus a random noise of 3% in amplitude, which is around 30 dB below the main disturbance signal. The convergence is achieved after only $\frac{1}{4}$ of a period, or seven floating data points. The cancellation is about 30 dB, which means that the main disturbance signal is eliminated. The purpose of this simulation is merely to check the precision and the convergence speed of the control algorithm. The actual convergence speed and the mean square error depend on many physical parameters such as the physical control path transfer function. The control algorithm is built on a dSPACE (DS1103 PPC) controller, which is a real-time system with multiple A/D and D/A channels, and a Motorola

PowerPC 604e microprocessor running at 333 MHz, which is connected to a personal computer through an ISA bus. A real time interface (RTI) is used to build the code downloaded to and executed on the dSPACE hardware. The rotational reference signal and the error microphone signal are sampled at 10 kHz, and the output analog signal is also constructed at an update rate of 10 kHz; both are the upper limit of the DS1103 PPC controller board.

The technique of the open-loop control was applied to the thrust noise control (Wang *et al.*, 2005) and it succeeded in giving a total of 10.8-dB rotary sound power reduction. Considering the prospect of practical implementation for a cooling fan, the open-loop control is also tested here to see how its performance compares with that of the closed-loop described above. In the open-loop control, the same reference signal from the electret microphone is used. The signal is simply band-pass filtered, phase delayed, and amplified to drive the secondary loudspeaker. The phase delay and amplification are manually adjusted by using an error microphone located on the rotational plane. Once the two parameters are tuned, they are fixed in the control algorithm. For the current application of drag noise control, two output channels are used to drive two loudspeakers in one pair. The primary output drives the horizontal loudspeaker while the channel with the time derivative as described earlier drives the vertical loudspeaker. Of course, the second output channel is fixed to the first, but it can have its own fixed amount of time delay and independent amplification to account for the differences in the loudspeaker responses. The finding of the second channel parameters has to rely on the use of a second error microphone located on the $+z$ axis where the first loudspeaker does not radiate sound. In fact, these compensatory parameters are also used in the closed-loop control. Since, in reality, two pairs of loudspeakers are used, one channel drives two horizontal loudspeakers and another does the two vertical loudspeakers.

The acoustic directivity and the integrated sound powers before and after the control are measured at a spherical radius of $r_0=0.5$ m by a second measurement microphone (B&K type 4187) not shown in Fig. 4(a). The signal from the measurement part of the instrumentation has no involvement in the control circuit. The acoustic directivity is measured on the central horizontal plane level with the fan center, $z=0$ or $\theta=\pi/2$. The synchronously averaged sound is used to calculate the far-field approximation of the sound intensity, $I \approx p_{\mathrm{rms}}^2/\rho_0 c_0$, where p_{rms} is the rms value of the BPF component of sound. A total of 36 points are measured with an angular interval of 10°, and the sound power is calculated by (Huang and Wang, 2005)

$$W = 2\pi r_o^2 \int_0^{\pi/2} I(\alpha)|\sin(\alpha)|d\alpha \approx \pi^2 r_o^2 (\Delta\alpha) \sum_{i=1}^{36} I(\alpha_i)|\sin(\alpha_i)|. \quad (7)$$

To evaluate the global effectiveness of the control, directivity measurement is also conducted for another two horizontal planes of $\theta=\pi/3$, $\pi/6$ together with the top point of $\theta=0$, cf. Fig. 1(a).

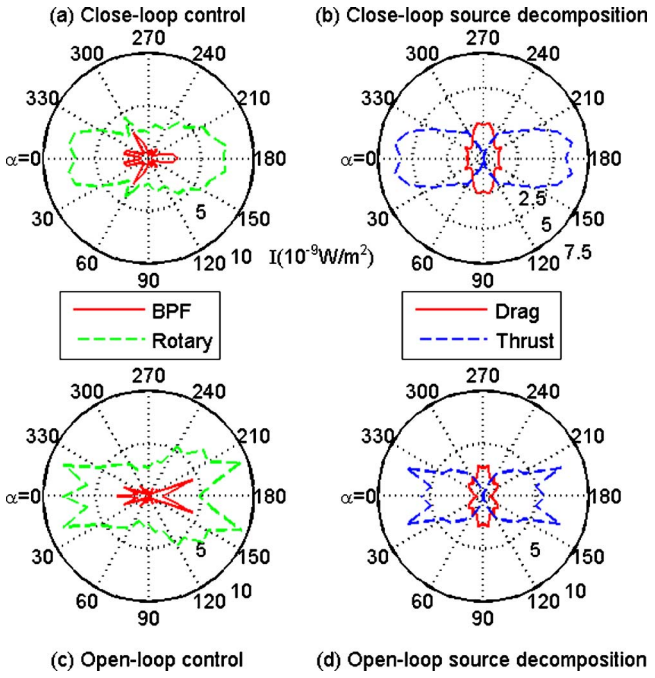


FIG. 6. Control-on sound intensity directivity for the close-loop (a,b) and open-loop (c,d) schemes with source decomposition analyses (b,d).

The results are presented in two parts in the next two subsections. In the first, directivity patterns for the control-on and control-off are compared for the central horizontal plane ($\theta=0, z=0$). The results for the closed-loop and the open-loop controls are compared. The comparison shows that the two control algorithms give very similar results. So, in the second part, the results of the three-dimensional measurement for the whole observer sphere is conducted for the open-loop control.

B. Directivity results and analyses

Figure 6 shows the results of the directivity measurement on the central horizontal plane, which should be studied together with the control-off measurements shown in Fig. 2. Figures 6(a) and 6(b) are for the closed-loop control and Figs. 6(c) and 6(d) are for the open-loop control. Figure 6(a) shows that the total rotary noise (outer dashed curve) is mainly aligned in the axial direction, which implies thrust noise and contrasts with the dominant drag noise pattern of Fig. 2(d). The BPF component, shown as the inner solid line in Fig. 6(a), is subject to control and it has a rather irregular shape, meaning that the residual noise is small. Assuming that there is a symmetry on the rotational plane, the sound power reductions, denoted as ΔSWL, from the control-off to the control-on states are found as follows,

$$\Delta SWL_{\mathrm{BPF}}^{(\mathrm{drag})} = 47.4 - 32.0 = 15.4 \text{ dB},$$

$$\Delta SWL_{\mathrm{All}}^{(\mathrm{drag})} = 48.0 - 37.4 = 10.6 \text{ dB},$$

$$\Delta SWL_{\mathrm{BPF}}^{(\mathrm{thrust})} = 31.4 - 32.5 = -1.1 \text{ dB},$$

$$\Delta SWL_{\mathrm{All}}^{(\mathrm{thrust})} = 39.1 - 38.2 = 0.9 \text{ dB},$$

where subscripts "All" imply all harmonics of the BPF. Note that the thrust noise is not subject to any active control and its level varies slightly between the control-on and control-off conditions. The results of the open-loop control shown in Figs. 6(c) and 6(d) are very similar to Figs. 6(a) and 6(b). Again, the residual rotary noise pattern is irregular, implying that the drag noise abatement by the active control is quite complete.

It can be concluded that the drag noise reduction for the BPF achieved in this study, which is around 15–16 dB, validates the basic principle pursued in this study, but it is well below the theoretical prediction of about 33 dB. The performance is apparently limited by factors not considered in the theoretical prediction. In the previous study of thrust noise control using a smaller cooling fan (Wang *et al.*, 2005), it was found that the effect of the variation of the rotational speed from one cycle to the next is negligible. This conclusion is reexamined for this study and is also found to be valid. The factor of mismatch between two loudspeakers used in a pair for rotating dipole is also excluded by the following considerations. Numerical simulation for the acoustic interference, Fig. 3(d), shows that at least 20 dB sound power reduction is achieved even when the two loudspeakers mismatch in their phase by about 20° or in their amplitudes by about 50%. Observations during experiment show that the two pairs of loudspeakers chosen have much less mismatches.

Two more clues for the ultimate performance limitation are analyzed below. The first is the variation of sound radiation by the cooling fan when its rotational speed is held absolutely constant. This variation may have its origin in turbulent flow aerodynamics, and it has also been analyzed previously (Wang *et al.*, 2005). The algorithm of all active fan noise control takes the signal from one blade passage to construct the antisound for the sound radiated in the time period of the next blade passage. The delay can be longer if the error microphone is placed in a distant far field. The apparently turbulent variation in sound radiation from the fan represents the part of uncontrollable noise. To analyze this effect, the time-domain signals from the whole central horizontal plane are analyzed in terms of the BPF amplitude variation from one blade passage to the next, and the result shows that, on average, $\sigma = 10.0\%$ of amplitude variation is found all around the fan. The difference between the controllable and uncontrollable sound power is thus estimated as

$$\Delta SPL = -20 \log_{10}(\sigma) \underset{\sigma=0.10}{\rightarrow} 20.0 \text{ dB}. \tag{8}$$

This value is very close to the number of sound pressure reductions achieved at the error microphone position, where, theoretically, the sound is supposed to be completely cancelled if all sound radiation is deterministic.

The second possible performance limitation is analyzed as follows. The sound pressure level reduction of around 20 dB at the error microphone position indicates that the basic control strategy is sound. The fact that both closed- and open-loop controls manage to achieve similar performance in this regard means that the system is essentially time stationary, and the additional capability of the adaptive control has not shown its potential benefit. The flaw in the control rig then must lie in the lack of perfect acoustic directivity match between the reality and the ideal distribution of $p_{D1} \propto \sin \phi$ for the leading mode drag noise. Factors of nonideal sound radiation include higher order drag noise featuring $p_{D2} \propto \sin^2 \phi$, cf. Eq. (2). Assume now that the measured drag noise is a sum of the two in the form of

$$p = p_1 \sin \phi + p_2 \sin^2 \phi, \tag{9}$$

where the magnitudes of each mode can actually be found by the source and mode decomposition method designed by Huang and Wang (2005), but their phase relation is not known. When the control is applied at one point, the sound power W becomes

$$\begin{aligned} W &= \int_0^{\pi} (2\pi r_0^2 \sin \phi) I d\phi \\ &= 2\pi r_0^2 (\rho_0 c_0)^{-1} \int_0^{\pi} \sin \phi (p_1 \sin \phi + p_2 \sin^2 \phi)^2 \, d\phi \\ &= \underbrace{2\pi r_0^2 (\rho_0 c_0)^{-1}}_{\equiv C_w} \int_0^{\pi} \\ &\quad \times (p_1^2 \sin^3 \phi + 2 p_1 p_2 \sin^4 \phi + p_2^2 \sin^5 \phi) \, d\phi \\ &= C_w \left(\frac{2}{3} p_1^2 + \frac{3\pi}{8} p_1 p_2 + \frac{8}{15} p_2^2 \right). \end{aligned} \tag{10}$$

When the error microphone is located at $\phi = \pi/2$ where the two modes add up, $p_{\phi=\pi/2} = p_1 + p_2$, the secondary sound field is given as $p_s = -p_{\phi=\pi/2} \sin \phi$ and the residual noise p_{res} and its sound power are found as follows,

$$\begin{aligned} p_{\mathrm{res}} &= p_s + p = -p_2 \sin \phi + p_2 \sin^2 \phi, \\ W_{\mathrm{res}} &= C_W \left(\frac{2}{3} - \frac{3\pi}{8} + \frac{8}{15} \right) p_2^2 = C_W 0.0219 p_2^2. \end{aligned} \tag{11}$$

When the error microphone is located where the first and second mode sounds cancel, $\phi = 3\pi/2$, the residual sound and its power, both denoted by a subscript "res," are found to be

$$\begin{aligned} p_{\mathrm{res}} &= p_s + p = +p_2 \sin \phi + p_2 \sin^2 \phi, \\ W_{\mathrm{res}} &= C_W \left(\frac{2}{3} + \frac{3\pi}{8} + \frac{8}{15} \right) p_2^2 = C_W 2.3781 p_2^2. \end{aligned} \tag{12}$$

The above two scenarios represent two extreme cases. The second case is the worst result, and its sound power compares with the second mode drag noise, denoted by a subscript "2," as follows,

$$\begin{aligned} \frac{W_{\mathrm{res}}}{W_2} &= \frac{C_W 2.3781 p_2^2}{C_W (8/15) p_2^2} = 4.4589, \\ SWL_{\mathrm{res}} - SWL_2 &= 10 \log_{10} (4.4589) = 6.5 \text{ dB}. \end{aligned} \tag{13}$$

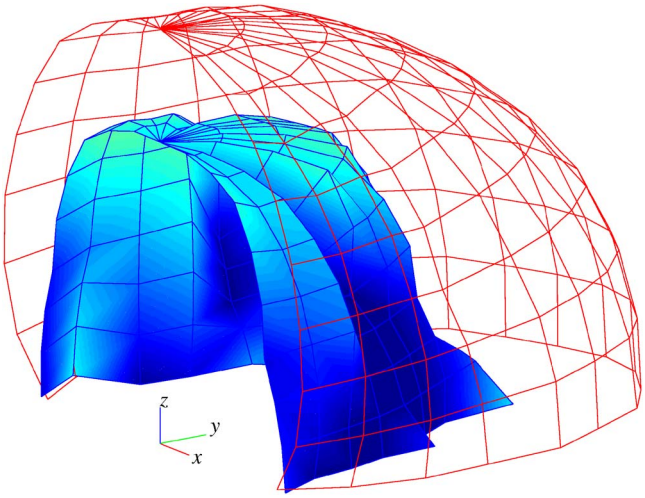


FIG. 7. (Color online) Comparison of the experimental data of the sound pressure level for the control off (outer wire-mesh) and control on (inner surface) configurations for one-quarter of the observation sphere.

The difference of 6.5 dB represents the intermodal coupling which cannot be tackled by the single rotating dipole control method, which only deals with the leading mode drag noise.

When the measured directivity of Fig. 2(d) is analyzed for the second-order drag noise, it is found that its sound power level is $SWL_{\mathrm{BPF}}^{(D2)}=31.2$ dB. The final residual drag sound power is 37.4 dB for the closed loop and 37.9 dB for the open loop. These represent 6.2 and 6.7 dB above $SWL_{\mathrm{BPF}}^{(D2)}$, which are, incidentally, very close to the upper limit of 6.5-dB intermodal coupling error. It must be emphasized, however, that the above analysis has many implicit assumptions about other aspects of the control rig, and the quantitative coincidence must be treated cautiously.

C. Global noise reduction

Three $\frac{1}{2}$-in. B&K microphones provide simultaneous directivity measurement for the three horizontal cross sections of $\theta=30°,60°,90°$ on the observer sphere of $r_0=0.5$ m, cf. Fig. 1(a). Due to the physical limitations, the error microphone is placed at a level slightly higher than the fan center, but it is still on the rotational plane. A total of 36 points are measured for each horizontal plane with an angular interval of $\Delta\alpha=10°$. The data of the three planes, $\theta=30°,60°,90°$, are used to calculate the total sound power radiation by the fan by assuming a perfect symmetry of the upper $(+z)$ and lower $(-z)$ hemisphere,

$$W=2\int_0^{\pi/2} r_0^2 \sin\theta\left[\int_0^{2\pi} I(\alpha,\theta)\,d\alpha\right]d\theta$$
$$\approx 2(\Delta\alpha\Delta\theta)r_0^2\sum_{n=1}^{3} T_n \sin\theta_n \sum_{m=1}^{36} I(\alpha_m,\theta_n), \tag{14}$$

where $T_n=1,1,0.5$ for $\theta_n=30°,60°,90°$ are the weighting coefficients for the numerical integration following the trapezoidal rule. Note that the plane of $\theta=0$ is reduced to one point on the top of the sphere with $\sin\theta=0$; it makes no contribution to the numerical summation and is left out. In order to plot the results in three-dimensional view smoothly, the sound intensity is further interpolated from three to nine horizontal mesh sections of $\theta=10°,20°,\ldots,90°$, where $\theta=0$ also provides one data point for interpolation. The comparison of the BPF drag noise for the conditions of control-on and control-off is made in Fig. 7, where only one quadrant of the 3D directivity for sound pressure level is given. This figure confirms that the noise suppression for the rotating dipole is global in nature. Note that the noise is not reduced along the rotational axis where thrust noise peaks and the applied control has, theoretically, no effect.

TABLE I. Changes in sound power levels (all in dB *re* 10^{-12} W).

Sound power level	Control off	Control on	Reduction
Random noise	49.6	49.8	–0.2
Rotary total	48.5	41.4	7.1
Rotary BPF	47.8	35.5	12.3
Drag noise BPF	47.8	34.8	13.0

The numerical comparisons for the sound power levels for various components are given in Table I. As shown in the first row of Table I, the random noise hardly changes; in fact, it increases by 0.2 dB. The focus is on the tonal noise, namely the rotary noise in the current context. Looking down the first column for the control-off state, it is found that the rotary noise is mainly dominated by the BPF component, which is in turn dominated by the drag noise. Note that there is a second decimal point difference between the rotary noise and the drag noise. For the control-on state, the second column shows less dominance by the drag noise in the total rotary noise since the drag noise is suppressed by the active control scheme. The direct objective of the control is the BPF component of the drag noise, which is reduced by 13.0 dB, as shown in the last row of the table. The total rotary noise is decreased by 7.1 dB, which is much less impressive than the drag noise reduction since it contains all frequencies and all noise mechanisms. In terms of the rotary noise for the first BPF, it is reduced by 12.3 dB, which is very close to the 13.0 dB reduction for the drag noise. Notice that the sound power reduction for the rotary BPF is about 3 dB more than that calculated by the data for the central horizontal plane alone. This could be attributed to the fact that the position of the error microphone has changed a little during the measurements.

Figure 8 gives the spectral comparison between the configurations of control-on and control-off. The three subfigures are for the three rotational plane points $(\phi=90°)$ on the three measurement planes of $\theta=30°,60°,90°$. The sound pressure level reductions for the BPF are 10.7, 18.6, and 20.6 dB for the three positions, respectively.

V. CONCLUSIONS

The findings of this work are summarized before comparison is made with the thrust noise control reported in Wang *et al.* (2005).

(1) A typical computer cooling fan available in the market is very noisy due to two gross features of the

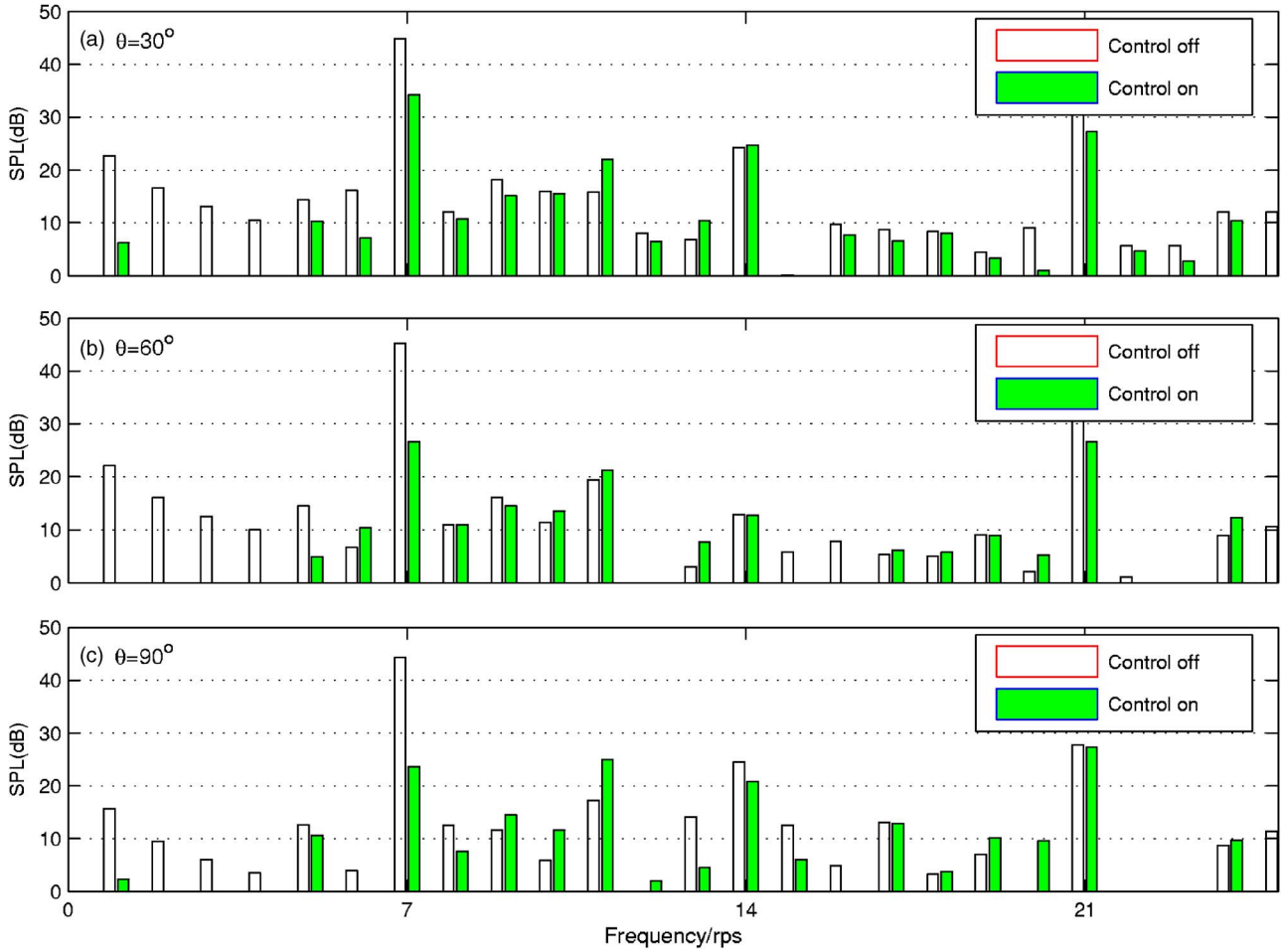


FIG. 8. (Color online) Spectral comparison for the control off (open bars) and control on (filled bars) for three points on the rotational plane ($\phi=\pi/2$).

structural design. One is the square frame which distorts the inlet flow to form a four-lobe pattern. The other is the wire-carrying strut which is a powerful noise source for both drag and thrust components. When these two features are corrected, the rotary sound power is reduced by about 4.2 dB. The measurement of the noise radiated by the improved fan shows a very ideal rotating dipole pattern, with an outstanding BPF peak at about 20 dB above the broadband of the spectrum.

(2) Based on the point force approximation, drag noise is radiated by the dipole source rotating with rotor blade around the rotational plane. The numerical simulation decomposes the rotating point force into two components with a rotating phase relation. Experimentally, a pair of two loudspeakers can be used to construct the rotating dipole when the two are installed perpendicular to each other with a 90° phase difference. In the current study, two such pairs are used in order to collocate the secondary sources with the fan center. The antisound approximates the primary fan noise very well, and a power reduction of 13.0 dB for the drag noise is achieved at the blade passing frequency. However, higher order drag noise also exists, and this component is beyond the scope of the current active control scheme.

(3) Fan noise radiation varies from one blade passage to the next in a random fashion even when the rotational speed is held constant, and the linear amplitude variation is found to be about 10%. Since the antisound is constructed from the signal input from the previous blade passage of the rotation, such random variation would cause incomplete noise cancellation. The measured maximum pointwise noise reduction of around 20 dB at the position of the error microphone is compatible with this hypothesis of the performance limitation. The mechanism might be rooted in the turbulent nature of the aerodynamic process.

(4) Higher order drag noise exists due to many factors, such as the difference between the fluctuation forces arising from the rotor-strut encounters between different rotor blades and struts due to either aerodynamic uncertainties or structural imperfections. The first higher order drag noise features a spinning pressure mode of $|\nu|=2$, and its sound pressure directivity is $\sin^2\phi$, which means that the sound pressure on the opposing sides on the rotational plane are the same, $\sin^2\phi=\sin^2(\phi+\pi)$. This differs from the leading mode drag noise with a pressure directivity of $\sin\phi$, which is antisymmetric across the rotational plane. The existence of a higher order drag noise component

would lead to the excess leading mode antisound determined by the pressure minimization procedure at the error microphone position. The coupling of this excess leading mode antisound and the higher order drag noise could lead to a maximum of 6.5-dB amplification of the higher order drag noise. The analysis of the control results indicates that the actual performance is close to this maximum overshoot, and this could be the ultimate limitation factor for the active drag noise control using the leading mode rotating dipole.

(5) Both open- and closed-loop (adaptive feedforward) controls are implemented, but no significant difference in performance is found between the two. This finding implies that, although the system under investigation could have one or more random variation factors, such as that described in (3), the processes are nevertheless stationary and the adaptive capability of the closed-loop control does not really have an opportunity to show its impact on the performance during a short-term experiment.

The present work adopts many common techniques used in a previous work on the thrust noise (Wang *et al.*, 2005) but there are also differences. The present work focuses on the rotating drag noise, while the previous one is on the thrust noise. In both cases, the global suppression of the dominant noise is achieved. The original fan is also modified in both cases before the active control is applied. In the case of the thrust noise control, a special coincident design of $B=S=7$ is used. That modification actually represents an increase of the noise radiation from the dominant source although the increase can be minimized if smaller strut size is adopted. In the present study, however, the modifications of the inlet bellmouth and the strut size equalization serve as a significant noise reduction from the original fan with the same number of struts. In other words, the current work begins with an already quiet design version of the most popular design configuration of $B=7$, $S=4$. In this sense, this study goes much beyond the previous. However, a rotating antisound is more difficult to construct, and more loudspeakers are used in the current study than in the previous. For a general case where both drag and thrust noises are present at the leading radiation modes, perhaps at different frequencies for each component, a minimum of three loudspeakers would be needed to construct the antisound for the three force components.

ACKNOWLEDGMENTS

The first author thanks the Hong Kong Polytechnic University for the Ph.D. research studentship during the early stages of the work reported here. The main support came from a University grant (G-U076) and the Research Grants Council of the Government of HKSAR (Grant No. PolyU 1/02C). Earlier work was also partially funded by the Intel Corporation.

Dowling, A. P. (**1998**). "Steady-state radiation from sources," in *Handbook of Acoustics*, edited by M. J. Crocker, (Wiley, New York), Chap. 8.

Gee, K. L., and Sommerfeldt, S. D. (**2004**). "Application of theoretical modeling to multichannel active control of cooling fan noise," J. Acoust. Soc. Am. **115**, 228–236.

Gerard, A., Berry, A., and Masson, P. (**2005**). "Control of tonal noise from subsonic axial fan. Part 2: active control simulations and experiments in free field," J. Sound Vib. **288**, 1077–1104.

Gerhold, C. H. (**1997**). "Active control of fan-generated tone noise," AIAA J. **35**, 17–22.

Huang, L. (**2003**). "Characterizing computer cooling fan noise," J. Acoust. Soc. Am. **114**, 3189–3199.

Huang, L., and Wang, J. (**2005**). "Acoustic analysis of a computer cooling fan," J. Acoust. Soc. Am. **118**, 2190–2200.

Lauchle, G. C., Macgillivray, J. R., and Swanson, D. C. (**1997**). "Active control of axial-flow fan noise," J. Acoust. Soc. Am. **101**, 341–349.

Lowson, M. V. (**1965**). "The sound field for singularities in motion," Proc. R. Soc. London, Ser. A **286**, 559–572.

Lowson, M. V. (**1970**). "Theoretical analysis of compressor noise," J. Acoust. Soc. Am. **47**, 371–385.

Neuhaus, L., Schulz, J., Neise, W., and Moser, M. (**2003**). "Active control of the aerodynamic performance and tonal noise of axial turbomachines," Proc. Inst. Mech. Eng., Part A **217**, 375–383.

Quinlan, D. A. (**1992**). "Application of active control to axial flow fans," J. Audio Eng. Soc. **39**, 95–101.

Rao, N. M., Feng, J. E., Burdisso, R. A., and Ng, W. F. (**2001**). "Experimental demonstration of active flow control to reduce unsteady stator-rotor interaction," AIAA J. **39**, 458–464.

Simonich, J., Lavrich, P., Sofrin, T., and Topol, D. (**1993**). "Active aerodynamic control of wake-airfoil interaction noise—experiment," AIAA J. **31**, 1761–1768.

Thomas, R. H., Burdisso, R. A., Fuller, C. R., and O'Brien, W. F. (**1993**). "Preliminary experiments on active control of fan noise from a turbofan engine," J. Sound Vib. **161**(3), 532–537.

Thomas, R. H., Burdisso, R. A., Fuller, C. R., and O'Brien, W. F. (**1994**). "Active control of fan noise from a turbofan engine," AIAA J. **32**, 23–30.

Tyler, J. M., and Sofrin, T. G. (**1962**). "Axial flow compressor noise studies," SAE Trans. **70**, 309–332.

Wang, J., Huang, L., and Cheng, L. (**2005**). "A study of active tonal noise control for a small axial flow fan," J. Acoust. Soc. Am. **117**, 734–743.

Orthogonal adaptation for active noise control

Jing Yuan[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong

(Received 12 January 2006; revised 7 April 2006; accepted 17 April 2006)

Many active noise control (ANC) systems apply the filtered-*x* least mean squares (FXLMS) algorithm for controller adaptation. The accuracy of path models is an important issue in these systems. Since parameter drifting in a noise field may cause model error between the secondary path and its prestored model in an ANC system, some ANC systems employ two adaptive processes for path modeling and controller adaptation respectively. In this paper, a new ANC system is proposed with adaptive path modeling and nonadaptive controller design. The proposed ANC system is noninvasive without persistent excitations. It avoids the slow convergence and inevitable estimation errors in controller adaptation. A rigorous analysis is presented to prove that the new ANC system will converge to an optimal one in the minimum H_2 norm sense. Experimental results are presented to verify the performance of the proposed ANC system.
 [DOI: 10.1121/1.2202908]

PACS number(s): 43.50.Ki, 43.50.Gf [KA] Pages: 204–210

I. INTRODUCTION

An important principle of active noise control (ANC) is to create quiet zones by generating destructive interference at sensed locations. Since magnitude distributions of sound fields are related to the wavelengths of sound signals, the range of destructive interference is proportional to the wavelengths of sound signals. This makes ANC attractive for controlling low-frequency noise.

The filtered-*x* least mean squares (FXLMS) algorithm is a popular tool for controller adaptation.[1,2] The algorithm is stable[3–5] if the phase error in the secondary path model is less than 90°. In many applications, environmental or boundary conditions of noise fields are not necessarily constant. There may be parameter drifting in noise fields. Online modeling of the secondary path is adopted in many ANC systems to keep the model as close to the real path as possible.

In most ANC systems, path models are finite impulse response (FIR) filters with many parameters. Accurate estimation of model parameters requires "persistent excitations"[6]—the invasive injection of probing signals into the ANC system.[7] Since the probing signals cannot be cancelled by the controller, some researchers investigate how to regulate the magnitudes of probing signals according to ANC performance,[8,9] others focus on noninvasive identification of secondary paths.[10–12]. An ANC system may perturb controller parameters to estimate the secondary path accurately. The FXLMS is used by existing noninvasive ANC systems for controller adaptation.

In this paper, a new ANC system is proposed with a single adaptation process for noninvasive path modeling. A major difference between the proposed and existing noninvasive ANC systems is the controller that is designed by the orthogonal principle in the new system. The new method only requires the estimated models to share a normal vector with the true paths. Path modeling is easier without persistent excitations. Stability and convergence of the new ANC system is proven via the Lyapunov approach.

With the controller designed by the orthogonal principle, the new ANC system avoids the slow convergence and estimation errors in controller adaptation. It will converge to an optimal ANC system in the minimum H_2 norm sense. The new ANC system uses the least squares (LS) algorithm that tends to converge faster than LMS counterparts.[13] Experimental results are presented to demonstrate the performance of the new ANC system.

II. PROBLEM STATEMENT

A typical ANC system may be described by a block diagram in Fig. 1(a), where $P(z)$, $S(z)$, $F(z)$, and $R(z)$ are the *p*rimary, *s*econdary, *f*eedback, and *r*eference paths, respectively. Path models and signals in Fig. 1 are in the Z-transform domain. Many researchers assume that path models can be approximated by FIR filters with negligible errors (assumption A1). This study is also based on such an assumption. If the primary source $n(z)$ is not available to the controller, a signal $x(z)$ is measured in the noise field to recover the reference $r(z)$. The controller must cancel feedback signals in $x(z)$ to ensure a stable closed-loop, which requires an estimated model $\hat{F}(z)$ to approximate $F(z)$.

A. Controller structure and objective

It is shown in Fig. 1(a) that $x(z)=R(z)n(z)+F(z)a(z)$. The reference is recovered in Fig. 1(b) as $r(z)=x(z)-\hat{F}(z)a(z)$; here $a(z)$ is the actuation signal. Eliminating $a(z)$ from the above three equations, one obtains

$$r(z)=\frac{R(z)}{1+\Delta F(z)G(z)}n(z)=\tilde{R}(z)n(z), \tag{1}$$

where $\Delta F(z)=\hat{F}(z)-F(z)$ is the model error of the feedback path. The closed-loop is stable if $\|\Delta F(z)G(z)\|_\infty<1$, which is

a)Electronic mail: mmjyuan@polyu.edu.hk

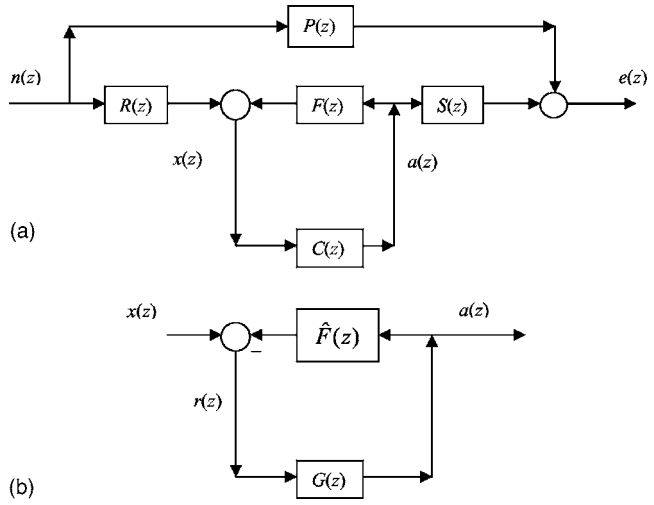


FIG. 1. (a) Block diagram of a typical ANC system with acoustic feedback. (b) Block diagram of ANC controller $C(z)$.

satisfied in most ANC systems. Many researchers assume $\Delta F(z)=0$ (assumption A2) and use $\tilde{R}(z)=R(z)$ in the analysis. The present study also adopts such an assumption since the presence of $\Delta F(z)\neq 0$ only causes unnecessary distractions without affecting the analytical results, as to be verified by experimental results.

In Fig. 1(b), the actuation signal is $a(z)=G(z)R(z)n(z)$ after substitution of $r(z)=R(z)n(z)$ by assumption A2. As a result, the error signal in Fig. 1(a) can be expressed as

$$e(z)=P(z)n(z)+S(z)a(z)=[P(z)+S(z)G(z)R(z)]n(z). \tag{2}$$

The ANC objective is $e(z)=0$ for broadband noise control. The above equation has an ideal solution

$$G(z)=\frac{-P(z)}{S(z)R(z)}. \tag{3}$$

Practically, the ideal controller may not be stable due to possible nonminimum phase (NMP) roots in $S(z)$ or $R(z)$.

For distributed-parameter systems to which the modal theory is applicable, only path transfer functions between collocated sources and sensors are surely minimum phase (MP). It means a MP $R(z)$ if the reference sensor is substantially collocated with the primary source. To avoid the near-field effects, however, the error sensor is usually placed far away from the secondary source and $S(z)$ is NMP without a stable inverse. A practically achievable objective of the ANC system is the minimization of

$$\|P(z)+S(z)G(z)R(z)\|_2 \tag{4}$$

in the H_2 norm sense. The proposed ANC system is intended to achieve this objective.

B. Online path modeling

When the ANC system recovers the reference signal $r(z)=R(z)n(z)$ to synthesize the actuation signal $a(z)=G(z)r(z)$ and estimate the path models, Eq. (2) may be rewritten as

$$e(z)=H(z)r(z)+S(z)a(z)=[H(z)+S(z)G(z)]r(z), \tag{5}$$

where an equivalent primary path $H(z)=P(z)/R(z)$ is introduced as if the noise field was caused by $r(z)$. If the reference sensor is substantially collocated with the primary source, $R(z)$ is MP with a stable inverse. It is assumed that the equivalent primary path $H(z)$ can be approximated by a FIR filter with negligible errors (assumption A3). The ANC objective is modified from the minimization Eq. (4) to the minimization of

$$\|H(z)+S(z)G(z)\|_2. \tag{6}$$

This is a practical approach due to the unavailability of $n(z)$. The proposed ANC system is able to minimize Eq. (4) if $n(z)$ is available.

The ANC system estimates the equivalent primary path $H(z)$ and the secondary path $S(z)$ simultaneously by minimizing estimation error

$$\begin{aligned}\|\varepsilon(z)\|_2&=\|e(z)-\hat{H}(z)r(z)-\hat{S}(z)a(z)\|_2\\&=\|\Delta H(z)r(z)+\Delta S(z)a(z)\|_2,\end{aligned} \tag{7}$$

where $\hat{H}(z)$ and $\hat{S}(z)$ are online models of $H(z)$ and $S(z)$, respectively. The model errors are $\Delta H(z)=H(z)-\hat{H}(z)$ and $\Delta S(z)=S(z)-\hat{S}(z)$. Let $\hat{h}=[\hat{h}_1\cdots\hat{h}_m]$, $h=[h_1\cdots h_m]$, $\hat{s}=[\hat{s}_1\ldots\hat{s}_m]$ and $s=[s_1\cdots s_m]$ denote coefficient vectors of $\hat{H}(z)$, $H(z)$, $\hat{S}(z)$, and $S(z)$, respectively. The time domain version of Eq. (7) would be

$$\begin{aligned}\sum\varepsilon_t^2&=\sum\left(e_t-\sum_{k=1}^{m}\hat{h}_k r_{t-k}-\sum_{k=1}^{m}\hat{s}_k a_{t-k}\right)^2\\&=\sum\left(\sum_{k=1}^{m}\Delta h_k r_{t-k}+\sum_{k=1}^{m}\Delta s_k a_{t-k}\right)^2,\end{aligned} \tag{8}$$

where $\{\Delta h_k=h_k-\hat{h}_k\}$ and $\{\Delta s_k=s_k-\hat{s}_k\}$. The summation of error squares is carried out in a sliding time window. Equation (8) may be written as

$$\begin{aligned}\sum\varepsilon_t^2&=\sum\left(e_t-\sum_{k=1}^{m}\hat{h}_k\psi_k(t)-\sum_{k=1}^{m}\hat{s}_k\varphi_k(t)\right)^2\\&=\sum\left(\sum_{k=1}^{m}\Delta h_k\psi_k(t)+\sum_{k=1}^{m}\Delta s_k\varphi_k(t)\right)^2\end{aligned} \tag{9}$$

by introducing $\psi_k(t)=r_{t-k}$ and $\varphi_k(t)=a_{t-k}$.

To aid the stability analysis, one may express model errors $\Delta H(z)$ and $\Delta S(z)$ in terms of coefficient vectors $\Delta_h^T=[\Delta h_1,\ldots,\Delta h_m]$, $\Delta_s^T=[\Delta s_1,\ldots,\Delta s_m]$ and $\Delta_\theta^T=[\Delta_h^T,\Delta_s^T]$. Similarly, the signal samples are represented by regression vectors $\psi^T(t)=[\psi_1(t),\ldots,\psi_m(t)]$, $\varphi^T(t)=[\varphi_1(t),\ldots,\varphi_m(t)]$, and $\phi^T(t)=[\psi^T(t),\varphi^T(t)]$. As a result, Eq. (9) may be seen as the square sum of a linear projection $\varepsilon_t=\phi^T(t)\Delta_\theta$. The convergence of $\varepsilon_t\to 0$ does not necessarily mean $\Delta_h\to 0$ or $\Delta_s\to 0$. For this reason, some researchers propose to perturb the ANC controller and hence perturb $a(z)$ and $\varphi(t)$. The projection of Δ_θ on a time-varying vector $\phi(t)$ becomes the projection on linearly independent vectors if $\varphi(t)$ is perturbed properly. This is exactly the idea of persistent excitation.[6]

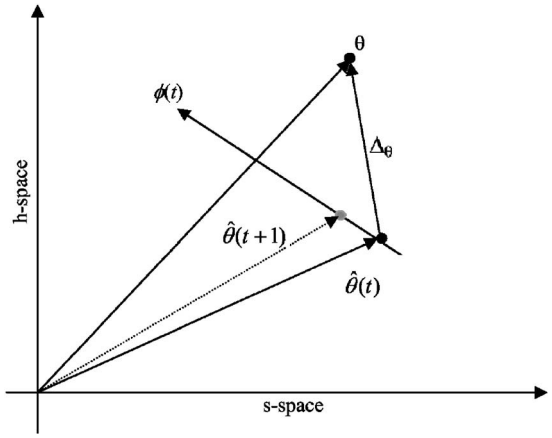


FIG. 2. Illustration of the orthogonal adaptation algorithm.

Identification accuracy depends analytically on the level of persistent excitation to the regression vector[14,15] whose elements are samples of $a(z)$ in an ANC system. Since controller perturbation is equivalent to persistent excitation of $a(z)$, it is still an invasive method. Only the persistent excitation is generated by a different method.

A significant difference between the proposed and existing noninvasive ANC systems is the absence of persistent excitations in the new system. Optimal ANC is possible if Eq. (9) is minimized. The remaining part of the paper is intended to explain how this is possible.

C. Orthogonal adaptation

A visual illustration of the proposed method is presented in Fig. 2, before a rigorous analysis in the next section. The vertical and horizontal axes of Fig. 2 represent, respectively, the h space and s space. These are vector spaces of the coefficients of $H(z)$ and $S(z)$, respectively.

By assumption A1, the true path transfer functions $H(z)$ and $S(z)$ are represented by coefficient vector $\theta^T=[h^T,s^T]$ in Fig. 2. Similarly, path models $\hat{H}(z)$ and $\hat{S}(z)$ are represented by coefficient vector $\hat{\theta}^T(t)=[\hat{h}^T,\hat{s}^T]$. It is reasonable to assume a constant θ since the drift of acoustical parameters is significantly slower than the adaptation rate. The purpose of Fig. 2 is to show that (*i*) $\hat{\theta}(t+1)$ will be closer to θ than $\hat{\theta}(t)$ and (*ii*) optimal ANC is possible without persistent excitations.

An important feature of the proposed ANC system is to design a controller $G(z)$ that forces $\phi(t)$ as orthogonal to $\hat{\theta}(t)$ as possible in every step of the adaptation. To keep a better focus, detailed design of $G(z)$ will be delayed to Sec. IV. The effect of this principle is illustrated here. Since perfectly orthogonal between $\phi(t)$ and $\hat{\theta}(t)$ may not be possible, $\phi(t)$ is not perfectly orthogonal to $\hat{\theta}(t)$ in Fig. 2. The estimation error is rewritten as

$$\varepsilon_t=\sum_{k=1}^{m}\Delta h_k\psi_k(t)+\sum_{k=1}^{m}\Delta s_k\varphi_k(t)=\phi^T(t)\Delta_\theta. \tag{10}$$

Online path modeling is updated by

$$\hat{\theta}(t+1)=\hat{\theta}(t)+\kappa\varepsilon_t\phi(t), \tag{11}$$

where $\kappa>0$ is a positive constant. Since $\hat{\theta}(t+1)$ is updated by adding a fraction of $\phi(t)$ to $\hat{\theta}(t)$, it will bias to a new point in Fig. 2, depending on the sign of $\varepsilon_t=\phi^T(t)\Delta_\theta$.

In the present form of Fig. 2, $\phi^T(t)\Delta_\theta>0$ and $\hat{\theta}(t+1)$ will bias in the same direction of ϕ to be closer to θ. If the direction of ϕ flips by 180°, then $\phi^T(t)\Delta_\theta<0$ and $\hat{\theta}(t+1)$ will bias in the opposite direction of ϕ to be closer to θ. In the next section, a rigorous stability analysis will prove the monotonous reduction of $\|\Delta_\theta\|_2$ until ε_t converges to zero. For the new ANC system, $\varepsilon_t\to 0$ is good enough to ensure stable ANC performance. The key measure is to force $\phi^T(t)\hat{\theta}(t)=0$ for all t. This condition and $\phi^T(t)\Delta_\theta=0$ mean $\phi^T(t)\theta=\Sigma_{k=1}^{m}h_kr_{t-k}+\Sigma_{k=1}^{m}s_ka_{t-k}=0$. As a result, the proposed ANC system converges to an optimal ANC controller such that $\|H(z)r(z)+S(z)a(z)\|_2=0$. A rigorous proof is presented in the next section.

III. STABILITY ANALYSIS

Since the proposed ANC system involves a single adaptation process, there are no coupling effects between adaptation processes. When the orthogonal condition $\phi^T(t)\hat{\theta}(t)\approx 0$ is enforced by the nonadaptive method of Sec. IV, the present section will focus on the adaptation process when analyzing the stability of the overall system.

A. Convergence of estimation error

All recursive identification algorithms depend on the regression vector $\phi(t)$ to update the parameter vector. The difference between algorithms is the multiplier to $\phi(t)$. Here a simple LS algorithm is adopted with a scalar multiplier $\kappa=1/\|\phi(t)\|^2$. As a result, Eq. (11) becomes

$$\hat{\theta}(t+1)=\hat{\theta}(t)+\frac{\varepsilon_t}{\|\phi(t)\|^2}\phi(t). \tag{12}$$

The convergence of Eq. (12) may be analyzed with the help of a positive definite function $V(t)=\Delta_\theta^T(t)\Delta_\theta(t)$. It can be verified that

$$V(t+1)-V(t)=[\Delta_\theta(t+1)-\Delta_\theta(t)]^T[\Delta_\theta(t+1)+\Delta_\theta(t)]. \tag{13}$$

Using Eq. (12) and $\Delta_\theta=\theta-\hat{\theta}$, one can write

$$\Delta_\theta(t+1)-\Delta_\theta(t)=\hat{\theta}(t)-\hat{\theta}(t+1)=-\varepsilon_t\phi(t)/\|\phi(t)\|^2, \tag{14}$$

and

$$\begin{aligned}\Delta_\theta(t+1)+\Delta_\theta(t)&=2\theta-\hat{\theta}(t)-\hat{\theta}(t+1)\\&=2\Delta_\theta(t)-\varepsilon_t\phi(t)/\|\phi(t)\|^2.\end{aligned} \tag{15}$$

Combining Eqs. (13)–(15), one obtains

$$V(t+1)-V(t)=\frac{-\varepsilon_t}{\|\phi(t)\|^2}\phi^T(t)\left[2\Delta_\theta(t)-\frac{\varepsilon_t}{\|\phi(t)\|^2}\phi(t)\right]. \tag{16}$$

The above equation becomes

$$V(t+1)-V(t)=-\frac{\varepsilon_t^2}{\|\phi(t)\|^2}\leqslant 0 \tag{17}$$

by substitution of $\varepsilon_t=\phi^T(t)\Delta_\theta$. Therefore $V(t)=\Delta_\theta^T(t)\Delta_\theta(t)$ decreases monotonously until $\varepsilon_t\to 0$.

B. Divide and conquer

The performance of the ANC system is judged by the error signal $e(z)$. The time domain expression of this signal is given by

$$e_t=\sum_{k=1}^{m}h_k r_{t-k}+\sum_{k=1}^{m}s_k a_{t-k}=\phi^T(t)\theta. \tag{18}$$

If one subtracts

$$\phi^T(t)\hat{\theta}(t)=\sum_{k=1}^{m}\hat{h}_k r_{t-k}+\sum_{k=1}^{m}\hat{s}_k a_{t-k} \tag{19}$$

from Eq. (18) and then adds it back, the result would be

$$e_t=\varepsilon_t+\sum_{k=1}^{m}\hat{h}_k r_{t-k}+\sum_{k=1}^{m}\hat{s}_k a_{t-k}=\varepsilon_t+\phi^T(t)\hat{\theta}(t). \tag{20}$$

The error signal is now divided into two parts to be conquered by two respective methods.

The first part of e_t is ε_t. It is conquered by online path modeling. In the previous subsection, a simple LS algorithm is proven able to drive the convergence of $\varepsilon_t\to 0$. There are many other LS or LMS algorithms suitable for this purpose. The difference is the convergence rate and computational load.[6]

The second part of e_t is $\phi^T(t)\hat{\theta}(t)$. Several methods will be discussed to force $|\phi^T(t)\hat{\theta}(t)|\approx 0$ in every step of the adaptation. The convolution of Eq. (19) implies the equivalence between $|\phi^T(t)\hat{\theta}(t)|\approx 0$ and $\|\hat{H}(z)+\hat{S}(z)G(z)\|_2\approx 0$ with coefficients of $\hat{H}(z)$ and $\hat{S}(z)$ taken from $\hat{\theta}(t)$.

C. Co-plane requirement versus persistent excitation

Since $\varepsilon_t=\phi^T\Delta_\theta$ is an inner product, its convergence to zero does not necessarily mean $\Delta_\theta\to 0$. However, if $|\phi^T(t)\hat{\theta}(t)|\approx 0$ for all t, $\varepsilon_t=\phi^T\Delta_\theta\to 0$ implies the co-plane relation of θ and $\hat{\theta}$. This is the combined result of two methods employed by the proposed ANC system to conquer the two parts of Eq. (20). In Fig. 2, one can see that forcing $|\phi^T(t)\hat{\theta}(t)|\approx 0$ and driving $\varepsilon_t=\phi^T\Delta_\theta\to 0$ is very cooperative. It is therefore possible to achieve optimal ANC performance without persistent excitations.

IV. OPTIMAL H_2 NORM CONTROLLER

The key feature of the proposed system is how to design a controller such that $\phi(t)$ is as orthogonal to $\hat{\theta}$ as possible. This is not a problem if $\hat{S}(z)$ is MP in every step of the adaptation process. The controller has a simple form $G(z)=-\hat{H}(z)/\hat{S}(z)$, and $\hat{\theta}$ is perfectly orthogonal to $\phi(t)$, because $\phi^T(t)\hat{\theta}=0$ is equivalent to

$$\hat{H}(z)r(z)+\hat{S}(z)a(z)=\left[\hat{H}(z)-\hat{S}(z)\frac{\hat{H}(z)}{\hat{S}(z)}\right]r(z)=0. \tag{21}$$

In practice, the error sensor is not collocated with the secondary source to avoid the near-field effects. $S(z)$ and $\hat{S}(z)$ are very likely to be NMP. Design methods must be studied to find $G(z)$ such that $\hat{\theta}$ is as orthogonal to $\phi(t)$ as possible, which is the focus of the present section. This is equivalent to minimizing $\|\hat{H}(z)+\hat{S}(z)G(z)\|_2$ by an optimal controller $G(z)$.

A. IIR solution

When $\hat{S}(z)$ is NMP, it is possible to achieve optimal performance in the minimum H_2 norm sense using an infinite impulse response (IIR) filter. The first step is the factorization of $\hat{S}(z)=S_m(z)S_n(z)$ where $S_m(z)$ and $S_n(z)$ are the MP and NMP parts of $\hat{S}(z)$, respectively. If $S_n(z)=\Sigma_{i=0}^{d}s_{ni}z^{-i}$, then $\widetilde{S}_n(z)=\Sigma_{i=0}^{d}s_{n(d-i)}z^{-i}$ can be obtained by using coefficients of $S_n(z)$ in the reversed order. Since roots of $S_n(z)$ are $\{|r_i|>1\}$, $\widetilde{S}_n^{-1}(z)$ is stable because roots of $\widetilde{S}_n(z)$ are $\{|1/r_i|<1\}$.[16] One may write

$$\hat{S}(z)=S_n(z)S_m(z)=\frac{S_n(z)}{\widetilde{S}_n(z)}\widetilde{S}_n(z)S_m(z), \tag{22}$$

where $|\widetilde{S}_n(e^{j\omega})|=|e^{-dj\omega}S_n(e^{-j\omega})|=|S_n(e^{j\omega})|$ for all ω[16] and $S_n(z)/\widetilde{S}_n(z)$ is a stable all-pass filter.

Equation (22) implies the factorization of $\hat{S}(z)=F_a(z)F_m(z)$ into the product of an all-pass filter $F_a(z)=S_n(z)/\widetilde{S}_n(z)$ and a MP filter $F_m(z)=\widetilde{S}_n(z)S_m(z)$. The objective is to minimize $\|[\hat{H}(z)+\hat{S}(z)G(z)]r(z)\|_2$, which is equivalent to the minimization of

$$\|\hat{H}+\hat{S}G\|_2=\|F_a(z)\|_2\|F_a^{-1}\hat{H}+F_mG\|_2=\|F_a^{-1}\hat{H}+F_mG\|_2, \tag{23}$$

where $\|F_a(z)\|_2=1$ since it is an all-pass filter. One may apply the long-division to obtain

$$F_a^{-1}(z)\hat{H}(z)=\frac{\widetilde{S}_n(z)\hat{H}(z)}{S_n(z)}=\frac{H_r(z)}{S_n(z)}+H_q(z), \tag{24}$$

where $H_q(z)$ and $H_r(z)$ are, respectively, the quotient and remainder polynomials. The two parts on the right-hand side of Eq. (24) are orthogonal to each other in the H_2 norm sense.

Since $H_r(z)/S_n(z)$ is the unstable part of Eq. (24), it cannot be cancelled by any stable feedforward controller. Optimal control, in the minimum H_2 norm sense, is the cancellation of the stable part of $F_a^{-1}(z)\hat{H}(z)$ by a controller in the form of

$$G(z) = -F_m^{-1}(z)H_q(z) = -\frac{H_q(z)}{\tilde{S}_n(z)S_m(z)}. \tag{25}$$

Substituting Eqs. (24) and (25) into Eq. (23), one minimizes $\|\hat{H}+\hat{S}G\|_2=\|H_r/\tilde{S}_n\|_2$ in the H_2 norm sense. Any other feedforward controller only increases $\|\hat{H}+\hat{S}G\|_2$ if it replaces $G(z)$. This method requires online root-finding and heavy computations. It is desired to design a suboptimal controller in a less expensive way, which will be discussed in the next subsection.

B. FIR solution

With some sacrifice of performance, it is possible to find a FIR filter controller $G(z)$ and minimize $\|\hat{H}(z)+\hat{S}(z)G(z)\|_2 = \|Q(z)\|_2$. When $G(z)$ is a FIR filter, $Q(z)$ is also a FIR filter with coefficients given by

$$q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ \\ \vdots \\ q_{2m-1} \end{bmatrix} = \begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \vdots \\ \hat{h}_m \end{bmatrix} + \begin{bmatrix} \hat{s}_1 & & & \\ \hat{s}_2 & \hat{s}_1 & & \\ \vdots & \hat{s}_2 & \ddots & \\ \hat{s}_m & \vdots & \ddots & \hat{s}_1 \\ & \hat{s}_m & \vdots & \hat{s}_2 \\ & & \ddots & \vdots \\ & & & \hat{s}_m \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}$$

$$= \hat{h} + M_s g, \tag{26}$$

where $q^T=[q_1,\ldots,q_{2m-1}]$ and $g^T=[g_1,\ldots,g_m]$ are coefficient vectors of $Q(z)$ and $G(z)$, respectively.

According to Parseval's theorem, minimization of $\|\hat{H}(z)+\hat{S}(z)G(z)\|_2=\|Q(z)\|_2$ is equivalent to minimization of $\|q\|^2$. The focus is now directed to

$$q^Tq = (\hat{h}+M_sg)^T(\hat{h}+M_sg) = \hat{h}^T\hat{h} + \hat{h}^TM_sg + g^TM_s^T\hat{h} + g^TR_sg, \tag{27}$$

where $R_s=M_s^TM_s$ is the autocorrelation matrix of the impulse response of $\hat{S}(z)$. Introducing $p=M_s^T\hat{h}$, one may substitute $g^TM_s^T\hat{h}=g^TR_sR_s^{-1}p$ into Eq. (27) to write

$$q^Tq = \hat{h}^T\hat{h} - p^TR_s^{-1}p + (R_sg+p)^TR_s^{-1}(R_sg+p), \tag{28}$$

where parameter vector g only affects $(R_sg+p)^TR_s^{-1}(R_sg+p)$. The minimization of q^Tq, subject to the constraint of a FIR $G(z)$, has a unique solution $R_sg=-p$.

The structure of R_s depends on the structure of M_s. Let m_i denote the ith column of M_s. An examination of Eq. (26) will enable one to see that m_i and m_j are related to each other by $|i-j|$ vertical shifts. Let $r_s(i,j)$ denote the (i,j)th element of R_s. Then

$$r_s(i,j) = m_i^Tm_j = m_j^Tm_i = \sum_{k=1}^{m-|i-j|} \hat{s}_k\hat{s}_{k+|i-j|}. \tag{29}$$

It turns out that R_s is a Toeplitz matrix. There are many fast algorithms for the solution of $R_sg=-p$ by taking advantages of the Toeplitz structure of R_s.[17] Most of them are available commercially in the Digital Signal Processing (DSP) block sets of MATLAB. The FIR controller is suboptimal if compared with its IIR counterpart, but it is less computationally expensive.

C. Iterative FIR solution

In many ANC applications, $H(z)$ and $S(z)$ must be approximated by FIR filters with sufficiently large numbers of coefficients. The degree m may be very large so that direct solution of $R_sg=-p$ is still expensive even using the available efficient algorithms. In that case, it is possible to consider iterative minimization of q^Tq.

Consider a positive definite function $J=0.5qTq$ and an iterative algorithm that keeps modifying g to minimize J. The time derivative of J is given by

$$\dot{J} = q^TM_s\dot{g} \tag{30}$$

where Eq. (26) has been used to link $\dot{q}$ to $\dot{g}$. The above equation suggests a very simple way to modify g. It is given by

$$\dot{g} = -\mu M_s^Tq. \tag{31}$$

where μ is a small positive constant. When coded in a high-level computer language, such as MATLAB, this is equivalent to an instruction "$g=g-\mu M_s^Tq\,\delta t$." Combining Eqs. (30) and (31), one can see that $\dot{J}=-\mu q^TM_sM_sq\leqslant 0$, which means $J=0.5qTq$ will be minimized by the simple modification rule of Eq. (31).

The only advantage of iterative solution is the reduced computations. It comes with a further sacrifice that $J=0.5qTq$ is not minimized instantly. If adaptation speed is not a critical issue, this method may be used to reduce the cost of the ANC system.

V. EXPERIMENTAL VERIFICATION

An experiment was conducted to verify the analytical results. A feedforward ANC was implemented in a duct with a cross-sectional area of 11×14.5 cm^2. The primary source was generated by a 4-in. loudspeaker placed at the upstream end of the duct. It was excited by the pseudo-random noise $n(z)$ that was not available to the ANC system. The secondary source was a 4-in. loudspeaker placed at the midpoint of the 2-m duct. A microphone sensor was placed in front of the primary source to measure the reference signal. The error sensor was another microphone placed 0.2 m downstream from the secondary source to avoid the near field effects of the secondary source. The sampling frequency of the system was 2.5 kHz, and all signals were low-pass filtered with a cutoff frequency 950 Hz.

The block diagram of the experimental system is the same as a typical ANC system in Fig. 1. Since the duct is a resonant system, the downstream reflection joins the secondary signal to contaminate the measured signal $x(z)=R(z)n(z)+F(z)a(z)$. Although the reference sensor was placed in front of the primary source, $R(z)\neq 1$ due to the resonant effects. The collocation of the reference sensor and the primary source causes a MP $R(z)$. Since the error signal was not collocated with the secondary source, the secondary path $S(z)$ was NMP. In the experiment, both $F(z)$ and $S(z)$

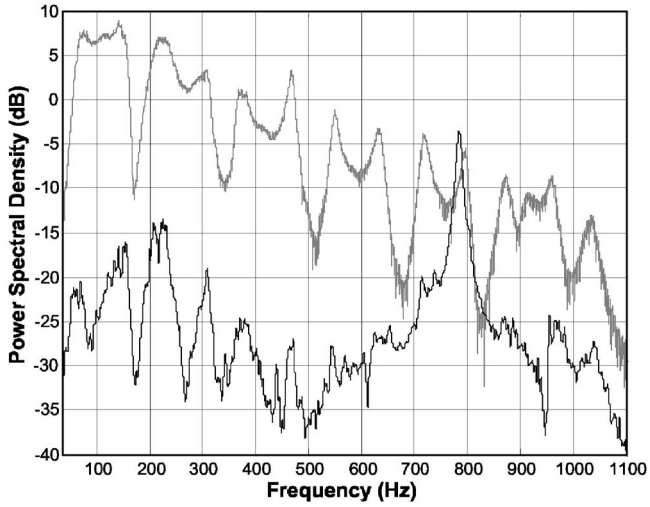


FIG. 3. Power spectral density plots of uncontrolled (gray) and controlled (black) noise.

were not available to the ANC system. A FIR controller with $m=400$ coefficients was implemented using the method of Sec. IV B. The optimal IIR controller, discussed in Sec. IV A, was not implemented due to its heavy computations for online root-finding.

The experimental results are plotted in Fig. 3 as two curves. The gray curve represents the normalized power spectral density (PSD) $|e(z)/r(z)|$ of the uncontrolled noise. It was obtained when the active controller was turned off. The black curve plots the normalized PSD $|e(z)/r(z)|$ of the controlled noise. It demonstrates significant noise suppression effects in most frequencies except a small region around 780 Hz. The ANC enhanced noise near 780 Hz instead of suppressing it. In the experiment, broadband noise was heard when the ANC system started. As the controller converged, audible noise became weaker and eventually reduced to a weak single tone noise.

The poor performance near 780 Hz was due to the NMP zero of $S(z)$. In Fig. 4, the magnitude responses of the primary and secondary paths are shown. The secondary path has zeros (dips) in a number of frequencies. When the error

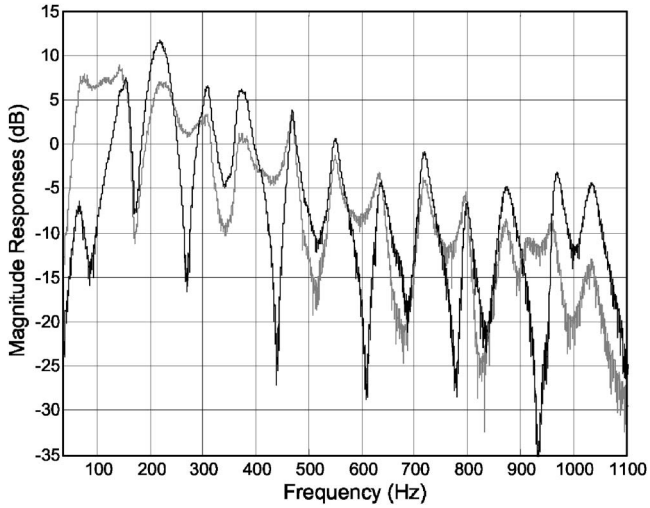


FIG. 4. Magnitude responses of primary (gray) and secondary (black) paths.

sensor was not too far away from the secondary source, most zeros of $S(z)$ were MP except one near 780 Hz. A NMP $S(z)$ is the problem of all ANC systems. Even if both $H(z)$ and $S(z)$ were available without errors, perfect cancellation of broadband noise would still be impossible for a single-source feedforward ANC system, since $-S^{-1}(z)H(z)$ is not stable. It is possible to design an IIR controller to match the stable part of $-S^{-1}(z)H(z)$ for optimal ANC performance. In simulations, the IIR controller neither enhances nor suppresses noise at the NMP zeros. Its FIR counterpart reduces online computations significantly with inevitable sacrifice in ANC performance that is most visible near the NMP zeros of $S(z)$. Since heavy online computations prevented the testing of the IIR controller at 2.5-kHz sampling rate, a FIR solution was tested to approximate $-S^{-1}(z)H(z)$. The experimental performance of the FIR controller is very similar to its simulation performance.

In Eq. (20), the ANC control error is divided into $e_t = \varepsilon_t + \phi^T(t)\hat{\theta}(t)$. All LS or LMS algorithms are able to drive the convergence of $\varepsilon_t \to 0$ regardless the zeros of $S(z)$. The NMP zeros of $\hat{S}(z)$ only hinder the minimization of $|\phi^T(t)\hat{\theta}(t)|$. The FIR controller is not able to minimize $|\phi^T(t)\hat{\theta}(t)|$ as well as an IIR controller, but it is stable and achieved the smallest $|\phi^T(t)\hat{\theta}(t)|$ achievable by a FIR controller in the H_2 norm sense.

The NMP problem is solvable by adding an extra secondary source. If the secondary paths are co-prime, it is possible to achieve perfect cancellation performance.[18,19] A further study is conducted to integrate the proposed method with the extra-speaker method. The main obstacle is online computations. If a controller involves the online inverse of an $m \times m$ matrix, then two controllers involve the online inverse of a $2m \times 2m$ matrix. This means substantial reduction of sampling rate and controller bandwidth. Faster methods are sought for online minimization of $|\phi^T(t)\hat{\theta}(t)|$.

VI. CONCLUSION

An orthogonal adaptation algorithm is proposed for noninvasive adaptive noise control. All available noninvasive adaptive algorithms consist of two adaptation processes: one for online path modeling and the other for controller tuning. The proposed method solves controller parameters to speed up system convergence. This is the first feature of the proposed method. Some of the noninvasive algorithms perturb controllers to avoid the probing signals. Those methods are not truly noninvasive since controller perturbation is an invasive measure. The second feature of the proposed method is the replacement of persistent excitation with the co-plane requirement. It leads to a truly noninvasive adaptive ANC system without controller perturbation or probing signals. All available noninvasive algorithms involve two adaptation processes; it is very difficult to analyze the coupling effects of the adaptation processes and derive rigorous stability proofs. For the proposed method a rigorous stability analysis is presented using the orthogonal projection method plus the Lyapunov approach. It is proven that the proposed controller

will converge to an optimal one in the minimum H_2 norm sense. Experimental results are presented to demonstrate the performance of the proposed method.

[1]C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (E and FN Spon, London, 1997).

[2]P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).

[3]M. A. Vaudrey, W. T. Baumann, and W. R. Saunders, "Stability and operation constraints of adaptive LMS-based feedback control," Automatica **39**, 595–605 (2003).

[4]E. Bjarnason, "Analysis of the filtered-x LMS algorithm," IEEE Trans. Speech Audio Process. **3**(3), 504–514 (1995).

[5]S. D. Snyder and C. H. Hansen, "The influence of transducer transfer functions and acoustic time delay on the LMS algorithm in active noise control systems," J. Sound Vib. **140**(3), 409–424 (1990).

[6]G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control* (Prentice-Hall, Englewood Cliffs, NJ, 1984).

[7]L. J. Eriksson and M. C. Allie, "Use of random noise for on-line transducer modeling in an adaptive active attenuation system," J. Acoust. Soc. Am. **85**, 797–802 (1989).

[8]M. Zhang, H. Lan, and W. Ser, "Cross-updated active noise control system with online secondary path modeling," IEEE Trans. Speech Audio Process. **9**, 598–602 (2000).

[9]M. Zhang, H. Lan, and W. Ser, "A robust online secondary path modeling method with auxiliary noise power scheduling strategy and norm constraint manipulation," IEEE Trans. Speech Audio Process. **11**(1), 45–53 (2003).

[10]W. C. Nowlin, G. S. Guthart, and G. K. Toth, "Noninvasive system identification for multichannel broadband active noise control," J. Acoust. Soc. Am. **107**, 2049–2060 (2000).

[11]X. Qiu and C. H. Hansen, "An algorithm for active control of transformer noise with on-line cancellation path modelling based on the perturbation method," J. Sound Vib. **240**(4), 647–665 (2001).

[12]B. J. Kim and D. C. Swanson, "Linear independence method for system identification/secondary path modeling for active control," J. Acoust. Soc. Am. **118**(3), 1452–1468 (2005).

[13]B. Widrow and E. Walach, "On the statistical efficiency of the LMS algorithm with nonstationary inputs," IEEE Trans. Inf. Theory **30**(2), 211–221 (1984).

[14]I. Markovsky, J. C. Willems, P. Rapisarda, and B. L. M. De Moor, "Algorithms for deterministic balanced subspace identification," Automatica **41**, 755–766 (2005).

[15]V. Saligrama, "A convex analytic approach to system identification," IEEE Trans. Autom. Control **50**(10), 1550–1566 (2005).

[16]P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control* (Dekker, New York, 1995).

[17]J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE **63**(4), 561–580 (1975).

[18]J. S. Vipperman and R. A. Burdisso, "Adaptive feedforward control of non-minimum phase structural systems," J. Sound Vib. **183**(3), 369–382 (1995).

[19]M. Miyoshi and Y. Kaneda, "Active control of broad-band random noise in a reverberant 3-dimensional space," Noise Control Eng. J. **36**(2), 85–90 (1991).

Spectral velocity estimation in ultrasound using sparse data sets

Jørgen Arendt Jensen
Ørsted•DTU, Building 348, Technical University of Denmark, DK-2800 Lyngby, Denmark

(Received 11 January 2006; revised 4 May 2006; accepted 4 May 2006)

Velocity distributions in blood vessels can be displayed using ultrasound scanners by making a Fourier transform of the received signal and then showing spectra in an *M*-mode display. It is desired to show a *B*-mode image for orientation, and data for this have to be acquired interleaved with the flow data. This either halves the effective pulse repetition frequency f_{prf} or gaps appear in the spectrum from *B*-mode emissions. This paper presents a technique to maintain the highest possible f_{prf} and at the same time show a *B*-mode image. The power spectrum can be calculated from the Fourier transform of the autocorrelation function, and it is shown that the autocorrelation function can be calculated for a sparse set of data where flow and *B*-mode emissions are interspaced. Both short deterministic sequences of emissions and full random sequences can be used. The dynamic range of the sparse sequence is reduced compared to a full sequence. Typically, a reduction of 20 dB is found when using 66% of the data compared to using all data. The theory of the method and examples from simulations of flow in arteries are presented. The audio signal can also be generated from the spectrogram. [DOI: 10.1121/1.2208428]

PACS number(s): 43.60.Hj [TDM] Pages: 211–220

I. INTRODUCTION

Medical ultrasound systems can be used for finding the blood and tissue velocity within the human body.[1–5] This is done by emitting a pulse consisting of a number of sinusoidal oscillations, and then measuring the scattered signal returned from the blood or tissue. The measurement is repeated a number of times, and data are sampled at the depth of interest in the tissue, yielding one sample per pulse emission. The frequency of the received sampled signal is proportional to the velocity of the object along the ultrasound beam, and is given by[5]

$$f_p = \frac{2|\mathbf{v}|\cos\theta}{c} f_0, \tag{1}$$

where $\mathbf{v}$ is the velocity vector, θ is the angle between the ultrasound beam and the velocity vector, c is the speed of sound, and f_0 is the center frequency of the emitted ultrasound pulse.

The velocity distribution for a given spatial position over time can be found by focusing the ultrasound beam at the point of interest. The received rf data are Hilbert transformed to give the in-phase and quadrature component. The data are sampled at the depth of interest to give the complex signal $y(i)$, where i is the pulse emission number. A Fourier transform on the sampled data gives the power spectrum, which corresponds to the velocity distribution, and the short-time Fourier transform displayed over time reveals the temporal variation of the velocity distribution.

The sampled data used for determining the velocity distribution have a sampling frequency of

$$f_{\mathrm{prf}} = \frac{c}{2d}, \tag{2}$$

where d is the depth of interrogation. The maximum frequency that can be correctly found is, thus, $f_{\max} \leqslant f_{\mathrm{prf}}/2$, and the maximum unambiguous velocity is

$$v_{\max} = \frac{c}{2\cos\theta} \cdot \frac{f_{\mathrm{prf}}}{2f_0}. \tag{3}$$

The Fourier transform of the data is performed on short segments of data consisting of usually 128 or 256 samples (pulse emissions) to capture the frequency variation over time of the signal. A Hanning window is often applied on the data, and a fast Fourier transform is then performed. An estimate of the power spectrum $\hat{P}_y(f)$ of the sampled complex signal $y(i)$ for a rectangular window is

$$\hat{P}_y(f) = \frac{1}{N}\left|\sum_{i=0}^{N-1} y(i)\exp(-j2\pi fi/f_{\mathrm{prf}})\right|^2, \tag{4}$$

where i is the sample number, and N is the number of samples in a segment.

The estimate has a significant variance given by[6,7]

$$\mathrm{Var}[\hat{P}_y(f)] \approx P_y^2(f)\left[1 + \left(\frac{\sin 2\pi f/f_{\mathrm{prf}}N}{N\sin 2\pi f/f_{\mathrm{prf}}}\right)^2\right], \tag{5}$$

where $P_y(f)$ is the true power spectrum. The variance is for $f \neq 0$, thus, on the order of the estimate itself, and this is seen as speckle noise in the resulting spectral display.

Often a *B*-mode image should be shown at the same time for orientation and selection of the point of interest, and time must be spent on acquiring this image. This can either

be done by acquiring the *B*-mode data interleaved with the velocity data or by acquiring a full *B*-mode image over a time interval. The first approach will only make every second emission useful for velocity estimation, and this will reduce the pulse repetition frequency by a factor of 2 and reduces the maximum velocity $v_{\max}$ by a factor of 2. The second approach introduces periods where no velocity estimation can be made since data are not acquired, and the true velocity variation therefore cannot be followed.

Several authors have addressed the problem. Kristoffersen and Angelsen[8] used data before the gap to design a filter with roughly the same frequency content as the gap. Using the filter on a Gaussian, random signal then generates data that can fill the gap. The method, however, has to assume that the flow is roughly constant, as a significant acceleration will change the frequency content. Klebæk *et al.*[9] used a neural network to predict the evolution of the mean frequency and the bandwidth of the spectrum, and used this to make a parametric model for filling the gap. Again, the prediction is based on previous data, and abrupt changes in frequency content will make the gap filling wrong. Also, the model might in certain instances not fit the data accurately. Other techniques that take the instantaneous frequency content into account are, thus, needed.

The components in the measured signal will lie in the audio range. Emitted frequencies f_0 of 3 to 5 MHz and velocities of 0.5 to 2 m/s at $\theta=45°$ give frequencies f_p of 1 to 9 kHz, which can be perceived by the human ear. The sound of the measured signal is, thus, often played. This is a problem in the second approach, where there are gaps in the audio stream. This will easily be perceived by the human ear, and the signal cannot be used for faithful audio reproduction.

II. VELOCITY ESTIMATION FOR SPARSE DATA SETS

The method devised here acquires a sparse sequence of sampled data, where flow and *B*-mode emissions are interspaced. It then uses an autocorrelation estimator and a Fourier transform for determining the velocity distribution. This makes it possible to keep the highest attainable velocity equal to the theoretical maximum, and at the same time acquire a *B*-mode image using part of the sparse data sequence. The method can also be used to reconstruct the audio signal as described in Sec. II E. The limit on maximum velocity can also be exceeded by using a cross-correlation estimator to find the mean velocity and then adjusting the velocity distribution according to this estimate. The details of the method are described in the subsequent sections.

A. Power spectrum estimation

The power spectrum of a stochastic signal $y(i)$ is formally calculated from the Fourier transform of the autocorrelation function $R_y(k)$ as

$$R_y(k) \leftrightarrow P_y(f) = \sum_{k=-\infty}^{\infty} R_y(k)\exp(-j2\pi fk\Delta T), \tag{6}$$

where ΔT is the sampling interval. An estimate of the autocorrelation can be calculated by

$$\hat{R}_y(k) = \frac{1}{N-|k|}\sum_{i=0}^{N-k-1} y(i)y^*(i+k), \tag{7}$$

when data are available for a segment of N samples and * denotes complex conjugate. The estimate of the power spectrum is then calculated by applying, e.g., a Hanning window on $\hat{R}_y(k)$ and then performing a Fourier transform. A tradeoff between spectral resolution and spectral estimate variance can be selected by using a window shorter than $2N-1$. The velocity spectrum can thus be found, if a proper estimate of the autocorrelation function can be determined.

B. Sparse data sequences

The autocorrelation calculated by (7) is found by correlating all samples in the signal segment $y(i)$ with a time-shifted version $y(i+k)$ of the signal. It is, however, possible to calculate the correlation estimate, even if some of the samples in the signal are missing. This would be the case if *B*-mode emissions were interleaved with velocity emissions. The correlation is then calculated with fewer values, and this will result in an increased standard deviation of the estimate. In general, the variance of the estimate is inversely proportional to the number of independent values, which here is proportional to $N-k$. Having $M(k)$ missing values will increase the variance by a factor $(N-k)/(N-k-M(k))$. Keeping $M(k)$ moderate compared to N will thus give a moderate increase in variance. The overall variance of the spectral estimate will be determined by the lag values with the highest variance, and therefore it should be ensured that $M(k)$ roughly has the same value for all k.

For a sparse sequence $M(k)$ will in general depend on the lag k, and it must be ensured that all lag values of $\hat{R}_y(k)$ can be calculated with a sufficient accuracy. The estimate of the autocorrelation function is then

$$\hat{R}_y(k) = \frac{1}{N-|k|-M(k)}\sum_{i=0}^{N-k-1} y(i)y^*(i+k), \tag{8}$$

where missing data in the signal are represented by a zero. This equation assumes that only a fixed segment of data is passed to the estimator.

It is also possible to use data from the next segment. The estimate of autocorrelation function is then

$$\hat{R}_y(k) = \frac{1}{N-M(k)}\sum_{i=0}^{N-1} y(i)y^*(i+k), \tag{9}$$

since data for $2N$ samples are available. It is then possible to get a more accurate estimate of higher lags in the autocorrelation function as more data are used, which improve the accuracy of the final velocity estimate. The drawback is a smoothing in time of the calculated power spectrum.

It should also be noted that only the autocorrelation function for positive lags needs to be calculated, since negative lags can be reconstructed from

$$\hat{R}_y(k) = \hat{R}_y^*(-k). \tag{10}$$

The power spectrum is then calculated using (6), and the final display is denoted the auto spectrogram.

The missing values in the sparse sequence can be used for, e.g., *B*-mode emissions so that a *B*-mode image can be acquired simultaneously with the velocity data. An example of a sequence is

$$v\ v\ b\ v\ v\ b \ldots\ ,$$

where v is a velocity emission, and b is a *B*-mode emission. Overlapping for the different lags k is illustrated by

$k=0$	$y(i)$	$v\ v\ b\ v\ v\ b\ \ldots$
	$y(i+0)$	$v\ v\ b\ v\ v\ b\ \ldots$
$k=1$	$y(i)$	$v\ v\ b\ v\ v\ b\ \ldots$
	$y(i+1)$	$v\ b\ v\ v\ b\ v\ \ldots$
$k=2$	$y(i)$	$v\ v\ b\ v\ v\ b\ldots$
	$y(i+2)$	$b\ v\ v\ b\ v\ v\ \ldots$
$k=3$	$y(i)$	$v\ v\ b\ v\ v\ b\ldots$
	$y(i+3)$	$v\ v\ b\ v\ v\ b\ldots$

For each lag k the top line is the received signal sequence and the next row is the lag-shifted version of the signal. A value different from zero in the autocorrelation sum can be calculated if a column contains vv. It can be seen that there is overlap for all lags between velocity data, and all autocorrelation values can therefore be calculated. For this sequence 66% of the time is spent on velocity data and 33% is spent on *B*-mode data acquisition. For imaging to a depth of 15 cm, a pulse repetition frequency of 5 kHz can be maintained, and this gives a frame rate of 15 images/s for images consisting of 100 emissions. Note that it is very important that two adjacent velocity emissions are found in the sequence, since this ensures that the lag 1 autocorrelation can be calculated and the maximum velocity range is thereby maintained.

The frame rate can be lowered by inserting more velocity emissions between each *B*-mode emission, and the *B*-mode frame rate can therefore easily be selected. Other sequences can put more emphasis on the *B*-mode imaging to increase frame rate at the drawback of an increased variance of the spectral estimate. Some other sequences are

B-mode	Flow
40%	60%: $v\ b\ v\ v\ b\ \ldots$
50%	50%: $v\ b\ v\ v\ b\ b\ \ldots$
57%	43%: $v\ b\ v\ v\ b\ b\ b\ \ldots$
62%	38%: $v\ b\ v\ v\ b\ b\ b\ v\ v\ b\ b\ b\ \ldots$

The interleaved emissions can also be used for color flow mapping, which also can be found from sparse sequences.[10] A 50%-50% sequence can also be used to make two spectral estimates at the same time with full velocity range. Hereby the change in flow waveform can be studied over, e.g., a stenosis.

It is also possible to use fully random sequences, where there is no deterministic repetition of the emission sequence. The sequence could for example be determined by using a white, random signal $x(n)$ with a rectangular distribution between zero and one. The determination of whether a *B*-mode or flow emission should be made is determined by

$$e(n) = (x(n) > P_f), \tag{11}$$

where $e(n)=1$ indicates a flow emission and $e(n)=0$ indicates a *B*-mode emission, and P_f is the probability of flow emission. The ratio between flow and *B*-mode emission is then determined by P_f and $P_B=1-P_f$, respectively. It has to be ensured that the autocorrelation can be found for all lags as explained above. The advantage of this approach is that noise appearing in the power spectrum due to a deterministic emission sequence can be spread out over the full spectrum for a random emission sequence. Another advantage is that the time division between flow estimation and *B*-mode imaging can be precisely determined using P_f.

C. Averaging rf data

The pulse emitted for velocity estimation will in general have a number of sinusoidal oscillations to keep the bandwidth small and increase the emitted energy. The received signal is then correlated over the pulse duration, and applying a matched filter to increase the signal-to-noise ratio will increase the correlation to a duration of roughly two pulse lengths. These data can also be used when calculating the autocorrelation as

$$\hat{R}_y(k) = \frac{1}{(N-|k|-M(k))N_r} \sum_{j=0}^{N_r-1} \sum_{i=0}^{N-k-1} y(j+J_d,i) \times y^*(j+J_d,i+k), \tag{12}$$

where j is the rf sample index, J_d is the index for the depth of the range gate start, and N_r is the number of rf samples to average over. Averaging over several rf samples will in general lower the variance of the estimated autocorrelation function and thereby of the spectral estimate.[11]

It is also possible to use data from the next segment. The estimate of the autocorrelation function is then

$$\hat{R}_y(k) = \frac{1}{(N-M(k))N_r} \sum_{j=0}^{N_r-1} \sum_{i=0}^{N-1} y(j+J_d,i) y^*(j+J_d,i+k), \tag{13}$$

since data for $2N$ samples are available. It is then possible to get a more accurate estimate of higher lags in the autocorrelation function as more data are used, which improves the accuracy of the final velocity estimate.

To get an unbiased estimator, it can be beneficial to compensate for the windowing of the received data in the estimate of the autocorrelation function. This is done by

$$\hat{R}_y(k) = \frac{1}{N_w(k)} \sum_{j=0}^{N_r-1} \sum_{i=0}^{N-1} y(j+J_d,i) y^*(j+J_d,i+k), \tag{14}$$

where $N_w(k)$ is the compensation factor given by

TABLE I. Standard parameters for transducer and femoral flow simulation.

Transducer center frequency	f_0	5 MHz
Pulse cycles	M	4
Speed of sound	c	1540 m/s
Pitch of transducer element	w	0.338 mm
Height of transducer element	h_e	5 mm
Kerf	k_e	0.0308 mm
Number of active elements	N_e	128
rf lines for estimation	N	256
rf samples for estimation	N_r	32
Corresponding range gate size		123 mm
Sampling frequency	f_s	20 MHz
Pulse repetition frequency	f_{prf}	15 kHz
Radius of vessel	R	4.2 mm
Distance to vessel center	Z_{ves}	38 mm
Angle between beam and flow	θ	60°

$$N_w(k) = \sum_{i=0}^{N-1} s(i)w(j,i)w(j,i+k)s(i+k). \tag{15}$$

Here, $w(j,i)$ is the two-dimensional window employed on the rf data and $s(i)$ is the sparse sequence which contains a 1 for a velocity emission and 0 for a B-mode emission. In this paper a separable window $w(j,i)$ is used, with a rectangular weighting in the axial direction and a symmetric Blackman window across pulse emissions.

D. Stationary echo canceling

The measured signal will often contain large signal components around low frequencies emanating from the tissue, especially near the vessel wall. This stationary signal must be removed, since it obscures the blood velocity signal and make its spectral visualization difficult. This can be done either in the time or the frequency domain. The first approach is to take the mean value of the signals and subtract that. The mean signal as a function rf sample number j is found from

$$y_{\mathrm{sta}}(j) = \frac{1}{(N - M(0))} \sum_{i=0}^{N-1} y(j,i), \tag{16}$$

where $y_{\mathrm{sta}}(j)$ is the estimated stationary signal. Missing rf signals are replaced by zeros in the sum. The estimated stationary signal is then subtracted from $y(j,i)$ to remove a fully stationary component. This should be done before the autocorrelation function is calculated.

This processing can also be performed in the frequency domain. Here, frequency components around $f=0$ Hz are set to zero in the spectrum to remove the stationary component. The cutoff frequency in the spectrum should be determined from the velocity of the tissue surrounding the blood vessel using (1). This can be used as a supplement to (16), since such tissue motion often is encountered for *in vivo* measurements.

For strong tissue motion in the surrounding tissue (16) might not give a satisfactory suppression of the low-frequency tissue signal. An increased attenuation can then be attained by fitting a higher order polynomial to the sparse data and then subtracting this from the data. A first-order approach was suggested in Ref. 12. Higher order polynomials of order N_p can be fitted using a least-squares approach, where the criterion

$$E_j = \sum_{i=0}^{N-1} \left(y(j,i) - \sum_{k=0}^{N_p} a_k \cdot i^k \right)^2 \tag{17}$$

is minimized like in MATLAB's *polyfit* routine for each depth corresponding to j. Here, a_k are the polynomial coefficients. The polynomial values are then subtracted from the sparse signal to remove the slowly varying tissue signal as

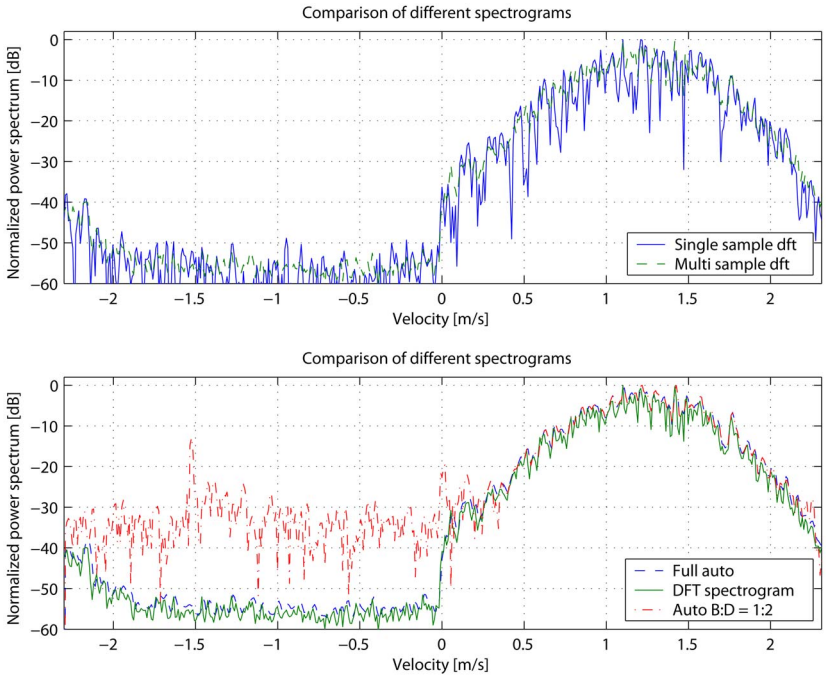


FIG. 1. (Color online) Comparison of different spectrograms for peak systole in the femoral artery at $T=0.1$ s in Fig. 2. The top graph shows the normal spectrogram calculated for a single rf sample per emission and for using averaging over two pulse lengths. The lower graphs also show the spectrogram for the autocorrelation method using the full data and a 1:2 [v v B] sequence.

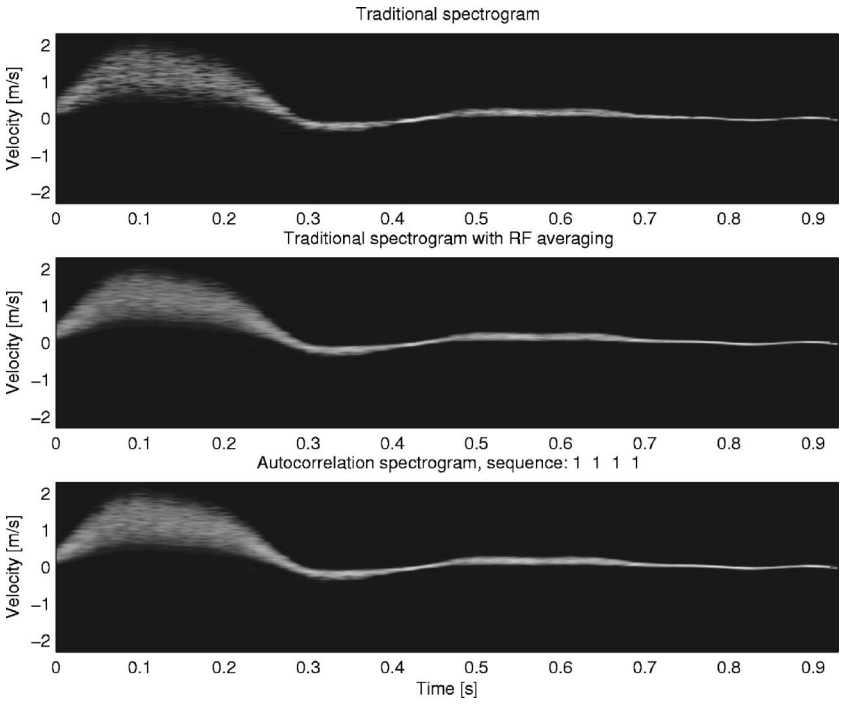


FIG. 2. Normal spectrogram (top) using a single sample per emission, rf-averaged spectrogram, and new method (bottom) for simulated flow in the femoral artery using the full data set.

$$y_{\text{can}}(j,i) = y(j,i) - \sum_{k=0}^{N_p} a_k \cdot i^k, \tag{18}$$

where $y_{\text{can}}(j,i)$ then is used in the estimation of the autocorrelation function. An example of this approach is shown in Sec. III A.

E. Audio reproduction

The audio signal can be regenerated from the estimated autocorrelation function. An appropriate model for the audio signal $y(n)$ is given by

$$y(n) = h(n;n) * e(n), \tag{19}$$

where $h(n;n)$ is a time-varying filter impulse response at time index n and $e(n)$ is a Gaussian, white random signal. Here, $e(n)$ models the many random and independent red blood cells in the vessel; $h(n;n)$ models the velocity spectrum at the given time. The filter is time varying, since the velocity and thereby frequency content varies over the cardiac cycle. The autocorrelation of this is

$$\begin{aligned} R_y(k;n) &= R_h(k;n) * R_e(k) = R_h(k;n) * P_e\delta(k) \\ &= P_e R_h(k;n) \leftrightarrow P_e|H(f;n)|^2, \end{aligned} \tag{20}$$

where P_e is the power of the blood scattering signal and $H(f;n)$ is the Fourier transform of $h(n;n)$. The linear phase impulse response of the filter can then be found from

$$h_l(k;n) = \mathcal{F}^{-1}\{\sqrt{\mathcal{F}\{R_y(k;n)\}}\} = \mathcal{F}^{-1}\{\sqrt{P_e}|H(f;n)|\}, \tag{21}$$

where $\mathcal{F}\{\,\}$ denotes Fourier transform and $\mathcal{F}^{-1}\{\,\}$ inverse Fourier transform. A window can be applied to the impulse response to reduce edge effects. It is also appropriate to mask out small amplitude values in the frequency domain, since this most probably is noise from the reconstruction process.

The phase of the filter is neglected and only a linear phase version is reconstructed. A minimum phase version could be reconstructed using a Hilbert transform, but this is of no consequence since it is a stochastic signal that needs to be made. The final signal is made by convoluting $h_l(k;n)$ with a Gaussian, white random signal as in Ref. 8. This will be the audio signal for a given time segment, and this signal should be added to signals from other segments properly time aligned. To avoid edge effects, a window is applied on the signal segment before addition.

F. Increasing the maximum velocity

The maximum velocity that can be estimated is restricted by (3) due to aliasing. This is really not a restriction on the maximum velocity, but on the widest spread of velocities, where the distance between the lowest and highest velocity at any given time must be less than

$$2v_{\max} = \frac{c}{2\cos\theta}\frac{f_{\text{prf}}}{f_0}. \tag{22}$$

Estimating the mean velocity and adjusting the spectrum to lie around this velocity can therefore increase the maximum velocity range as suggested in Ref. 13.

The mean velocity can be estimated by using the cross-correlation approach developed in Refs. 14 and 15. Two or more rf lines are then cross correlated and the shift in time between them found. This will yield the mean velocity of the

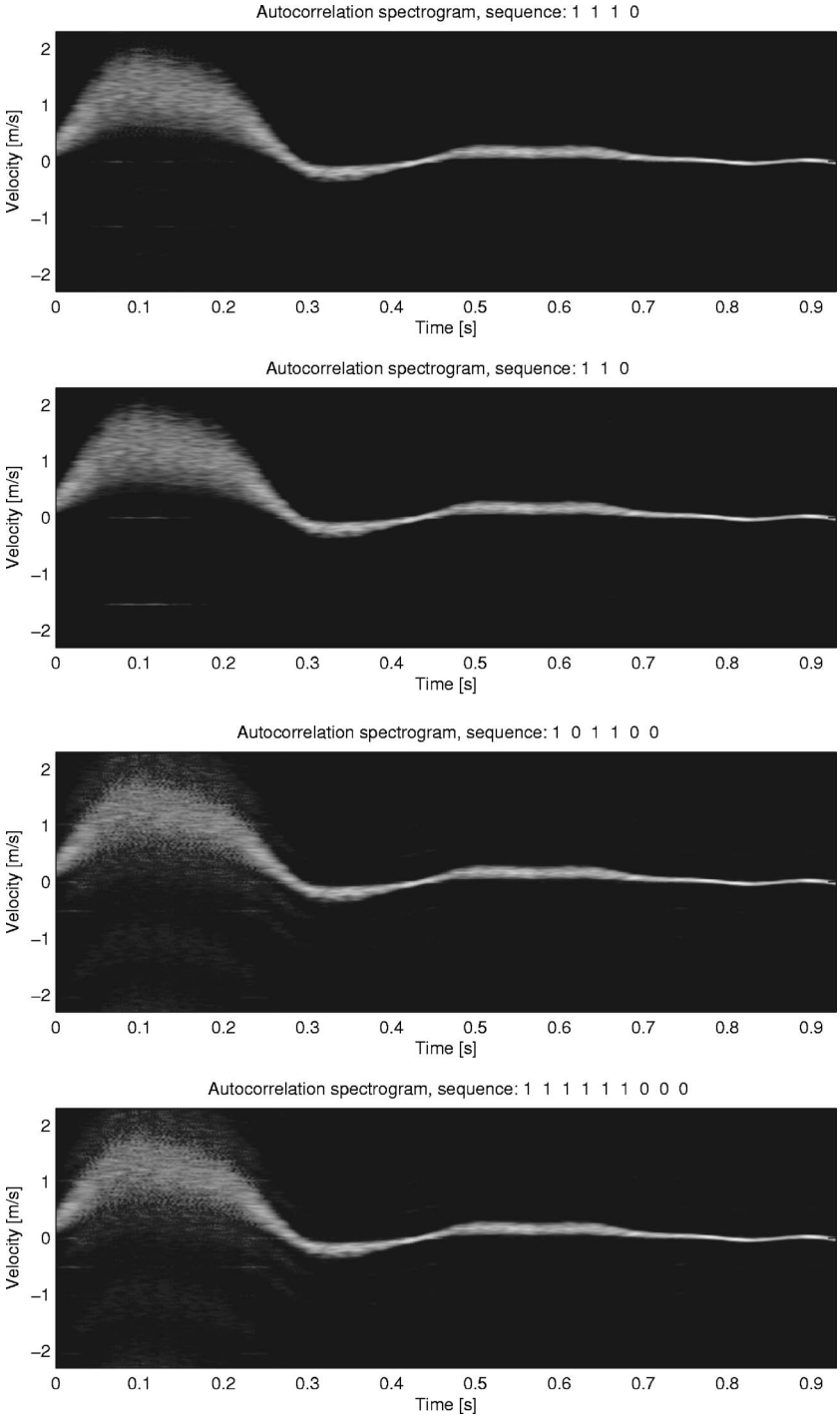


FIG. 3. Deterministically sampled spectrograms using different ratios between *B*-mode and velocity emissions. The emission sequence can be seen in the title, where 1 denotes a flow emission and 0 a *B*-mode or missing emissions. The graphs from top to bottom show results, when reducing the time spent on velocity emissions.

flow. The center of the spectrum is then offset to lie around this mean frequency.

The same data as for the spectral estimation can be used if a narrow pulse is emitted. The spectrum will be widened due to the wide bandwidth of the pulse, but this can be avoided by filtering the received rf data with a narrow-band pulse before calculating the autocorrelation function. This will narrow the bandwidth and the velocity spectrum width.

G. Directional focusing

Data beam formed along the flow direction as described in Ref. 16 can also be used for the flow estimation. The received data then track the movement of the scatterers, and a single or narrow distribution of velocities is then found. This will give a spectrum that is narrower than for taking data out at a range of depths.

III. RESULTS

The method is investigated using simulated data, where the exact result of the velocity estimation is known. Hereby both the traditional spectrogram and the new auto spectrogram can be calculated.

The FIELD II program[17,18] was used for the simulation.[19] The Womersley model[20,21] for pulsating flow in a vessel was used for generating realistic flow data from the femoral artery. This artery was selected since the flow is highly pulsat-

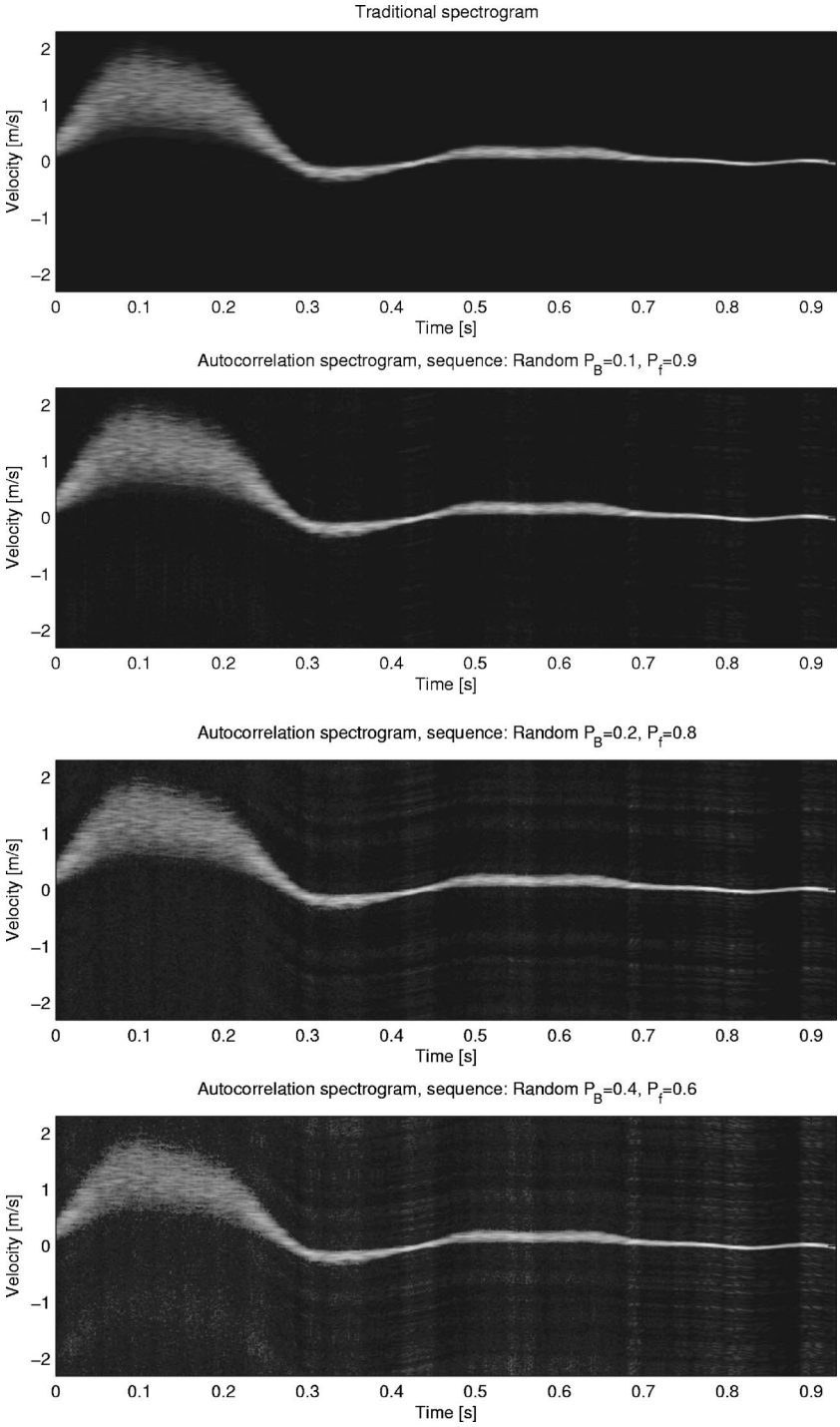


FIG. 4. Randomly sampled spectrograms using different ratios between *B*-mode and velocity emissions. P_f denotes the relative time spent on velocity emissions. The top graph shows the normal spectrogram when using all data. The other graphs show results when reducing the time spent on velocity emissions using random sampling.

ing, and it will therefore reveal if the estimator has problems with following rapid variations in velocity. A linear array transducer with 128 elements was used with a Hamming apodization in both transmit and receive. Other parameters for the simulation can be seen in Table I. The number of point scatterers was 43 468 and the stationary tissue outside the vessel had a scattering amplitude 100 times higher than inside the vessel to mimic the higher scattering of tissue. The sampling frequency for the simulation was 100 MHz, but the data were subsequently decimated to a sampling frequency of 20 MHz to reduce memory demands. A Hilbert transformation was then performed on the rf data to yield the in-phase and quadrature component in $y(i)$. A sparse set of data is emulated by inserting zeros for missing data.

The result of the processing is shown in Fig. 1. A reference spectrogram is shown in the top graph. It is calculated by

$$\hat{P}_y(f) = \frac{1}{N_r}\sum_{j=0}^{N_r-1}\left|\frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{2}\left(1-\cos\left(\frac{2\pi i}{i+1}\right)\right)y(j,i) \times \exp(-j2\pi f i/f_{\mathrm{prf}})\right|^2, \quad (23)$$

where the data are weighted by a Hanning window, Fourier transformed, and averaged over the range gate duration N_r. A

spectogram without averaging of rf samples is also shown in the top graph. It is clearly shown how the rf averaging reduces the standard deviation of the estimate and makes it more smooth. The lower graph also shows two spectrograms calculated using the autocorrelation method. The blue line shows the result, when the full data sequence is available. The auto spectrogram is calculated using (14), where the autocorrelation function is averaged over a range gate duration of two pulse lengths to emulate the function of a matched filter on the data. A symmetric Blackman window weighted the data across pulse emissions and a rectangular window in the axial direction. Echo canceling is performed using (16) on the sparse data set. A Blackman window was multiplied onto the autocorrelation function before the power spectrum was found. The estimate is very close to the direct spectral estimate, with roughly the same standard deviation of the estimates. The red curve shows the result from using a 2:1 $[v\ v\ b]$ sequence, with two velocity emissions and one *B*-mode emission. The velocity spectrum is accurately estimated, but the noise from positive velocities has been increased from roughly −55 dB to around −30 dB. It is thus possible to use the method for images with a dynamics range of roughly 30 dB. The level will depend on the sparseness of the sequence, and the level will in general be increased with increasing sparseness as shown in the following plots.

In Fig. 2 the process is repeated continuously and the spectra are displayed as a gray-scale image as a function of time and velocity. The display has been compressed to a dynamic range of 40 dB, and the spectrum is calculated for 256 samples for every 2.1 ms. It can be seen that the new method yields a spectrum closely corresponding to the traditional method.

In Fig. 3 the top graph shows the result from using 25% of the time on *B*-mode acquisitions ($v\ v\ v\ B$ sequence), where every fourth received signal was replaced by zeros. The autocorrelation estimate was calculated as described above. It can be seen that a slightly more smooth spectrum is found, although 25% of the data is missing. In the next graph 33% of the time is spend on *B*-mode acquisition ($v\ v\ b$ sequence), and then 50% of the time ($v\ b\ v\ v\ b\ b$ sequence) in the next graph. The last sequence can also be used for interleaving two auto spectrograms in different directions with full velocity range. The noise in the spectrograms is progressively increased, when more time is spent on *B*-mode acquisition, but only for the last plot is the noise becoming significant, especially at the systolic phase in the cardiac cycle. For the $v\ v\ b$ sequence 33% of the time can be used for *B*-mode imaging and $0.33 f_{\mathrm{prf}} = 0.33 \cdot 15 \cdot 10^3 = 4950$ lines/s can be acquired for *B*-mode imaging. This corresponds to 24 images/s consisting of 200 lines, which is a normal *B*-mode frame rate. The pulse repetition frequency must be reduced to 5 kHz if imaging is performed to a depth of 15 cm. The *B*-mode frame rate is then reduced to 8 Hz, which in many applications is still acceptable. The last graph in Fig. 3 shows a longer sequence of velocity emissions and three *B*-mode emissions, so that 33% of the time is spent on *B*-mode imaging. This sequence can be used to make small blocks of *B*-mode emissions and reduce the influence between *B*-mode and velocity emissions. The spectrum at peak systole, however, gets significantly more blurred and this might preclude the automatic detection of peak velocity or other derived quantities from the spectrum.

Figure 4 shows the employment of full random sampling as described in Sec. II B using (11). The top graph is the reference spectrogram made using (23) and the second graph uses random sampling with $P_f = 0.9$. Ten percent of the time is thus used for *B*-mode emissions. The next graph uses 20% and the last 40%. Again, a progressive increase in the amount of noise is seen with an increase in time spent on *B*-mode imaging, and this makes the last case with $P_f = 0.6$ unacceptable. Little noise is seen for $P_f = 0.9$, and here $P_B f_{\mathrm{prf}} = 0.1 \cdot 15 \cdot 10^3 = 1500$ lines/s are acquired for the *B*-mode images. This corresponds to 15 images/s consisting of 100 lines, which is sufficient to follow fairly rapid tissue motion.

A. Carotid artery with strong tissue motion

The previous section did not include a strong tissue motion, and it is easy here to remove the stationary component just by subtracting the mean value of the input signal for a given depth. To include a significant, time-varying tissue component, a simulated example for the carotid artery with a strong tissue motion is shown in this section. The tissue motion is derived from the pulsating flow described by the Womersley theory. It is calculated as the derivative of the

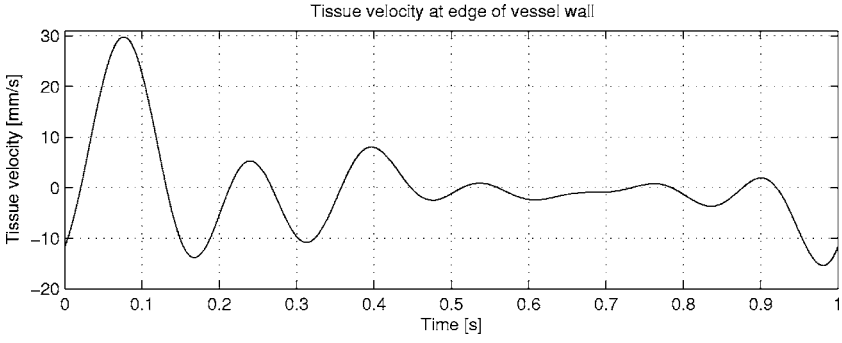


FIG. 5. Velocity of the tissue signal in the radial direction at the edge of the simulated carotid artery.

TABLE II. Standard parameters for carotid flow simulation.

Radius of vessel	R	6 mm
Distance to vessel center	Z_{ves}	40 mm
Angles between beam and flow	θ	60°
Velocity scaling factor	k_t	0.001
Decay factor	τ_t	2 mm

blood velocity at a radial position of $r=0.95R$, where R is the radius of the vessel. The tissue velocity $v(t)$ is calculated as

$$
\begin{aligned}
v(t,) &= \sum_{m=1}^{\infty} k_t \omega |V_m| |\psi_m(0.95, \tau_m)| \cos(m\omega t - \phi_m \\
&\quad + \chi_m(0.95, \tau_m)), \\
\psi_m(r/R, \tau_m) &= \frac{\tau_\alpha J_0(\tau_m) - \tau_\alpha J_0(r/R\tau_m)}{\tau_\alpha J_0(\tau_m) - 2J_1(\tau_m)}, \\
\chi_m(r/R, \tau_m) &= \angle\, \psi(r/R, \tau_m), \\
\tau_m &= j^{3/2} R \sqrt{\frac{\rho}{\mu}\omega_m},
\end{aligned}
\tag{24}
$$

where $J_n(x)$ is the nth-order Bessel function, $\angle\psi(r/R, \tau_m)$ denotes the angle of the complex function ψ, and $|\psi|$ denotes its amplitude. The function ψ is dependent on radial position in the vessel, angular frequency, and fluid properties.[5,21] The variables V_m and ϕ_m are the amplitude and phase, respectively, of the Fourier components describing the pulsating flow as given in Refs. 5 and 21. The constant k_t scales the tissue velocity to lie in the range of mm/s as measured in Ref. 22. The tissue velocity in the radial direction at the vessel boundary is shown in Fig. 5. The peak velocity is chosen to be higher than normally encountered in the patient (30 mm/s) to show a worst-case example. The tissue scatterers are moved with the calculated velocity at the vessel edge and the motion is then exponentially attenuated in the radial direction, so that the motion gets progressively smaller further away from the vessel. The tissue velocity as a function of radial distance is given as

$$
v_t(t, r_t) = v(t)\exp(-(r_t - R)/\tau_t), \tag{25}
$$

where r_t is the radius from the vessel center, R is the vessel radius, and τ_t is the decay constant. The Fourier components V_m and ϕ_m for the velocity profile are taken from Ref. 23 and the other parameters are given in Table II. The scattering of the tissue is assumed to be 40 dB stronger than the blood scatterers.

The spectrograms obtained for this data are shown in Fig. 6. The top graph shows the spectrogram when using mean subtraction for echo canceling as given by (16). The two lower graphs use a third-order polynomial fit as described by (18). All components below 120 Hz in the spectrum have been set to zero before display.

It can be seen that a satisfactory spectrum can be obtained, although the data contain a significant stationary component. The polynomial canceling gives a slightly better suppressed stationary signal, and the method is therefore better suited for strong tissue signals. There is still a small stationary component present, but this can be removed in the frequency domain. The spectrogram in the lowest graph can be used for either a high frame rate B-mode system or it can be used for having two simultaneous spectral measurements at the same time, so that, e.g., the velocity distribution before and after a stenosis can be evaluated.

IV. AUDIO GENERATION EXAMPLE

The audio signals for the examples in the last section was generated using the method described in Sec. II E and the examples are stored on the EPAPS website

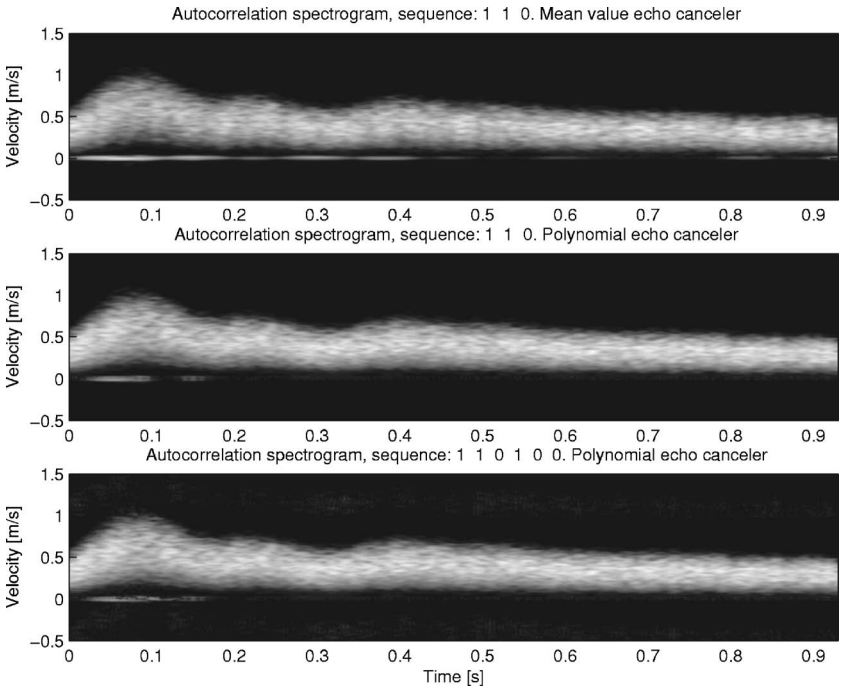


FIG. 6. Spectrograms for the carotid artery with tissue motion. The top graph shows the spectrogram when using mean subtraction for echo canceling and the two lower graphs use a third-order polynomial fit.

(http://www.aip.org/pubservs/epaps.html).[24] The reference signal generated from all data is stored in the file named reference_audio.wav. The sound generated using the auto spectrogram and the full data is in auto_full.wav. To reduce noise all components in the spectrum that have an amplitude less than 2% of the spectrum peak amplitude are set to zero. One can hear that the two files are nearly indistinguishable. Sound files for the sparse sequences are stored in the files auto_bd_1_3.wav, auto_bd_1_2.wav, and auto_bd_3_3.wav, where bd_1_2 denotes the $v\ v\ B$ sequence with two velocity emissions and one B-mode emission. The bd_1_3 is nearly indistinguishable from the reference file, whereas the bd_1_2 and bd_3_3 files contain progressively more noise. The random sampling files are contained in files named auto_random_pb_X_X.wav, where pb_X_X denotes the factional time spent on B-mode imaging. Files are found for the fractions 0.1, 0.2, and 0.4. The 0.1 file is nearly indistinguishable from the reference file, whereas the other two files contain progressively more noise. The noise seems more prominent than for the deterministic sampling sequences for the same amount of time spent on B-mode acquisitions. Other ways of reconstructing the audio signal might change this.

V. CONCLUSION

A method for preserving the full velocity range in duplex ultrasound systems has been presented. The method samples both velocity and B-mode emissions interleaved in either a deterministic or random order and the full velocity spectrum can be determined by estimating the autocorrelation function from the sparse data set. The full velocity range can be preserved, if consecutive velocity emissions are performed at some point in the sequence. The accuracy of the estimated spectrum and the noise in it is determined from the fraction of time spent on velocity emissions. A higher fraction gives a better estimate, but also a lower frame rate for the B-mode image. It has also been shown how the audio data can be recovered from the sparse sequence of data.

[1]S. Satomura, "Ultrasonic Doppler method for the inspection of cardiac functions," J. Acoust. Soc. Am. **29**, 1181–1185 (1957).

[2]D. W. Baker, "Pulsed ultrasonic Doppler blood-flow sensing," IEEE Trans. Sonics Ultrason. **SU-17**, 170–185 (1970).

[3]P. N. T. Wells, "A range gated ultrasonic Doppler system," Med. Biol. Eng. **7**, 641–652 (1969).

[4]F. E. Barber, D. W. Baker, A. W. C. Nation, D. E. Strandness, and J. M. Reid, "Ultrasonic duplex echo-Doppler scanner," IEEE Trans. Biomed. Eng. **BME-21**, 109–113 (1974).

[5]J. A. Jensen, *Estimation of Blood Velocities Using Ultrasound: A Signal Processing Approach* (Cambridge University Press, New York, 1996).

[6]G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications* (Holden-Day, San Francisco, 1968).

[7]J. S. Bendat and A. G. Piersol, *Random Data. Analysis and Measurement Procedures*, 2nd ed. (Wiley, New York, 1986).

[8]K. Kristoffersen and B. A. J. Angelsen, "A time-shared ultrasound Doppler measurement and 2-D imaging system," IEEE Trans. Biomed. Eng. **BME-35**, 285–295 (1988).

[9]H. Klebæk, J. A. Jensen, and L. K. Hansen, "Neural network for sonogram gap filling," in *Proceedings IEEE Ultrasonic Symposium*, Vol.2, pp. 1553–1556 (1995).

[10]W. Wilkening, B. Brendel, and H. Ermert, "Fast, extended velocity range flow imaging based on nonuniform sampling using adaptive wall filtering and cross correlation," in *Proceedings IEEE Ultrasonic Symposium*, pp. 1491–1494 (2002).

[11]T. Loupas, J. T. Powers, and R. W. Gill, "An axial velocity estimator for ultrasound blood flow imaging, based on a full evaluation of the Doppler equation by means of a two-dimensional autocorrelation approach," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 672–688 (1995).

[12]A. P. G. Hoeks, M. Hennerici, and R. S. Reneman, "Spectral composition of Doppler signals," Ultrasound Med. Biol. **17**, 751–760 (1991).

[13]P. Tortoli, "A tracking FFT processor for pulsed Doppler analysis beyond the Nyquist limit," IEEE Trans. Biomed. Eng. **36**, 232–237 (1989).

[14]O. Bonnefous and P. Pesqué, "Time domain formulation of pulse-Doppler ultrasound and blood velocity estimation by cross correlation," Ultrason. Imaging **8**, 73–85 (1986).

[15]S. G. Foster, *A pulsed ultrasonic flowmeter employing time domain methods*, Ph.D. thesis, Dept. Elec. Eng., University of Illinois, Urbana IL, 1985.

[16]J. A. Jensen, "Directional velocity estimation using focusing along the flow direction. I. Theory and simulation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 857–872 (2003).

[17]J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).

[18]J. A. Jensen, "FIELD: A program for simulating ultrasound systems," Med. Biol. Eng. Comput.10th Nordic-Baltic Conference on Biomedical Imaging, **4**, Supplement 1, Part 1, 351–353 (1996).

[19]More information about the program can be found on its website at http://www.es.oersted.dtu.dk/staff/jaj/field/. It can also be freely downloaded from the site.

[20]J. R. Womersley, "Oscillatory motion of a viscous liquid in a thin-walled elastic tube. I. The linear approximation for long waves," Philos. Mag. **46**, 199–221 (1955).

[21]D. H. Evans, "Some aspects of the relationship between instantaneous volumetric blood flow and continuous wave Doppler ultrasound recordings. III," Ultrasound Med. Biol. **8**, 617–623 (1982).

[22]M. Schlaikjer, S. Torp-Pedersen, and J. A. Jensen, "Simulation of rf data with tissue motion for optimizing stationary echo canceling filters," Ultrasonics **41**, 415–419 (2003).

[23]D. H. Evans, W. N. McDicken, R. Skidmore, and J. P. Woodcock, *Doppler Ultrasound, Physics, Instrumentation, and Clinical Applications* (Wiley, New York, 1989).

[24]See EPAPS Document No. E-JASMAN-120-050607 for audio and wave files. This document can be reached through a direct link in the online article's HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).

Matched-field geoacoustic inversion with a horizontal array and low-level source

Dag Tollefsen
Norwegian Defence Research Establishment (FFI), Box 115, 3191 Horten, Norway

Stan E. Dosso and Michael J. Wilmut
School of Earth and Ocean Sciences, University of Victoria, Victoria, British Columbia, Canada V8W 3P6

(Received 10 January 2006; revised 24 April 2006; accepted 24 April 2006)

This paper applies geoacoustic inversion to acoustic-field data collected on a bottom-moored horizontal line array due to a continuous-wave towed source at a shallow water site in the Barents Sea. The source transmitted tones in the frequency band of 30–160 Hz at levels comparable to those of a merchant ship, with resulting signal-to-noise ratios of 9–15 dB. Bayesian inversion is applied to cross-spectral density matrices formed by averaging spectra from a sequence of time-series segments (snapshots). Quantifying data errors, including measurement and theory errors, is an important component of Bayesian inversion. To date, data error estimation for snapshot-averaged data has assumed either that averaging reduces errors as if they were fully independent between snapshots, or that averaging does not reduce errors at all. This paper quantifies data errors assuming that averaging reduces measurement error (dominated by ambient noise) but does not reduce theory (modeling) error, providing a physically reasonable intermediary between the two assumptions. Inversion results in the form of marginal posterior probability distributions are compared for the different approaches to data error estimation, and for data collected at several source ranges and bearings. Geoacoustic parameter estimates are compared with data from supporting geophysical measurements and historical data from the region.
[DOI: 10.1121/1.2205132]

PACS number(s): 43.60.Pt, 43.30.Pc [AIT] Pages: 221–230

I. INTRODUCTION

Geoacoustic characterization of the seabed by inversion of acoustic data has been subject to extensive research, with perhaps the widest attention given to matched-field inversion (MFI) methods that exploit the spatial coherence of the acoustic field as measured at an array of sensors, e.g., Refs. 1–18. In much of the initial experimental work, MFI methods were typically applied to high-level broadband or transient-signal data from controlled sources received at a vertical line array of sensors. Inversion of horizontal line array (HLA) data has been studied in the context of towed arrays using towed sources[10–12] and a ship-noise source.[13] Practical considerations, such as array stability, covertness, ease of deployment, and sustainability to shipping activity, can make a HLA laid on the seabed the preferable choice for a deployable sensor system. The use of a bottom-moored HLA for matched-field geoacoustic inversion is a topic of more recent interest.[14–17]

This paper considers the matched-field geoacoustic inversion of acoustic data recorded on a bottom-mounted HLA due to a low-level continuous-wave (cw) source in the context of a shallow water experiment conducted by the Norwegian Defence Research Establishment (FFI) in the Barents Sea. An acoustic source was towed along radial tracks at ±30° bearing relative to array endfire and transmitted low-frequency (30–160 Hz) cw tones at sound levels comparable to those of merchant ships, with resulting signal-to-noise ratio (SNR) at the array of 9–15 dB. Thus, the experiment emulated conditions (apart from source depth) closer to those of using noise from "ships of opportunity" for MFI rather than those of more-traditional inversion experiments using high-level broadband sources. The HLA measurements also differ from previous reported experiments in that the array is longer (900 m) but sparser (18 elements), and in that the experimental site was located in a continental shelf environment characterized by relatively compact (glacigenic) sedimentation. In addition to the acoustic data, a number of types of geophysical data (seismic reflection and refraction, bottom-penetrating sonar, gravity core) were collected at the experiment site to provide independent information on bottom properties for comparison with the MFI results.

The Bayesian geoacoustic inversion applied in this paper consists of estimating parameter values, uncertainties, one- and two-dimensional marginal probability distributions, and interparameter correlations for a layered seabed model using a nonlinear numerical approach based on Gibbs sampling.[7–9] Inversion is applied to cross-spectral density matrices formed by averaging spectra from a sequence of time-series segments (data snapshots). Defining the data uncertainties, including both measurement and theory errors, is an important component of Bayesian analysis. To date, data error estimation for MFI of snapshot-averaged acoustic data has been based on one of two assumptions: Either that the averaging procedure reduces the errors as if they are fully independent from snapshot to snapshot (an optimistic assumption),[5] or that the averaging procedure does not reduce the error at all (a pessimistic assumption).[13] This paper develops a new approach to quantifying errors for snapshot-averaged data

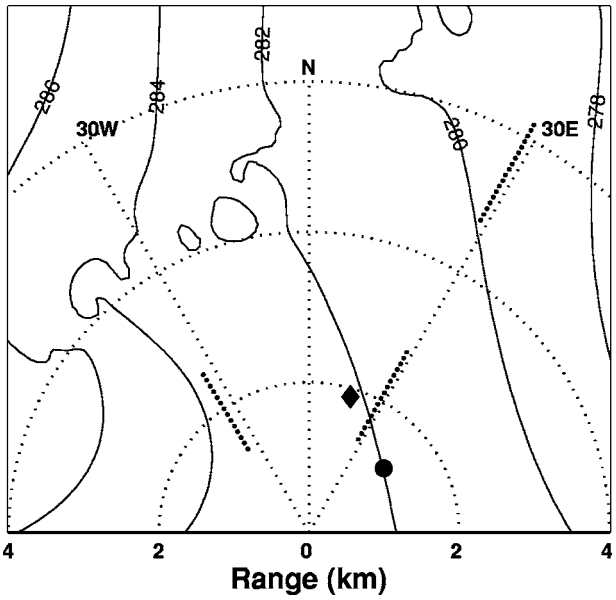


FIG. 1. Experiment region. The HLA was laid north-south with the north end at the origin of the coordinate system. Solid lines are water depth (m) contours; dotted lines indicate range (km) and azimuth (°) to the array; heavy dotted lines indicate source tow tracks. The diamond indicates the location of the gravity core; the filled circle indicates the sonobuoy location for the seismic refraction experiment.

based on the assumption that the averaging procedure reduces measurement error (dominated by ambient noise which may reasonably be assumed independent from snapshot to snapshot), but does not significantly reduce theory error (dominated by modeling errors, which are generally not independent). This approach provides a physically reasonable intermediate result between the optimistic and pessimistic assumptions.

The remainder of this paper is organized as follows. Section II describes the Barents Sea experiment, acoustic data, and supporting environmental measurements. Section III outlines the Bayesian inversion, including error estimation for snapshot-averaged data. Section IV presents the MFI results, including an examination of the effects of error estimation and source range and bearing for the experimental data, and a comparison with reference geophysical data. Finally, Sec. V summarizes and discusses this work.

II. EXPERIMENT, DATA, AND MODEL

A. Acoustic measurements and data processing

The acoustic experiment analyzed here was conducted in June 2003, in a little-surveyed area of the southwestern Barents Sea, representing a typical high-latitude continental-shelf environment. A 900-m-long HLA was deployed in a north-south orientation on the (relatively flat) seabed at a depth of approximately 282 m, as shown in Fig. 1. For matched-field processing of HLA data, a long array is desirable to increase the spatial sampling of the acoustic field. A length requirement can be formulated quantitatively in terms of the effective vertical aperture of the horizontal array.[19] For the scenarios considered in this paper, the effective vertical aperture of the HLA is equivalent to approximately two-thirds of the water depth. The array was comprised of 18 sensors (hydrophones), with 7 sensors spaced at 10-m intervals at the north end of the array (nearest the source tows), and sensor separation increasing over the remaining 11 elements to a maximum of 160 m at the south end. The position and orientation of the HLA were determined using travel-time measurements from an airgun source towed in a circle of a 1 km radius around the center of the array. Results indicated that the array did not deviate significantly from the nominal linear configuration, and a straight HLA was assumed for the geoacoustic inversions.

The acoustic source was towed at nominal depth of 54 m and speed of 5 kts by the FFI vessel R/V H U SVERDRUP II along radial tracks oriented at 30° angles east and west of endfire to the HLA (henceforth referred to as the east and west tracks), as illustrated in Fig. 1. The source-receiver ranges along the tracks extended from 1.4–2.8 km and 4.7–6.2 km. The source transmitted cw tones at five frequencies within the band 30–160 Hz.

Acoustic pressure-time series at the array were digitized at 3 kHz and transmitted from a surface buoy to the receiving ship via radio-frequency data link where they were recorded on computer disk. The recorded time series were fast Fourier transformed and from resulting spectra, the frequency bin of maximum power at or near each nominal source frequency was selected for analysis. Cross-spectral density matrix (CSDM) estimates $\hat{\mathbf{R}}_f$ were formed by averaging over the complex pressures corresponding to K time-series data segments (snapshots):

$$\hat{\mathbf{R}}_f = \frac{1}{K}\sum_{k=1}^{K} \mathbf{d}_{f,k}\mathbf{d}_{f,k}^{\dagger}, = \langle \mathbf{d}_f \mathbf{d}_f^{\dagger} \rangle. \tag{1}$$

In Eq. (1), $\mathbf{d}_{f,k}$ represents the vector of complex-pressure data along the array at the fth frequency corresponding to the kth snapshot and $\langle\cdot\rangle$ represents snap-shot averaging († indicates conjugate transpose). The processing sequence consisted of averaging $K=5$ snapshots each of length 6.6 s, with a Hamming windowing function applied and 50% snapshot overlap. The total averaging time was 19.8 s, over which the source moved approximately 50 m. Ambient noise levels were estimated using the Order-Truncate-Average algorithm[20] in small frequency bands surrounding each processing frequency. The average SNR at each frequency is defined by

$$\mathrm{SNR}_f = 10\log_{10}\frac{\langle|\mathbf{d}_f|^2\rangle - \langle|\mathbf{n}_f|^2\rangle}{\langle|\mathbf{n}_f|^2\rangle}, \tag{2}$$

where $\mathbf{n}_f$ represents noise vectors. The estimated SNR values vary from 9 to 15 dB for the data considered in this paper.

B. Environmental data

As part of the experiment, several types of supporting oceanographic and geophysical measurements were made from the R/V H U SVERDRUP II, as summarized in Fig. 2. Water-column temperature and salinity profiles were measured using a conductivity-temperature-depth probe at the beginning of each source track, and by expendable bathythermograph casts from the source ship along the

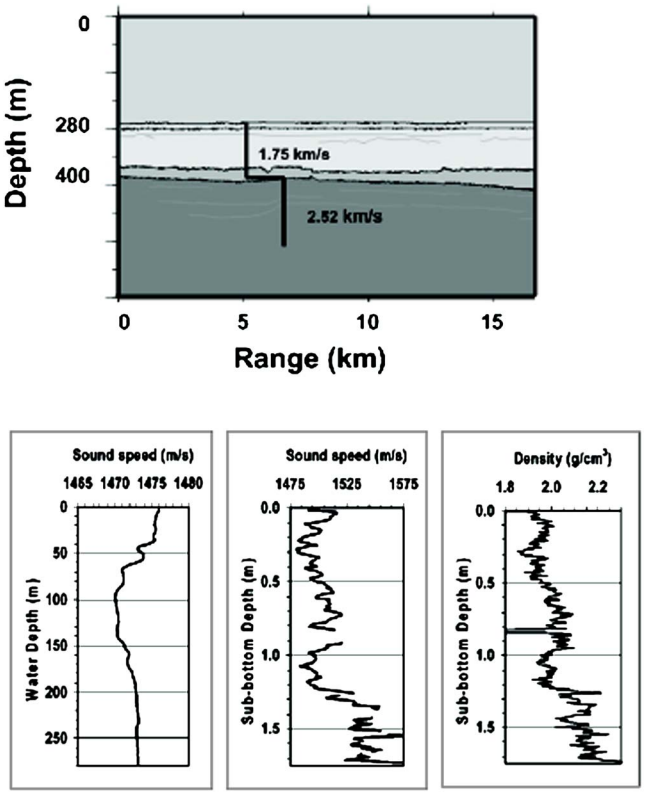


FIG. 2. (Color online) Environmental data. Upper plot: Interpreted seismic section, with inset sound-speed profile in seabed from wide-angle bottom refraction measurements; HLA located at zero range. Lower plots: Water-column sound-speed profile (left panel), and surficial sediment sound-speed (middle) and density (right) profiles from gravity core measurements.

tracks. The sound-speed profiles calculated from these measurements exhibited little variation over time or position, and indicated a higher-speed (warmer) surface layer of approximately 1475 m/s extending to 40 m depth; below this layer, the sound speed decreased abruptly to 1470 m/s then increased to 1473 m/s at the seabed (see Fig. 2 for a representative profile). In the inversions described in the following sections, each acoustic track was analyzed using a sound-speed profile measured during that track.

To obtain an indication of the sub-bottom structure, seismic-reflection and bottom-penetrating sonar data were collected simultaneously along a track that closely approximated the east track of the acoustic experiment. The seismic reflection survey employed a source array of two 50-cubic -in. airguns recorded on a 10-m single-channel towed streamer. The reflection data indicate the seabed consists of an upper layer, approximately 120–140 m thick, which is interpreted to be Quaternary sediments, overlying several layers of consolidated material (upper and lower Triassic sediments).[21] Data from a Kongsberg TOPAS PS018 parametric sub-bottom profiling sonar (operated with a frequency-modulated sweep at frequencies of 2–4 kHz) suggested some internal structure in the sediments near the seafloor and a possible weak reflector at approximately 10–20 m depth. The seismic reflection and sonar data, interpreted in Fig. 2, indicate essentially range-independent structure within the Quaternary sediments along the track.

Further estimates of the geophysical properties were obtained from a shallow gravity core (penetration depth: 1.7 m) and a wide-angle bottom refraction survey based on recordings of the airgun source on a sonobuoy hydrophone (see Fig. 1). Analysis of the core indicated a silty-clay sediment composition, with sound speed and density of approximately 1500 m/s and 2.0 g/cm^3, respectively, over the top 0.8 m, increasing to about 1520 m/s and 2.1 g/cm^3 over the lower 0.9 m.[22] The relatively high sound speed and density values from the core measurement are consistent with values reported for glacigenic sediments of this type from other sites in the Barents Sea.[23,24] Standard slope-intercept analysis of the refraction data[25] indicated an overall sound speed of 1745 m/s for the Quaternary sediments and 2520 m/s for the Triassic sediments (analysis assumes constant-speed layers).

C. Geoacoustic model

A simple geoacoustic model of the seabed, consistent with the geophysical data, was developed for the purposes of inversion (Sec. IV). The model consists of an upper layer with depth-dependent properties overlying a homogeneous half-space. The two layers are designed to represent geoacoustic variations within the upper few tens of meters of the Quaternary sediment layer, since a sensitivity study indicated that the acoustic data were not sensitive to properties at a greater depth within the Quaternary sediments or to the underlying Triassic sediments (Fig. 2). The parameters that describe the upper model layer are the layer thickness, h, sound speed at the top and bottom of the layer, c_1 and c_2, attenuation coefficient at the top and bottom of the layer, α_1 and α_2, and a depth-independent density ρ_1. The lower layer is described by (constant) sound-speed and attenuation values that are identical to those at the base of the upper layer (c_2 and α_2, respectively), and by an independent density ρ_2. This parametrization provides continuous profiles for sound-speed and attenuation in the upper sediments where strongly discontinuous layering is not expected. Constant densities are assumed in the upper and lower layers since low-frequency acoustic-field data generally show little sensitivity to density gradients within layers.[26]

In addition to the seabed geoacoustic parameters described above, the model determined by MFI also includes several geometric parameters of the experiment which are not otherwise known to sufficient accuracy (i.e., the geometric parameters are essentially nuisance parameter in the inversion). These parameters include the water depth, D, source depth and bearing, z and b, and a correction to the source-receiver range, dr. The inversion search bounds applied for the geoacoustics and geometric parameters throughout this paper are given in Table I.

III. INVERSE THEORY

A. Bayesian geoacoustic inversion

This section briefly summarizes a Bayesian approach to geoacoustic inversion;[4–9] more complete treatments of Bayesian theory can be found elsewhere.[27,28] Let **m** and **d** repre-

TABLE I. Model parameters and search bounds used in inversions.

Parameter and units	Lower bound	Upper bound
h (m)	1.0	40
c_1 (m/s)	1450	1900
c_2 (m/s)	c_1	c_1+30h
ρ_1 (g/cm^3)	1.40	3.00
ρ_2 (g/cm^3)	1.40	3.00
α_1 (dB/m kHz)	0.01	1.00
α_2 (dB/m kHz)	0.01	1.00
D (m)	278	288
dr (km)	–0.10	0.15
z (m)	50	58
b(°)	27	30

sent model and data vectors, respectively, with elements considered to be random variables. Bayes' rule may be expressed

$$P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})P(\mathbf{m}), \tag{3}$$

where $P(\mathbf{m}|\mathbf{d})$ is the posterior probability density (PPD), $P(\mathbf{d}|\mathbf{m})$ represents data information, and $P(\mathbf{m})$ is prior information. Interpreting $P(\mathbf{d}|\mathbf{m})$ as a function of $\mathbf{m}$ for the measured data defines the likelihood function $L(\mathbf{m})$, which can generally be expressed $L(\mathbf{m}) \propto \exp[-E(\mathbf{m})]$, where $E(\mathbf{m})$ is the data error function (considered in Sec. III B). The PPD becomes

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m})]}{\int \exp[-\phi(\mathbf{m}')]\mathrm{d}\mathbf{m}'}, \tag{4}$$

where a generalized error function (combining data and prior) is defined

$$\phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m}), \tag{5}$$

and the domain of integration spans the parameter space.

The multidimensional PPD is interpreted in terms of properties defining parameter estimates and uncertainties, such as the maximum *a posteriori* (MAP) estimate, the posterior mean estimate, the model covariance matrix, parameter mean deviations (MD), and marginal probability distributions defined, respectively, as

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max}\{P(\mathbf{m}|\mathbf{d})\}, \tag{6}$$

$$\bar{\mathbf{m}} = \int \mathbf{m}P(\mathbf{m}|\mathbf{d})\mathrm{d}\mathbf{m}, \tag{7}$$

$$\mathbf{C} = \int (\mathbf{m}-\bar{\mathbf{m}})(\mathbf{m}-\bar{\mathbf{m}})^T P(\mathbf{m}|\mathbf{d})\mathrm{d}\mathbf{m}, \tag{8}$$

$$\mathrm{MD}_i = \int |m_i' - \bar{m}_i| P(\mathbf{m}'|\mathbf{d})\mathrm{d}\mathbf{m}', \tag{9}$$

$$P(m_i|\mathbf{d}) = \int \delta(m_i' - m_i)P(\mathbf{m}'|\mathbf{d})\mathrm{d}\mathbf{m}', \tag{10}$$

where δ is the Dirac delta function and higher-dimensional (joint) marginal distributions are defined in a manner similar to Eq. (10). Parameter correlations are quantified by normalizing the covariance matrix to produce the correlation matrix,[27,28]

$$S_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}. \tag{11}$$

Elements S_{ij} are within $[-1,1]$, with a value of $+1(-1)$ indicating perfect correlation (anticorrelation) between m_i and m_j.

For nonlinear problems, analytic solutions to Eqs. (6)–(10) are generally not available, and numerical approaches must be applied. Computing the MAP estimate requires minimizing ϕ, which is typically carried out using global-search or hybrid optimization schemes, such as adaptive simplex simulated annealing.[6] However, for unimodal probability distributions, the posterior mean can provide parameter estimates that better represent the parameter uncertainty distribution and is used in this paper. The integrals in Eqs. (7)–(10) are solved here using the Markov-chain Monte Carlo method of fast Gibbs sampling.[7–9] The priors employed consist of a uniform distribution for each parameter on a bounded interval; hence, the remainder of this paper considers only the likelihood function, with the understanding that the prior is also included as per Eq. (5).

B. Likelihood function and snapshot averaging

Specifying the data uncertainty (error) distribution defines the likelihood function and is an important aspect of Bayesian inversion. Data uncertainties, which include both measurement and theory errors, are generally not known *a priori*, and physically reasonable assumptions are required; for example, independent Gaussian-distributed errors. These assumptions should be examined *a posteriori* by applying appropriate statistical tests to the data residuals.[18] For clarity, this section first considers defining the likelihood function for the case of a single data snapshot and known data variance. Three approaches are then described for treating variances that incorporate multiple-snapshot data (one of which is new). Finally, methods of treating unknown variances are summarized.

Consider first the case of a single data snapshot, $\mathbf{d}_f$, representing complex acoustic pressure vectors at N sensors for each of $f=1, F$ frequencies. Assuming that the data errors are complex zero-mean Gaussian-distributed random variables uncorrelated over space and frequency with variance ν_f at the fth frequency, the likelihood function is given by

$$L(\mathbf{m},S) = \prod_{f=1}^{F} \frac{1}{(\pi\nu_f)^N} \exp[-|\mathbf{d}_f - S_f\mathbf{d}_f(\mathbf{m})|^2/\nu_f], \tag{12}$$

where $\mathbf{d}_f(\mathbf{m})$ represents replica data predicted for model $\mathbf{m}$ and S_f is the (complex) source strength at the fth frequency. When explicit information on the source spectrum is unavailable (as in the experiment considered in this paper), the like-

lihood can be maximized analytically over S_f leading to likelihood function[5]

$$L_1(\mathbf{m}) = \prod_{f=1}^{F} \frac{1}{(\pi\nu_f)^N} \exp[-B_f(\mathbf{m})/\nu_f] \tag{13}$$

and corresponding error function

$$E_1(\mathbf{m}) = \sum_{f=1}^{F} B_f(\mathbf{m})/\nu_f. \tag{14}$$

In Eqs. (13) and (14), $B_f(\mathbf{m})$ represents the Bartlett mismatch, which may be written

$$B_f(\mathbf{m}) = \mathrm{Tr}\{\hat{\mathbf{R}}_f\} - \frac{\mathbf{d}_f^\dagger(\mathbf{m})\hat{\mathbf{R}}_f\mathbf{d}_f(\mathbf{m})}{|\mathbf{d}_f(\mathbf{m})|^2}, \tag{15}$$

where $\mathrm{Tr}\{\cdot\}$ represents the matrix trace operation and, for a single data snapshot,

$$\hat{\mathbf{R}}_f = \mathbf{d}_f\mathbf{d}_f^\dagger. \tag{16}$$

Next consider the case of $K>1$ data snapshots, $\mathbf{d}_{f,k}$, $k=1, K$, with the variance of all snapshots at the fth frequency given by ν_f. Two approaches appear to have been applied to date to incorporate multiple data snapshots in Bayesian MFI. In one approach,[13] an estimate of the CSDM given by Eq. (1) is used in place of Eq. (16) in the likelihood function, error function, and Bartlett mismatch given by Eqs. (13)–(15). That is, single snapshot data are replaced by snapshot averaged data without otherwise altering likelihood function L_1 or error function E_1. The second approach[5] is based on assuming that the errors on the K data snapshots are independent, so that the multiple snapshots are incorporated by multiplying the corresponding single snapshot probabilities, leading to

$$L(\mathbf{m},S) = \prod_{k=1}^{K}\prod_{f=1}^{F} \frac{1}{(\pi\nu_f)^N} \exp[-|\mathbf{d}_{f,k} - S_{f,k}\mathbf{d}_f(\mathbf{m})|^2/\nu_f]. \tag{17}$$

Employing maximum likelihood (ML) estimates for source terms $S_{f,k}$ leads to

$$L_2(\mathbf{m}) = \prod_{f=1}^{F} \frac{1}{(\pi\nu_f)^{NK}} \exp[-B_f(\mathbf{m})/(\nu_f/K)], \tag{18}$$

$$E_2(\mathbf{m}) = \sum_{f=1}^{F} B_f(\mathbf{m})/(\nu_f/K), \tag{19}$$

where the Bartlett mismatch, Eq. (15), is defined with the snapshot-averaged CSDM given by Eq. (1).

The two approaches to incorporating multiple snapshot data described above differ in the variance applied to the Bartlett mismatch: In the first approach, the variance applied in L_1 and E_1 is equal to the single snapshot variance ν_f, while in the second approach, the variance applied in L_2 and E_2 is ν_f/K. The two approaches represent end-case interpretations: The first approach assumes that snapshot averaging results in no reduction in the effective data variance (i.e., that errors on multiple snapshots are not random), while the second approach assumes the maximum reduction in variance (i.e., the reduction for fully independent random snapshot errors). In general, applying the first approach will lead to model parameter uncertainties that are overly pessimistic, while the second approach leads to parameter uncertainties that are overly optimistic.

This paper proposes an *effective variance* estimate for snapshot-averaged data based on (approximately) apportioning the single snapshot data variance ν_f as the sum of a measurement-error component ν_f^{ME} and a theory-error component ν_f^{TE}:

$$\nu_f = \nu_f^{ME} + \nu_f^{TE}. \tag{20}$$

Measurement error is often dominated by ocean ambient noise, and may reasonably be assumed to be independent from snapshot to snapshot. On the other hand, theory error is generally dominated by modeling errors due to the simplified model parametrization and approximate physics of the forward problem. These errors are generally not independent from snapshot to snapshot (e.g., identical modeling errors over snapshots are expected for fixed source/receiver geometry and stable environment). Hence, in deriving the effective variance estimate for snapshot averaged data, $\bar{\nu}_f$, it is reasonable to assume that the measurement-error variance component is reduced by a factor of $1/K$, while the theory-error component is not reduced at all:

$$\bar{\nu}_f = \nu_f^{ME}/K + \nu_f^{TE} \tag{21}$$

(a straightforward derivation of this result is given in the Appendix). Taking into account snapshot averaging effects in this manner, the likelihood and error functions become

$$L_3(\mathbf{m}) = \prod_{f=1}^{F} \frac{1}{(\pi\bar{\nu}_f)^N} \exp[-B_f(\mathbf{m})/\bar{\nu}_f], \tag{22}$$

$$E_3(\mathbf{m}) = \sum_{f=1}^{F} B_f(\mathbf{m})/\bar{\nu}_f, \tag{23}$$

where $B_f(\mathbf{m})$ is based on the cross-spectral data average, Eq. (1).

To estimate the single snapshot measurement and theory variance components, consider the definition of the SNR as the ratio of the average received signal level to the average ambient noise level given by Eq. (2). Assuming the ambient noise $\mathbf{n}_f$ is due to an independent, complex Gaussian-distributed random process with variance ν^{ME} leads to

$$\mathrm{SNR}_f = 10\log_{10}\frac{\langle|\mathbf{d}_f|^2\rangle - \nu_f^{ME}N}{\nu_f^{\mathrm{ME}}N}. \tag{24}$$

Equation (24) can be solved for the measurement variance component in terms of the SNR:

$$\nu_f^{\mathrm{ME}} = \frac{\langle|\mathbf{d}_f|^2\rangle}{N(10^{\mathrm{SNR}_f/10}+1)}, \tag{25}$$

and the theory variance component ν_f^{TE} can then be estimated from Eq. (20).

The development so far has assumed that the data variances ν_f are known *a priori*, which is often not the case.

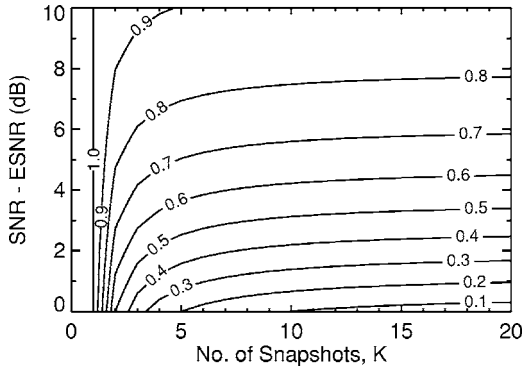


FIG. 3. Variance reduction factor A as a function of number of snapshots K and the SNR–ESNR difference (assuming ESNR=5 dB).

However, under the assumption of Gaussian-distributed errors, variances can be estimated from the data: Maximizing likelihood L_1 (single snapshot form) over ν_f, yields,[4]

$$\hat{\nu}_f = B_f(\mathbf{m})/N. \tag{26}$$

Explicit variance estimates are obtained by evaluating Eq. (26) at the ML model estimate $\hat{\mathbf{m}}$ obtained by minimizing the error function

$$E_4(\mathbf{m}) = N\sum_{f=1}^{F} \log_e B_f(\mathbf{m}) \tag{27}$$

[this error function results from substituting Eq. (26) back into Eq. (13)]. Once $\hat{\nu}_f$ is estimated, $\hat{\nu}^{\mathrm{TE}}$ is calculated using Eq. (20) with $\hat{\nu}_f$ replacing ν_f. These variances can be used as fixed estimates in Bayesian MFI by Gibbs sampling $E_3(\mathbf{m})$, as given by Eqs. (23) and (21), with $\hat{\nu}_f^{TE}$ replacing ν_f^{TE}.

It is interesting to consider the reduction in variance over the pessimistic case achieved by the effective variance estimate, Eq. (21), which can be quantified as

$$A_f = \frac{\bar{\nu}_f}{\hat{\nu}_f} = \frac{\hat{\nu}_f^{ME}/K + \hat{\nu}_f^{TE}}{\hat{\nu}_f}. \tag{28}$$

The variance-reduction factor A can also be formulated in terms of the SNR and the effective signal-to-noise ratio (ESNR), which provides a measure of the signal level compared to all sources of error (measurement and theory) defined[8]

$$\mathrm{ESNR}_f = 10\log_{10}\frac{\langle|\mathbf{d}_f|^2\rangle - \hat{\nu}_f N}{\hat{\nu}_f N} \tag{29}$$

(note that SNR $\geqslant$ ESNR by definition). The variance reduction can be written

$$A_f = 1 - \left(1 - \frac{1}{K}\right)\frac{10^{\mathrm{ESNR}_f/10}+1}{10^{\mathrm{SNR}_f/10}+1}. \tag{30}$$

The two extremes are SNR= ESNR (i.e., negligible theory error) where $A_f = 1/K$, and SNR $\gg$ ESNR (negligible measurement error) where $A_f \to 1$. Figure 3 shows the variance reduction factor A as a function of the number of snapshots K and the SNR–ESNR difference (assuming ESNR=5 dB, representative of the data considered in this paper). The reduction is most pronounced for low SNR and large K, and hence is valuable in this paper as well as in low-SNR applications, such as ship-noise or ambient-noise inversion.

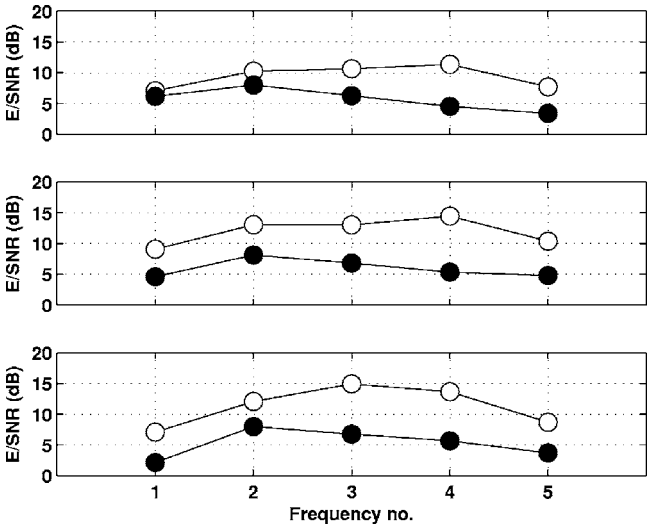


FIG. 4. Estimated SNR (open circles) and ESNR (filled circles) for various source frequencies for the experimental data. Upper: Short-range west-track data (1.51 km); middle: short-range east-track data (1.58 km); lower: Long-range east-track data (5.17 km).

IV. INVERSION RESULTS

A. Variance estimation

This section examines the effect of the different variance estimates described in Sec. III B for experimental HLA data as a prelude to considering inversion results in detail in the following section. HLA data for the source at a range of 1.58 km (east track) were selected for this analysis. ML single-snapshot variances were estimated at each of the five source frequencies using Eqs. (26) and (27), with ASSA applied to compute the ML model estimate. The ESNR, which compares the signal level to the total error, was then computed using Eq. (29). The SNR was estimated using Eq. (2), and the measurement-error and theory-error variances were computed using Eqs. (25) and (20). SNR and ESNR estimates at the five processed frequencies are shown in Fig. 4 (middle panel). For the inversions, fast Gibbs sampling was applied to error function E_3 of Eq. (23). Replica acoustic fields were generated using the normal-mode propagation model ORCA[29] using a complex-plane mode search to accurately model near-field effects due to leaky (continuous) modes. Numerical parameters of the ORCA model were set to values recommended in Ref. 30.

Before presenting inversion results for different data variance estimates, it is worth considering the underlying assumptions on the data errors. The assumptions that the errors are Gaussian distributed and spatially uncorrelated (random) can be examined by applying statistical tests to the data residuals (difference between measured data and data computed for the optimal model) at each frequency.[18] In particular, the Kolmogorov-Smirnov (KS) test was applied to test the residuals for Gaussianity and the runs test (median-

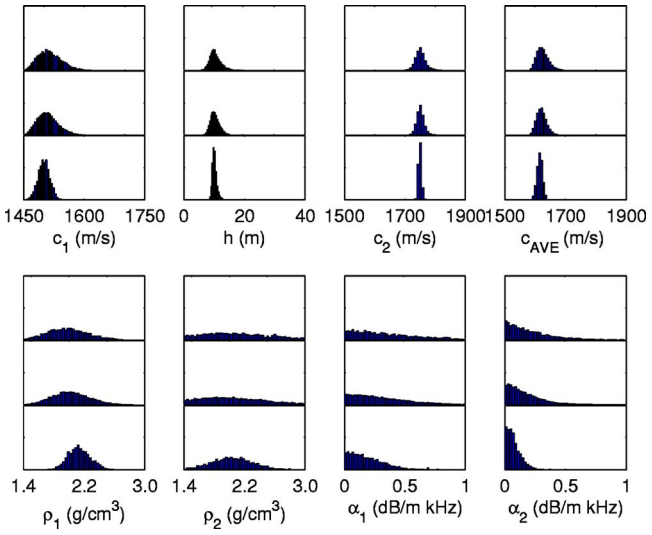


FIG. 5. (Color online) Marginal PPDs for different approaches to data variance estimation. Upper distributions correspond to pessimistic variance estimates, middle distributions to effective variance estimates, and lower distributions to optimistic variance estimates.

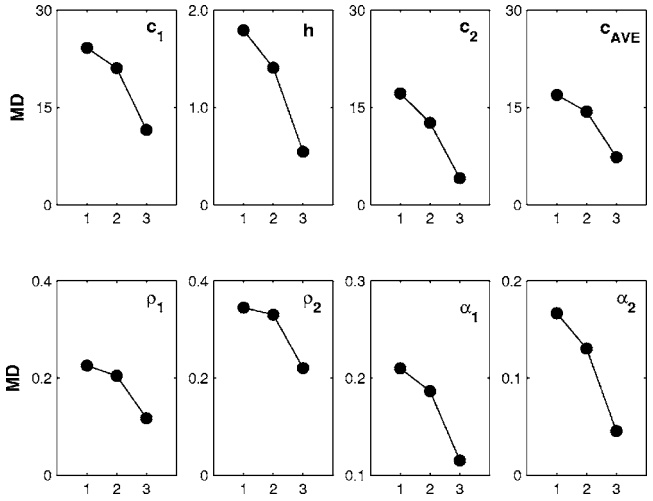


FIG. 6. Geoacoustic posterior uncertainty estimates quantified in terms of mean absolute deviations for: 1—Pessimistic variance estimates, 2—Effective variance estimates, and 3—Optimistic variance estimates. Parameter units vary between plots and are given in Table I.

delta test) was applied to test for randomness. The KS test indicated no significant evidence against Gaussianity at the 95% confidence level for any of the data snapshots considered in this paper, while the runs test provided no evidence against randomness at the 95% level for approximately 95% of the snapshots. Hence, there is no significant evidence against the error assumptions underlying the Bayesian analysis.

Geoacoustic inversion results for the HLA data are presented in Fig. 5 for the three approaches to variance estimation for snapshot-averaged data using $K=5$ snapshots. The variance reduction factor, Eq. (30), for the effective variance estimate, averaged, over frequency, is $A=0.75$. Marginal PPDs are presented for the seven geoacoustic model parameters described in Sec. II C and for an additional geoacoustic parameter consisting of the average sound speed in the upper layer defined by

$$c_{\mathrm{AVE}} = 2c_1c_2/(c_1+c_2). \tag{31}$$

The average layer sound speed (which accounts for the $1/c^2$-linear sound-speed gradient applied by ORCA) was not a parameter in the inversion but is readily computed from the model samples collected by the Gibbs sampler, and provides another indicator of geoacoustic information content (e.g., in some problems, c_{AVE} can be better determined than c_1 and c_2 indicating that the average sound speed is more important acoustically than the sound-speed gradient). Note that for some parameters, the plot limits in Fig. 5 are narrower than the prior bounds used in the inversions (Table I).

As expected, the width of the marginal PPDs for all parameters decrease systematically in going from the pessimistic variance estimate to the effective variance estimate to the optimistic variance estimate. This can be quantified in terms of parameter MDs, Eq. (9), a robust measure of uncertainty that is not as sensitive to a small number of sample outliers as the standard deviation. The MDs for the three variance estimates are compared in Fig. 6. The MDs for the effective-variance estimates are reduced by 5–30% over those for the pessimistic estimates, while the optimistic MDs are reduced a further 30–50% over the effective-variance MDs. It is worth emphasizing here that the goal of Bayesian inversion is not to compute the smallest parameter uncertainties, but rather the correct parameter uncertainties (i.e., uncertainties that correctly quantify the geoacoustic information content of the data). As discussed in Sec. III B, the effective variance estimates represent a physically reasonable compromise between the (overly) optimistic and pessimistic variance estimates.

B. Source range and bearing effects

This section considers geoacoustic inversion results for experimental data collected for three different source positions at the Barents Sea site, including short-range (1.58-km) and long-range (5.17-km) data along the east track and short-range (1.51-km) data along the west track (see Fig. 1). The source moved radially inward (toward the HLA) along the west track and outward along the east track. Data processing involved averaging over $K=5$ snapshots and inversions were carried out using effective variance estimates. SNR and ESNR values at the five source frequencies are shown in Fig. 4 for the three source positions (note that the source level was 6 dB higher for the long-range data than for the short-range data). Frequency-averaged variance reduction factors were approximately 0.75 for all three data sets. Geoacoustic inversion results for the three data sets are shown in Fig. 7 in terms of marginal PPDs and summarized in Table II in terms of mean parameter estimates with one MD uncertainties.

The parameters defining the sound-speed profile in the seabed, h, c_1, c_2, and c_{AVE}, are reasonably well determined, particularly for the short-range data sets. Considering these parameters first, Table II indicates that the posterior-mean parameter estimates are consistent between the three data sets (within estimated uncertainties). There is particularly close agreement between the parameter estimates and the uncertainties for the two short-range data sets. The differ-

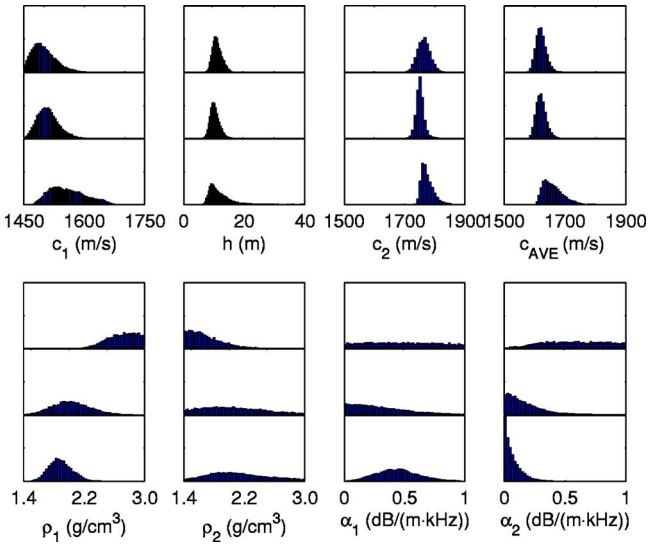


FIG. 7. (Color online) Marginal PPDs for data at 1.51 km along west track (upper distributions), 1.58 km along east track (middle distributions), and at 5.17 km along east track (lower distributions).

ences between the parameter estimates for the long- and short-range data are significantly larger than the differences between the two short-range data sets. In addition, the parameter uncertainties are substantially larger for the long-range data set, indicating that these data are less informative for seabed sound speed. In particular, the uncertainty estimates for the half-space sound speed c_2 and layer thickness h are approximately two and three times larger than for the short-range data sets, respectively. This is likely because the higher-order modes of the acoustic field are more strongly attenuated with range, leaving the lower-order modes which propagate at smaller grazing angles with shallower bottom penetration, and hence are less sensitive to deeper structure. The surficial sound speed c_1 is also not as well determined for the long-range data (mean deviation almost twice as large as for short-range data). This may also be due to the attenuation of higher-order modes and the consequential loss of resolution of surficial structure provided by the short vertical wavelengths of these modes. For all data sets, the uncertainty of the halfspace sound speed c_2 is significantly smaller than that for the surficial sound speed c_1 or the average layer sound speed c_{AVE}.

TABLE II. Geoacoustic parameter estimates (mean with mean-deviation uncertainties) from inversion of experimental data at the indicated ranges r and source track. Also included are approximate values from the supporting geophysical measurements.

Parameter and units	r=1.58 km east	r=1.51 km west	r=5.17 km east	Geophysical data
h (m)	11.2±1.4	11.7±1.2	14.0±4.4	10–20
c_1(m/s)	1510±21	1501±24	1559±37	1500–1520
c_2(m/s)	1753±13	1763±17	1783±26	1745
c_{AVE} (m/s)	1623±14	1621±13	1663±29	
ρ_1 (g/cm^3)	2.03±0.20	2.7±0.16	1.89±0.13	2.0–2.1
ρ_2 (g/cm^3)	2.06±0.33	1.71±0.21	2.12±0.31	
α_1 (dB/m kHz)	0.32±0.18	0.50±0.23	0.45±0.14	
α_2 (dB/m kHz)	0.21±0.13	0.57±0.21	0.10±0.06	

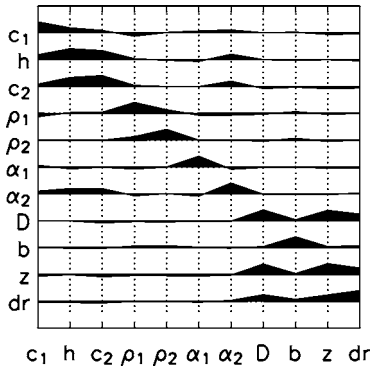


FIG. 8. Parameter correlation matrix for short-range east-track data (similar results for other data sets).

Figure 7 and Table II indicate that the seabed density and attenuation parameters are generally not well determined, as is commonly the case in MFI studies. While the density estimates for the two east-track data sets are reasonable, the combination of a relatively high ρ_1 estimate and low ρ_2 estimate for the west-track data is unlikely to be correct. It is also worth noting that attenuation estimates from geoacoustic inversion typically account for a variety of loss mechanisms, including the effects of bottom roughness and inhomogeneity, and are often higher than the intrinsic attenuation of the seabed. The density and attenuation parameters are generally better determined for the long-range data than for the short-range data. This is in agreement with other reported inversions of long-range acoustic data (e.g. Refs. 9 and 15), and appears to be due to the accumulated effect of bottom interaction with range.

To further investigate parameter uncertainties, it is useful to consider interparameter correlations and joint marginal distributions. Figure 8 shows the correlation matrix for the short-range east-track data (correlation matrices for the other data sets are similar). Significant positive correlations are evident between h and both c_1 and c_2, indicating that the acoustic data have a limited ability to discern between a thinner slower layer and a thicker faster layer. In terms of the geometric parameters, strong positive correlations exist between the water depth D, source depth z, and range r (particularly between D and z). Since correlations between geoacoustic and geometric parameters are small, uncertainties in the experiment geometry do not significantly affect geoacoustic uncertainties in this case. This is in contrast to inversion results of vertical-array data for a low-speed seabed in the Mediterranean Sea[8] which exhibited a strong negative correlation between water depth and sediment thickness, indicating, in that case, uncertainty in water depth degraded geoacoustic knowledge. Figures 9 and 10 show joint marginal PPDs for the short- and long-range data, respectively, for selected pairs of parameters (both correlated and uncorrected). The joint marginals for h and c_1, h and c_2, and D and r illustrate how the above correlations increase parameter uncertainties. The larger uncertainties for most parameters for the long-range data are evident in comparing Figs. 9 and 10.

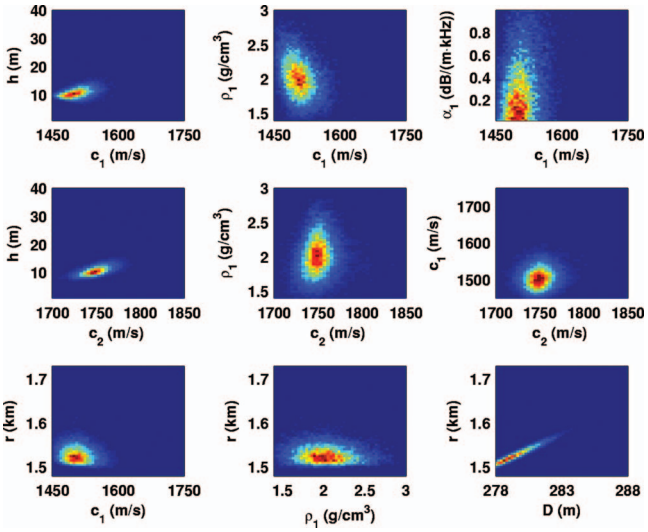


FIG. 9. Selected joint marginal PPDs for short-range east-track data.

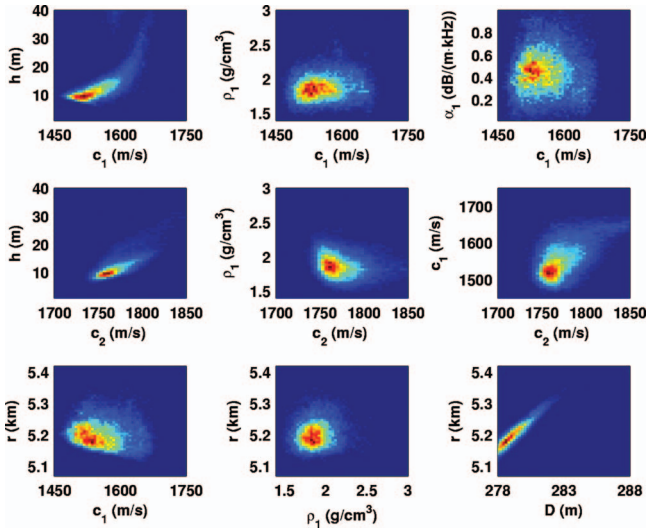


FIG. 10. Selected joint marginal PPDs for long-range data.

C. Comparison to geophysical measurements

It is important to compare the geoacoustic inversion results to the geophysical measurements described in Sec. II B, which are summarized in Table II. The surficial sound-speed estimates of $c_1 = 1510 \pm 21$, 1501 ± 24, and 1559 ± 37 m/s for the short- and long-range data are in good agreement with the gravity core (Fig. 2), which indicated an average sound speed of approximately 1500–1529 m/s over the top 1.7 m of sediments. Further, the half-space sound-speed estimates of $c_2 = 1753 \pm 13$, 1763 ± 17, and 1783 ± 26 m/s are in reasonable agreement with the seismic-refraction survey, which indicated an overall sound speed of the Quaternary sediments (120–140-m thick) of 1745 m/s. It is also interesting to note that the upper layer thickness estimates of $h = 11.2 \pm 1.4$, 11.7 ± 1.2, and 14.0 ± 4.4 m are in general agreement with evidence from the bottom-penetrating sonar of a possible weak reflector at 10–20 m (however, this may be fortuitous as the geoacoustic model was designed to represent continuous gradients and not discontinuous structures). Although poorly determined, the upper-layer density estimates of $\rho_1 = 2.03 \pm 0.20$ and 1.89 ± 0.13 g/cm^3 along the east track are in reasonable agreement with the gravity core which indicated a density of 2.0–2.2 g/cm^3 over the top 1.7 m of the sediments. Finally, it is noteworthy that the geoacoustic inversion results obtained here are in good general agreement with historic geophysical survey results published for glacigenic seabed sediments of the south-western Barents Sea,[24] which indicated an average sound speed of 1550 m/s for depths <4 m (based on data from 11 cores) and 1780 m/s for depths >10 m (based on 33 boreholes).

V. SUMMARY AND DISCUSSION

This paper presented results of Bayesian matched-field inversion applied to acoustic data collected on a bottom-moored horizontal line array due to a towed source that transmitted low-frequency tones at levels comparable to a merchant ship. Inversions were based on matching cross-spectral density matrices computed by averaging data snapshots. An approach was developed for estimating the effective variance of snapshot-averaged data based on the assumption that measurement errors (dominated by ambient noise) are independent and are therefore reduced by averaging, while theory (modeling) errors are not independent and not reduced. This approach provides a physically reasonable intermediate result between optimistic and pessimistic variance estimates based on assumptions of fully independent and dependent errors, respectively. The variance reduction factor for the effective variance estimate over pessimistic estimates improves with an increasing number of snapshots and with decreasing SNR. For the data considered here, the use of effective variance estimates provided significantly smaller geoacoustic uncertainties than pessimistic variance estimates; for inversion applications with low SNR and/or stationary sources that allow large snapshot averages, the effect could be even more pronounced.

Geoacoustic inversions were carried out for three source positions, including short-range (1.58-km) and long-range (5.17-km) data along a source track at 30° bearing east of HLA endfire and short-range (1.51-km) data along a track at 30° west of endfire. The sound-speed profile in the seabed was reasonably well determined by inversion, particularly for the short-range data sets. Recovered sound speeds and layer thicknesses were consistent within uncertainties for all three data sets, with close correspondence in estimates and uncertainties for the short-range data. The estimated surficial sediment sound speed was in good agreement with gravity core measurements and the half-space sound speed agreed with results from a wide-angle seismic refraction survey carried out at the experiment site. The surficial and half-space sound speeds also agreed with published results for the southwest Barents Sea based on a survey of core and borehole measurements. Density and attenuation parameters were not as well determined via geoacoustic inversion; however, there was general agreement between inversion results and local core measurements of density for two source positions. While the upper and lower sound speeds of the geoacoustic model agree well with the geophysical measurements and historical data, the sound-speed gradient over the upper sedi-

ments is relatively high and could include the effects of a discontinuous change at a layer boundary. However, including a discontinuity in the model parametrization did not substantially change the geoacoustic inversion results, indicating that if a discontinuity exists, the acoustic data cannot resolve its presence.

The overall consistency of the recovered parameters for the sediment sound-speed profile and the good agreement with independent geophysical measurements indicates that the use of data from a bottom-moored HLA and the inversion methodology developed here produce reliable and repeatable results for the Barents Sea site.

APPENDIX

Consider K realizations of an error (noise) process n_k, $k=1, K$ consisting of measurement error n_k^{ME} and theory error n^{TE} (assumed constant over k)

$$n_k = n_k^{ME} + n^{TE}, \quad k = 1, K, \tag{A1}$$

where n_k^{ME} and n^{TE} are zero-mean independent random processes with variances ν^{ME} and ν^{TE}, respectively. Defining $\langle\cdot\rangle$ as the expectation operator in this Appendix , the variance of the mean of these realizations is given by

$$\begin{aligned}\bar{\nu} &= \left\langle \left[\frac{1}{K}\sum_{k=1}^{K} n_k^{ME} + n^{TE} \right]^2 \right\rangle \\ &= \frac{1}{K^2}\sum_{k=1}^{K}\sum_{j=1}^{K} \langle n_k^{ME} n_j^{ME}\rangle + \langle n^{TE} n^{TE}\rangle + \langle n_k^{ME} n^{TE}\rangle + \langle n_j^{ME} n^{TE}\rangle \\ &= \frac{1}{K^2}[K\nu^{ME} + K^2\nu^{TE}] = \nu^{ME}/K + \nu^{TE},\end{aligned} \tag{A2}$$

as assumed in Eq. (21).

[1] M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92**, 2770–2783 (1992).
[2] S. E. Dosso, M. L. Yeremy, J. M. Ozard, and N. R. Chapman, "Estimation of ocean-bottom properties by matched-field inversion of acoustic field data," IEEE J. Ocean. Eng. **18**, 232–239 (1993).
[3] P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–781 (1994).
[4] P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[5] C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **8**, 259–270 (2000).
[6] S. E. Dosso, M. J. Wilmut, and A. L. Lapinski, "An adaptive hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).
[7] S. E. Dosso, "Quantifying uncertainties in geoacoustic inversion I: A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[8] S. E. Dosso and P. L. Nielsen, "Quantifying uncertainties in geoacoustic inversion II: Application to a broadband shallow-water experiment," J. Acoust. Soc. Am. **111**, 143–159 (2002).
[9] S. E. Dosso and M. J. Wilmut, "Quantifying data information content in geoacoustic inversion, "IEEE J. Ocean. Eng. **27**, 296–304 (2002).
[10] A. Caiti, S. M. Jesus, and Å. Kristensen, "Geoacoustic seafloor exploration with a towed array in a shallow water area of the Strait of Sicily," IEEE J. Ocean. Eng. **21**, 355–366 (1996).
[11] M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).
[12] M. R. Fallat, P. L. Nielsen, S. E. Dosso, and M. Siderius, "Geoacoustic characterization of a range-dependent ocean environment using towed array data," IEEE J. Ocean. Eng. **30**, 198–206 (2005).
[13] D. J. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. L. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[14] D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).
[15] R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources," J. Acoust. Soc. Am. **117**, 626–637 (2005).
[16] D. Tollefsen, M. J. Wilmut, and N. R. Chapman, "Estimates of geoacoustic model parameters from inversions of horizontal and vertical line array data," IEEE J. Ocean. Eng. **30**, 764–772 (2005).
[17] R. M. S. Barlee, N. R. Chapman, and M. J. Wilmut, "Geoacoustic model parameter estimation using a bottom moored hydrophone array," IEEE J. Ocean. Eng. **30**, 773–783 (2005).
[18] S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).
[19] C. W. Bogart and T. C. Yang, "Source localization with horizontal arrays in shallow water: Spatial sampling and effective aperture," J. Acoust. Soc. Am. **96**, 1677–1686 (1994).
[20] R. O. Nielsen, *Sonar Signal Processing* (Ardent House, London, 1990).
[21] C. E. Solberg, "Geoacoustic models for the 2003 antenna experiment area compiled from shallow seismic data," Technical Report 2004/01602, Norwegian Defence Research Establishment, Kjeller, Norway, 2004.
[22] A. Lepland, "Results of analytical tests on FFI 2003 sediment cores," Technical Report 2004.019, Geological Survey of Norway, Trondheim, Norway, 2004.
[23] T. H. Orsi and D. A. Dunn, "Correlations between sound velocity and related properties of glaciomarine sediments: Barents Sea," Geo-Mar. Lett. **11**, 79–83 (1991).
[24] J. Sættem, L. Rise, and D. A. Westgaard, "Composition and properties of glacigenic sediments in the southwestern Barents Sea," Marine Georesources & Geotechnology **10**, 229–255 (1991).
[25] W. M. Telford, L. P. Geldart, and R. E. Sheriff, *Applied Geophysics* (Cambridge University Press, Cambridge, 1990).
[26] S. R. Rutherford and K. E. Hawker, "The effects of density gradients on bottom reflection for a class of marine sediments," J. Acoust. Soc. Am. **63**, 750–757 (1978).
[27] A. Tarantola, *Inverse Problem Theory* (Elsevier, Amsterdam, 1987).
[28] M. Sen and P. L. Stoffa, *Global Optimization Methods in Geophysical Inversion* (Elsevier, Amsterdam, 1995).
[29] E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," J. Acoust. Soc. Am. **100**, 3631–3645 (1996).
[30] E. K. Westwood and R. A. Koch, "Elimination of branch cuts from the normal-mode solution using gradient half spaces," J. Acoust. Soc. Am. **106**, 2513–2523 (1999).

Consistency and reliability of geoacoustic inversions with a horizontal line array

Laurie T. Fialkowski, T. C. Yang, Kwang Yoo, Elisabeth Kim, and Dalcio K. Dacol
Naval Research Laboratory, Washington, DC 20375

(Received 15 December 2005; revised 5 May 2006; accepted 5 May 2006)

Geoacoustic inversions with a towed horizontal array are of interest for rapidly characterizing sediment properties over changing regions. To be of practical value, inversions must yield consistent and reliable results for consecutive or neighboring data. A method of determining inversion reliability *a priori* is delineated using an empirical approach and confirmed with inversion results in terms of consistency. Geoacoustic parameter hierarchy and resolvability are empirically analyzed using two different methods: one requires knowledge of the source function and the other does not. Inversion results using the two methods are compared using both synthetic data and experimental data from MAPEX2000. The inversions employ a global optimization technique which navigates the parameter space in directions aligned with valleys of the cost function, increasing inversion algorithm efficiency and disclosing parameter correlations and hierarchy.
 [DOI: 10.1121/1.2208453]

PACS number(s): 43.60.Pt, 43.30.Wi, 43.30.Pc [AIT] Pages: 231–246

I. INTRODUCTION

Geoacoustic inversion has been traditionally investigated using a vertical line array (VLA), e.g., Refs. 1–5. The source is nominally a few kilometers away, allowing low grazing angle interactions of the acoustic energy with the ocean bottom, so that results are relevant for long-range propagation modeling. Consequently, these inversion results represent average bottom properties over the long source-receiver range. For reasons related to localization and operational advantages, there exists a high level of interest in using towed horizontal line arrays (HLAs) for geoacoustic inversions.[6–13] With a towed array, there is the possibility of using ship self-noise as the source. These inversions yield bottom properties of the local area, rather than an average over the traditionally longer source-receiver range. In addition, the moving platform makes it possible to cover large regions of interest while exhibiting the locally changing properties of the bottom, allowing one to build a large application-specific database *in situ*.[11,12,14]

The purpose of this paper is to examine the issues of reliability and parameter resolvability for towed HLA inversions, especially inversions suitable for using self-noise as the probing source. The primary focus of the analysis is the robustness and consistency of the inversion results. Comparisons to ground truth data are valuable when the intended purpose is to build a database suitable for any frequency band. For the more specific purposes of acoustic localization and acoustic system performance predictions, a sufficient and also valuable inversion result is a consistent "effective" geoacoustic environment. This effective geoacoustic environment is useful over a wide frequency band for the intended applications (e.g., source localization) and can thus be considered a representation of the "true" environment. HLA inversion results are sought which yield consistent effective environments as the platform traverses an area of similar bottom properties. Inversion results should also clearly expose where the platform crosses from one bottom type to another.

Parameter robustness and consistency addressed in this paper are not to be confused with parameter accuracy (i.e., resolution or, conversely, uncertainty) addressed by others using the Bayesian approach based on the *a posteriori* probability distribution, e.g., Ref. 15. Consider, for example, matched-field source localization. One can argue that the determination of range and depth accuracies (resolutions) requires *a posteriori* estimation of the range and depth probability distribution. That information, although useful, does not guarantee that range and depth estimation for a moving source (in practice) will be robust and consistent as the source traverses over different ranges. In fact, a lack of consistency has been considered as the major deficiency in the application of matched-field localization to real data. For practical applications, obtaining representative, consistent, and repeatable estimations is often more important than obtaining an accurate estimation or knowing the uncertainty of the estimation. Likewise, for geoacoustic inversions, our interest lies in how to obtain a robust and consistent inversion for a moving platform, rather than estimating the parameter resolution in an absolute sense.

In order to examine parameter resolvability, a rotated coordinate method based on the cost function is used to expose *a priori* the geoacoustic parameter hierarchy and correlations. The parameter hierarchy identifies the parameters, or the combination(s) of parameters, that the cost function is most sensitive to. The importance of the source and receiver geometry in relation to the sediment parameters is clearly evident from the rotated coordinate analysis, as is the change in parameter hierarchy due to a change in cost function. *A posteriori* plots of inverted parameter values as a function of parameter window size give an empirical measure of parameter resolvability as a function of window size, and also fur-

ther validate the parameter hierarchy and correlations exposed by the rotated coordinate method.

Matched-field inversions (MFIs) involve the optimization of a matched-field cost function. To date, the conventional phone-coherent cost function has been applied to multi-frequency vertical array data, where matched-field correlations between data and replica phone vectors are summed incoherently over frequency.[2–5,16] On the other hand, recent analyses of horizontal array data have had most success implementing the frequency-coherent cost function, where matched-field correlations between data and replica frequency vectors are summed incoherently over range (phones).[10,11,13,16] The phone-coherent MFI requires precise knowledge of phone positions. The frequency-coherent MFI requires knowledge of the source spectrum, as frequency correlation is the same as correlation of the time series by the convolution theorem. When using sources-of-opportunity (e.g., surface ships), or ship self-noise, since the source spectrum is commonly unknown, it is more convenient to use the traditional phone-coherent cost function, where knowledge of the source spectrum is not necessary.

HLA inversions implementing both the phone-coherent and frequency-coherent cost functions are examined, with an emphasis on determining the necessary criteria for successful HLA inversions with the phone-coherent cost function. To this end, both synthetic data and experimental data are analyzed. The synthetic data are produced with some distortion to the HLA, as one would expect to encounter with a towed array. For the experimental data analysis, a portion of the MAPEX2000 data is used; these data were collected in 2000 by SACLANTCEN (now NURC). The data was previously analyzed by Siderius *et al.* and Fallat *et al.*[10,11,13] These analyses applied an MFI technique for a broadband source using the frequency-coherent cost function. The phone-coherent method has been previously implemented by Battle *et al.* for HLA MFI of narrow-band data, and their array deformation model is also implemented here.[12]

Previous analysis indicates that the frequency-coherent method is insensitive to the array shape for small deviations from a straight line (described in terms of bow and tilt), but the method requires precise knowledge of the source-receiver range.[17] Hence, the frequency-coherent method is preferred if the array shape is unknown. The phone-coherent method, on the other hand, is shown here to be more sensitive to the array shape, and practically insensitive to the source-receiver range. This insensitivity to source-receiver range indicates that the phone-coherent method is more applicable to geoacoustic inversions using ship self-noise, because ship noise originates presumably from a distributed source, corresponding to various source-receiver ranges for each individual receiver. Due to the method's sensitivity to array shape, the caveat for the phone-coherent method is that array deformation needs to be approximated by the model in order to achieve reliable results.

We empirically demonstrate the impact parameter search windows have on achieving consistent and reliable inversion results using a towed line array. The analysis presented has value in guiding future experimental measurements, and for aiding the design of a towed-array geoacoustic inversion system. The parameter windows are quantified using empirical standard deviations that are intended only as a measure of the inversion performance. Using parameter constraints guided by this analysis, consistent inversion results are obtained for the MAPEX2000 towed array data. The inversion procedure presented can be generalized to other inversion scenarios.

This paper is organized as follows: In Sec. II, the inversion method is described, and the phone and frequency-coherent cost functions are defined. In Sec. III, the parametrization of the environment and the source-receiver geometry for all inversions are defined. In Sec. IV, the rotated coordinate technique is applied to synthetic data and used to identify the differences in parameter hierarchy for the two cost functions, especially in the context of whether or not it is important to include array deformation in the parametrization. Section V demonstrates the impact the parameter window size has on inversion results using synthetic data. Sections VI and VII present HLA MFI results when the phone-coherent inversion method is applied to portions of MAPEX2000 data; Sec. VI presents an analysis of selected transmissions from the beginning, middle, and end of one track of data, and Sec. VII presents a summary of the inversion results for the whole track of data. Finally, Sec. VIII summarizes the method and results.

II. INVERSION METHOD

The inversion method is most easily described in terms of its three main components: the global optimization method, the forward propagation method, and the cost function. For all analyses presented in this paper, the inversions implement a modified simulated annealing method for the global optimization, a wave integration technique for the forward propagation method, and both phone-coherent and frequency-coherent matched-field cost functions.

The global optimization method is a modified simulated annealing technique described in Ref. 18, with a fast cooling schedule implemented as in Ref. 19. There are other modified simulated annealing methods, such as fast simulated annealing (FSA)[19] and the adaptive simplex simulated annealing (ASSA);[16] these have been compared in the context of algorithm-induced variability in geoacoustic inversion.[16] The purpose of this paper is not for algorithm comparison; however, previous concerns with mitigating the initial estimate dependence of a simulated annealing search are addressed by implementing the rotated coordinate method.[18] As described in Ref. 18, the parameter search space is traversed along rotated coordinates during the optimization. The coordinate rotation is determined by the gradient of the cost function for a given data set, so that the coordinates are aligned with the most prominent features in the parameter landscape. This optimal coordinate rotation is estimated prior to inverting for geoacoustic parameters and increases the efficiency of navigating the parameter landscape. The rotated coordinate method incorporates cost function gradient information into the geoacoustic inversion search algorithm, an inversion element which previous publications have reported as being important and necessary for success.[20,21]

The parameter search space is defined by $\Omega=\{\mathbf{x}|a_i<x_i<b_i\}$ for $1\leq i\leq n$, where $\mathbf{x}$ is the set of n parameter values defining the geoacoustic environment, and a and b are the bounds which define the parameter search windows. The rotated coordinate system is defined by the eigenvectors, $\mathbf{v}_j$, of the covariance matrix K, where

$$K=\int_{\Omega}\nabla C(\nabla C)^t\,d\Omega, \tag{1}$$

and C is any cost function parameterized by $\mathbf{x}$.[18] For high-dimensional parameter spaces, the integral for K can be estimated with a Monte Carlo method using several hundred sample points.[18] The original geoacoustic parameter vector $\mathbf{x}$ is a linear transformation of the eigenvectors with coefficients $\{y_j\}$:

$$\mathbf{x}=\sum_j y_j\mathbf{v}_j. \tag{2}$$

The inversion algorithm converges to an optimal set of coefficients $\{y_j\}$, which in turn define an optimal set of geoacoustic parameters. In addition to improving search efficiency, the eigenvectors $\mathbf{v}_j$ disclose the correlations between parameters, and ordering these eigenvectors by the size of each associated eigenvalue of K illustrates the parameter hierarchy. The eigenvectors associated with the eigenvalues of larger magnitude identify the parameter combinations that most significantly affect the gradient of the cost function; these are the most resolvable parameter combinations for the data set in question. For a more detailed discussion on this topic, see Ref. 22.

There are several forward propagation models that are suitable for geoacoustic inversions; often the choice of which model to use is dictated by the complexity of the environment in question, as well as the source and receiver geometry. For the examples presented in this paper, a wave number integration technique was chosen as the forward propagation model. *A priori* tests with normal mode propagation models using KRAKEN[23–25] allowing for more complex sound speed profiles showed that both the water and sediment profile could be approximated by a series of iso-velocity layers, as required by the most straightforward implementation of a wave number integration technique. The wave number integration method is efficient for HLA inversions because only a few receiver depths are needed, and range independence is sufficient at the short ranges typically involved (<1 km). Propagation is simplified so that the solution is analytic in each layer, with coefficients determined by boundary conditions.[26] This method of undetermined coefficients has also been applied with other propagation methods, with the environment generalized to iso-velocity layers to gain efficiency.[27,28] Note that an inhomogeneous water column sound speed profile can also be easily incorporated into this propagation model as described in Ref. 28; the method could also be extended to a simple linearization of the sediment profiles with efficiency maintained for a small number of sediment layers.

For the cost functions to be optimized, two matched-field methods are implemented here: a phone-coherent cost function and a frequency-coherent cost function. The phone-coherent cost function used, C_P, is defined as in Ref. 10:

$$C_P=\frac{1}{N_{fr}}\sum_{j=1}^{N_{fr}}(1-B_{Pj}), \tag{3}$$

where

$$B_{P_j}=\frac{|\sum_{i=1}^{N_{ph}}p_{ij}q_{ij}^*|^2}{\sum_{i=1}^{N_{ph}}|p_{ij}|^2\sum_{i=1}^{N_{ph}}|q_{ij}|^2}. \tag{4}$$

Here, p_{ij} is the data pressure on the ith phone at the jth frequency, and q_{ij} is the modeled replica of the data pressure of the same form. The numbers of phones and frequencies are N_{ph} and N_{fr}, respectively. Similarly, the frequency coherent cost function is defined, also as in Ref. 10:

$$C_F=\frac{1}{N_{ph}}\sum_{j=1}^{N_{ph}}(1-B_{F_j}), \tag{5}$$

where

$$B_{F_j}=\frac{|\sum_{i=1}^{N_{fr}}p_{ji}q_{ji}^*|^2}{\sum_{i=1}^{N_{fr}}|p_{ji}|^2\sum_{i=1}^{N_{fr}}|q_{ji}|^2}. \tag{6}$$

The performance of these two cost functions with HLAs is examined using synthetic and experimental data in Secs. IV and VI.

III. PROBLEM DEFINITION AND PARAMETRIZATION

The same geoacoustic parametrizations are used for the synthetic and for the experimental data inversions; these were motivated by one of the environments encountered during the MAPEX2000 experiment. For all inversions, a 64-element HLA with 4-m phone spacing is modeled at 11 frequencies equally spaced from 250 to 750 Hz. The nominal source and HLA depths are 55 and 60 m, respectively, with the closest receiver at a range of 300 m from the source. For all inversions, the replicas q_{ij} used for the matched-field cost function are the closest computed value in range and depth over an equally spaced grid. The 0.5-m grid resolution in range and the 0.25-m grid resolution in depth are sufficient for this closest point approach. The maximum number of depth values due to array deformation fitted onto an equally spaced grid is much less than the total number of elements on the HLA. This closest point approach results in considerable computational time savings.

The environment contains a thin sediment layer with a slow compressional sound speed over a reflecting bottom layer. The water column sound speed profile (SSP) is assumed known *a priori* and is a four-iso-velocity-layer approximation to one of the sound speed profiles measured during MAPEX2000 [Fig. 1(a)]. The normal mode method does not require a layered profile, and therefore the KRAKEN normal mode model[23] was used in the creation of broadband synthetic data and replicas for the *a priori* tests: data were created using the experimentally measured SSP, and replicas were created using the iso-velocity layered SSP. The resulting cost function values matching the data to the

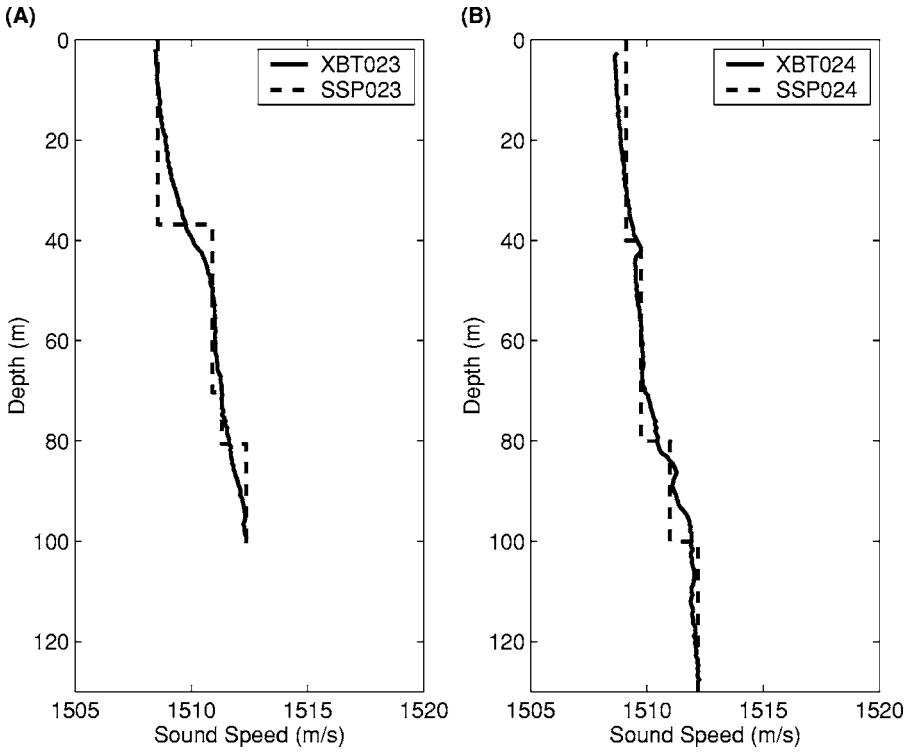


FIG. 1. Experimentally measured sound speed profiles and their iso-layer approximations used for inversions. (a) The solid line is XBT023, measured at 08:07Z, 36°32.45′ N, 14°49.20′ E; the dashed line is SSP023, the four-layer approximation of XBT023. (B) The solid line is XBT024, measured at 09:11Z, 36°27.34′ N, 14°46.47′ E; the dashed line is SSP023, the four-layer approximation of XBT024.

replicas are $C_P=0.0118$ and $C_F=0.0122$; this corresponds to an approximate 0.05-dB expected cost function degradation due to using the iso-velocity layer approximation for the water SSP.

The sediment is also parametrized as a series of iso-layers; the layered environmental model is suitable because *a priori* parameter hierarchy analysis of the MAPEX2000 data shows the sediment sound speed gradient is not one of the most critical parameters. Using a normal mode propagation model, and the KRAKEN normal mode program[23] (which allows for gradients in the sediment sound speed), the covariance matrix K [Eq. (1)] was computed with cost functions C_P and C_F. The eigenvectors of these matrices exposed parameter hierarchies which consistently place the sediment sound speed gradient in the bottom half of the parameter hierarchy, indicating that—for this environment and data set—the sediment sound speed gradient has minor impact on both cost functions in question.

The geoacoustic parameters and parameter windows that define the search space for all synthetic data inversions are detailed in Table I. The environmental parameters included in the search space are the depth of the water/sediment interface (z_w); the sediment thickness (h), sound speed (c_1), density (ρ_1), and attenuation (α_1); and the half-space sound speed (c_2), density (ρ_2), and attenuation (α_2). In order to enforce realistic behavior of sound speeds and densities with depth (i.e., mostly increasing), the parametrization of these involves the ratios $c_1/c_w, c_2/c_1, \rho_1/\rho_w$, and ρ_2/ρ_1. For the synthetic data, note that the *true* sound speed ratio between the sediment and the water ($x_3=c_1/c_w$) is less than 1.0, which represents one of the environments during MAPEX2000.[11]

The geometric parameters sought in the inversions are the horizontal array deformation parameters, tilt (θ) and bow (b); the nominal HLA depth, z_0; the range between the source and the first element of the HLA, r_0; and the source depth, z_s. The deformation of the HLA is parametrized parabolically as in Ref. 12. We choose z_0 as the nominal depth of the HLA, so that each phone depth is defined by $z_i=z_0+z_{\Delta i}$, where $z_{\Delta i}=b[1-(2d_i/L)^2]-d_i \sin\theta$. Here, b is the bow, or the height of the array below its midpoint; θ is the tilt of the array with the rotation point at the first phone; L is the length of the array; and d_i is the signed distance from phone i to the midpoint of the array $[d_i=(i-1)\Delta ph-L/2$, with Δph as the phone spacing]. Although the change in range for each phone

TABLE I. Parametrization for geoacoustic inversion of synthetic data.

i	Parameter (x_i)	True	Min	Max
	Sediment layer			
1	z_w, Water depth (m)	98.0	95.0	101.0
2	h, Thickness (m)	9.7	1.0	25.0
3	c_1/c_w, Sound speed ratio	0.98	0.95	1.05
4	ρ_1, Density (g/cm^3)	1.2	1.00	1.80
5	α_1, Attenuation (dB/λ)	0.10	0.00	0.60
	Reflecting layer			
6	c_2/c_1, Sound speed ratio	1.15	1.05	1.20
7	ρ_2/ρ_1, Density ratio	1.20	1.00	1.80
8	α_2, Attenuation (dB/λ)	0.10	0.05	1.00
	Geometric parameters			
9	θ, Array tilt (deg)	−2	−2.5	2.5
10	b, Array bow (m)	5	−7.0	7.0
11	z_0, Array depth (m)	60.0	50.0	65.0
12	r_0, Sor-Rec range(m)	300.0	285.0	305.0
13	z_s, Source depth (m)	55.0	50.0	60.0

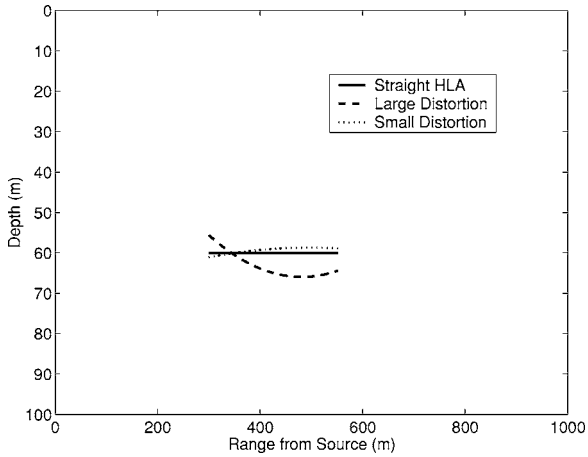


FIG. 2. Effect of bow and tilt on HLA. The large distortion has a bow of 5 m and a tilt of −2°, the same deformation parameters used for synthetic data. The small distortion has a bow of −1 m and a tilt of 0.5°, which is similar to the array distortion encountered in MAPEX2000.

of the HLA is small, we similarly define each phone position in range by $r_i = r_0 + r_{\Delta i}$, where $r_{\Delta i} = d_i \cos\theta + (L/2)$. The synthetic data is computed with an HLA bow of $b = 5$ m and tilt of $\theta = -2°$ (see Fig. 2).

IV. PARAMETER HIERARCHY, RESOLVABILITY, AND CONSISTENCY

The resolvability or uncertainty of a geoacoustic parameter is determined, in theory, by the standard deviation of the parameter distribution, for example, as defined by the Bayesian marginal distribution.[15,29,30] For practical purposes, parameter resolvability can be empirically determined by the standard deviation of a parameter distribution based on cost function values for a sufficient number of search paths.[16,31,32] A third method, presented here, determines the parameter hierarchy and relative resolvability for a given data set and chosen parametrization through the use of rotated coordinates determined by the covariance matrix K [Eq. (1)]. The coordinate rotation identifies the parameters, or the combination(s) of parameters, that the cost function is most sensitive to. Higher sensitivity implies higher resolvability. For practical purposes, one is interested in the hierarchy of the resolvability for assessing the confidence level of the results. The standard deviation discussed below does not imply uncertainty of the parameters in an absolute sense. This method will be used to compare parameter hierarchies for two different inversion methods: frequency coherent and phone coherent. Additionally, this approach is verified by examining geoacoustic parameter resolvability as defined by empirically obtained standard deviations for the parametrization. These standard deviations are based on various optimal parameter estimates resulting from several inversion search paths, each path determined by a different seed to the random number generator in the simulated annealing algorithm.

For the synthetic data case, the environment described in the previous section is used, and a source is placed at a range of $r_0 = 300$ m from the first element of the HLA. The nominal source and receiver depths are $z_s = 55$ m and $z_0 = 60$ m, respectively. The HLA has a tilt of $\theta = -2°$ and a bow of $b = 5$ m. The water depth is $z_w = 98$ m, and the sediment thickness is $h = 9.7$ m. The first sound speed ratio is $c_1/c_w = 0.98$, which corresponds to a sediment sound speed of 1481.3 m/s. The second sound speed ratio is $c_2/c_1 = 1.15$, which corresponds to a sound speed of 1703.5 m/s in the bottom sublayer. The first density ratio, $\rho_1/\rho_w = 1.2$, corresponds to a sediment density of 1.2 g/cm^3; the second density ratio, $\rho_2/\rho_1 = 1.2$, corresponds to a bottom sublayer density of 1.44 g/cm^3. Both the sediment and the bottom sublayer have attenuations of 0.1 dB/λ.

Using the synthetic data, the covariance matrix K [Eq. (1)] is estimated using 600 sample points. The eigenvectors of K define an optimal coordinate rotation and are examined to determine the parameter hierarchy and resolvable parameter combinations. Results using cost functions C_P and C_F are examined; for both cost functions, cases which include the bow and tilt in the parameter search space are compared to those assuming a perfectly straight HLA (i.e., the synthetic data p_{ij} has nonzero bow and tilt, and the replica g_{ij} has zero bow and tilt). Figure 3 displays four sets of eigenvectors which define optimal coordinate rotations for the synthetic data case, each applying a different cost function and parameter search space combination. Figure 3(a) shows the eigenvectors when the cost function C_P is used, and the HLA is incorrectly assumed perfectly straight; Fig. 3(b) shows the eigenvectors when the cost function C_F is used with the same parameter search space as for Fig. 3(a). Figure 3(c) shows the eigenvectors when the cost function C_P is used and the array deformation parameters (bow and tilt) are incorporated into the parameter search space, and Fig. 3(d) shows the eigenvectors when C_F is used with the same parameter search space as for Fig. 3(c). In all four plots, the associated eigenvalues are displayed to the right of each eigenvector line plot, and the most resolvable eigenvector is the lowest on the vertical axis. Examining the relative sizes of the eigenvalues, the first five or six eigenvectors in each case are considered resolvable. These resolvable eigenvectors represent resolvable parameter combinations. Parameters which consistently appear at the top of the parameter hierarchy, regardless of parameter search space and cost function, are the sediment thickness (h), the source and receiver depths (z_s and z_0), and the sediment sound speed ratio (c_1/c_w). These parameter combinations are therefore resolvable by all inversion methods presented here for this synthetic data.

When comparing Figs. 3(b) and 3(d) (cost function C_F), note that there is not much change in the first two eigenvectors when the bow and tilt are incorporated into the parameter search space. Contrast this with the change at the top of the parameter hierarchy when the cost function C_P is used, and the bow and tilt are included in the parameter space [Figs. 3(a) and 3(c)]. Also note that, for both parametrizations using C_F, the most resolvable parameter is the source-receiver range, r_0, and it is practically uncoupled from all other parameters, indicating that the range r_0 is the most resolvable parameter. This is in contrast to the parameter hierarchy when the cost function C_P is used, where the range r_0 appears very low in the hierarchy, indicating it is much less likely to be consistently resolved.

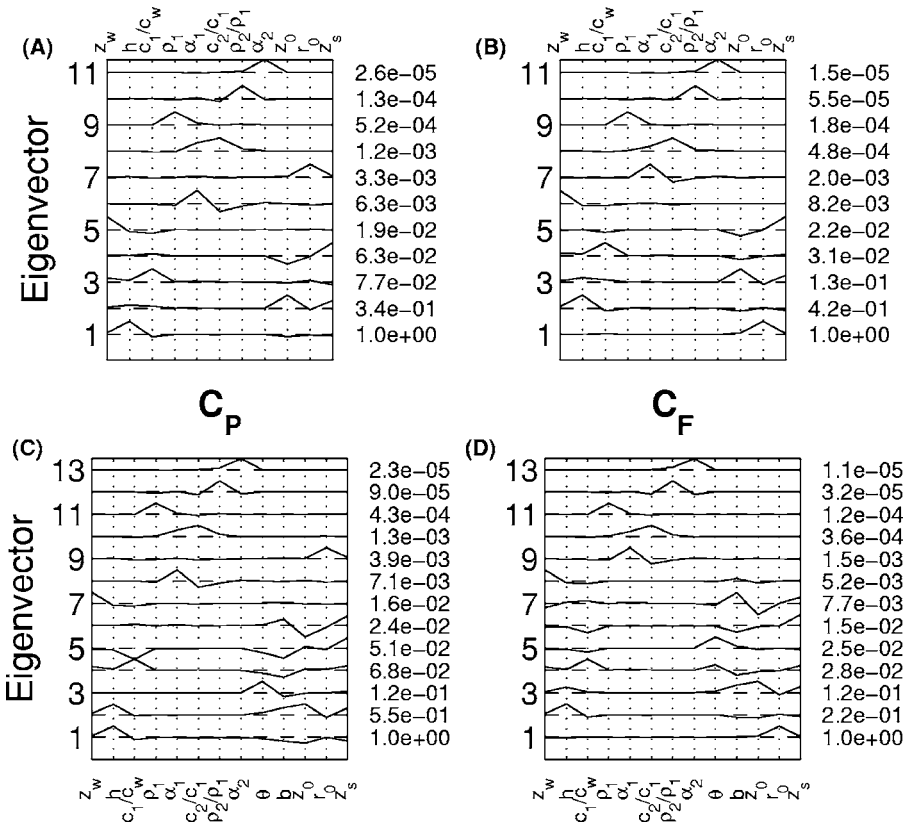


FIG. 3. Eigenvectors of K for synthetic data; associated eigenvalues are printed to the right of each eigenvector. (a) C_P cost function; bow and tilt assumed zero and not included in parameter search space. (b) C_F cost function; bow and tilt assumed zero and not included in parameter search space. (c) C_P cost function; bow and tilt included in parameter search space. (d) C_F cost function; bow and tilt included in parameter search space.

The corollary to the above finding is many-fold. Incorrect windowing of the source-receiver range will likely yield incorrect geoacoustic parameter values for inversions using the frequency-coherent method. More importantly, the phone-coherent method's insensitivity to the source-receiver range will likely allow for successful broadband geoacoustic inversion where the ship self-noise sources are distributed with different ranges to an individual receiver. However, the phone-coherent method is more sensitive to array deformation than the frequency-coherent method, and therefore it is important to account for array deformation with phone-coherent inversions. These differences between the two methods are not surprising: Frequency-coherent correlation is identical to time-domain correlation by the convolution theorem, and the relative multi-path arrival time is sensitive to the source-receiver range. Phone-coherent correlation, on the other hand, is identical to matched-field beamforming, and the relative phase between the phones is sensitive to the phone position, and therefore array shape is important.

In order to determine resolvability and consistency of the geoacoustic parameters, inversion results are analyzed in conjunction with the eigenvectors. Using the four sets of rotated coordinates defined by the eigenvectors plotted in Fig. 3, four different inversions were implemented. Figure 4 shows the values of the eigenvector coefficients $\{y_j\}$ at every iteration for each inversion, illustrating their convergence; the true coefficient values are represented by the horizontal dashed lines on each plot. Figures 4(a) and 4(b) are for C_P and C_F, respectively, when the replica q_{ij} is computed with no array distortion; Figs. 4(c) and 4(d) are for C_P and C_F, respectively, when the replica q_{ij} is computed with array distortion. Without bow and tilt in the parameter search space, the phone coherent inversion obtained an optimal cost function value of $C_P=0.56$, while the frequency coherent inversion obtained an optimal cost function value of $C_F=0.42$.

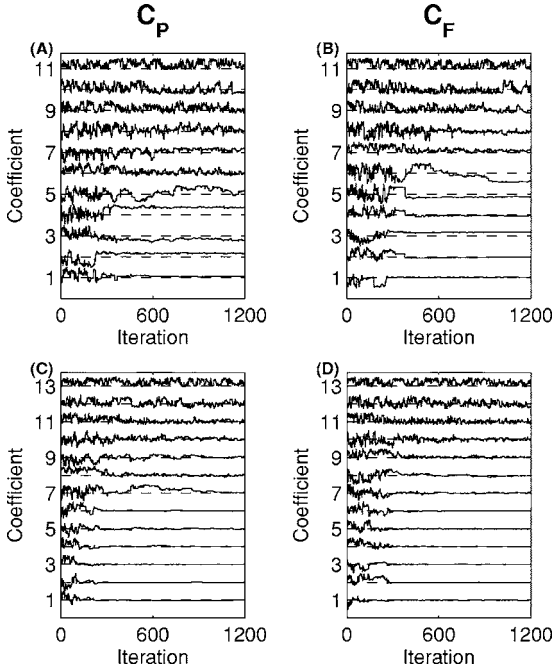


FIG. 4. Geoacoustic inversion coefficient history for synthetic data case. The dashed lines represent the true coefficients, y_j. (a) C_P cost function; b and θ assumed zero and not included in parameter search space. (b) C_F cost function; b and θ assumed zero and not included in parameter search space. (c) C_P cost function; b and θ included in parameter search space. (d) C_F cost function; b and θ included in parameter search space.

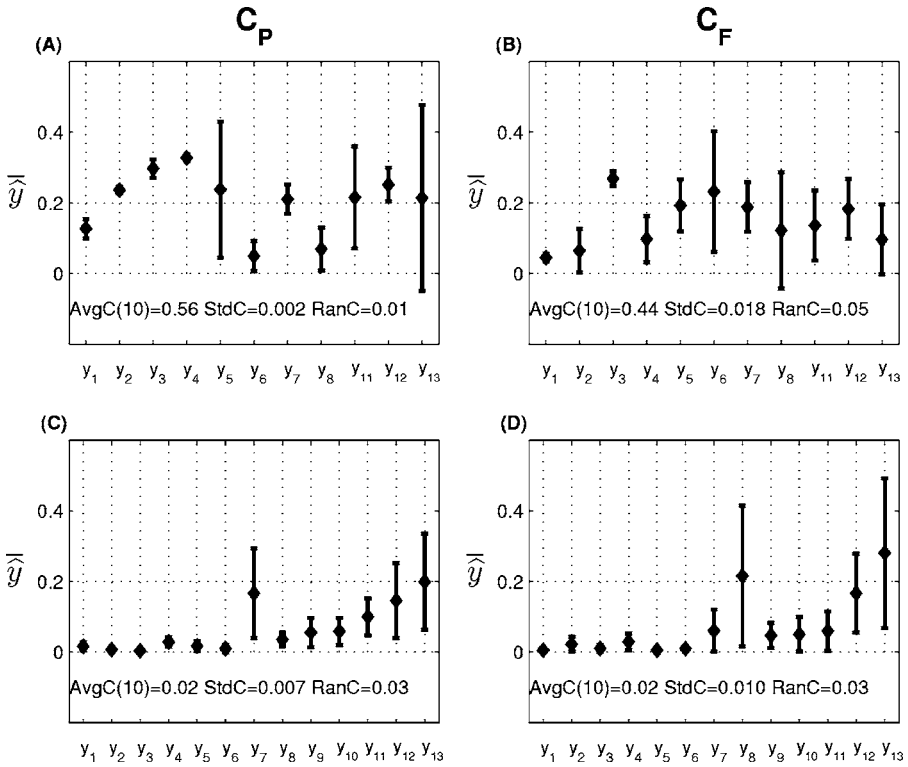


FIG. 5. Average and standard deviation of coefficient error, $\hat{\mathbf{y}}$, for the synthetic data case. Each plot represents an average of the ten inversion results; the average optimum cost function value of these ten inversions (AvgC) is printed at the bottom of each plot, along with the standard deviation (StdC) and range (RanG) of the cost function. (a) C_P cost function, assuming $b=\theta=0$. (b) C_F cost function, assuming $b=\theta=0$. (c) C_P cost function; b and θ included in the parameter search space. (d) C_F cost function; b and θ included in the parameter search space.

Both of these cost function values are far from the optimal $C_P=C_F=0$. Figures 4(a) and 4(b) show incorrect convergence for all but one and three of the eigenvector coefficients, respectively. The very poor optimal cost function value in these inversions is an indicator that the parametrization used does not allow for a true realization of the geoacoustic environment. Incorporating the bow and tilt into the parameter search space improved cost function values to $C_P=7.4\times10^{-3}$and $C_F=6.8\times10^{-3}$. Accompanying this significant cost function value improvement is good convergence for a large number of eigenvalue coefficients, as seen in Figs. 4(c) and 4(d).

The goal of an inversion is to resolve the original geoacoustic parameters, $\{x_i\}$. It is well known that, for large parameter search spaces, there is more than one set of parameters $\{x_i\}$ that will produce an acceptable cost function value. For this reason, it is desirable to have a measure of each parameter's resolvability. The eigenvectors which define the optimal rotation offer a tool for determining the hierarchy of parameter resolution. The parameters highest in the parameter hierarchy are consistently resolved close to the true value, while those lowest in the hierarchy are not often reliably resolved. The parameters in the middle of the hierarchy are often resolvable, but with a substantial amount of expected error. Following, we present an empirical measure of parameter resolvability in terms of consistency by using the standard deviation of several inversion results. These standard deviations will also show that the inversion results are "seed independent," that is, they are independent of the initial parameter values, which are randomly selected within the parameter windows using different arbitrary seeds for the random number generator.

To analyze results of several inversions initialized with a different seed for the random number generator, we examine the average normalized error of the eigenvector coefficients, $\bar{\hat{\mathbf{y}}}$, where

$$\bar{\hat{y}}_j = \frac{1}{N_{\text{run}}} \sum_{k=1}^{N_{\text{run}}} \frac{|y_{\text{best}}(j,k) - y_{\text{true}}(j)|}{2}. \tag{7}$$

Here, $y_{\text{true}}(j)$ is the coefficient used for generating the synthetic data [y_j in Eq. (2)], and $y_{\text{best}}(j,k)$ is the jth optimal coefficient (estimate of y_j) for the kth inversion of N_{run} inversions. This normalization was chosen because usually $-1\leq y_j\leq 1$, and the normalization allows for ease in plotting y_j. Figure 5 is a plot of the average of the normalized error, $\bar{\hat{\mathbf{y}}}$, for the best 10 out of 20 inversions. The error bars in Fig. 5 represent their standard deviation from $\mathbf{y}_{\text{true}}$. Figures 5(a) and 5(b) are for inversions using the cost functions C_P and C_F, respectively, when the HLA is assumed perfectly straight, and Figs. 5(c) and 5(d) are for inversions using the cost functions C_P and C_F when the parameter search space includes HLA bow and tilt. Overall, there are larger errors with inversions using C_P when bow and tilt are not included in the parameter search space.

Recall from Eq. (2) that the geoacoustic parameters $\mathbf{x}$ are a linear transformation of the eigenvectors $\mathbf{v}_j$, with coefficients y_j; Fig. 6 is a plot of the average and standard deviation of the normalized parameter error, $\bar{\hat{\mathbf{x}}}$, for the same ten inversion runs, where

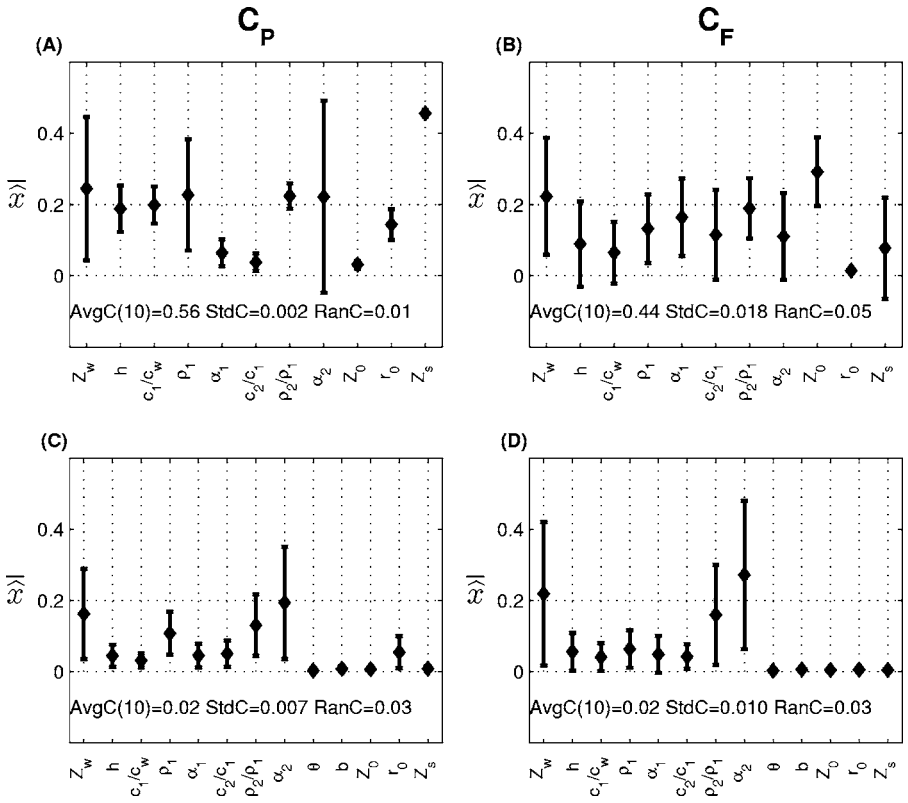


FIG. 6. Average and standard deviation of parameter error, $\hat{\mathbf{x}}$, for the synthetic data case. Each plot represents an average of the ten inversions; the average optimum cost function value of these ten inversions is printed at the bottom of each plot (AvgC), along with the standard deviation (StdC) and range (RanC) of the cost function. (a) C_P cost function, $b=\theta=0$. (b) C_F cost function, $b=\theta=0$. (c) C_P cost function; b and θ included in the parameter search space. (d) C_F cost function; b and θ included in the parameter search space.

$$\bar{\hat{x}}_i = \frac{1}{10}\sum_{k=1}^{10} \frac{|x_{\text{best}}(i,k) - x_{\text{true}}(i)|}{b_i - a_i}. \tag{8}$$

Here $x_{\text{best}}(i,k)$ is the ith optimal parameter x_i for the kth inversion run, $x_{\text{true}}(i)$ is the parameter value used when generating the synthetic data, and $[a_i,b_i]$ are the parameter windows, or constraints. The overall large errors in the coefficients $\{y_j\}$ for C_P inversions without HLA bow and tilt are reflected in the average errors $\bar{\hat{x}}$ in Fig. 6(a). As expected, Figs. 6(b) and 6(d) show that the parameter r_0 is very confidently resolved when C_F is used as the cost function, regardless of whether or not bow and tilt are included in the parameter search space.

V. EFFECTS OF PARAMETER CONSTRAINTS

Generally it is expected that tighter parameter constraints will yield more favorable geoacoustic inversion results, as long as the optimal solution(s) are within the parameter search space (Ω), and the inversion results are not biased by the inversion algorithm employed.[16] The previous section showed that the inversions presented here are not biased by the randomly generated initial parameter values, a bias to consider with a simulated annealing method. The windows $[a_i,b_i]$ chosen for the inversions in the previous section, as detailed in Table I are reasonable assumptions for a typical scenario: generally it is possible to have tighter constraints on geometric parameters than on environmental parameters. However, because parameters are correlated in complicated ways, it is not obvious what impact parameter window size has on the resolvability of each parameter. In this section we explore the effect of narrowing the parameter constraints.

Figure 7 is a plot of optimal parameter values as a function of normalized window size, as determined by 300 separate inversions, each of which achieved an optimal cost function value of $C_P \leq 0.05$. In these plots we see a reflection of the parameter hierarchy: the most resolvable parameters have points clustered around the true value, while the least resolvable parameters have points scattered throughout, regardless of window size. The most resolvable geometric parameters are θ, b, z_0, and z_s, followed by the resolvable environmental parameters h and c_1/c_w. These parameters were identified as being in the top half of the parameter hierarchy in similar parametrization in Sec. IV. Also resolvable to a lesser degree are $\rho_1, \alpha_1, c_2/c_1$, and the geometric parameter r_0. These parameters are approximately in the middle of the parameter hierarchy in similar parametrization in Sec. IV.

For all of these inversions, the cost function C_P was used, as well as the synthetic data from the previous section. The window size for each parameter in each inversion is randomly determined: the largest possible window for each parameter is limited by $[A_i,B_i]$, and each inversion uses the randomly determined window $[a_i,b_i]$ defined by

$$a_i(\nu) = \nu(A_i - x_i) + x_i, \tag{9}$$

$$b_i(\nu) = \nu(B_i - x_i) + x_i. \tag{10}$$

Here, $\{x_i\}$ is the set of geoacoustic parameter values used to compute the synthetic data, $[A_i,B_i]$ define the maximum geoacoustic parameter windows possible, and ν is the normalized window size, randomly generated and uniformly distributed between 0.1 and 1.0 A separate random ν is used for each parameter within one inversion run, so that each inversion is constrained by different normalized window

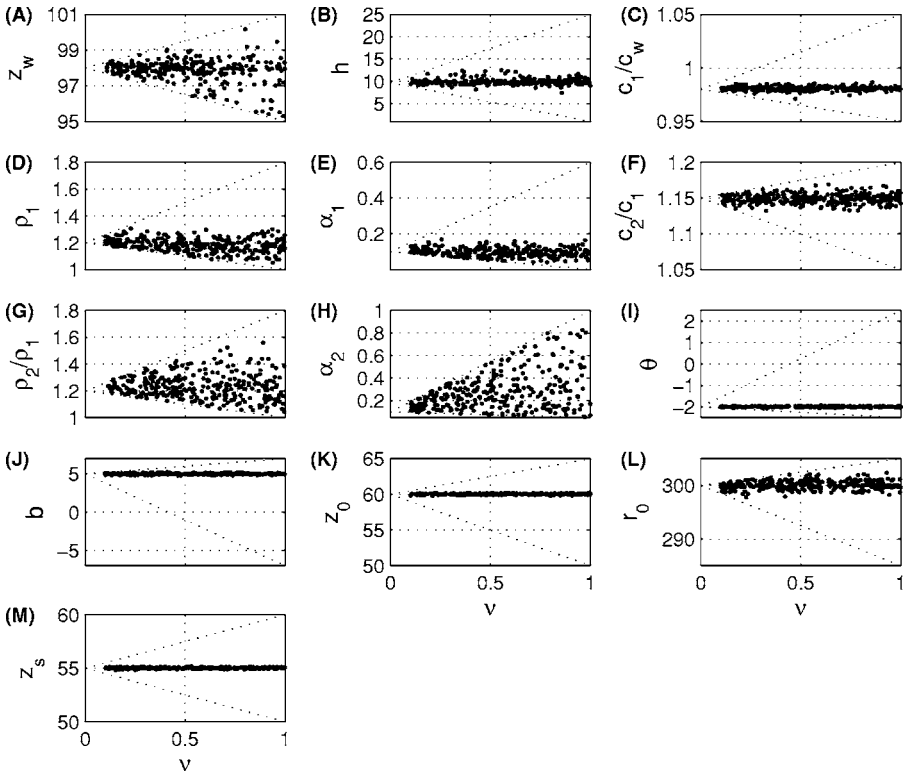


FIG. 7. (a)–(m) Optimal parameter estimates as a function of normalized window, size, ν The full vertical axis in (a) through (m) is the maximum possible window, $[A_i, B_i]$. The dashed lines on each plot denote the windows for each parameter as a function of normalized window size, ν. (a) Water depth (z_w), (b) sediment thickness (h), (c) sediment-water sound speed ratio, (c_1/c_w), (d) sediment density (ρ_1), (e) sediment, attenuation (α_1), (f) bottom-sediment sound speed ratio (c_2/c_1), (g) bottom-sediment density ratio (ρ_2/ρ_1), (h) bottom attenuation (α_2), (i) HLA tilt (θ), (j) HLA bow (b), (k) HLA nominal depth (z_0), (l) range from source to first element of HLA (r_0), and (m) source depth (z_s).

sizes for each parameter. The full vertical axes in Figs. 7(a)–7(m) are the maximum windows considered for each parameter, $[A_i, B_i]$, and these are the same as the limits in Table I. On each plot in Fig. 7, the transverse dashed lines are $a_i(\nu)$ and $b_i(\nu)$, delineating the region in which possible values can be found. Note that a new coordinate rotation is defined for each new set of constraints, $[a_i(\nu), b_i(\nu)]$, prior to the inversions. As the parameter constraints vary drastically from those presented in Sec. IV, it is possible for the parameter hierarchy to change.

Figure 8 shows a summary of the same inversion results as in Fig. 7 in the form of average values and standard deviations. For each parameter, inversion results were grouped by values of ν in increments of 0.10. The average for each group is displayed as a function of ν, and the standard deviation is displayed as error bars. The dashed line on each

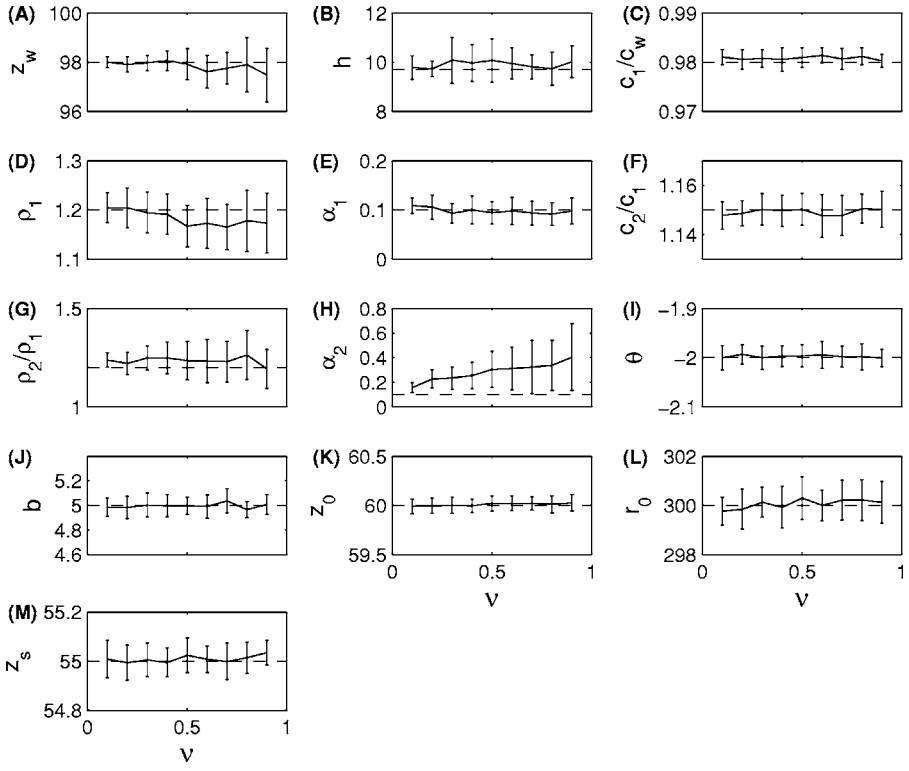


FIG. 8. Average and standard deviation of optimum parameter estimates as a function of normalized window size, ν. Inversion results are grouped by values of ν in increments of 0.1; the average and standard deviation for each group is plotted for each parameter. These are the same inversion runs as represented in Fig. 7. (a) Water depth (z_w), (b) sediment thickness (h), (c) sediment-water sound speed ratio, (c_1/c_w), (d) sediment density (ρ_1), (e) sediment attenuation (α_1), (f) bottom-sediment sound speed ratio, (c_2/c_1), (g) bottom-sediment density ratio (ρ_2/ρ_1), (h) bottom attenuation (α_2), (i) HLA tilt (θ), (j) HLA bow (b), (k) HLA nominal depth (z_0), (l) range from source to first phone of HLA (r_0), and (m) source depth (z_s).

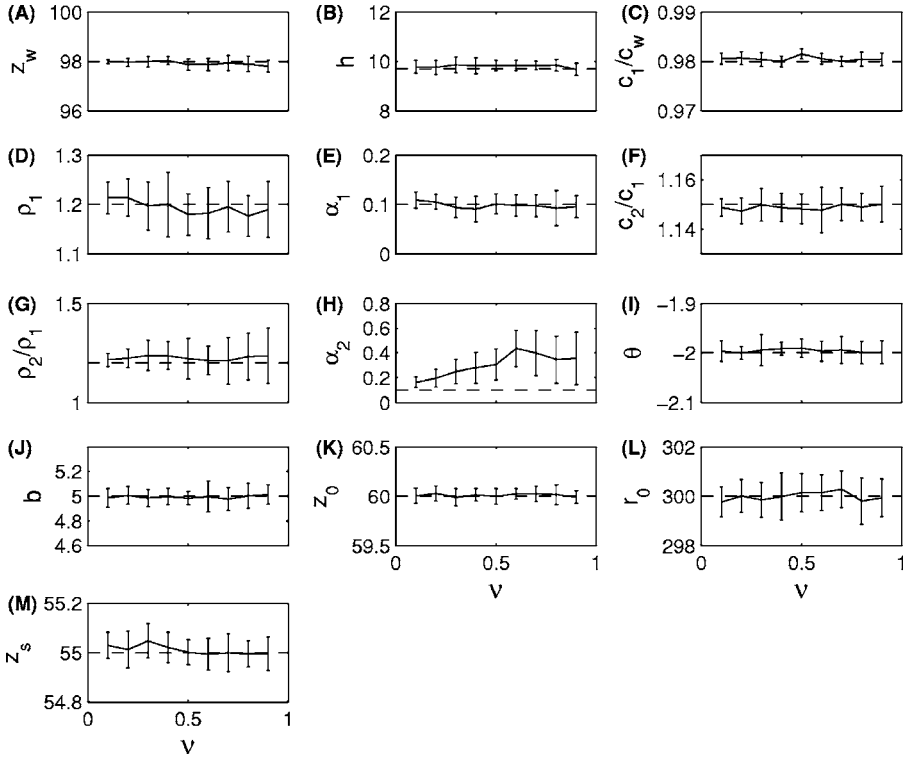


FIG. 9. Average and standard deviation of optimum parameter estimates as a function of normalized window size, ν. These plots are similar to those in Fig. 8 for a different set of inversions, where the maximum window size for water depth (z_w) is narrowed to ±1; all other parameters' maximum window size is the same as for inversions represented in Fig. 8. The vertical axes here are held the same as those in Fig. 8 for ease in comparing the change in standard deviations.

plot represents the true value for that parameter. Note that the vertical axes in Fig. 8 are different from those in Fig. 7, so that the error bars are visible for the most resolvable parameters. In Fig. 8, we see that the standard deviation for the most resolvable parameters remains almost constant with increasing window size ν. Just as interesting are the practically unresolvable parameters, ρ_2/ρ_1 and α_2; in Fig. 7 we see that almost any value for these parameters will give an acceptable cost function. In Fig. 8, these are the parameters whose standard deviation grows with the window size ν, further confirming that these parameters are not resolved by the inversion method implemented, and the parameter values resulting from the inversions are arbitrary within the window specified.

When examining Figs. 7 and 8, note that these offer a single slice through a multi-dimensional parameter space. If one of the maximum windows $[A_i,B_i]$ were to change for just one parameter, then the results could also change because the parameters are correlated in complicated ways—further emphasizing the need to traverse the parameter landscape through a rotated coordinate system. We illustrate the complex effect of correlated parameter constraints by viewing similar results after reducing the maximum window $[A_1,B_1]$ for z_w to [97,99]; the results for this new set of inversions is plotted in Fig. 9. The axis limits in Fig. 9 are the same as those in Fig. 8, to illustrate the decrease in standard deviation for the resolvable parameters, specifically h and c_1/c_w. Previously unresolvable parameters ρ_2/ρ_1 and α_2 remain unresolvable.

Conversely, when we increase the window for r_0 to ±45 m, then r_0 moves up in the parameter hierarchy for C_P (results not shown here). The range parameter, r_0, becomes more resolvable relative to the environmental parameters, but still with an expected error much larger than the same problem using C_F as the cost function. This change in parameter hierarchy is expected since the matched field cost function C_P will be significantly altered when the ranges between the source and receivers change by a substantial amount; C_P is insensitive to r_0 only for a small change in r_0 in this HLA geometry because the source is near endfire.

In summary, the effect parameter constraints have on parameter resolvability is on the amount of expected error in the results, illustrated by the standard deviation here. The expected error in the inversion results appears constant for the resolvable parameters (those highest in the hierarchy), but this constant value is tied to the maximum possible windows, $[A_i,B_i]$; this was seen by the narrowing of just one maximum window leading to a decrease on the expected error for other resolvable geoacoustic parameters. It is important to note that drastic changes in parameter constraints can lead to a change in parameter hierarchy. For inversion problems with large constraint windows for the geometric parameters, it is possible to apply sequential annealing to constrain geometric parameters to smaller windows.[33] The smaller window constraints will decrease the expected error for the other resolvable geoacoustic parameters. Quantifying the relationship between a resolvable parameter's expected error and the maximum parameter constraints $[A_i,B_i]$ requires more investigation than is presented here. However, the results presented do provide a qualitative representation of the relationship between parameter constraints and parameter resolvability.

VI. EXPERIMENTAL HLA DATA: PARAMETER HIERARCHY

In this section, inversion results are analyzed for several points along a track of MAPEX2000 data, all collected on 7

TABLE II. Parametrization and optimum parameter estimates for geoacoustic inversions of MAPEX2000 data measured on 7 March 2000 at 08:05Z, 09:06Z, and 10:24Z.

i	Parameter (x_i)	08:05Z	09:06Z	10:24Z	Min	Max
	Sediment layer					
1	z_w, Water depth (m)	101.9	124.45	136.6	$z_{\text{echo}}-3.0$	$z_{\text{echo}}+3.0$
2	h, Thickness (m)	5.43	20.3	15.2	1.0	25.0
3	c_1/c_w, Sound speed ratio	0.98	1.03	1.04	0.95	1.05
4	ρ_1, Density (g/cm^3)	1.05	1.54	1.61	1.00	1.80
5	α_1, attenuation (dB/λ)	0.08	0.17	0.08	0.00	0.60
	Reflecting layer					
6	c_2/c_1, Sound speed ratio	1.14	1.09	1.15	1.05	1.20
7	ρ_2/ρ_1, Density ratio	1.00	1.04	1.02	1.00	1.80
8	α_2, Attenuation (dB/λ)	0.37	0.25	0.92	0.05	1.00
	Geometric parameters					
9	θ, Array tilt (deg)	0.50	0.48	0.34	−0.5	0.5
10	b, Array bow (m)	−1.10	−0.44	−0.88	−3.0	3.0
11	z_0, Array depth (m)	55.1	55.6	55.1	55.0	61.0
12	r_0, Sor-Rec range (m)	293.6	293.7	293.6	290.0	294.0
13	z_s, Source depth (m)	55.1	53.8	55.1	52.0	58.0

March 2000: 08:04Z, 08:05Z, 09:06Z, and 10:24Z. All of the signals analyzed are broadband acoustic linear frequency-modulated (LFM) sweeps from 150 to 850 Hz. Detailed information on the signals broadcast and on the source and receiver configurations during this portion of the experiment are available in Ref. 10. As with the synthetic in data in previous sections, 11 frequency bins corresponding to 250–750 Hz in 50-Hz increments were used in the inversions. The entire HLA is 254 m in length, with 128 elements spaced at 2 m; for the analysis presented here, only the even numbered elements were used, giving a nominal phone spacing of 4 m and an effective 64-phone HLA. These are the same HLA attributes used in the synthetic data examples presented in the previous sections. The nominal source and receiver depths are 55 and 60 m, respectively.

The water sound speed profile (SSP) was assumed known, and a four iso-velocity layer approximation of the experimentally obtained profile was used in the inversions. The two water SSPs used in the inversions are plotted in Fig. 1: Fig. 1(a) is the SSP measured at 08:07Z (solid line) and its four-layer approximation (dashed line); Fig. 1(b) is the SSP measured at 09:11Z (solid line) and its four-layer approximation (dashed line). The four-layer SSP in Fig. 1(a) is the same profile used for the synthetic data inversions and was used for inversions of all experimental data sets prior to 08:40Z. The four-layer SSP in Fig. 1(b) (SSP024) was used for inversions of all experimental data sets from 08:40Z and later.

The parametrization used for the geoacoustic inversions of the MAPEX2000 data sets is detailed in Table II. Most of these parameter windows are the same as those used for the simulations presented in Sec. IV, although some geometric windows were shifted or narrowed, based on *a priori* tests; the tests showed the array bow and tilt to be small, and provided smaller windows for the source and receiver range and depths. Table II also displays the estimates of the parameter values resulting from phone-coherent inversions of the experimental data.

For each of the time periods 08:05Z, 09:06Z, and 10:24Z, Figs. 10–12 display (a) the eigenvectors ($\mathbf{v}_j$), (b) the inversion history for the coefficients (y_i), (c) the average and standard deviation of the normalized coefficients ($\tilde{y}_i$), and (d) the average and standard deviation of the normalized original parameters ($\tilde{x}_i$). The averages and standard deviations in Figs. 10, 11, and 12(a)–12(d) result from ten separate inversions with different randomly generated initial parameter values for each case; the average cost function value $\bar{C}_P$ is printed on these plots, as well as the standard deviation and the range of values of C_P. The parameter values ($\tilde{y}_i$) and ($\tilde{x}_i$) are normalized as follows:

$$\tilde{\mathbf{y}} = \frac{\mathbf{y}}{2}, \tag{11}$$

$$\tilde{x}_i = \frac{x_i - (a_i + b_i)/2}{b_i - a_i}. \tag{12}$$

As expected, the eigenvectors in Figs. 10, 11, and 12(a) are similar to each other, and similar to those in the simulation of Sec. IV. By examining the coefficient inversion histories in conjunction with the eigenvectors, one can see that at least five coefficients are resolved; the resolution of these coefficients is reflected in their small standard deviations in Figs. 10, 11, and 12(c). It follows that the parameter combinations in the first five eigenvectors are expected to be resolved. The first five eigenvectors are linear combinations of the sediment thickness (h), receiver bow (b), receiver depth (z_0), source depth (z_s), water depth (z_w), and sediment sound speed ratio (c_1/c_w). The resolution of these parameters is

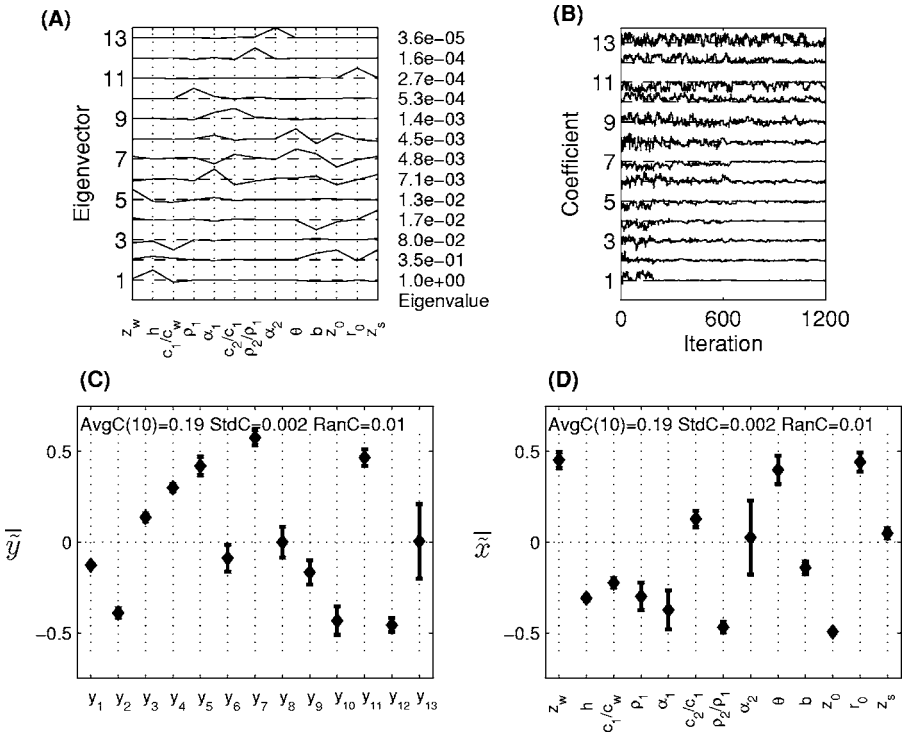


FIG. 10. Analysis of inversion results for MAPEX2000 data collected on 7 March at 08:05Z. (a) The eigenvectors and eigenvalues of K, (b) the coefficient history for a typical inversion (all y_j as a function of iteration number), (c) average and standard deviation of normalized coefficient values $\tilde{\mathbf{y}}$ for ten inversions, and (d) average and standard deviation of normalized parameter values $\tilde{\mathbf{x}}$ for the same ten inversions. The average cost function value (AvgC) for these ten inversions is printed at the top of (c) and (d), along with the standard deviation (StdC) and range of values (RanC).

reflected in their small standard deviations in Figs. 10, 11, and 12(d). Similarly, the parameters low in the parameter hierarchy and contained in eigenvectors 9–13 are not expected to be sufficiently resolved, and these are $\rho_1, \rho_2/\rho_1, \alpha_2$, and r_0; parameters middle to low in the hierarchy have questionable resolution, and these are α_1 and c_2/c_1.

As stated in Sec. IV, it is expected that both the phone-coherent and frequency-coherent inversion methods will reliably resolve the parameters $h, c_1/c_w, z_0$, and z_s. The main expected difference between the two methods is in the resolution of the source-receiver range, r_0. These expected differences are confirmed with experimental data: Fig. 13 compares the averages and standard deviations of $\tilde{x}$ [Eq. (12)] resulting from inversions of MAPEX2000 data taken at 08:04Z, using C_P and C_F, each with and without bow and tilt included in the parameter search space. Each symbol in Fig. 13(a) represents an average of ten inversion results, and each symbol in Fig. 13(b) represents the associated standard deviation: the symbol $\times$ is for inversions using C_P as the cost function with no array deformation in the parameter search

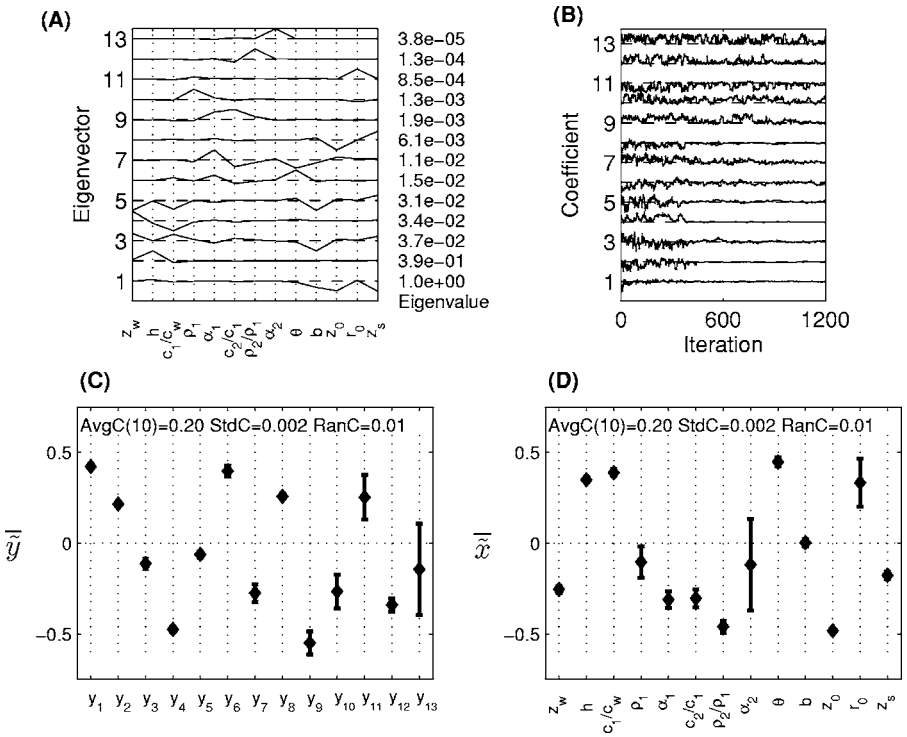


FIG. 11. Analysis of inversion results for MAPEX2000 data collected on 7 March at 09:06Z. (a) The eigenvectors and eigenvalues of K, (b) the coefficient history for a typical inversion (all y_j as a function of iteration number), (c) average and standard deviation of normalized coefficient estimates $\tilde{\mathbf{y}}$ for ten inversions, and (d) average and standard deviation of normalized parameter estimates $\tilde{\mathbf{x}}$ for the same ten inversions. The average cost function value (AvgC) is printed at the top of (c) and (d), along with its standard deviation (StdC) and range of values (RanC).

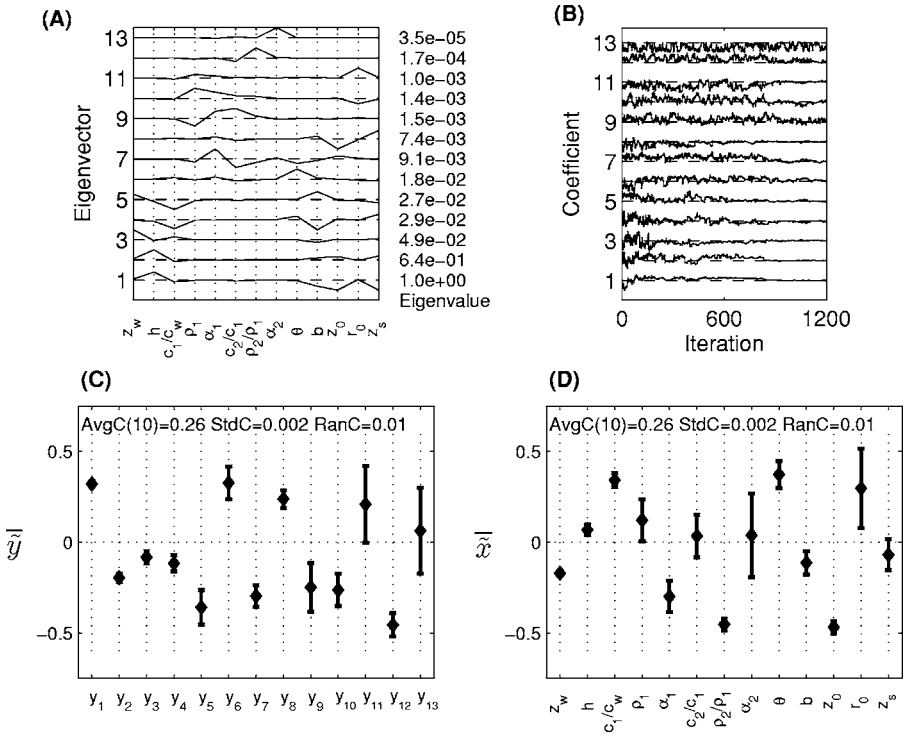


FIG. 12. Analysis of inversion results for MAPEX2000 data collected on 7 March at 10:24Z. (a) The eigenvectors and eigenvalues of K, (b) the coefficient history for a typical inversion (all y_j as a function of iteration number), (c) average and standard deviation of normalized coefficient estimates $\tilde{\mathbf{y}}$ for ten inversions, and (d) average and standard deviation of normalized parameter estimates $\tilde{\mathbf{x}}$ for the same ten inversions. The average cost function value (AvgC) is printed at the top of (c) and (d), along with its standard deviation (StdC) and range of values (RanC).

space, the symbol **o** is for inversions using C_P with array bow and tilt included in the parameter search space, the symbol + is for inversions using C_F and no array deformation, and the symbol $\diamond$ is for inversions using C_F with bow and tilt in the parameter search space. The parameter windows for these inversions are the same as those detailed in Table I, and are the same parameter constraints used for the synthetic data inversions in Sec. IV. Note the standard deviation of the geometric parameter r_0 in Fig. 13(b): regardless of whether array deformation parameters are included in the parameter search space, when the cost function C_F is used, the standard deviation is very small. This is very different from when the cost function C_P is used; the standard deviation for r_0 is relatively large. Overall, the average inverted parameter values [Fig. 13(a)] are similar for parameters at the top of the parameter hierarchy, and these average values have a rela-

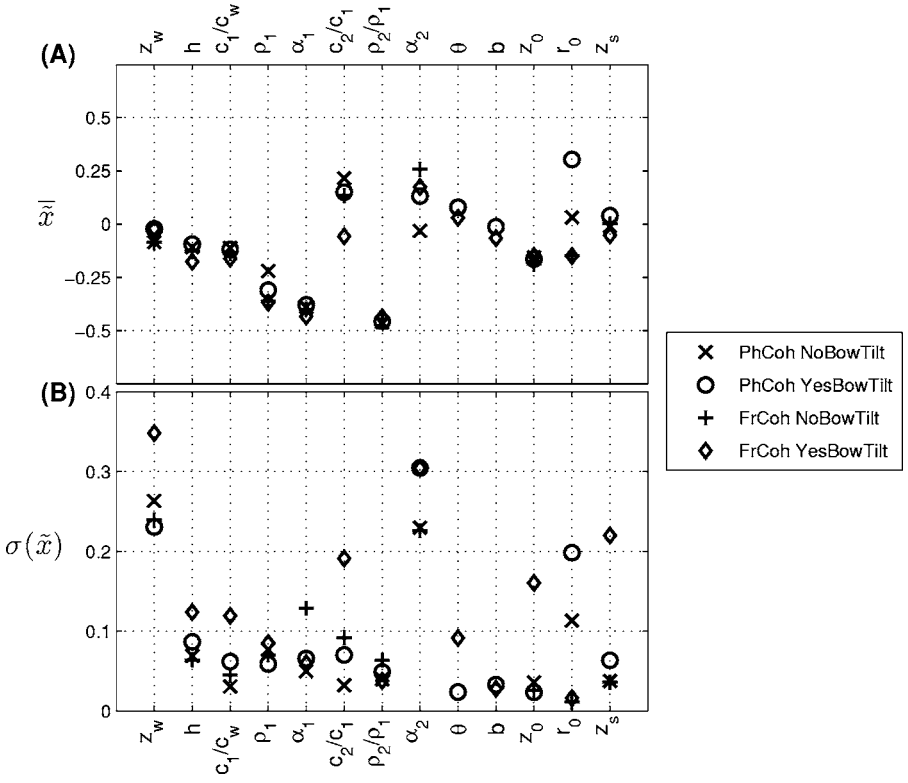


FIG. 13. Comparison of (a) the average and (b) the standard deviations of normalized parameter estimates, $\tilde{\mathbf{x}}$, resulting from ten inversions of MAPEX2000 data collected on 7 March at 08:04Z. The limits of the parameter search windows are the same as in Table I. The symbol $\times$ is for inversions using the cost function C_P and no bow and tilt in the parameter search space; the symbol **o** is for inversions using the same cost function and including bow and tilt in the parameter search space. The symbol + is for inversions using the cost function C_F and no bow and tilt in the parameter search space, and the symbol $\diamond$ is for inversions using the same cost function and including bow and tilt in the parameter search space.

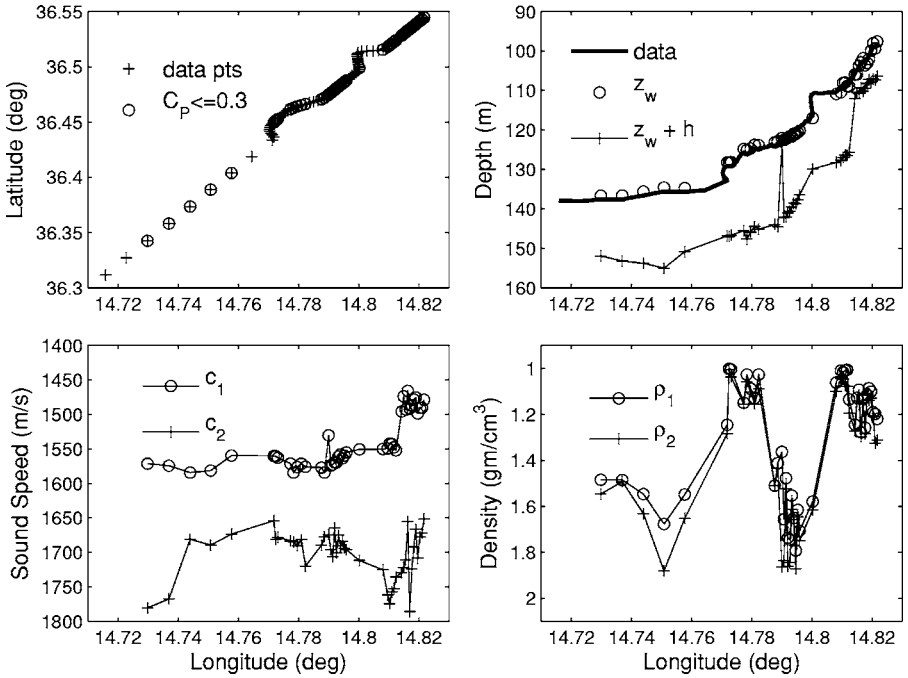


FIG. 14. MAPEX2000: 7 March track from 08:04Z through 10:44Z. Only inversion results where the cost function ≤ 0.3 are displayed. (a) + symbols indicate data frames processed, and **o** symbols indicate frames where a cost function ≤ 0.3 was obtained. (b) Solid line indicates water depth as measured by navigation during the experiment, **o** symbols indicate depth estimates resulting from geoacoustic inversion, and + symbols indicate depth estimates of sediment layer resulting from geoacoustic inversions. (c) **o** symbols indicate sediment sound speed estimates (c_1), + symbols indicate bottom sublayer sound speed estimates (c_2) resulting from geoacoustic inversions. (d) **o** symbols indicate sediment density estimates (ρ_1), and + symbols indicate bottom sublayer density estimates (ρ_2) resulting from geoacoustic inversions.

tively small standard deviation [Fig. 13(b)]. The parameters at the bottom of the parameter hierarchy (e.g., $\alpha_2, c_2/c_1$) are considered unresolvable, and the standard deviation for these is relatively large.

VII. EXPERIMENTAL HLA DATA: GEOACOUSTIC INVERSION OF A TRACK

Phone-coherent geoacoustic inversions were performed along a track of experimental data measured on 7 March 2000; data was processed every minute from 08:04Z to 09:22Z, and every 10 min from 09:24Z to 10:44Z. The broadcast signals are broadband acoustic LFM sweeps from 150 to 850 Hz; the data analyzed in Sec. VI is a subset of this data. As in previous sections, 11 frequency bins corresponding to 250–750 Hz in 50-Hz increments were used in the inversions, and the same 64 of 128 HLA elements were used. The nominal source and receiver depths are 55 and 60 m, respectively. The parameter constraints are the same as those detailed in Table II, except the water depth was constrained to ±1 m from the echo sounder output. Figures 14 and 15 provide a summary of the geoacoustic inversion results from the entire track of data processed. Figure 14(a) shows the latitude and longitude locations of the data points processed (+ symbol), and indicates which of those data frames provided inversion results with a cost function $C_P \leq 0.3$ (**o** symbol). A cost function value of 0.3 corresponds to

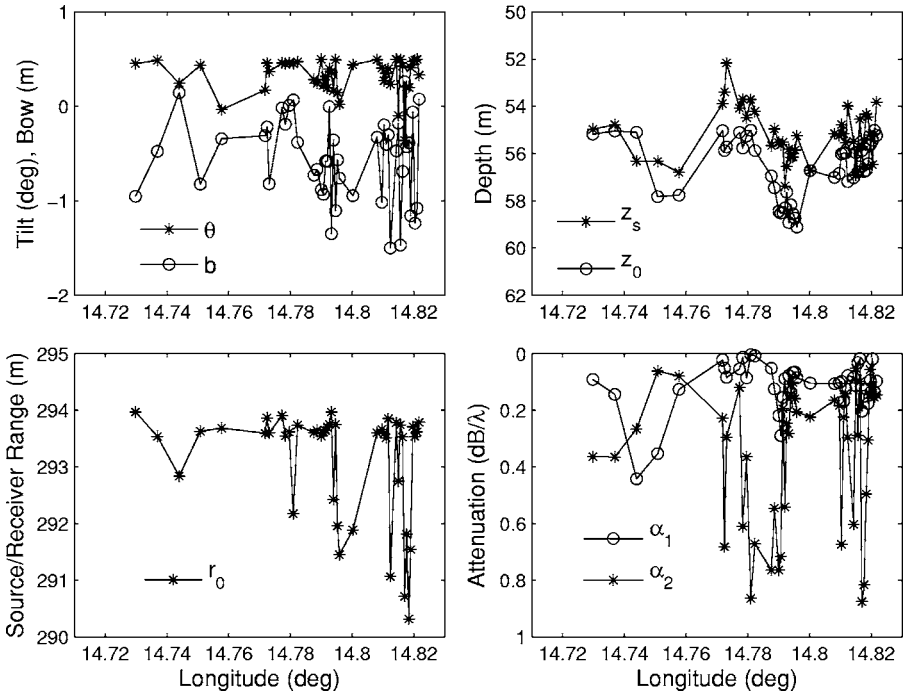


FIG. 15. MAPEX2000: 7 March track from 08.04Z through 10.44Z. Only the inversion results where the cost function ≤ 0.3 are displayed. (a) * symbols indicate tilt estimates (θ), and **o** symbols indicate bow estimates (b) resulting from geoacoustic inversions. (b) * symbols indicate estimates of source depth (z_s), and **o** symbols indicate estimates of nominal HLA depth (z_0) resulting from geoacoustic inversions. (c) * symbols indicate estimates for range between the source and the first phone of the HLA (r_0) resulting from geoacoustic inversions. (d) **o** symbols indicate sediment attenuation estimates (α_1), and * symbols indicate bottom sublayer attenuation estimates (α_2) resulting from geoacoustic inversions.

a matched-field value of 0.7, or 1.5 dB of degradation. Acceptable cost function values were not obtained at times when the tow ship made sharp turns.

Typical parameter hierarchies for data along this track were seen in the previous section, illustrated by the eigenvectors in Figs. 10, 11, and 12(a). The most resolvable parameters are $h, c_1/c_w, b, z_0$, and z_s, with the three source-receiver geometry parameters being highly correlated to each other (b, z_0, and z_s). The inversions are expected to resolve h and c_1/c_w with a small amount of error, as well as a combination of b, z_0, and z_s. Parameters in the middle of the hierarchy are $\alpha_1, c_2/c_1$, and θ; these parameters are resolvable, but with a larger amount of expected error. The parameters at the bottom of the hierarchy, and therefore considered unresolvable by the method implemented, are $\rho_1, \rho_2/\rho_1, \alpha_2$, and r_0. The consistency in the results, as illustrated in Figs. 14 and 15, confirm these parameter hierarchies and resolvabilities.

Figure 14(b) is a comparison of the experimentally obtained water depth (solid line) with that resulting from the inversions (symbol **o**). These coincide well because the parameter window for z_w only allows for a 1-m deviation from the measured depth. The more interesting curve in Fig. 14(b) is the sediment depth (line with + symbol). Inversions indicate there is a significant change in sediment thickness around 14.81°E. The change in inverted sediment thickness coincides with a change in inverted sediment sound speed (c_1) [line with **o** symbol in Fig. 14(c)]. Inspection of the eigenvectors in Figs. 10, 11, and 12(a) indicates that these two parameters (h and c_1/c_w) are not highly correlated to each other, and therefore a concerted change in both parameters is not merely an artifact of parameter correlation. Different bottom properties from the two distinct regions along this track also resulted from inversions in Ref. 10, and the change in bottom properties at approximately 14.81°E was also observed in inversion results of the same data track in Refs. 11 and 13; the change is validated by the high-resolution seismic reflection profile obtained during the MAPEX2000 experiment of the same region.[11] In the region where current inversions resolve a consistently thin sediment layer (6.9±1.5 m), c_1 is also consistently slower than the water column sound speed, on average 1485 m/s, ±10.0 m/s. In the region from 14.71°E to 14.81°E, inversions resolve a thicker sediment layer (18.5±2.0 m), and a faster sediment sound speed, on average 1564 m/s, ±14 m/s. All of these results are in keeping with the most recent inversions by Fallat *et al.* in Ref. 13, where inversions of the data from the thin sediment region resulted in average sound speeds of 1484 m/s, and inversions from the thicker sediment region yielded average sound 1562 m/s. In the same publication, the values were shown to coincide closely with cores of the top sediment layer.[13]

Figure 14(c) also has a plot of c_2 values resulting from inversions; recall that the parameter c_2/c_1 is resolvable, but with a substantial amount of expected error. The average value of c_2 over the entire track is 1703 m/s, with a standard deviation along the track of 35 m/s. Note in Fig. 14(c) that there is a larger variability of the parameter c_2 between 14.81°E and 14.82°E; without including that portion of the track for the c_2 ensemble, the average drops to 1688 m/s, with a smaller standard deviation of 16 m/s. Both of these standard deviation measures for c_2 are within reason of the expected error, as predicted by the standard deviation of repeated inversions of one data point, illustrated in Fig. 12(d): the standard deviation for c_2 resulting from inversions of data from 10:24Z is 23 m/s. The inverted values of c_2 averaged over the track (1703 or 1688 m/s)are similar to previous inversions;[11,10,13] the most recent inversions by Fallat *et al.* resolve c_2 to be approximately 1691 m/s.

In Fig. 15(a) are the tilt and bow parameters; we see that the tilt remains fairly constant, with an average of 0.3° along the track (and a standard deviation of 0.2°), and the bow is generally between −2 and 0 m, with an average of −0.6 m and a standard deviation of 0.5 m. Figure 15(b) summarizes the source and nominal HLA depths resulting from the inversions; variations in these depths are in concert with each other and with the HLA bow. These three parameters (z_s, z_0, and b) are correlated to each other, as seen in the eigenvector plots of Figs. 10, 11, and 12(a), and in this case the relationship between the source-receiver geometry parameters is more resolvable than the individual parameter values.

Figure 15(c) is a plot of source-receiver ranges (r_0) resulting from the inversions, showing that the inversions were unable to conclusively resolve this parameter within the 4-m parameter window, especially at the higher longitudes. This window for r_0 was narrowed based on *a priori* inversions of a few data points using C_F. Because r_0 is at the bottom of the parameter hierarchy when cost function C_P is used, it is not expected to be resolved, especially within the small 4-m window. Also at the bottom of the parameter hierarchy for C_P (and not expected to be resolvable) are ρ_1 and ρ_2/ρ_1, seen in Fig. 14(d), and α_2, seen in Fig. 15(d).

VIII. SUMMARY

Geoacoustic inversions using two matched-field cost functions are applied to broadband synthetic and experimental data. The frequency-coherent cost function, which requires prior knowledge of the source spectrum, is shown to be very sensitive to the source-receiver range, but relatively insensitive to deformations to the HLA. The phone-coherent cost function, which does not require knowledge of the source spectrum, is not sensitive to the source-receiver range, making it a suitable choice for self-noise geoacoustic inversions where the probing source is likely to be distributed throughout the tow ship. The phone-coherent cost function is shown to be sensitive to HLA shape, with larger errors in parameter resolution resulting when the HLA is incorrectly assumed to be completely horizontal. The phone-coherent cost function is successfully applied in HLA inversions with array deformation parameters incorporated into the parameter search space.

The global optimization method used in all the geoacoustic inversions presented traverses the parameter landscape in directions that are aligned with valleys of the cost function. These directions are defined by the eigenvectors of the covariance matrix of the gradient of the cost function over the parameter space.[18] The eigenvectors disclose param-

eter correlations and hierarchy *a priori* to inversions, making it possible to predict the resolvability of the geoacoustic parameters. The parameter correlations, hierarchy, and resolvability are empirically confirmed using both synthetic and experimental data. Additionally, the robustness of parameter resolvability to increasing parameter window size is demonstrated using synthetic data.

In summary, this paper conducts a systematic study of a geoacoustic inversion method using a frequency-incoherent approach applicable to future inversions using ship self-noise as the probing signal. Its performance is compared to that of a frequency-coherent method previously used to analyze LFM sweeps broadcast during MAPEX2000.[10,11,13] Reliability and consistency are necessary if this method is to be used (in practice) to estimate the geoacoustic properties of the bottom over a large area. To assure reliability and consistency, the geometric parameters of the source and receiver array need to be known to a certain accuracy; this accuracy is determined in this study using an empirical approach. The result serves as a guide for the specification and design of a towed-array based geoacoustic inversion system. We demonstrate the methodology, but note that specific values will be case dependent, specifically in terms of water depth, source-receiver separation, and water sound speed profile.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors wish to acknowledge the engineering staff and crew of the *NRV Alliance*, and NURC, under whose direction the MAPEX2000 data was collected. The authors would also like to thank Martin Siderius for his help in accessing and processing the MAPEX2000 data presented here.

[1]D. F. Gingras and P. Gerstoft, "Inversion for geometric and geoacoustic parameters in shallow water: Experimental results," J. Acoust. Soc. Am. **95**, 770–782 (1994).
[2]J. P. Hermand and P. Gerstoft, "Inversion of broadband multi-tone acoustic data from the Yellow Shark summer experiments," IEEE J. Ocean. Eng. **2l**(4), 324–346 (1996).
[3]M. Siderius, M. Snellen, D. G. Simons, and R. Onken, "An environmental assessment in the Strait of Sicily: Measurement and analysis techniques for determining bottom and oceanographic properties," IEEE J. Ocean. Eng. **25**(3), 364–386 (2000).
[4]C. F. Huang and W. S. Hodgkiss, "Matched-Field Geoacoustic Inversion of Low-Frequency Source Tow Data From the ASIAEX East China Sea Experiment," IEEE J. Ocean. Eng. **29**(4), 952–963 (2004).
[5]K. Yang, Y. Ma, C. Sun, J. H. Miller, and G. R. Potty, "Multistep Matched-Field Inversion for Broad-Band Data from ASIAEX2001," IEEE J. Ocean. Eng. **29**(4), 964–972 (2004).
[6]G. V. Frisk and J. F. Lynch, "Shallow water waveguide characterization using the Hankel transform," J. Acoust. Soc. Am. **76**, 205–216 (1984).
[7]W. A. Kuperman, M. F. Werby, K. E. Gilbert, and G. J. Tango, "Beam forming on bottom-interacting tow-ship noise," IEEE J. Ocean. Eng. **10**(3), 290–298 (1985).
[8]S. M. Jesus and A. Caiti, "Estimating geoacoustic bottom properties from towed array data," J. Comput. Acoust. **4**(3), 273–290 (1996).
[9]A. Caiti, S. M. Jesus, and A. Kristensen, "Geoacoustic seafloor exploration with a towed array in a shallow water area of the Strait of Sicily," IEEE J. Ocean. Eng. **21**(4), 355–366 (1996).
[10]M. Siderius, P. L. Nielsen, and P. Gerstoft, "Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array," J. Acoust. Soc. Am. **112**, 1523–1535 (2002).
[11]M. R. Fallat, P. L. Nielsen, and M. Siderius, "The characterization of a range-dependent environment using towed horizontal array data from the MAPEX 2000 experiment," SACLANTCEN Report SM-402 (2002).
[12]D. J. Battle, P. Gerstoft, W. A. Kuperman, W. S. Hodgkiss, and M. Siderius, "Geoacoustic inversion of tow-ship noise via near-field matched-field processing," IEEE J. Ocean. Eng. **28**(3), 454–467 (2003).
[13]M. R. Fallat, P. L. Nielsen, S. E. Dosso, and M. Siderius, "Geoacoustic characterization of a range-dependent ocean environment using towed array data," IEEE J. Ocean. Eng. **30**(1), 198–206 (2005).
[14]M. Siderius, P. L. Nielsen, and P. Gerstoft, "Performance comparison between vertical and horizontal arrays for geoacoustic inversion," IEEE J. Ocean. Eng. **28**, 424–431 (2003).
[15]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[16]M. R. Fallat, S. E. Dosso, and P. L. Nielsen, "An investigation of algorithm-induced variability in geoacoustic inversion," IEEE J. Ocean. Eng. **29**, 78–87 (2004).
[17]K. Yoo, T. C. Yang, L. Fialkowski, D. Dacol, J. Perkins, M. Fallat, P. Nielsen, and M. Siderius, "Sensitivity to array tilt and bow for broadband geoacoustic inversion using a towed array," J. Acoust. Soc. Am. **113**, 2217 (2003).
[18]M. D. Collins and L. Fishman, "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**, 1637–1644 (1995).
[19]H. Szu and R. Hartley, "Fast simulated annealing," Phys. Lett. A **122**, 157–162 (1987).
[20]H. Schmidt and A. B. Baggeroer, "Physics-imposed resolution and robustness issues in seismo-acoustic parameter inversion," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht, 1995), pp. 85–90.
[21]M. R. Fallat and S. E. Dosso, "Geoacoustic inversion via local, global and hybrid algorithms," J. Acoust. Soc. Am. **105**, 3219–3230 (1999).
[22]L. T. Fialkowski, J. F. Lingevitch, J. S. Perkins, D. K. Dacol, and M. D. Collins, "Geoacoustic inversion using a rotated coordinate system and simulated annealing," IEEE J. Ocean. Eng. **28**, 370–379 (2003).
[23]M. Porter, "The KRAKEN normal mode program," SACLANTCEN SM-254, La Spezia, Italy, SACLANT Undersea Research Centre (1991).
[24]M. B. Porter and E. L. Reiss, "A numerical method for bottom interacting ocean acoustic normal modes," J. Acoust. Soc. Am. **77**, 1760–1767 (1985).
[25]M. B. Porter and E. L. Reiss, "A numerical method for ocean acoustic normal modes," J. Acoust. Soc. Am. **76**, 244–252 (1984).
[26]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).
[27]J. F. Lingevitch and M. D. Collins, "Estimating elastic sediment properties with the self-starter," Wave Motion **31**, 157–163 (2000).
[28]D. K. Dacol, M. D. Collins, and J. F. Lingevitch, "An efficient parabolic equation solution based on the method of undetermined coefficients," J. Acoust. Soc. Am. **106**, 1727–1731 (1999).
[29]S. E. Dosso and P. E. Nielsen, "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband shallow-water data," J. Acoust. Soc. Am. **111**, 142–159 (2002).
[30]D. J. Battle, P. Gerstoft, W. S. Hodgkiss, and W. A. Kuperman, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[31]P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).
[32]P. Gerstoft and C. F. Mecklenbrauker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[33]T. B. Neilsen, "An iterative implementation of rotated coordinates for inverse problems," J. Acoust. Soc. Am. **113**, 2574–2586 (2003).

Point-to-point underwater acoustic communications using spread-spectrum passive phase conjugation

Paul Hursky,[a] Michael B. Porter, and Martin Siderius
Heat, Light, and Sound Research Inc., 12730 High Bluff Drive, San Diego, California 92130

Vincent K. McDonald
Space and Naval Warfare Systems Center, San Diego, 53560 Hull Street, San Diego, California 92152-5001

(Received 13 December 2005; revised 19 April 2006; accepted 19 April 2006)

The ocean is often a complex multipath channel and progress has been made in developing equalization algorithms to overcome this. Unfortunately, many of these algorithms are computationally demanding and not as power-efficient as one would like; in many applications it may be better to trade bit rate for longer operational life. In 2000 the U.S. Navy was developing an underwater wireless acoustic network called Seaweb, for which a number of modulation schemes were being tested in a series of SignalEx experiments. This paper discusses two modulation schemes and associated receiver algorithms that were developed and tested for Seaweb applications. These receiver designs take advantage of time reversal (phase conjugation) and properties of spread spectrum sequences known as Gold sequences. Furthermore, they are much less complex than receivers using adaptive equalizers. This paper will present results of testing these signaling and receiver concepts during two experiments at sea. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2203602]

PACS number(s): 43.60.Tj, 43.60.Dh, 43.30.Re [EJS] Pages: 247–257

I. INTRODUCTION

Acoustic signaling for wireless digital communications in the undersea environment can be a very attractive alternative to both radio telemetry systems (vulnerable to weather, rough seas, and pilfering) and cabled systems (vulnerable to commercial trawling). However, time-varying multipath and often harsh ambient noise conditions characterize the underwater acoustic channel, often making acoustic communications challenging. Much effort has been directed at developing channel equalizers and adaptive spatial processing techniques so that coherent phase modulation can be used to achieve the desired high spectral efficiencies.[1,2] These techniques are computationally demanding with many parameters needing to be set, requirements that are not especially well suited for applications where autonomy, adaptability, and long-life battery operation are being contemplated.

Time reversal (or phase conjugation, in the frequency domain) was demonstrated[3–5] in the early 1960s as a means of refocusing sound that had been spread in time by propagation through the ocean. More recent work on this concept[6–8] has seen further experimental validation and the development of a number of applications. In particular, passive phase conjugation can be used for pulse compression,[6] using a vertical line array receiver so that both spatial as well as temporal focusing is achieved, which addresses the difficulties posed by multipath for acoustic communications.[9,10]

In 2000, the U.S. Navy was developing an underwater wireless networking system, called Seaweb, using acoustic communications as the physical transport layer.[11] We were tasked to investigate and test a variant of pulse position modulation (PPM) for use as an alternative modulation scheme in this system. In the course of adapting the original PPM scheme to Seaweb, we realized how this modulation technique was connected to evolving work on time reversal (phase-conjugation),[8] and ended up modifying it to better exploit these new techniques for the Seaweb system.[12] Comparing our work to other reported passive phase conjugate methods,[9,10,13] we use a single-element source and a single-element receiver, without relying on an aperture at the source or receiver for spatial focusing. To compensate, we rely upon the gain from despreading direct-sequence spread-spectrum (DSSS) sequences.[14,15] DSSS is a form of code division multiple access or CDMA. As in terrestrial wireless CDMA systems, this modulation scheme can accomodate multiple users if different orthogonal codes are assigned to different users. In addition, using spread-spectrum sequences renders this modulation scheme more difficult to detect[16] for applications where covertness is desired.

We review time reversal and phase conjugation in Sec. II. Section III presents the details of our modulation scheme. In particular, we show how varying the parameters of PPM to increase its spectral efficiency pushes us to smaller alphabets, leading us to abandon PPM in favor of differential phase-shift keying (DPSK), which has the smallest possible alphabet. In Sec. IV, we present the results of testing these modulation techniques and receiver algorithms in several sea experiments.

Previously, work in both terrestrial wireless and underwater acoustics considered a modulation technique called code shift keying[17] or sequence position modulation[18] (both variants of pulse position modulation or PPM) that also uses

[a]Electronic address: paul.hursky@hlsresearch.com

spread-spectrum sequences.[14,15] As we will outline, we favor DPSK over PPM, but we note that code shift keying can potentially benefit from time reversal (phase conjugation) as well. We will also comment on the similarity of our DPSK scheme to a Rake receiver (a type of receiver often used in spread spectrum systems[14]).

II. HOW TIME REVERSAL (PHASE CONJUGATION) BENEFITS ACOUSTIC COMMUNICATIONS

A communication system consists of a transmitter that sends a data-modulated waveform through a channel (in our case, the ocean) and on to a receiver which must perform some processing to recover the transmitted data. Typically, the channel introduces distortion that limits the receiver's ability to recover the transmitted information. This distortion includes attenuation, time spreading or multipath, and Doppler shifts and spreads. The Doppler effects are due to transmitter and receiver motion, as well as the motion of the ocean boundaries and the ocean itself. If time spreading is present in the channel (due to transmitted signals arriving along multiple paths), previously transmitted symbols may corrupt the detection of the current symbol, a problem known as intersymbol interference.

Various receiver algorithms have been developed to despread the multipath arrivals (or equalize them, viewed in the frequency domain). In some cases, the diversity provided by multiple arrivals is exploited so successfully that a net processing gain is achieved, thus turning a problematic channel property into an asset. Reducing intersymbol interference is the "holy grail" of communications, because it enables symbols to be transmitted at very high rates. One approach for coping with intersymbol interference is to use an adaptive filter[19] to adjust a set of filter coefficients to minimize the mean squared difference between the filtered output and either a known training sequence or the closest known discrete symbol value in a "decision-directed" mode.[2] Although fast algorithms for such equalizers have been the subject of much research, this approach still requires great care with respect to computational load, algorithm stability, and automated selection of adaptive filter parameters for an unknown and often time-varying channel.

The time reversal (phase-conjugation) approach that we are exploring in this paper avoids the explicit recovery of the channel and its subsequent equalization via signal processing and its associated algorithmic complexity. Instead, this approach implicitly recombines the multiple arrivals signal instead of trying to invert the channel.[9,10] To review how this focusing is achieved, Fig. 1 illustrates two ways of implementing time reversal, or phase conjugation (its frequency domain equivalent). We have labeled these two configurations active phase conjugation,[7,8] or APC, and passive phase conjugation,[6] or PPC. Both have been experimentally validated in the ocean. The channel impulse response (CIR) function is $h(t)$ and its Fourier transform is $H(\omega)$. $H(\omega)$ is the Green's function for the particular ocean waveguide and source and receiver locations, although we omit the dependence on locations in our notation. Recall that, in the frequency domain, the convolution of $h(t)$ and $s(t)$ is $H(\omega)S(\omega)$.

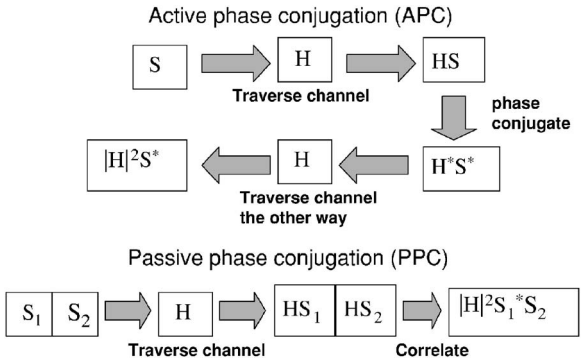


FIG. 1. Active and passive phase conjugation.

Similarly, the correlation of $s_1(t)$ and $s_2(t)$ is $S_1(\omega)S_2^*(\omega)$, where the asterisk indicates complex conjugation. In Fig. 1 and the text below, we use the frequency domain equivalents of all waveforms, and drop the dependence on ω.

In the active configuration (APC):

- The left-hand station transmits a waveform S, which travels through the channel H (from left to right in Fig. 1), and is recorded on the right-hand station as HS.
- The right-hand station time reverses the received waveform, or equivalently, phase conjugates it, producing H^*S^*, and retransmits it back to the left-hand station (from right to left in Fig. 1). The retransmitted waveform can carry either a sign or a phase to convey information (back) to the left-hand station. Assuming the channel has not changed, the time-reversed waveform H^*S^* travels back through the same channel, and is convolved with H again, producing $|H|^2S^*$ at the original left-hand station, the time-reversed version of the original signal S convolved with the autocorrelation of H. The left-hand station is the information receiver in this configuration.

In the passive configuration (PPC), shown in the lower part of Fig. 1:

- The left-hand station transmits S_1, which travels through the channel H, and is observed on the right-hand station as HS_1.
- The left-hand station transmits S_2, which is observed on the right-hand station as HS_2 (again, if the channel has not changed).
- The right-hand station cross-correlates HS_1 and HS_2, producing $|H|^2S_1S_2^*$, the correlation of S_1 and S_2, convolved with the autocorrelation of the CIR H. The right-hand station is the information receiver in this configuration.

The basic idea in phase conjugation is that the autocorrelation of the CIR $|H|^2$ tends to reconcentrate or focus the multipath arrivals at zero time lag. The term $|H|^2$ is the time reversal or phase conjugation focusing operator.[20] However, depending on the distribution of the multipath arrivals in $h(t)$, the autocorrelation may also have temporal sidelobes that result in residual intersymbol interference, even after focusing. Other researchers[9,21] have used arrays of receivers

or transmitters to average down these temporal sidelobes. The different transmitter-receiver focused terms (i.e., $|H|^2S$ or $|H|^2S_1S_2^*$, depending on the configuration) are typically aligned along their main peak, prior to averaging. As a result, they share a common main peak, but have sidelobes at different locations. Upon averaging, all elements contribute to the same main peak, but spread their sidelobes wherever they may fall.

In our technique, we avoid the cost and complexity of transmit or receive arrays by using spread spectrum sequences. This relies upon despreading gain and temporal focusing alone,[12] although this does not yield bit rates as high as can be produced with source and receiver arrays. This focusing, or multipath recombination, is achieved at each "symbol" by forming an inner product of the current and predecessor snapshots of the channel, where each channel snapshot is the output of a correlator matched to the known spread spectrum sequence (we cycle through a known set of orthogonal sequences). Such a receiver structure is similar in function to and has the (low) computational complexity of a linear equalizer,[19] although it is approximating the channel inverse by its adjoint (our receiver forms the inner product of H and its adjoint H^*, to form the focusing operator $|H|^2$).

Figure 1 shows that the APC configuration focuses the original waveform S (producing $|H|^2S$ at its receiver station), while the PPC configuration focuses $S_1S_2^*$ (producing $|H|^2S_1S_2^*$ at its receiver station), the correlation of the two consecutively transmitted waveforms (S_1 and S_2). Therefore, in our PPC configuration, the message must be encoded in the correlation of the two consecutively transmitted waveforms S_1 and S_2. Encoding information in the correlation of two waveforms is not a typical signaling scheme and may provide some advantages, although we have not been particularly creative in pursuing this, other than to try the simple variations described below.

III. WAVEFORM DESIGN FOR POINT-TO-POINT ACOUSTIC COMMUNICATIONS USING PASSIVE PHASE CONJUGATION

We will describe two signaling schemes, one based on pulse position modulation (PPM), the other on differential phase shift keying (DPSK).[19] We have implemented these two modulation schemes to take advantage of passive phase conjugation (the PPC configuration in Fig. 1). These waveform designs rely upon transmitting two waveforms, s_1 and s_2, in which information bits have been embedded, with the expectation that both s_1 and s_2 will propagate through the same channel h, so that they are received as HS_1 and HS_2 (in the transform domain). As discussed in the previous section, the passive phase conjugation (PPC) is realized by correlating these received waveforms to produce $|H|^2S_1S_2^*$. This operation produces a despread or focused $S_1S_2^*$ because $|H|^2$ combines the multipath arrivals.

As already mentioned in Sec. II, because we are working with single hydrophones (no arrays), we have more of an intersymbol interference problem than in configurations where arrays are used. To compensate for this, we rely upon families of sequences called Gold codes,[14,15] designed to minimize the correlation between the sequences in each family. Gold codes are similar to maximal-length sequences, or m-sequences, that are often used in the underwater acoustic community for their autocorrelation properties.[22,23] Gold sequences, like m-sequences, are bipolar sequences with values -1 and 1. Their special property is that any two different Gold sequences from the same family have very low (circular) cross-correlation values (i.e., at all lags). For integers $m>2$ at which a *preferred pair of m-sequences* can be found, a family of Gold sequences can be derived whose cross-correlation spectrum is three-valued. There are $2^{m-1}+2$ Gold sequences in each family. When transmitted, each Gold sequence modulates the phase of a carrier (i.e., using binary phase-shift keying or BPSK modulation).

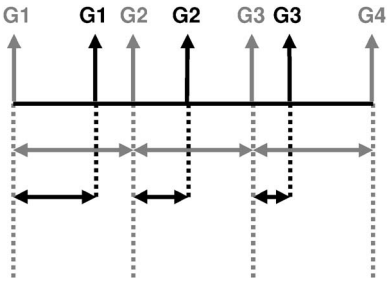


FIG. 2. This diagram illustrates the waveform design for PPC-PPM. Each arrow indicates the onset (in time) of a Gold sequence. Each G_i indicates that the ith Gold sequence is being used (from a particular family of sequences). Each symbol consists of a gray reference pulse and a black position pulse (so three symbols are shown above). The information is conveyed by the separation between the reference pulse and the position pulse (shown by the horizontal black arrows). All sequences have the same length—the interval between consecutive reference pulses (shown by the horizontal gray arrows). Gold sequences corresponding to the position pulses of consecutive symbols may overlap.

Our modulation cycles through a series of Gold sequences (i.e., from the same family), using a different Gold sequence for each symbol. The particular order of the Gold sequences being transmitted is known at the receiver, so the appropriate matched filter can be applied to each received symbol. When symbols overlap due to multipath, the low cross-correlation property of the Gold sequences ensures that the different matched filters do not let through much of the interfering symbols. To accomodate multiple users simultaneously, different subsets of Gold sequences can be assigned to different users—each user must have enough sequences so that the time it takes to cycle through this user's subset exceeds the time spreading in the channel. However, note that we do not fully exploit the touted low cross-correlation property of these sequences, because in the presence of multipath, the correlation that is being performed on all the different multipath arrivals is not circular.

A. Pulse position modulation using Gold sequences (PPC-PPM)

Figure 2 shows how we implement (PPM) to take advantage of passive phase conjugation (PPC). Here G_i indicates the ith Gold sequence from a family of 2^m+1 sequences, each a sequence of 2^m-1 bipolar symbols, or chips, where the start of each sequence is indicated by a single arrow. Arrows indicating the onset of each Gold sequence rather than the full-length sequences are shown to illustrate

the scheme. Each sequence is actually a bipolar sequence modulated by a carrier (i.e., BPSK modulated) of length indicated by the horizontal gray arrows. Each G_i is repeated, with the first G_i (indicated in gray) setting a reference position, relative to which the position of the second G_i (indicated in black) is measured. The varying distances between reference and second positions are indicated by the horizontal black arrows. The position of the second G_i is purposely varied to convey the information bits being transmitted, hence the name PPM. If the time interval between reference G_i's, indicated by the horizontal gray arrows, from G_i to G_{i+i}, is divided into N resolvable time slots, each pair of G_i's will carry $\log_2 N$ bits of information. To relate this design to the notation in the introduction above, s_2 is identical to s_1, but they overlap and the time between them is varied to set the pulse position.

At the receiver, a matched filter tuned to G_i performs a pulse compression on each G_i and reproduces the multipath arrival structure associated with both instances of G_i. The arrivals associated with G_j (j not equal to i) are to a large extent suppressed (by the matched filter tuned to G_i), since the different Gold sequences have low cross correlation. After this pulse compression, the matched filter output contains two copies of the CIR, overlapped and delayed with respect to one another, corresponding to each of the two G_j. At this point it is possible to decode the information from the relative positions of the dominant arrivals only, but this would not take advantage of the additional signal energy available in the other arrivals. Instead, the concept is applied to our pulse-compressed pair of G_i receptions. Each is spread by what is probably the same channel, since there has been little time for the channel to have changed during this interval, so if we auto-correlate the matched filter output (tuned to G_i), we implicitly autocorrelate the CIR H by which each of the two G_i receptions have been spread, realizing a filter consisting of the time reversal (phase conjugation) operator $|H|^2$ as discussed in the previous section.

It is interesting to note that this refocusing could also be exploited in the code shift keying work, if a reference sequence and a sequence to indicate position were somehow incorporated [we need to correlate two copies of the $h(t)$ to form the autocorrelation of $h(t)$ which provides the focusing]. Our modulation scheme differs from those described in the code shift keying and sequence position modulation work,[16,18] in that we do not circularly permute the waveform that sets the position (for PPM), but instead merely delay it.

B. Differential phase shift keying (DPSK) using Gold sequences (PPC-DPSK)

With a bandwidth B, we can resolve PPM time slots spaced at intervals of $1/B$. A symbol period of T seconds will have room for BT time slots (or positions), or $\log_2 BT$ bits per symbol, with a bit rate of $\frac{\log_2 BT}{T}$. The bit rate can be increased by increasing B or reducing T. Because the denominator T grows faster than numerator $\log_2 T$, increasing the number of PPM slots by lengthening T actually reduces the bit rate. So, although we were originally motivated to use PPM to pack more bits into each symbol, we find instead that

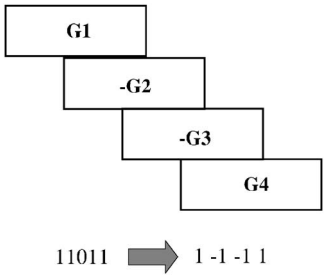


FIG. 3. PPC-DPSK waveform design.

reducing the number of PPM slots is what increases the bit rate. Ultimately, if we want a maximum bit rate, we should use the smallest number of PPM slots that we can, which suggests an alphabet of size two. In this section, we back off from using PPM entirely, and adopt a DPSK framework, encoding our information bits in the relative polarity of neighboring Gold sequences.

Figure 3 shows an alternative PPC modulation waveform design. In this case, we cycle through the 2^m+1 Gold sequences, spacing them at regular intervals, but varying their sign. The transmitted information is recovered at the receiver by comparing the sign of the current G_i with the sign of its predecessor G_{i-1}, in effect realizing a DPSK modulation. In this case, s_1 and s_2 are different Gold sequences, G_{i-1} and G_i, with the relative polarity indicating the information bit being transmitted.

Figure 4 shows how each pair of consecutively transmitted Gold sequences, G_i and G_{i-1}, are processed. S_1 and S_2 are the waveforms at the receiver corresponding to G_{i-1} and G_i, with S_2 following immediately after S_1. At the transmitter, a sign change may be applied to either G_{i-1} and/or G_i, depending on the information bit being transmitted. It is this relative sign that carries the information, and which must be recovered by the receiver. After traveling through the channel, S_1 is HG_{i-1} and S_2 is HG_i, each with a possible information bearing sign change. S_1 and S_2 overlap and their start times are calculated by a symbol timing process (to be described later). At the receiver, a matched filter tuned to the appropriate Gold sequence (G_{i-1} for S_1 and G_i for S_2) is applied to each of the two waveforms, producing $H|G_{i-1}|^2\text{sign}(G_{i-1})$ and $H|G_i|^2\text{sign}(G_i)$. The $|G|^2$ factors are pulse compressions of the Gold sequences. Both of these matched filter outputs are essentially estimates of the channel H with the polarity originally applied at the transmitter to carry the information

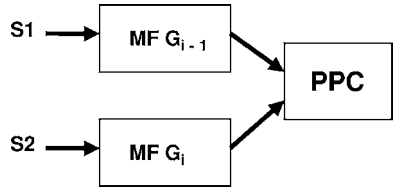


FIG. 4. This is the PPC-DPSK receiver design, in which every pair of consecutively arriving signals S_1 and S_2 is processed by a pair of matched filters, S_1 being correlated with Gold sequence G_{i-1}, and S_2 being correlated with Gold sequence G_i. The two matched filter outputs are correlated (in block PPC) to compare the phases of S_1 and S_2. It is the phase difference between S_1 and S_2 that carries the information bit. We have only tested phase differences of 180 deg (i.e., changing the sign of S_2 relative to S_1 or not).

bit. Although we could compare the signs of individual peaks from the two channel estimates, we want to combine all of the multipath arrivals before we do that. As in the PPC-PPM modulation, we correlate the two matched filter outputs, producing waveform proportional to $|H|^2, |G_{i-1}|^2, |G_i|^2$, and the product of the signs of the two transmitted waveforms. Again, we end up filtering our information bearing waveform, in this case $\text{sign}(G_{i-1})\text{sign}(G_i)$, by the time reversal (phase conjugation) focusing operator $|H|^2$. This implicitly recombines the multipath. The information bit is recovered from the sign of this final correlation.

The PPC-DPSK modulation scheme is similar to spread-spectrum schemes that use a Rake receiver for recombining multipath arrivals.[24] In both cases, a matched filter is applied to a spreading sequence to isolate multipath arrivals. In the case of a Rake receiver, multiple matched filters are applied at a number of delays, or "fingers" (of the Rake), prior to combining the contributions from each "finger" so that they are all aligned in time. In our PPC-DPSK, phase conjugation or the inner product of $H|G_{i-1}|^2\text{sign}(G_{i-1})$ and conj $[H|G_i|^2\text{sign}(G_i)]$ is used to focus the multipath arrivals (this inner product contains the focusing operator $|H|^2$). In the case of the Rake receiver, some decision must be made as to where to position the multiple "fingers" of the Rake receiver. In Sec. IV B, we will show how we have borrowed an idea from the Rake receiver to greatly improve our demodulation results for the PPC-DPSK scheme.

C. Discussion

The usual dilemma in APC and PPC, at least in how it has been implemented previously, is that the channel estimate must be periodically refreshed to keep up with a time-varying channel. This is done by interrupting the flow of information bits to send a probe pulse to recalibrate the channel and waiting for the channel to clear before reinitiating information bits. An alternative to "clearing the channel" in this way has been investigated by other authors,[13] in which the channel estimate is continually refreshed, using the current block of detected symbols to estimate the channel for the next block of symbols. The channel is estimated by finding the best fit (in a least-squares sense) to the received data and the decoded symbols, although this seems to get away from the minimalist implementation that is usually cited as PPC's main attraction. Our two transmission schemes (PPC-PPM and PPC-DPSK), described in the previous section, do not require an explicit channel estimate since they implicitly refresh the channel state information by cross-correlating waveforms corresponding to consecutive pulses (PPM) or symbols (DPSK).[12]

The spacing between consecutive sequences determines the bit rate

$$R = 1/T_{\text{spacing}} \tag{1}$$

but we cannot reduce this spacing indiscriminately, because these sequences are not perfectly orthogonal, especially after being convolved with a CIR function. Their touted good (i.e., low) cross correlation properties are based upon periodic (or circular) cross correlation and accurate "framing" of the sequence. Neither condition is realized in our modulation schemes, because the various multipaths arrive at different times and because the different sequences are staggered in our modulation scheme. As a result, there is more interference between the overlapped sequences than is predicted for these sequences under perfect conditions. Increasing the bit rate by reducing the symbol spacing creates more overlap between sequences and makes the interference worse.

Gold sequences have lengths of 2^m-1 for a particular choice of m. The longer the sequence, the lower its autocorrelation sidelobes and correlations with other Gold sequences from the same family. However, longer sequence lengths mean more of an overlap with following sequences, if the spacing is kept the same. As a result, the length must be chosen as a compromise between these two competing considerations.

The chip rate should ideally be matched to the bandwidth. A chip rate of 4 kHz results in spectral nulls at the edges of our 8 kHz band, and is thus optimal for our bandwidth. We also tested chip rates of 8 and 16 kHz in order to gain their resulting higher bit rates, knowing they are not matched to our 8–16 kHz band and that their cross-correlation properties would be degraded as a result of losing out-of-band information.

When consecutive Gold sequences are overlapped, an unacceptably high peak to average power ratio can result, as in multiple carrier modulation schemes, such as orthogonal frequency-division multiplexing or OFDM. A set of constructively interfering Gold sequences can produce an unusually large and isolated peak. If we scale the transmitted waveform according to this single isolated peak, our transmitted waveform will have an unacceptably high peak-to-average-power ratio, resulting in very low average transmitted power. To avoid this, we clipped isolated peaks to minimize this loss in transmitted power, recognizing that this results in some degradation in the receiver matched filter, since some of the transmitted waveform was lost to this clipping. The incidence of unusually large and isolated peaks is proportional to the number of overlapping Gold sequences, so this too limits our bit rates.

IV. RESULTS OF TESTING DURING SIGNALEX EXPERIMENTS

The U.S. Navy has supported a series of sea tests under the SignalEx program[25] to measure channel effects upon different underwater acoustic signaling schemes, including our passive phase conjugate technique. These tests have been performed in a variety of environments using lightweight, modular hardware units, called Telesonar Testbeds, shown in Fig. 5. These testbeds, developed at SPAWAR Systems Center, are unique, high-fidelity, modular, reconfigurable, autonomous, wide-band instruments for high-frequency acoustic propagation and communication research. They consist of a PC-104, single-board computer in a deployable bottle, including hard drives for recording received waveforms or sourcing waveforms for transmission. The testbed can record waveforms on a four-channel receive array (spaced for diversity, with elements separated by five wavelengths at 12 kHz) and transmit in three bands, 8–16, 14–22, 25–50 kHz. A very

FIG. 5. Telesonar Testbed used to collect data.

accurate clock allows for time-division multiplexed transmissions from multiple testbeds. The testbed units are also equipped with a commercial acoustic modem for status checking and remote control. During a typical experiment, two testbed units are deployed. One is moored to the sea floor and the other is deployed over the side of a small boat, which is either anchored or allowed to drift. One testbed transmits a preprogammed sequence of waveforms stored on disk, and the other testbed records the received waveforms (to disk), after they have traveled through the ocean waveguide. The transmission sequences and power levels can be controlled remotely via the commercial modem. We will present results for fixed-drifting configurations at two sites.

A. Results from SignalEx-E: Ship Island, off Gulfport during AUV/Modem Fest 2001

The SignalEx-E experiment was performed near Ship Island on 24-25 October 2001, during ModemFest/AUVFest off the coast of Florida. The source was deployed at a depth of roughly 2 m over the side of a boat that was allowed to drift at roughly 0.6 m/s. The receiver was moored to the sea floor. The bathymetry along the transmission path was nearly constant at a depth of 5 m. Figure 6 shows the drift track, starting at a range of about 1 km and ending at a range of roughly 5 km.

Figure 7 shows the CIR measured by applying a matched filter to a series of hyperbolic frequency-modulated (HFM) chirps, each 50 ms long, sweeping from 8 to 16 kHz, transmitted every 250 ms. There is virtually no multipath in this extremely shallow water channel (the horizontal axis spans only 6 ms). The bottom was silt (from examining the receiver moorings), but even if it were more reflective, acoustic paths having even a slight grazing angle would interact with it so many times over the ranges we were operating at that they would be absorbed. The few bottom-interacting paths that would get through at very shallow grazing angles would have virtually the same travel time as a direct or surface reflected path (and would contribute to the response shown in Fig. 7).

Both the PPC-PPM and PPC-DPSK modulation schemes were tested at this site, each at two different rates. The transmitted packets for each of these rates were roughly 15 s long.

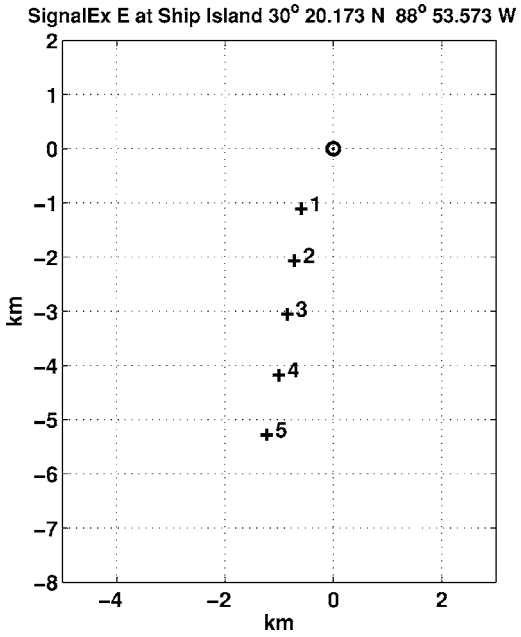


FIG. 6. SignalEx E configuration by Ship Island near Gulfport during AUV/Modem Fest 2001, showing the receive Telesonar Testbed at the origin, and five locations of the transmitter during its drift away from the receiver, when our packets were transmitted.

The PPC-PPM modulation used $m=7$ Gold sequences, each Gold sequence being repeated, as discussed in Sec. III A. Rates of 126 and 188 bps were tested. The lower-rate modulation used Gold sequences with a chip rate of 4 kHz and 4 bits per PPM symbol (16 PPM slots). The higher-rate modulation used Gold sequences with a chip rate of 8 kHz with 3 bits per PPM symbol (8 PPM slots).

The PPC-DPSK modulation used $m=7$ Gold sequences with a chip rate of 8 kHz (i.e., 8000 sequence bits per second). Rates of 160 and 264 bps were tested by varying the symbol periods (i.e., the spacing between sequences, or equivalently, between DPSK symbols).

Figure 8 shows a diagnostic output of our modem at the

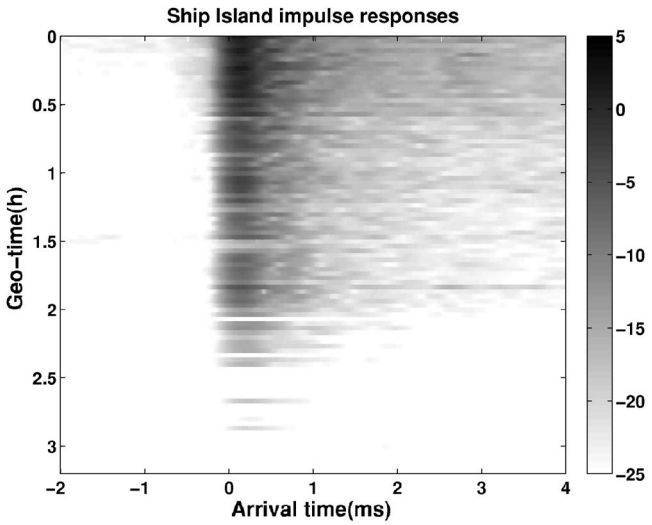


FIG. 7. Channel impulse response (CIR) measured as the transmitter drifted away from the receiver during SignalEx E at Ship Island near Gulfport, during ModemFest 2001.

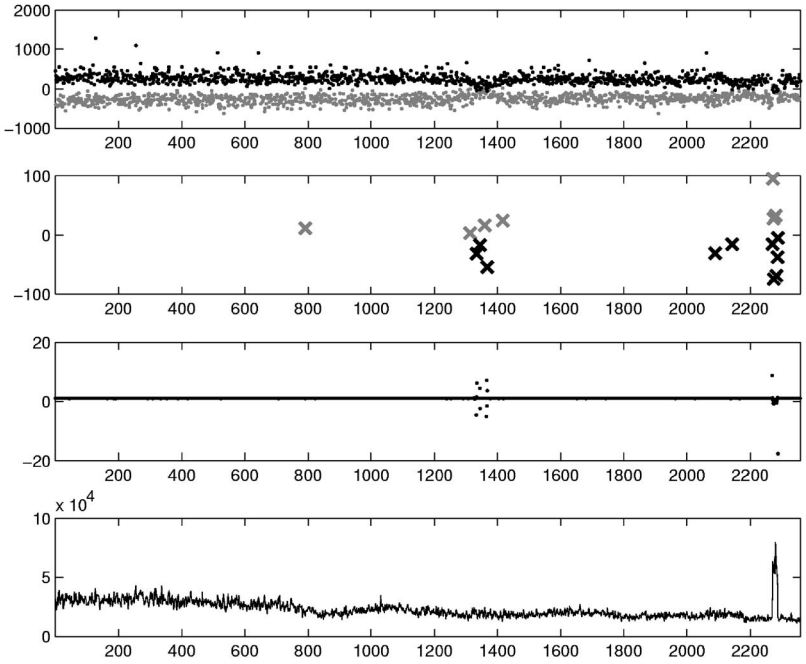


FIG. 8. Diagnostic display from testing the 264 bps PPC-DPSK modulation at 5 km range. All four plots are synchronized along a common x axis that shows the 2359 symbols that make up the packet. The first plot shows values for all the symbols in the packet (the sign of these values is used to "detect" which information bit was transmitted). The second plot shows only the symbols that were incorrectly labeled (note the different y-axis scale, compared to the first plot). The third plot shows deviations of estimated symbol times from expected symbol times (the symbol timing errors near the middle of the packet coincide with the first group of bit errors). The fourth plot shows the matched filter energy output by the symbol timing calculation (the energy spike near the end of the packet coincides with another group of bit errors).

5.4 km range for the 160 bps PPC-DPSK transmissions. This is the packet at which the most bit errors were observed. All four plots are synchronized along a common x axis that indicates symbol number. The upper two plots show symbol values over the length of a 2359 bit packet (i.e., values of $|H|^2 S_1 S_2^*$ from Sec. I), with the first plot showing values for *all* the symbols in the packet, and the second plot showing only the values at which bit errors occurred. The third plot shows the deviations of the estimated symbol times from symbol times calculated by adding the known symbol period to the previous symbol time. The purpose of this plot is to identify which bit errors are due to symbol timing errors. In this case, symbol synchronization loss accounts for some of the bit errors near the middle of the packet. The fourth plot shows the matched filter energy from the symbol timing calculation, and is intended to identify when bit errors are caused by low SNR. In this case, there is an energy spike near the end of the packet, caused by a loud broadband transient of unknown origin (that is apparent in the spectrogram of this data). The errors that occur at this time are probably due to this transient. The separation between positive and negative symbol values in the first plot is also a good indicator of the quality of the information bits being detected. When this gap closes, there is not much signal excess for ambient noise to overcome to cause a bit error.

Table I summarizes the results of testing the PPC-DPSK and PPC-PPM schemes.

B. Results from SignalEx-F: off the coast of La Jolla in San Diego, California

Figure 9 shows the bathymetry and drift track (top plot) and the sound speed, source, and receiver depths, and the bottom composition (bottom plot) from the SignalEx-F test performed off the coast of La Jolla in San Diego, California, on May 10, 2002.

Figure 10 shows how the ocean channel varied with range (from about 500 m to 6 km) as the source platform drifted away from the receive platform over a five hour time interval. Each scan line is the result of averaging the matched filter outputs from forty 50-ms chirps, repeated at 250 ms intervals. The matched filter outputs were aligned by circularly shifting each scan line relative to its predecessor, so that the maximum correlation of each two consecutive scan lines is shifted to time lag zero.

A set of PPC-DPSK packets, one at each of the three rates being tested, was transmitted every half hour during this test. Each packet contained 2240 bits. Each data packet was preceded by a 100–ms LFM chirp sweeping from 8 to 16 kHz, which was used to determine the start of a packet (i.e. initial synchronization).

We transmitted 127-chip Gold sequences ($m=7$) in the 8–16 kHz band at chip rates of 4, 8, and 16 kHz (having lengths of 32, 16, and 8 ms), spaced at intervals of 12, 6, and 3 ms (for rates of 80, 160, and 320 bps). There are 129 Gold sequences at $m=7$, so unless the channel spread is exceptionally long (387 ms for the 3 ms spacing), there are enough sequences to ensure that cycling through them will not result in a situation where a previous transmission of a particular sequence interferes with a current transmission of that same sequence, due to the channel spread.

TABLE I. Bit error rates from PPC-DPSK and PPC-PPM modulation schemes during SignalEx-E (at Ship Island) at five transmitter-receiver ranges (see Fig. 6).

	DPSK-1	DPSK-2	PPM-1	PPM-2
Range(km)	160 (%)bps	264 (%)bps	126 (%)bps	188 (%) bps
1.2	0.0	0.7	0.4	7.0
2.2	0.0	0.5	0.6	2.0
3.2	0.4	1.3	1.4	3.0
4.3	0.0	3.0	0.7	5.0
5.4	0.8	9.3	3.1	Synch failed
	2359 bits/ packet	3899 bits/ packet	1880 bits/ packet	2790 bits/ packet

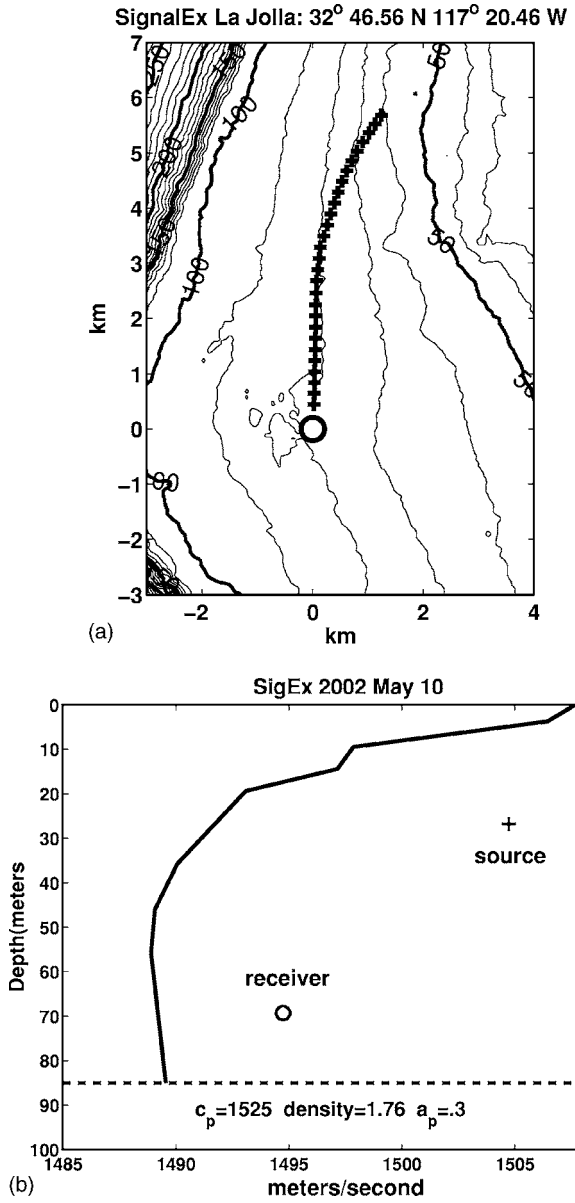


FIG. 9. The top plot shows the SignalEx F configuration in La Jolla, 2002, with the fixed receiver at the origin, and a series of transmitter locations (indicated by plus markers) as the source platform drifted away from the receiver. The bottom plot shows the source and receiver depths, the sound speed profile, and the bottom properties. (a) Bathymetry contours and source track. (b) Sound speed profile and bottom properties.

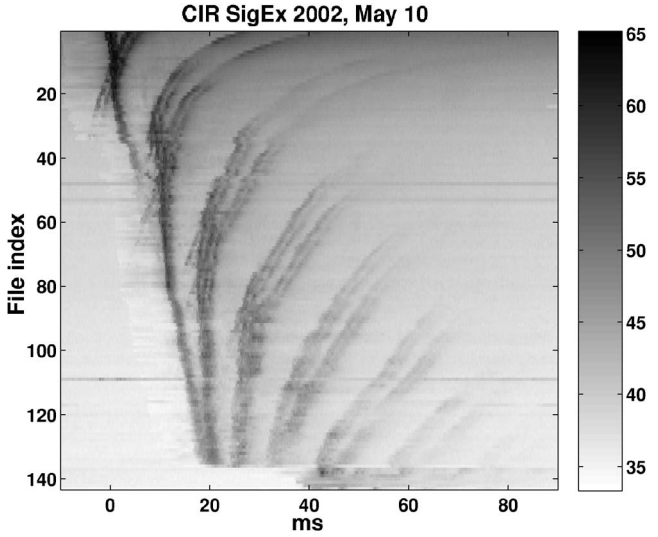


FIG. 10. Channel impulse response (CIR) measured during SignalEx F (off La Jolla in San Diego, CA, in 2002), using dedicated LFM channel probes. This shows how the CIR varies as a function of range (from 500 m to 6 km). These channel measurements correspond to periodic probe pulse transmissions (every two minutes) as the source platform drifted away from the receiver.

This data was initially processed using the PPC receiver described above, but showed poor results in this channel, compared to previous tests and other modulation schemes [CDMA with a Rake receiver and Multi-frequency shift keying or MFSK]. After reviewing the data in great detail, we modified the receiver algorithm to threshold the matched filter outputs, so that contributions whose values were less than 25% the value of the tallest peak were discarded. Only the values exceeding this 25% threshold were used to calculate the polarity of each symbol relative to its predecessor.

Apparently, due to the greater multipath in this environment, the contribution from the sidelobes of the Gold sequence auto and cross correlations and of the CIR autocorrelations was overwhelming the information in the multipath arrivals. This is not surprising, given that we were (1) pushing the chip rate to 8 and 16 kHz, twice and four times the 4 kHz chip rate supported by our 8–16 kHz band and (2) overlapping the Gold sequences to increase the information bit rate. Thresholding the matched filter outputs so that only the high amplitude multipath arrivals were "counted" resulted in greatly improved bit error rates.

This idea is also used in Rake receivers for modulation schemes, where contributions from Rake "fingers" are also thresholded and only the higher amplitude contributions are used to form the detection statistic.[24]

Figure 11 shows various diagnostic displays from our PPC-DPSK receiver for the 160 bps rate at 1.8 km range (packet 3). The upper left plot shows the CIR measured from a 100 ms LFM chirp, used as a synchronization marker roughly 30 ms ahead of the information bits. The upper right plot, a gray scale image, shows channel measurements made using the information-carrying Gold sequences. Each column of this image contains a channel estimate, with multipath time of arrival in milliseconds along the y axis, and symbol time (slow time) in seconds along the x axis. Each Gold sequence is roughly 16 ms long, and overlaps its immediate predecessor by 10 ms and its earlier predecessor by 4 ms, since the sequences are transmitted every 6 ms. The low cross correlation between different Gold sequences minimizes the "cross talk" between consecutive, overlapping Gold sequences, and provides frequent channel measurement updates (every 6 ms). The lower left plot looks very similar to the previous plot (upper right), but it is a black and white dot plot, with the dots indicating the subset of CIR waveform samples that were used to calculate the phase difference for each two consecutive Gold sequences (i.e., for PPC-DPSK). The same strong multipath arrival pattern is visible in both the upper right image and the lower left dot plot. The dots correspond to those samples whose values exceed the 25%

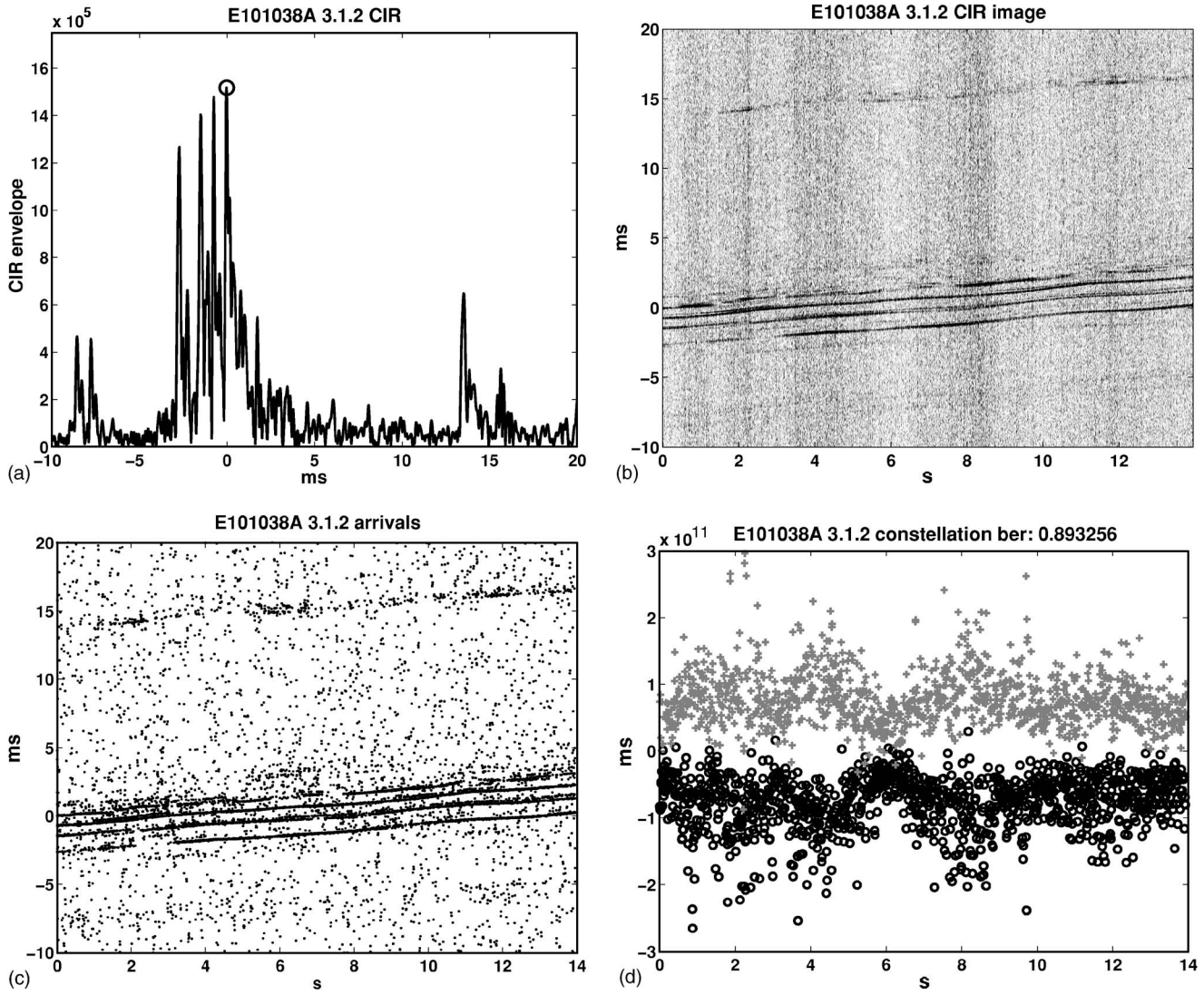


FIG. 11. Measured channel impulse response (upper left), gray-scale image of channel measurements using Gold sequences (upper right), dot plot showing subset of samples that exceed the 25% threshold (lower left), and constellation from the 160 bps packet 3 (lower right). (a) Initial measurement of channel impulse response (from probe). (b) Channel measurements, one per symbol, made from Gold sequences. (c) Multipath arrivals detected from the envelope of matched filter output (on each Gold sequence). Each dot indicates a sample used to form the symbol detection statistic. (d) DPSK constellation versus time.

threshold set relative to the tallest peak in each channel estimate. The lower right plot shows the PPC-DPSK constellation values for the entire 14 s packet.

Table II summarizes the results of testing our PPC-DPSK modulation scheme at three different rates during the SignalEx F experiment. The uncoded error rates for both the 80 and 160 bps data sets were low enough that a convolutional decoder could completely recover the data without errors. However, the bit error rates observed for the 320 bps rate were probably catastrophic at all ranges.

V. SUMMARY AND CONCLUSIONS

PPC implicitly equalizes the channel by refocusing channel spread. We have shown how PPC temporal compression, in a point-to-point configuration (i.e., without arrays of sources or receivers), can be augmented by using Gold sequences to implement receivers based on both PPM and DPSK modulation schemes. We have shown that the PPC-DPSK modulation is more spectrally efficient than PPC-PPM. These PPC-based modulations and their associated receiver algorithms were tested at two SignalEx experiments, one at Ship Island, Mississippi, a very shallow water site with a depth of 5 m and very little channel spread (both PPC-PPM and PPC-DPSK were tested there), and another off La Jolla, California, a moderately shallow water site with a depth of 75 m and moderate channel spread (only PPC-DPSK was tested there). For PPC-PPM, bit rates of 126 and 188 bps were reliably demodulated at the Ship Island site (reliable meaning uncoded bit error rates were low enough that a convolutional decoder would be able to completely recover from them). For PPC-DPSK, bit rates of 160 and 264 bps at Ship Island, and bit rates of 80 and 160 bps at the La Jolla site, all in a 8–16 kHz band, were reliably demodulated. These experiments were performed at environments and ranges that are being considered for underwater wireless networks.

After obtaining relatively poor results (high bit error rates) in the La Jolla data with an initial PPC-DPSK receiver

TABLE II. Bit error percentages from PPC-DPSK modulation at SignalEx F (off La Jolla in San Diego in 2002) for three bit rates at transmitter-receiver ranges of 500 m to 6.7 km. Each table entry represents four transmitted packets of 2240 bits each, corresponding to four receivers vertically separated by 14 in. (differences in bit error rate among the four receivers were negligible). Two packets, before and after the packet at 6.7 km, experienced some sort of recording failure, because no data was evident in the expected time interval (why the range increases by 1.2 km from 5.4 to 6.7, by twice the increment as for all the other packets, where the increment was roughly 0.6 km).

Range (km)	80 bps	160 bps	320 bps
0.6	0.00	0.16	16.20
1.2	0.00	0.10	17.98
1.8	0.04	0.85	26.81
2.4	0.00	0.34	13.84
3.1	0.00	0.04	10.84
3.7	0.06	3.07	29.35
4.3	0.16	6.96	31.30
4.9	0.14	10.01	35.37
5.4	0.07	4.08	26.01
6.7	0.23	9.64	36.53

design, we modified our receiver algorithm to omit low-amplitude contributions to the final phase comparison. At low amplitudes, most of the energy is due to sidelobes of the various correlation processes and to ambient noise. Restricting the phase comparison to higher amplitude contributions ensured that only those samples most likely to be multipath arrivals (and not spurious sidelobes) would contribute to the detection statistic. This is similar to what is sometimes done in Rake receivers for CDMA systems, where individual multipath arrivals are not included in the multipath recombination unless they exceed a minimum threshold. This modification resulted in greatly reducing the bit error rates at the 80 and 160 bps PPC-DPSK modulations. However, even with these improvements, the PPC-DPSK design at 320 bps consistently failed in the La Jolla test data. This was due to the level of multipath in this environment which exacerbated the degradation caused by compromises made to transmit at this rate, such as transmitting only part of the signal bandwidth needed to support the 16 kHz chip rate and clipping the signal to maintain its peak to average power ratio.

Thresholding the channel impulse response, so that only the strong arrivals contribute to the subsequent processing, can perhaps also improve the performance of other channel-estimate based signaling methods, including those based on multichannel time reversal (phase conjugation).

The novel modulation schemes considered in this paper have some interesting properties which may recommend them for use in operational systems. They are very simple, yet have a mechanism for dealing with intersymbol interference. Given their use of spread-spectrum sequences, they afford some potential for covert and multiuser applications. However, because their structure does not fully exploit the correlation properties of Gold sequences (due to multipath and because circular correlations cannot be used), they only mitigate intersymbol interference to a certain extent. As a result, the rates provided by these modulation schemes are relatively limited, compared to phase-coherent schemes in which more sophisticated receiver algorithms are typically used.

ACKNOWLEDGMENTS

SignalEx testing was sponsored by ONR 322OM (Tom Curtin and Al Benson). Additional support was provided by ONR 321OA. This work originated as part of the 6.2 Telesonar Technology program funded by ONR 321SS (Don Davison). We thank Joe Rice for originally suggesting this research.

[1]Stojanovic, M., Catipovic, J. A., and Proakis, J. G. (**1993**). "Adaptive multi-channel combining and equalization for underwater acoustic communications," J. Acoust. Soc. Am. **94**(3), 1621–1631.
[2]Stojanovic, M., Catipovic, J. A., and Proakis, J. G. (**1994**). "Phase-coherent digital communications for underwater acoustic channels," IEEE J. Ocean. Eng. **19**(1), 100–111.
[3]Parvulescu, A. (**1961**). "Signal detection in a multipath medium by M.E.S.S. processing," J. Acoust. Soc. Am. **33**(1), p. l674.
[4]Parvulescu, A. (**1995**). "Matched-signal (MESS) processing," J. Acoust. Soc. Am. **98**(2), 943–960.
[5]Parvulescu, A., and Clay, C. (**1965**). "Reproducibility of signal transmission in the ocean," Radio Electron. Eng. **29**, 223–228.
[6]Dowling, D. R. (**1994**). "Acoustic pulse compression using passive phase-conjugate processing," J. Acoust. Soc. Am. **95**(3), 1450–1458.
[7]Jackson, D. R., and Dowling, D. R. (**1991**). "Phase conjugation in underwater acoustics," J. Acoust. Soc. Am. **89**(1), 171–181.
[8]Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., Ferla, C., and Jackson, D. R. (**1998**). "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**(1), 25–40.
[9]Edelmann, G. F., Akal, T., Hodgkiss, W. S., Kim, S., Kuperman, W. A., and Song, H. C. (**2002**). "An initial demonstration of underwater acoustic communication using time reversal," IEEE J. Ocean. Eng. **27**(3), 602–609.
[10]Rouseff, D., Jackson, D. R., Fox, W. L. J., Jones, C. D., Ritcey, J. A., and Dowling, D. (**2001**). "Underwater acoustic communication by passive-phase conjugation: Theory and experimental results," IEEE J. Ocean. Eng. **36**(4), 821–831.
[11]Rice, J., Creber, B., Fletcher, C., Baxley, P., Rogers, K., McDonald, V. K., Rees, D., Wolf, M., Merriam, S., Mehio, R., Proakis, J., Scussel, K., Porta, D., Baker, J., Hardiman, J., and Green, D. (**2000**). "Evolution of seaweb underwater acoustic networking," in *OCEANS, 2000. MTS/IEEE Conference Proceeding* (IEEE, Piscataway, N.J.), Vol. **3**, pp. 2007–2017.
[12]Hursky, P., Porter, M. B., Rice, J. A., and McDonald, V. K. (**2001**). "Passive phase-conjugate signaling using pulse-position modulation," in *OCEANS, 2001. MTS/IEEE Conference Proceedings* (IEEE, Piscataway, N.J), Vol. **4**, pp. 2244–2249.
[13]Flynn, J. A., Ritcey, J. A., Fox, W. L. J., Jackson, D. R., and Rouseff, D. (**2001**). "Decision-directed passive phase conjugation: equalisation performance in shallow water," Electron. Lett. **37**(25), 1551–1553.
[14]Peterson, R. L., Ziemer, R. E., and Borth, D. E. (**1995**). *Introduction to Spread Spectrum Communications* (Prentice-Hall, Upper Saddle River, NJ).
[15]Sarwater, D. V., and Pursley, M. B. (**1980**). "Crosscorrelation properties of pseudorandom and related sequences," Proc. IEEE **68**(5), 593–619.
[16]Dillard, G. M., Reuter, M., Zeiddler, J., and Zeidler, B., (**2003**). "Cyclic code shift keying: A low probability of intercept communication technique," IEEE Trans. Aerosp. Electron. Syst. **39**(3), 786–798.
[17]Ritcey, J. A., and Griep, K. R. (**1995**). "Code shift keyed spread spectrum for ocean acoustic telemetry," in *OCEANS, 1995. MTS/IEEE Conference Proceedings* 9–12 October in San Diego, CA (IEEE, Piscataway, N.J.) Vol. **3**, pp. 1386–1391.
[18]Sanchez, C., Koski, P., Brady, D., and Massa, D. (**1999**). "Sequence position modulation for surf-zone underwater acoustic communications," in *Proceedings of the 11th International Symposium on Unmanned Untethered Submersible Technology 23–25 August* (IEEE, Piscataway, N.J.), Vol. **11**, pp. 270–279.
[19]Proakis, J. (**2001**). *Digital Communications* (McGraw-Hill, NY).
[20]Yang, T. C. (**2003**). "Temporal resolutions of time-reversal and passive-

phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**(2), 229–245.
[21]Rouseff, D., Fox, W. L. J., Jackson, D. R., and Jones, C. D. (**2001**). "Underwater acoustic communication using passive phase conjugation," in *OCEANS, 2001. MTS/IEEE Conference Proceedings 5–8 November Honolulu, Hawaii* (IEEE, Piscataway, N.J.), Vol. **4**, pp. 2227–2230.
[22]Kwon, H. M., and Birdsall, T. G. (**1991**). "Digital waveform acoustic codings for ocean telemetry," IEEE J. Ocean. Eng. **16**(1), 56–65.
[23]Heard, G. J., and Schumacher, I. (**1996**). "Time compression of m-sequence transmissions in a very long waveguide with a moving source and receiver," J. Acoust. Soc. Am. **99**(6), 3431–3438.
[24]Sozer, E. M., Proakis, J. G., Stojanovic, M., Rice, J. A., Benson, A., and Hatch, M. (**1999**). "Direct sequence spread spectrum based modem for under water acoustic communication and channel measurements," in *OCEANS '99 MTS/IEEE. Riding the Crest into the 21st Century* (IEEE, Piscataway, N.J.), Vol. **1**, pp. 228–233.
[25]Porter, M. B., McDonald, V. K., Baxley, P. A., and Rice, J. A. (**2000**). "Signalex: linking environmental acoustics with the signaling schemes," in *OCEANS, 2000. MTS/IEEE Conference* (IEEE, Piscataway, N.J.), Vol. **1**, pp. 595–600.

The effect of superior-canal opening on middle-ear input admittance and air-conducted stapes velocity in chinchilla

Jocelyn E. Songer[a)]
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Speech and Hearing Bioscience and Technology, Health Science and Technology, Harvard-MIT, Cambridge, Massachusetts 02138

John J. Rosowski[b)]
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, Speech and Hearing Bioscience and Technology, Health Science and Technology, Harvard-MIT, Cambridge, Massachusetts 02138 and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02114

(Received 6 January 2006; revised 17 April 2006; accepted 19 April 2006)

The recent discovery of superior semicircular canal (SC) dehiscence syndrome as a clinical entity affecting both the auditory and vestibular systems has led to the investigation of the impact of a SC opening on the mechanics of hearing. It is hypothesized that the hole in the SC acts as a "third window" in the inner ear which shunts sound-induced stapes volume velocity away from the cochlea through the opening in the SC. To test the hypothesis and to understand the third window mechanisms the middle-ear input admittance and sound-induced stapes velocity were measured in chinchilla before and after surgically introducing a SC opening and after patching the opening. The extent to which patching returned the system to the presurgical state is used as a control criterion. In eight chinchilla ears a statistically significant, reversible increase in low-frequency middle-ear input admittance magnitude occurred as a result of opening the SC. In six ears a statistically significant reversible increase in stapes velocity was observed. Both of these changes are consistent with the hole creating a shunt pathway that increases the cochlear input admittance.
 [DOI: 10.1121/1.2204356]

PACS number(s): 43.64.Ha, 43.64.Tk, 43.80.Lb [BLM] Pages: 258–269

I. INTRODUCTION

Superior semicircular canal dehiscence (SCD) syndrome is a recently defined clinical syndrome (Minor *et al.*, 1998) associated with both auditory and vestibular symptoms (Brantberg *et al.*, 2000; Cremer *et al.*, 2000; Mikulec *et al.*, 2004; Minor, 2000; Minor *et al.*, 2003, 1998) and caused by an opening or dehiscence in the bone separating the superior semicircular canal (SC) from the cranial cavity. The clinical presentation suggests that alterations in the state of the SC can affect auditory function. Previous work in our laboratory has demonstrated that introducing a hole in the chinchilla SC leads to decreased auditory sensitivity to air-conducted sound as measured by cochlear potential (Songer and Rosowski, 2005). The "third window" hypothesis has been suggested as an explanation for the auditory symptoms associated with SCD syndrome (Minor *et al.*, 1998; Rosowski *et al.*, 2004). The hypothetical existence of a third window in the normal inner ear has been discussed by Bekesy (1960), Barany (1938), and Tonndorf (1972) among others.

Our invocation of the third window hypothesis proposes that the SC opening acts as a third window in the inner ear, which provides an additional low-acoustic impedance pathway for perilymph volume velocity in the inner ear. This pathway shunts volume velocity away from the cochlea, resulting in decreased stimulus magnitude to the cochlea and a resultant decrease in auditory sensitivity. In addition to predictions of auditory sensitivity, the third window hypothesis predicts changes in inner ear mechanics as a result of the SC opening. One of the predictions is an increase in inner ear input admittance, which would cause an increase in middle-ear input admittance (Y_{ME}) at frequencies where the admittance of the inner ear influences the Y_{ME}. Previous work in chinchilla shows that the Y_{ME} is dominated by the admittance of the inner ear for frequencies between 80 and 500 Hz (Rosowski and Ravicz, 2001; Rosowski *et al.*, 2006), so we expect the greatest effects related to SC opening within this frequency range. Another prediction of the third window hypothesis is a frequency-dependent increase in stapes velocity due to the increased inner ear admittance.

In this study we will test the third window hypothesis by looking at changes in responses to sound caused by experimental manipulations of the SC. These manipulations are intended to mimic the acousto-mechanical consequences of a SC opening such as those observed in SCD syndrome.

II. MATERIALS AND METHODS

A. Animal preparation

Thirteen anesthetized chinchillas were used in this study. The chinchilla was chosen as an animal model for this work because its range of hearing is similar to that of humans

[a)]Electronic mail: jocelyns@mit.edu
[b)]Electronic mail: john_rosowski@meei.harvard.edu

 0001-4966/2006/120(1)/258/12/$22.50

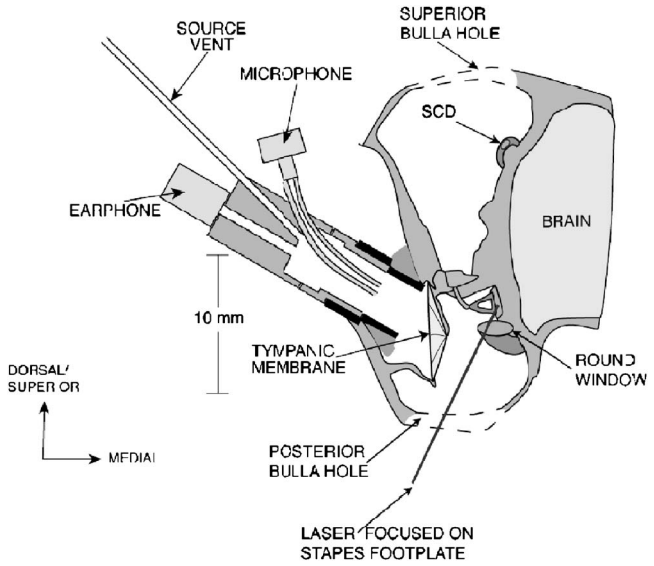


FIG. 1. A schematic of the chinchilla middle-ear cavity (coronal view) illustrating both the superior and posterior bulla approaches in addition to the stapes, round window, superior semicircular canal, sound source, and laser.

[100 Hz to 30 kHz at 30 dB above best thresholds (Miller, 1970)], it has an easily accessible SC, and it has been used in previous studies to evaluate the effect of SC manipulations on vestibular and auditory function (Carey *et al.*, 2000; Hirvonen *et al.*, 2001; Songer *et al.*, 2004; Songer and Rosowski 2005). The chinchillas were anesthetized with intraperitoneal (IP) injections of pentobarbital (50 mg/kg) and ketamine (40 mg/kg) with boosters every 2 h or sooner as needed. The experiments were carried out in accordance with the Massachusetts Eye and Ear Infirmary Animal Care and Use Guidelines.

After the animal was anesthetized and tracheotemized, the bone of the superior bulla was exposed, and a hole was cut into it so that both the medical surface of the tympanic membrane and the SC could be accessed (Fig. 1). The tendon of the tensor tympani was severed with a small knife and the stapedius muscle was immobilized by sectioning the facial nerve medial to the stapedial branch (Songer and Rosowski 2005). The middle-ear muscles were immobilized because previous work (Rosowski *et al.*, 2006) has demonstrated middle-ear muscle contractions in chinchilla despite anesthesia.

A hole was then introduced into the bone of the ipsilateral posterior bulla through which the round-window and lenticular process of the incus were observed. Both the superior and posterior bullar holes were left open throughout the measurements. The bony wall to which one end of the stapedius tendon attaches (the other end attaches to the stapes) was then resected between the tendon attachment, the horizontal semicircular canal, and the round window in order to visualize the stapes footplate and crura. Subsequently, three to six autoreflective beads (50 μ in diameter and approximately 0.07 μg each) were placed on the stapes footplate. Fluid in or around the footplate was removed using fine paper points prior to measurements of stapes velocity (Vs) to prevent fluid-based signal artifact and ensure a proper reflection from the beads.

The pinna and cartilaginous ear canal were resected and the bony ear canal shortened by surgically removing the bone. A probe-tube microphone and calibrated sound source were then coupled to a short brass tube which was cemented to the remnant of the bony ear canal. A log-chirp with frequency components at 11-Hz intervals from 11 Hz to 24 kHz was used as the stimulus; such chirps have a low-frequency emphasis. Both the sound pressure at the tympanic membrane (P_{TM}) and Vs were measured in response to repeated chirp stimuli at three different stimulus levels (covering a 20-dB stimulus range) to check for linearity.

After a series of baseline measurements was conducted, a large hole (0.4 mm by 0.6 mm; the diameter of the canal is 0.4 mm) was introduced into the SC, about 3 mm superior to the ampulla, using a fine chisel to remove the bone overlying the canal. This resulted in the fluid within the canal being exposed to the air-filled middle-ear cavity. A series of measurements was conducted with the SC open. Finally, the SC hole was covered (patched) with cynanoacrylic glue (Superglue®) and a third series of measurements was completed. The Superglue® provided a rigid, water-tight covering over the SC hole.

B. Instrumentation

1. Source and admittance measurements

All of the data in this paper are presented as transfer functions. Measurements of microphone voltage and laser output were made at three different stimulus levels corresponding to 94, 84, and 74 dB SPL. For frequencies below 500 Hz some nonlinear growth in the middle-ear input admittance (Y_{ME}) was observed (see Sec. III A). Such nonlinear growth has been documented previously in chinchilla (Kim *et al.*, 1980; Rosowski *et al.*, 2006; Ruggero *et al.*, 1996) and is not seen in passive loads, indicating that it is not due to a nonlinearity in the stimulus system. Over the remaining frequency range both the microphone signal and the laser signal responded linearly to varied stimulus levels and had a high (>10) dB signal-to-noise ratio across the range of levels tested.

The microphone voltage, in conjunction with calibrated source measurements, was used to calculate Y_{ME}. The measured microphone voltage was converted to sound pressure using a previously determined calibration factor. Following this, the volume velocity of the equivalent ideal velocity source for the calibrated source (U_S) was divided by the sound pressure measured at the tympanic membrane (P_{TM}) and the admittance of the source (Y_S) was subtracted from the total admittance in order to yield the Y_{ME} (Lynch *et al.*, 1994; Songer *et al.*, 2005):

$$Y_{ME} = \frac{U_S}{P_{TM}} - Y_S. \quad (1)$$

The admittance was then corrected for the residual ear canal and earphone coupler space using a transmission-line correc-

tion (Lynch *et al.*, 1994) where the length of the residual canal and sound coupler was estimated to be 6.0 mm and the average radius of the tube was 2.4 mm. Lynch *et al.* (1994) and Huang *et al.* (2000) demonstrate that the accuracy of admittance measurements and canal corrections are good (±30% in magnitude and ±0.3 rad at frequencies below 10 kHz).

2. Laser Doppler vibrometry

Laser Doppler vibrometry (LDV) was used to measure the sound-induced velocity of the stapes based on the Doppler shift of light reflected from autoreflective plastic beads on the stapes footplate. We used a "single point LDV" from Polytech PI (OFV 5000) to measure the velocity of the stapes motion over the frequency range of 11 Hz to 10 kHz. The proportionality between LDV output voltage and velocity was checked using a calibrated accelerometer mounted on a shaker and found to be equal to the manufacturer's specification for the frequencies of interest (between 50 Hz and 10 kHz). A micromanipulator was used to focus and direct the laser beam on the stapes footplate. The velocity data will be presented as transfer functions where Hp is the middle-ear transfer admittance (Vs/P_{TM}).

3. Inclusion criterion: Does closure of SC opening return measurements to intact values?

The Y_{ME} and Hp were measured with the labyrinth intact, with a SC hole, and with the hole patched. The measure we use for describing the effectiveness of the patch in recreating the intact state is the rms magnitude difference,

$$\text{rms difference} = \sqrt{\frac{1}{n}\sum_{0}^{n}\left[1 - \frac{|M_{\text{Intact}}|}{|M_{\text{Patched}}|}\right]^2}, \qquad (2)$$

where n refers to the number of data points between 100 and 5000 Hz, M_{Patched} is the measured Y_{ME} or Hp after the hole is patched, and M_{Intact} is the Y_{ME} or Hp with the SC intact. If there was a perfect match between the intact and patched measurements, the rms difference would equal zero. Data sets were included in the study if the rms difference for the Hp and/or the rms difference for the Y_{ME} was less than an arbitrary threshold set at 0.3. This criterion was designed to reject data from experiments in which either the inner ear was damaged or methodological problems occurred. A similar methodology had been used previously to evaluate changes in hearing sensitivity associated with SC opening (Songer and Rosowski, 2005).

III. RESULTS

A. Linearity

In the measurements of both middle-ear input admittance (Y_{ME}) and middle-ear transfer admittance (Hp) in the intact ear, a notch was observed between 150 and 200 Hz that exhibited nonlinear behavior (level dependence). Figure 2(a) shows the mean Y_{ME} magnitude and phase for five ears (ears 8, 9, 10, 11, and 12) with measurements at ear canal sound pressures of 94, 84, and 74 dB SPL. The five ears were chosen because measurements at all levels were available for both Y_{ME} and Hp in each of the states. A level dependence was observed in each of the five ears for frequencies below 300 Hz. Above 1 kHz there are no differences among the mean Y_{ME} measurements at different stimulus levels. Between 400 and 1000 Hz variations are seen in Y_{ME} that are attributed to noise in the sound pressure measurement at the lowest stimulus level. Below 400 Hz, however, differences in magnitude and phase are apparent for the three stimulus levels. The measurement at 94 dB SPL exhibits the most prominent notch near 165 Hz (both in magnitude and phase). In response to the 84 dB stimulus, the notch is shallower in both magnitude and phase. The shallowest notch in both magnitude and phase is observed in response to the 74 dB SPL stimulus. The measurement in response to the 74 dB SPL stimulus is noisier (has more irregularities in both magnitude and phase) for frequencies below 600 Hz than the measurements at 84 and 94 dB SPL. As the stimulus level decreases the signal-to-noise ratio worsens, leading to increased sensitivity to noise. The differences in the notch depth at different stimulus levels indicate a nonlinearity.

The Hp measurements also exhibit a notch in magnitude and phase near 165 Hz as illustrated in Fig. 2(b). Measurements of Hp at three different stimulus levels (74, 84, and 94 dB SPL) were made to determine if the notch in Hp was also nonlinear. Once again, the notch is deepest in response to the 94 dB SPL stimuli and shallowest in response to the 74 dB SPL stimulus. The measurements in response to the 74 dB SPL stimuli also have more irregularities in both magnitude and phase than the measurements at the other stimulus levels.

The notch we observed near 165 Hz in our Y_{ME} and Hp measurements is consistent with previous measurements of Vs (Ruggero *et al.*, 1990) and Y_{ME} (Rosowski *et al.*, 2006) with open bullae. The similarity between the Y_{ME} and Hp nonlinearities suggests that the nonlinearity is the result of mechanical processes that affect the sound pressure produced by our source.

In addition to the level dependence of the notch near 165 Hz, nonlinear behavior is observed in Y_{ME} and Hp for frequencies below 100 Hz. This apparent nonlinearity is thought to arise due to the deterioration of the signal-to-noise ratio as stimulus level is decreased. For the remainder of this paper all of the data presented will be in response to the 94 dB SPL stimulus.

B. Middle-ear input admittance (Y_{ME})

The Y_{ME} measured before and after the introduction of a SC hole was evaluated in eight ears corresponding to ear numbers: 5, 6, 7, 9, 10, 11, 12, and 13. Ears 1, 3, and 8 were excluded from the study because they did not meet the inclusion criteria (rms difference exceeded 0.3). Ears 2 and 4 were excluded due to a procedural problem: a leak between the acoustic coupler and the earphone introduced errors into the Y_{ME} measurements.

1. Y_{ME}: SC intact

The measured Y_{ME} with the SC intact in the eight ears is plotted in Fig. 3. There is a peak in $|Y_{ME}|$ near 100 Hz. At

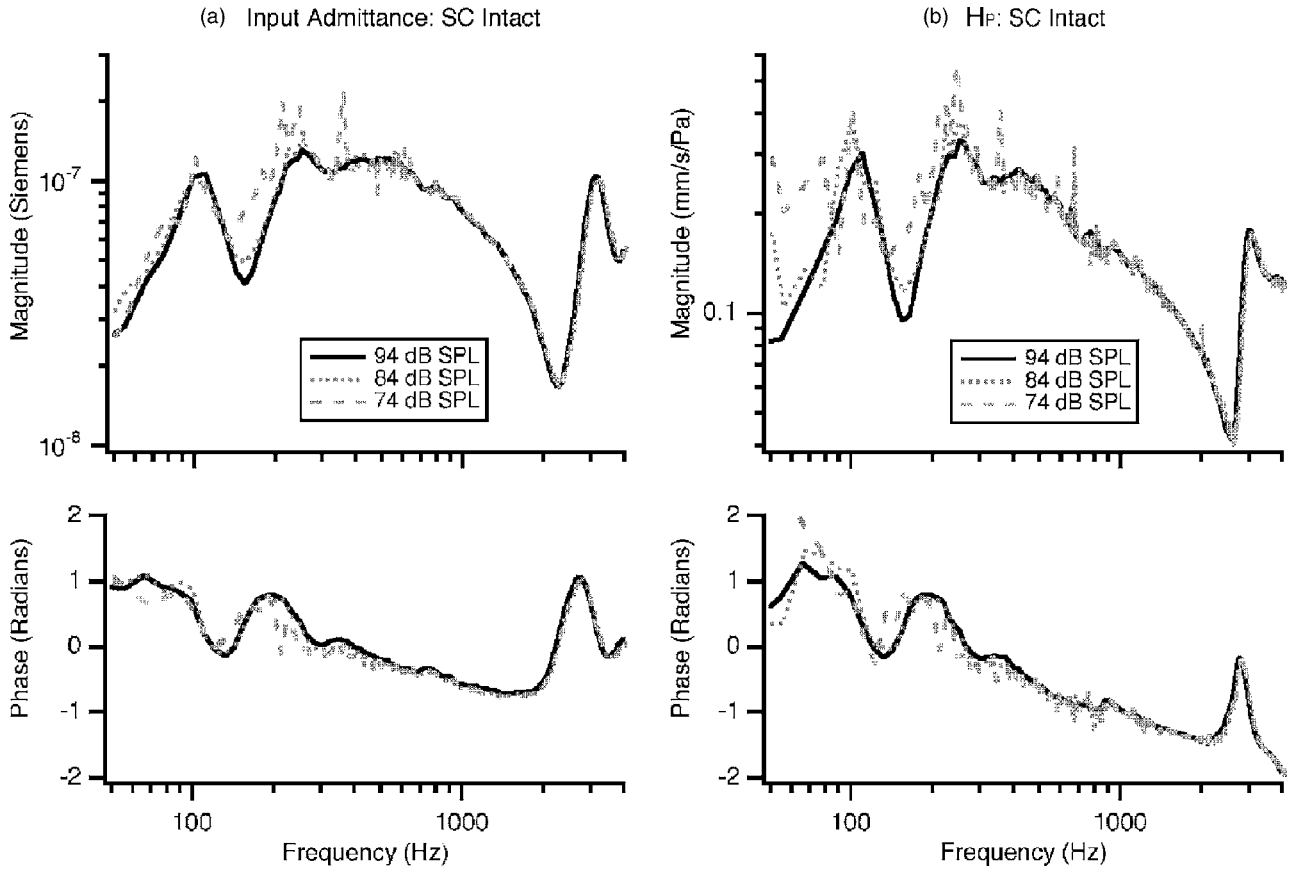


FIG. 2. The mean ($n=5$) magnitude and phase of (a) Y_{ME} and (b) Hp at three different stimulus levels with the SC intact. Differences in the depth of the notch between 100 and 200 Hz are observed with different stimulus levels. Additionally, the measurements in response to the stimulus at 74 dB are noisier than those observed at 84 and 94 dB SPL.

lower frequencies Y_{ME} is roughly compliant in six of the eight ears: the mean $|Y_{ME}|$ grows with frequency below 100 Hz, though the log-log slope is 1.7 (equivalent to 10.4 dB/oct), somewhat larger than the proportionality of Y_{ME} and frequency (log-log slope of 1 or 6 dB/oct) expected of a compliance, and the observed phases range between 0.8 and 1.5 rad, where a pure compliance would have a phase angle of $\pi/2$ (1.57 rad).[1]

There is a dip in both the magnitude and phase of Y_{ME} in all eight ears between 100 and 250 Hz, with a local minima near 165 Hz. Between 200 and 220 Hz there is a peak in magnitude and the phase approaches zero.

Between 200 and 660 Hz both the Y_{ME} magnitude and the phase angle decrease. The phase goes from near zero to −0.5 rad as the frequency increases, and the magnitude decreases with a log-log slope of −0.5 (−3.1 dB/oct). The changes in both magnitude and phase are consistent with the system becoming less resistive and more masslike. In this frequency range ear 5 is an outlier in that its magnitude and phase decrease more rapidly than the others with the phase appearing to change quickly from positive to negative near 250 Hz.

Between 1 and 2 kHz the mean $|Y_{ME}|$ decreases rapidly with a slope of −1.5 (−9.2 dB/oct). All eight ears then exhibit a minimum in the $|Y_{ME}|$ and a jump in the Y_{ME} phase of $>\pi/2$ radians between 2 and 3 kHz. The depth of the $|Y_{ME}|$ minimum and its precise frequency vary among ears; ear 5 shows the least prominent minimum. This minimum has previously been associated with an antiresonance produced by the interaction of the compliance of the middle-ear cavity and the radiation impedance of the cavity holes (Ravicz *et al.*, 1992; Rosowski *et al.*, 2006). At frequencies above this antiresonance (above 3500 Hz) the mean Y_{ME} magnitude and phase angle show a maximum and then are roughly frequency independent. The approximately constant mean magnitude and phase angle between 0 and $\pi/4$ (0.79 rad) at higher frequencies are consistent with resistive behavior, though the angles of some of the individuals are quite variable and look more compliant as they approach $\pi/2$.

Similar frequency dependencies in the magnitude and angle of Y_{ME} with open bullar holes were seen in a previous study, including: the approximately compliant behavior of the Y_{ME} at frequencies less than 100 Hz, the maxima and minima in Y_{ME} magnitude and angle near 200 Hz, and the large notch in magnitude and sudden phase change associated with bullar-hole resonances (Rosowski *et al.*, 2006).

2. Y_{ME}: SC opening

Figure 4 illustrates the Y_{ME} of a representative ear with the SC intact, with a SC hole, with the SC hole patched, and with the patch removed. This ear meets the reversibility requirement in that the intact and the patched measurements are very similar and have an rms difference of less than 0.3.

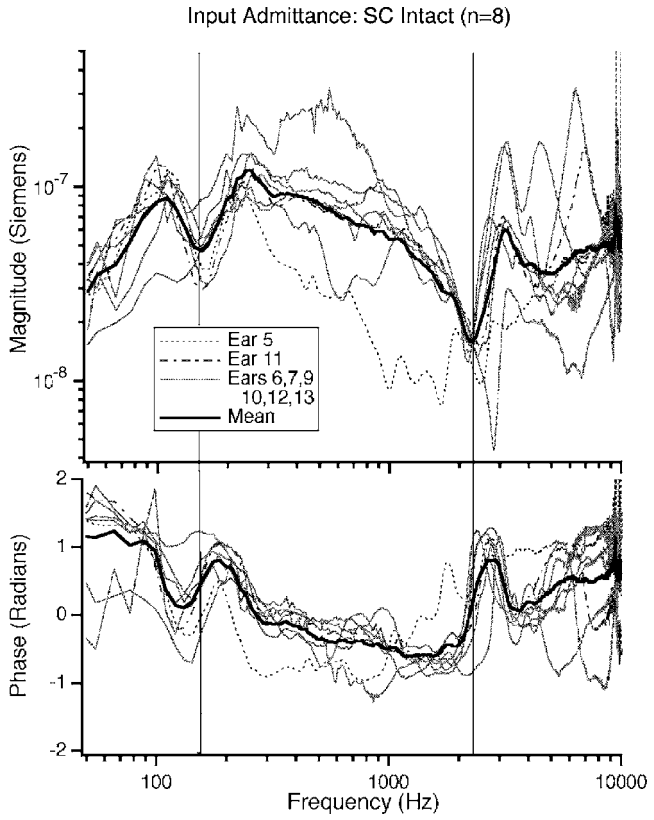


FIG. 3. The middle-ear input admittance in eight ears with intact SC as well as the mean of the eight ears. Vertical lines are drawn at the mean frequency of the two most prominent magnitude notches, one near 165 Hz and the other near 2600 Hz.

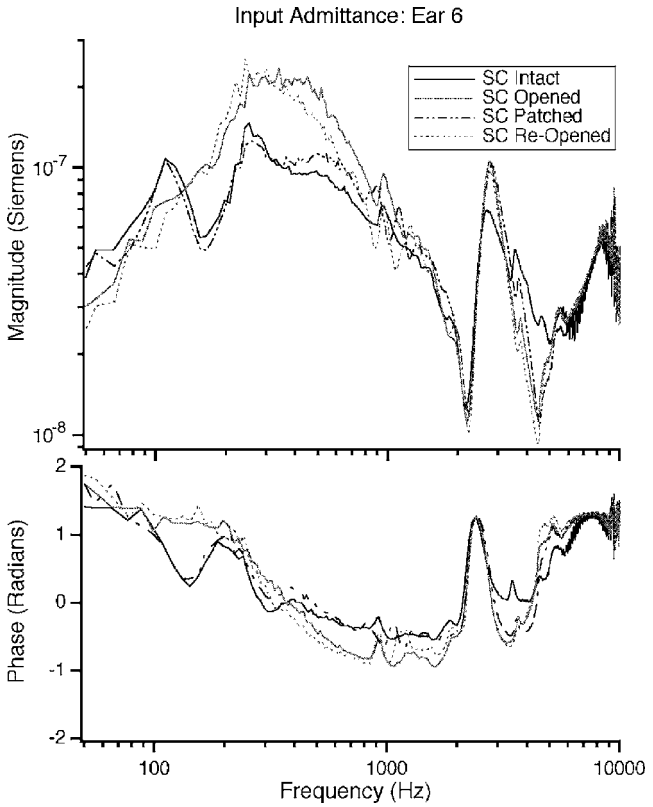


FIG. 4. Data from a representative ear illustrating the Y_{ME} with SC intact, SC open, SC patched, and when the patch is removed. A general increase in Y_{ME} is observed for frequencies between 110 Hz and 1 kHz as a result of the SC opening. The Y_{ME} after patching the SC hole approximates the Y_{ME} in the intact state.

As a result of the SC hole, there is an increase in $|Y_{ME}|$ for frequencies between 110 Hz and approximately 1 kHz, and a decrease in $|Y_{ME}|$ for frequencies below 110 Hz. The low-frequency notch at 165 Hz observed in the intact case is no longer evident after the SC is opened. The SC hole also causes changes in phase angle, especially near 100 Hz where the notch disappears. At frequencies greater than 800 Hz, the $|Y_{ME}|$ measured in all conditions is similar, though the phase angles in the SC hole and unpatched states are more negative than in the intact and patched states. The large dip observed in $|Y_{ME}|$ near 2 kHz in all the ears is likely the result of a cavity-bulla hole antiresonance. The domination of the antiresonance in this frequency region would account for the invariance of the Y_{ME} in these regions in all measurement conditions: intact, SC hole open, patched, and unpatched.

The SC hole-open Y_{ME} for each of the eight ears is illustrated in Fig. 5. Below 100 Hz the Y_{ME} is clearly dominated by a compliance more than in the intact state: the magnitude is increasing monotonically with a log-log slope of 1.1 (6.4 dB/oct), and the phase is roughly constant at a value between $\pi/4$ and $\pi/2$. There is no dip in $|Y_{ME}|$ between 100 and 250 Hz as seen in the intact state; rather, a maximum occurs between 200 and 400 Hz. Ear 5 is an outlier: it exhibits a peak magnitude and phase decrease at a lower frequency than the other ears, with its magnitude peak centered near 150 Hz.

Between 200 and 600 Hz there is a plateau in the $|Y_{ME}|$ in most ears and a decrease in the phase. The mean magnitude of the plateau from 250 to 450 Hz is 1.9×10^{-7} acoustic Siemens ($1S=1\ \mathrm{m^3-s^{-1}-Pa^{-1}}$). Between 600 Hz and 2 kHz there is a monotonic decrease in the $|Y_{ME}|$ with a mean log-log slope of -1.4 (-8.3 dB/oct) and a constant phase near $-\pi/2$. The magnitude and phase in this region are consistent with the system being mass dominated.

The persistence of a magnitude dip between 2 and 3 kHz in all eight ears is consistent with its being a consequence of a bullar hole-cavity antiresonance, which is unaffected by the SC hole opening.

3. The change in Y_{ME} produced by opening the SC

The change in Y_{ME} that results from opening the SC hole can be described by the ratio of the two conditions, ΔY_{ME}. The magnitude and angle of this ratio are plotted in Fig. 6, where the dB magnitude of ΔY_{ME}, is defined as

$$|\Delta Y_{ME}| = 20 \times \log_{10} \frac{|Y_{ME}^{\mathrm{scopen}}|}{|Y_{ME}^{\mathrm{intact}}|} \tag{3}$$

and the angle of ΔY_{ME} is the difference in phase angle between the open and intact state:

$$\angle \Delta Y_{ME} = \angle\, Y_{ME}^{\mathrm{scopen}} - \angle\, Y_{ME}^{\mathrm{intact}}. \tag{4}$$

The largest changes in the value $|\Delta Y_{ME}|$ related to SC opening occur between 100 Hz and 1 kHz; the magnitude increases from 2 to 14 dB, generally with two peaks (Fig. 6),

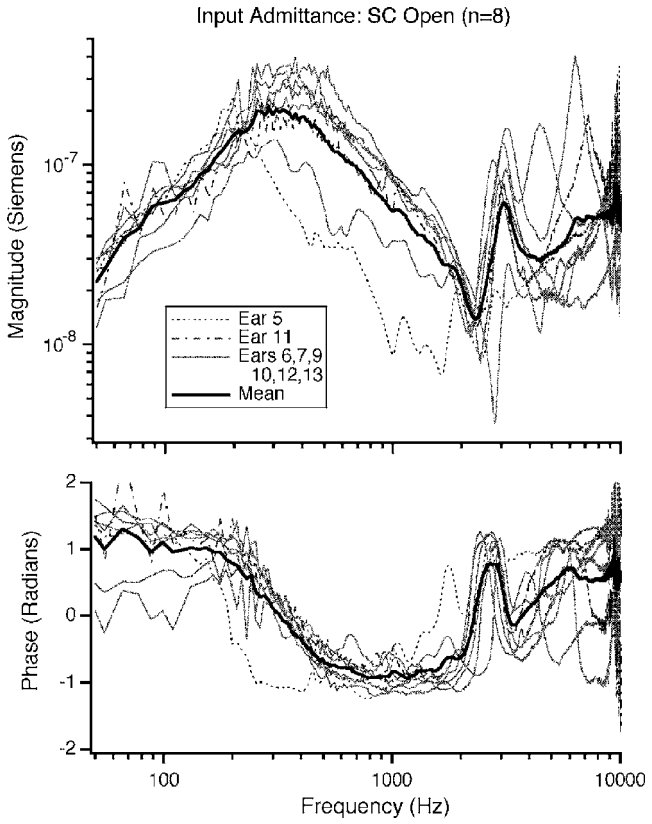


FIG. 5. Y_{ME} in eight ears after the SC is opened. A broad peak is observed in $|Y_{ME}$ between 200 and 600 Hz after the SC is opened. Below the peak, the admittance is more compliant than that seen in the intact ear. A minima occurs in all of the ears near 2600 Hz which is attributed to a bulla-hole antiresonance.

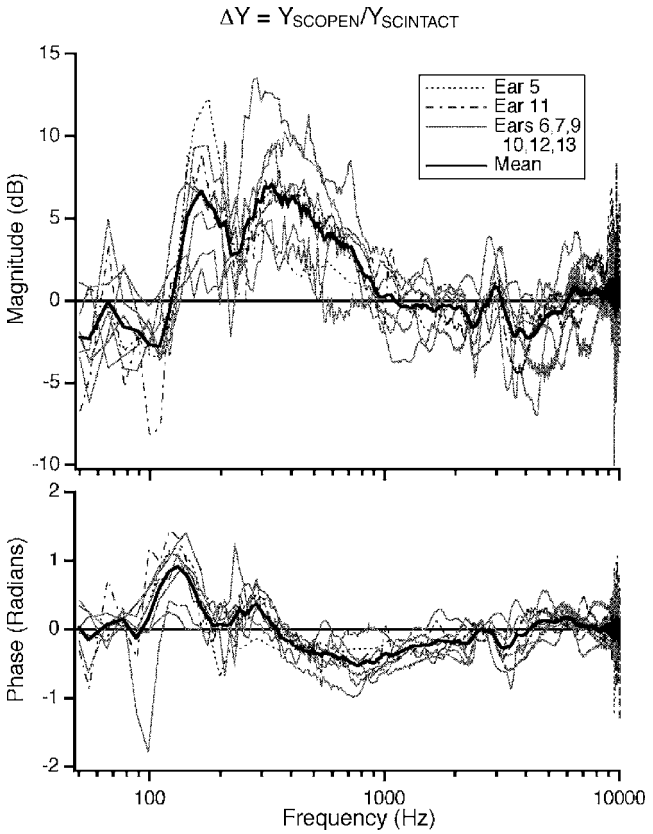


FIG. 6. The ratio between the Y_{ME} for the SC open and the SC intact states in all eight ears as well as the mean change in Y_{ME} are illustrated. The largest changes are observed for frequencies between 100 Hz and 1 kHz, with increases between 2 and 14 dB.

one at 165 Hz, and one between 300 and 450 Hz. The increase in the mean $|\Delta Y_{ME}|$ is statistically significant (the probability that the mean dB difference was 0 dB is less than 0.05) for $140 \leqslant f \leqslant 235$ Hz and $250 \leqslant f \leqslant 750$ Hz. Above 750 Hz and below 140 Hz the mean is not statistically different from zero. Two peaks, near 165 and 260 Hz, are also apparent in the phase of ΔY_{ME} but the mean change in the ΔY_{ME} phase is not statistically significant at any frequency.

C. Middle ear transfer function: *Hp*

In addition to measurements of Y_{ME}, we also measured stapes velocity (Vs). The Vs measurements will be presented in terms of a middle-ear transfer admittance (Hp), where $Hp \equiv Vs/P_{TM}$ with units of $\text{mm}-\text{s}^{-1}-\text{Pa}^{-1}$.

1. Hp: SC intact

The Hp measured for six ears is plotted in Fig. 7. For frequencies below 100 Hz, the mean $|Hp|$ increases with frequency with a log-log slope of 2.3 (13.6 dB/oct). There is a local minimum near 165 Hz between peaks at approximately 100 and 200 Hz (log-log slope of the mean in the range of 110–165 Hz is −3.0 or −17.8 dB/oct, and the log-log slope of the mean in the range of 165–220 Hz is 3.1 or 18.7 dB/oct). Between 300 Hz and 1 kHz the mean $|Hp|$ decreases with a log-log slope of −0.7 (−4.4 dB/oct). The log-log slope of the mean from 1100 to 2200 Hz is −0.9 (−5.2 dB/oct) and there is a minimum near 2600 Hz. As frequency increases above 2600 Hz the $|Hp|$ peaks before decreasing again. The Hp phase shows a dip of 1.5 rad at 165 Hz, and then a slow decline between 200 Hz and 2 kHz. At 2600 Hz there is a peak in the phase of approximately 2 rad and then the phase decreases with a greater slope out to 10 kHz. The total phase change between 60 and 10 000 Hz is between 4 and 6 rad. Ear 5 is an outlier: the decrease in $|Hp|$ above 250 Hz is more abrupt than that in the other ears, and $|Hp|$ in ear 5 does not exhibit a deep minima near 2600 Hz. The Hp measurements have many features in common with the Y_{ME} measurements (Fig. 3). Both Y_{ME} and Hp have two low-frequency peaks in magnitude with similar frequency dependence (one near 100 Hz and the other near 250 Hz); they also both exhibit a sharp dip in magnitude and peak in phase near 2600 Hz.

2. Hp: SC hole open

Figure 8 shows the Hp measured in ear 10 with the SC intact, with a SC hole, and with the hole patched. The rms difference between the intact and patched states was less than 0.3 and therefore met the inclusion criteria. The Hp measurements made in the SC intact and SC open states have much in common with the Y_{ME} measured in the same conditions (Fig. 4). After a SC hole was introduced, the $|Hp|$ increases for frequencies between 120 Hz and 1 kHz and decreases for frequencies below 120 Hz. Other features in common between the SC open Y_{ME} and Hp data include

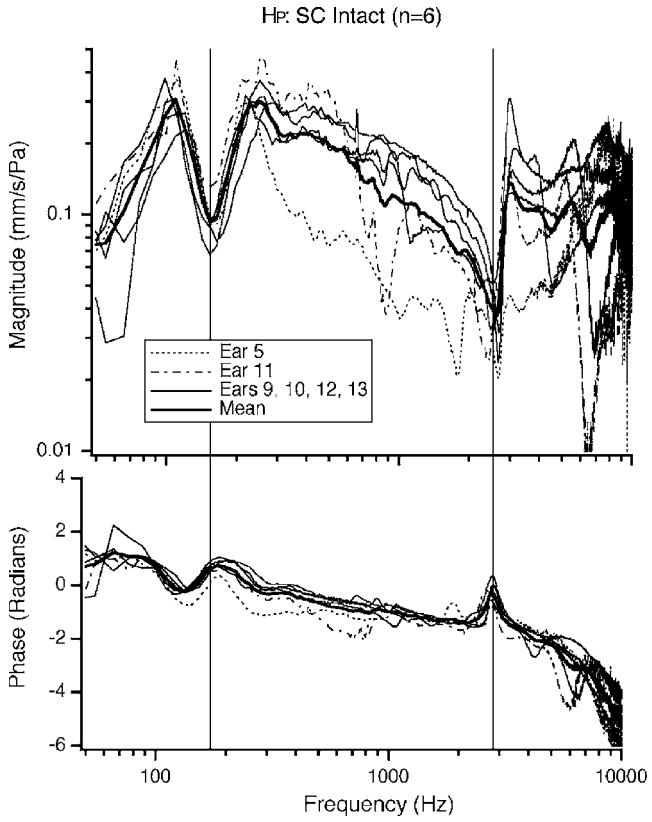


FIG. 7. The middle-ear transfer function (Hp) with the SC intact. Vertical lines are drawn at the mean frequency of the two most prominent magnitude notches, one near 165 Hz and the other near 2600 Hz. The measurements of Hp with the SC intact have many features in common with Y_{ME} measurements in the intact ear.

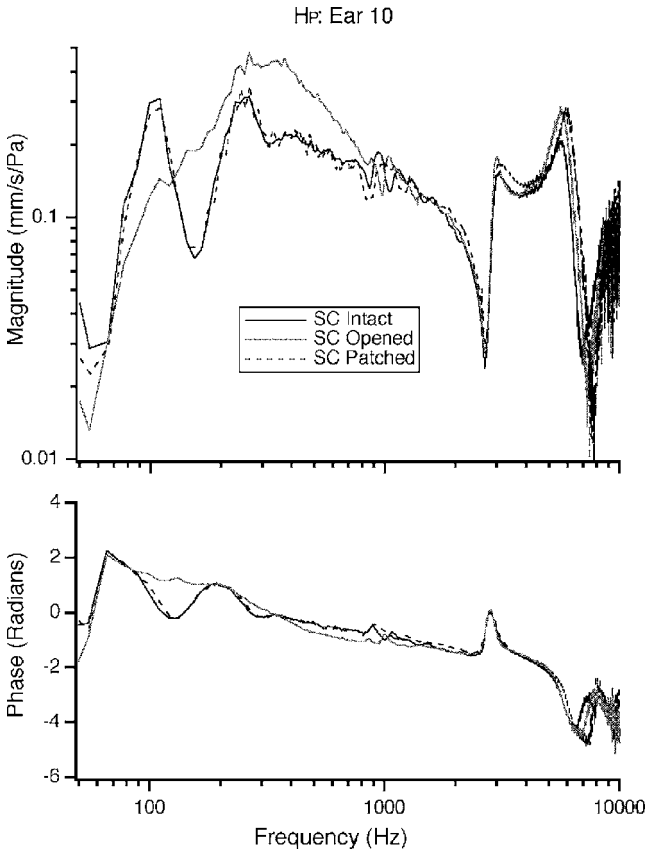


FIG. 8. Hp in an example ear with the SC intact, after opening the SC, and after patching a SC opening. As a result of SC opening there is a change in Hp for frequencies below 1 kHz with an increase observed between 140 Hz and 1 kHz. Patching the SC hole causes the measured Hp to approximate that seen in the intact state.

compliancelike behavior at frequencies below 200 Hz, a broad peak near 300 Hz, a masslike behavior between 300 Hz and 2000 Hz, and an antiresonance near 2600 Hz.

Figure 9 illustrates the SC open Hp magnitude and phase in each individual ear in addition to the mean. The $|Hp|$ increases monotonically until reaching a peak near 300 Hz in all six ears, and then decreases to a sharp minimum near 2600 Hz. The phase at 300 Hz is near $\pi/2$ rad and changes to near $-\pi/2$ between 500 and 2600 Hz. The sharp minimum in $|Hp|$ and peak in phase near 2600 Hz are consistent with the SC intact Hp measurements as well as the Y_{ME} measurements in which we attribute the dip in magnitude to a bullar cavity hole antiresonance. Ear 5 is the clear outlier: the peak in $|Hp|$ occurs at a lower frequency than the others and the Hp phase angle transitions from near $\pi/2$ to near $-\pi/2$ at a lower frequency.

3. The change in Hp produced by the SC opening

The change in Hp produced by opening the SC can be defined as the ratio of Hp_{scopen} to Hp_{intact} (ΔHp) with a magnitude calculated similarly to that defined in Eq. (3). The ΔHp magnitude and phase for each individual ear as well as the mean is illustrated in Fig. 10. The frequency dependence of ΔHp is similar to that of ΔY_{ME} (Fig. 6). There is a decrease in the mean $|\Delta Hp|$ from 95 to 118 Hz which is statistically significant for frequencies between 100 and 118 Hz. This decrease is followed by a sharp increase in the $|\Delta Hp|$ with a peak of 8.2 dB at 187 Hz, followed by a notch and another broader peak of 7.2 dB at 341 Hz. Above 341 Hz there is a steady decrease in $|\Delta Hp|$ to 1000 Hz, where it plateaus prior to a small peak near 2900 Hz. The $|\Delta Hp|$ and $|\Delta Y_{ME}|$ responses have many features in common; however, at high frequencies (above 1 kHz) there are statistically significant increases in $|\Delta Hp|$ which differ from the $|\Delta Y_{ME}|$ responses. While $|Y_{ME}|$ is only significant for $f<1000$ Hz the increase in $|\Delta Hp|$ is statistically significant for frequencies from 145 to 900 Hz, 990 to 2600 Hz, 2800 to 3600 Hz, and 4800 to 5800 Hz. Between 5800 and 7000 Hz there are 11 small frequency ranges of statistical significance; above 7000 Hz the $|\Delta Hp|$ is not statistically different from zero.

4. The velocity transfer ratio (G_{ME})

Both the intact $|Hp|$ and the intact, $|Y_{ME}|$ appear similar in frequency dependence as do the SC open $|Hp|$ and $|Y_{ME}|$. One way to evaluate the degree of similarity is to look at the velocity transfer ratio of the middle ear (G_{ME}),

$$G_{ME} = \frac{Hp}{Y_{ME}} = \frac{V_s}{U_{TM}}, \tag{5}$$

where U_{TM} is the volume velocity of the tympanic membrane and G_{ME} has units of m^{-2}. By comparing the G_{ME} with the SC intact and the G_{ME} with the SC open we can

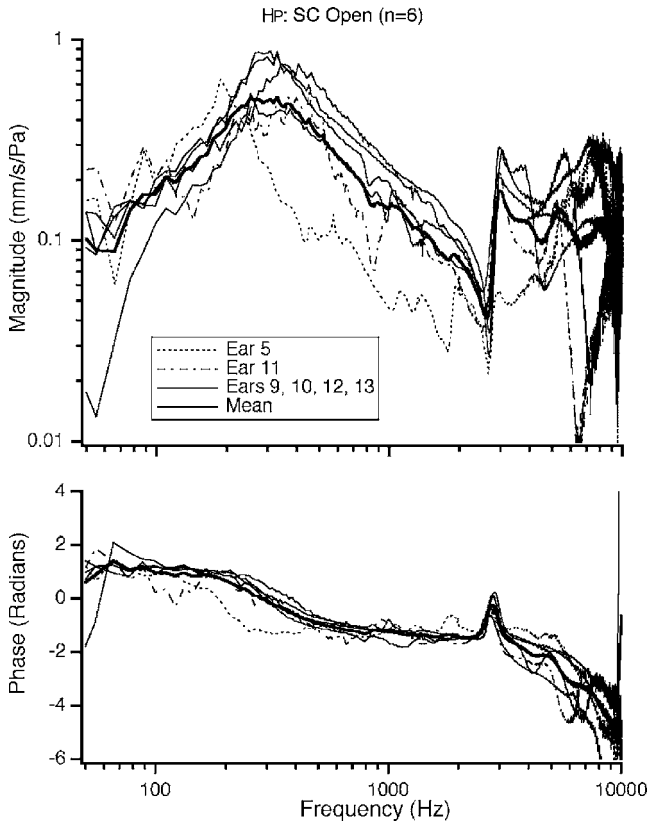


FIG. 9. *Hp* in six ears after opening the SC along with the mean *Hp*. After introducing a SC hole there is a peak in *Hp* magnitude near 300 Hz which is not observed in the intact state. A minima is observed near 2600 Hz that is attributed to a bulla-hole antiresonance.

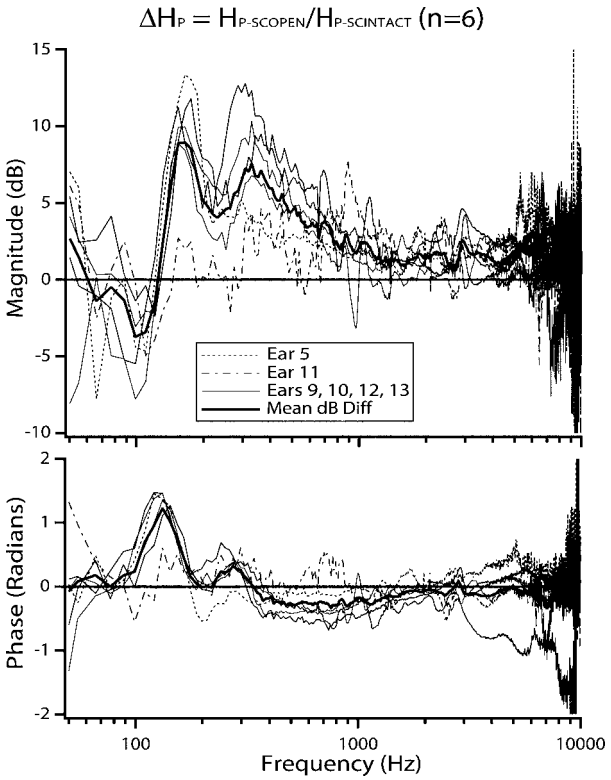


FIG. 10. The ratio of *Hp* between the SC open and the intact case in addition to the mean. A mean increase in $|Hp|$ is observed for frequencies above 140 Hz as a result of opening the SC hole. Between 95 and 118 Hz there is a mean decrease in the observed *Hp*.

determine if the velocity transfer function ratio of the middle ear is affected by the SC hole.

Figure 11(a) illustrates the mean $|G_{ME}|$ ($n=6$) with the SC intact and with the SC open along with the 95% confidence intervals around the mean. There are some small differences in G_{ME} between the two states: for frequencies above 800 Hz, the mean $|G_{ME}|$ after the SC is opened is greater than the mean $|G_{ME}|$ for the intact SC. Despite the overall increase in $|G_{ME}|$ for frequencies greater than 800 Hz, the 95% confidence intervals for the two sets of data show a large degree of overlap. Figure 11(b) illustrates the ratio between the SC open $|G_{ME}|$ and the intact $|G_{ME}|$ in terms of a dB difference and the associated 95% confidence intervals. The dB difference in $|G_{ME}|$ is statistically significant from 1615 to 1910 Hz, 2100 to 2960 Hz, 3075 to 4060 Hz, and 4775 to 5835 Hz. Thus, opening the SC causes small increases in the velocity transfer ratio in the mid-frequencies (1500–6000 Hz).

IV. DISCUSSION

In this study we have demonstrated reversible changes in both Y_{ME} and *Hp* in response to a SC opening. These findings demonstrate changes in middle-ear mechanics that result from inner-ear manipulations. The frequency dependence of these changes and the implications for middle-ear mechanics will be explored, and we will propose that the SC opening acts as a third window into the cochlea and explore the implications of this hypothesis. Our measurements and findings will then be compared with existing work. Finally, the clinical implication of our findings will be discussed in the context of superior semicircular canal dehiscence syndrome.

A. The effect of opening SC on middle-ear mechanics

1. Low-frequency effects (*f*<800)

SC opening has several low-frequency effects including increases in both Y_{ME} and *Hp* as well as reductions in nonlinearities and alterations in notches.

Prior to the SC opening both the Y_{ME} and *Hp* responses for $f<800$ Hz have a complex frequency response with maxima and minima in magnitude and a rapidly changing phase angle. The complex frequency dependence of Y_{ME} and *Hp* also varies with stimulus level. After the SC is opened, both Y_{ME} and *Hp* appear as a linear damped resonance at $f<800$ Hz with a clear compliance domination for $f<200$ Hz and masslike for $f>400$ Hz with a resistivelike plateau in between with an angle of zero (Fig. 4).

These alterations in both Y_{ME} and *Hp* after the SC is opened are consistent with the removal of a low-magnitude series admittance from the middle ear as a result of the hole in the SC. Similar observances of this middle-ear resonance on removing the cochlear load have been made in cat, rabbit, and human experiments (Moller, 1965; Puria and Allen, 1998; Rosowski *et al.*, 2006).

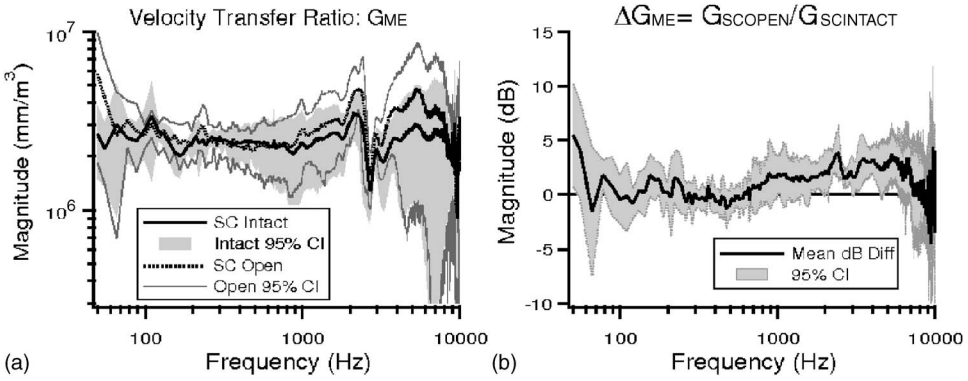


FIG. 11. (a) The mean (n=6) middle ear velocity transfer ratio (G_{ME}) with the SC intact and the SC open with 95% confidence intervals. (b) The mean dB difference between the G_{ME} with the SC intact and the SC open with associated 95% confidence intervals. The difference is statistically significant for most frequencies between 1500 and 6000 Hz.

2. High-frequency effects (*f*>800)

For frequencies above 800 Hz, both the Y_{ME} and Hp are dominated by the antiresonance (admittance zero) near 2600 Hz. Introducing a hole in the SC has little effect on Y_{ME} in this frequency range, but it does lead to significant increases in the Hp magnitude. The G_{ME} measurements also show increases in this frequency range as a result of SC opening. These statistically significant differences in high-frequency G_{ME} imply a difference in the coupling of volume velocity at the tympanic membrane to the stapes after the SC is opened. A potential reason for the difference in coupling at the high frequencies is that the mechanics of the tympanic membrane may dominate Y_{ME} at these higher frequencies whereas the decrease in load on the stapes after the SC is opened still leads to increased Vs.

B. Mechanical changes and the third window hypothesis

The third window hypothesis proposes a mechanism for the increases in Y_{ME} and Hp that we observed after introducing a SC hole in this study as well as decreases in cochlear sensitivity observed previously (Songer and Rosowski, 2005). In a healthy/intact ear there are two windows into the inner ear, the oval window and the round window. Assuming that the cochlear fluids are incompressible, in the healthy ear there is one main fluid pathway, where volume velocity is introduced at the oval window, propagates through the cochlea, and is dissipated at the round window. In the pathological case where there is a SC opening or hole, the SC hole acts as a third window into the inner ear, producing an additional fluid pathway that alters the mechanics of the inner ear. Using the third window hypothesis, one can predict the influence of a SC hole on Y_{ME} and Hp, as well as on cochlear sensitivity to auditory stimuli.

The anatomical location of the hole between the superior canal and the brain case suggests that the new window results in a fluid pathway that is in parallel with the normal fluid pathway through the cochlea. We hypothesize that the admittance associated with sound flow through the SC hole (third window) is mass dominated, where the mass can be estimated from the length and diameter of the SC in addition to the size of the hole. This mass domination causes the pathway through the SC hole to be highly admittant at low frequencies. At these frequencies the SCD acts to shunt sound energy away from the cochlea and reduce the cochlear admittance. As frequency increases, the admittance through the SC hole decreases, leading to a high-frequency cochlear admittance similar to that in the intact state. These predictions are consistent with the observations illustrated in Fig. 4. The increase in inner ear admittance as a result of the fluid pathway through the SC is also expected to lead to an increase in Hp in the low frequencies (Fig. 8). Despite the increased admittance of the inner ear, the low-frequency sensitivity of the cochlea to ear-canal sound pressure is expected to decrease because the stimulus is effectively being shunted through the SC hole, reducing the stimulus level reaching the cochlea. The predicted SCD-induced decrease in low-frequency sensitivity to auditory stimuli in chinchilla has been demonstrated previously (Songer and Rosowski, 2005). A codification of the third window hypothesis as a mechano-acoustic lumped-element circuit model is being developed. This model will allow us to quantitatively evaluate the predictions of the third window hypothesis as they relate to SC-opening-induced changes in both auditory sensitivity and mechanics in chinchilla.

One of the major differences between our SC opening and dehiscences observed in patients with SCD syndrome is that our dehiscence opens into the chinchilla middle-ear air space whereas in humans the dehiscence is open to the dura and cerebral spinal fluid of the middle cranial fossa. If we assume that the impedance of the cranial cavity is effectively that of a large fluid-filled cavity, then we would expect it to act as a large compliance in series with the dehiscence. The addition of such a compliance is likely to affect Y_{ME} and Vs at very low frequencies where it would add an additional resonance (between it and the masslike impedance of the superior canal), but is unlikely to qualitatively alter the results in the frequency range reported here. The impact of the addition of the cranial cavity and the frequencies where its effects would become important are being evaluated using a mechano-acoustic model of SCD for both chinchilla and human patients that is in development.

C. Comparison to previous work

1. Comparison of intact measurement to results from Ruggero et al.

The mean Vs measured with the SC intact and with open bullae are compared to that presented by Ruggero *et al.* (1990) in Fig. 12. Their data have been converted from peak velocity responses to 100 dB SPL tones to rms velocity per

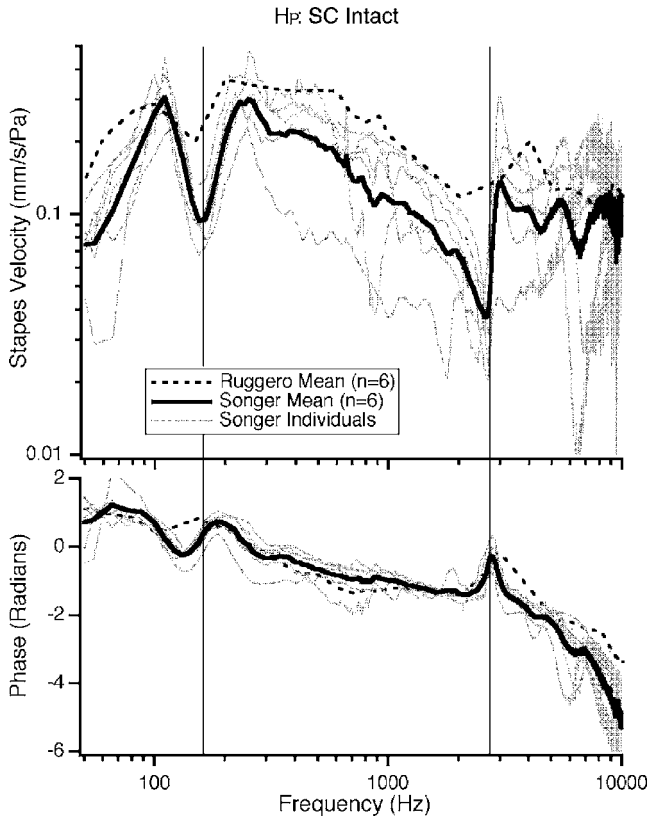


FIG. 12. Comparison of the Ruggero *et al.* (1990) data and the *Hp* data presented in this report. The Ruggero *et al.* data were originally reported in terms of peak velocity for a 100 dB SPL stimulus and have been converted to $mm-s^{-1}-Pa^{-1}$ as described in the text.

Pascal by dividing by $2\sqrt{2}$ (where 100 dB SPL=2 rms Pa and the $1/\sqrt{2}$ converts his peak value to rms) assuming linearity.

The data presented in this study as well as those presented by Ruggero *et al.* exhibit a *Vs* of 0.4 mm/s near 300 Hz followed by a decay out to near 2 kHz. Above 2 kHz there are differences between the two studies. This is likely due to the higher frequency resolution in this study, the lack of smoothing algorithms in this study, and differences in the notch near 2.6 kHz based on differences in bulla hole opening procedures. Despite these differences, the magnitude and general shape of the frequency responses are similar in both studies, which confirms the Ruggero *et al.* assumption that the patched slit in the TM in their recordings was not exerting a large influence on the recorded *Vs*. The overall magnitude of the Ruggero *et al. Vs* is greater that that in our study by a few dB, which may be due to differences in absolute calibration of the measurement systems.

2. The low-frequency notch

In both the intact Y_{ME} and *Hp*, a notch is observed in both the magnitude and phase between 150 and 200 Hz. Previous work (Rosowski *et al.*, 2006) has suggested an inner-ear-dependent nonlinearity which affects this notch. The data in this paper are consistent with the reports of a low-frequency nonlinear notch in the Y_{ME} of chinchilla.

A notch near 150 Hz in the chinchilla cochlear microphonic sensitivity functions has been observed by others

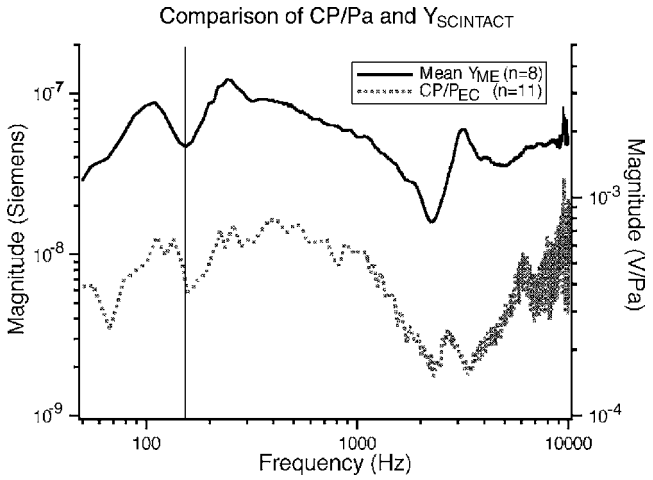


FIG. 13. A comparison of the cochlear potential normalized by ear canal sound pressure (Songer and Rosowski, 2005) and Y_{ME}. In both the measurements of Y_{ME} and CP normalized by ear canal sound pressure a low-frequency notch is observed near 165 Hz.

(Dallos, 1970); a similar notch has been observed in cochlear potential measurements in our laboratory (Fig. 13). Our Y_{ME} data have many features in common with cochlear potential (CP) measurements normalized by ear canal sound pressure (Songer and Rosowski, 2005). The CP and Y_{ME} data appear related by a single constant of proportionality, though the CP data are noisier for frequencies below 80 Hz and have more fine structure at higher frequencies than the Y_{ME} data. Dallos hypothesized that the notch is due to the influence of the helicotrema on the low-frequency cochlear impedance in chinchilla. He developed a low-frequency cochlear model that supports this view.

After introducing a SC hole, the low-frequency notch is no longer apparent in the Y_{ME} and *Hp* measurements. The presence of the notch in the intact state and its elimination after opening the SC can be explained in terms of the third window hypothesis affecting a cochlea with an impedance such as that proposed by Dallos. According to the third window hypothesis the SCD shunts volume velocity away from the cochlea and through the dehiscent canal. As argued previously, such a shunt might be expected to be most active at low frequencies where the acoustic admittance of the fluid in the canal is expected to be largest. If that is the case, the SC opening could "short-out" the helicotrema and the cochlea at low frequencies. The absence of the notch in the SC open Y_{ME} and *Hp* therefore support a proposed cochlear origin of the notch (Dallos, 1970). Similar effects of inner-ear manipulations on the notch and its level dependence are reported in another publication (Rosowski *et al.*, 2006).

3. The notch near 2.6 kHz

All of the measurements we made (Y_{ME} and *Hp*) in each condition, intact, SC open, SC patched, and with the patch removed, have a notch near 2600 Hz, which is visible in both magnitude and phase (Figs. 4 and 8). The notch is due to a resonance in response to introducing holes into the bullae. Work by Rosowski *et al.* demonstrates that this notch is eliminated by closing the bullar holes (Rosowski *et al.*,

2006). Open-hole resonances have been observed in other species e.g., cat (Guinan and Peake, 1967; Moller, 1965) and gerbil (Ravicz *et al.*, 1992).

To see if the notch frequency is consistent with a simple model of the cavity resonances, the frequency of the notch can be estimated by evaluating the lumped-element parameters (acoustic mass and compliance) of the hole and the bulla:

$$f_{\text{notch}} = \frac{\sqrt{1/MC}}{2\pi} = 2.6 \text{ kHz}. \quad (6)$$

We can calculate the acoustic compliance of the airspace, C, using the estimated volume of the cavity and the bulk modulus of air. Using a value for the volume of the chinchilla middle ear of 1.52 ml or 1.52×10^{-6} m^3 (Vrettakos *et al.*, 1988), thus $C=\text{volume}/\beta_{air}=1.52\times10^{-6}$ m^3/1.4*e*5 Pa$=1.09\times10^{-11}$ m^3/Pa. Rearranging Eq. (6) and inserting the values for C and f_{notch} we can solve for M,

$$M = \frac{1}{C\times(f_{\text{notch}}\times 2\pi)^2}, \quad (7)$$

the mass component of the system yielding a value of 344 kg/m^4. We then assume that the mass component is a radiation impedance where $M=0.8\rho_o/(\pi a)$ and ρ_o is the density of air. As noted previously, there are actually two holes introduced, which can be considered to be in parallel. Since both holes have approximately the same radius we will assume that the acoustic mass for the two holes is the same, thus the radius of each hole can be calculated as

$$a = \frac{2\times 0.8\rho_o}{\pi M} = 1.48 \text{ mm}. \quad (8)$$

This radius is comparable to the radius of the bulla holes (2–4 mm), though a bit of an underestimate. The underestimation may result from the middle-ear volume in our chinchillas being larger than that reported by Vrettakos *et al.* or from the simplifying assumption that only three (one compliance and two equal valued masses) lumped element parameters are sufficient to model the complex cavity system of the chinchilla.

D. Clinical significance

This study demonstrates that in chinchilla both Y_{ME} and the *Vs* produced in response to air-conducted sound increase after the introduction of a SC opening. These changes are similar to effects on umbo and stapes velocity measurements observed in SCD patients and in human temporal bone preparations (Chen *et al.*, 2006; Rosowski *et al.*, 2004).

The increases in both Y_{ME} and *Vs* magnitude after the SC is opened have clinical relevance. Superior semicircular canal dehiscence (SCD) syndrome is a disease in which a pathological hole in the SC is observed. Clinical results suggest that such a hole can cause changes in the effective stimulus to the inner ear (Mikulec *et al.*, 2004; Minor *et al.*, 2003). This study demonstrates that a SC opening increases the stimulus to the inner ear (sound-induced stapes motion), while previous work suggests that SC opening reduces the input to the cochlea leading to decreased auditory sensitivity to air-conducted sound stimuli (Songer and Rosowski, 2005). These two effects are consistent with the third window hypothesis described above. Clinical measures of the effect of SCD on hearing are variable and anatomical and physiological differences in the location of the dehiscence and in the middle ear of patients will affect the magnitude of the effect of SCD on the sensitivity of the cochlea leading to variations in the presentation of auditory symptoms.

The mechanical changes observed in the chinchilla middle ear have led to a number of proposed noninvasive measures to help diagnose SCD syndrome in human patients. The changes in Y_{ME} as a result of SC opening have raised the question of whether tympanometry may be an effective tool for diagnosing SCD syndrome. This seems unlikely because the measured changes in Y_{ME} are small and are not likely to fall outside of the range of Y_{ME} values seen in normal ears. Similarly, the small increases in umbo velocity that have been measured in patients with SCD syndrome (Chen *et al.*, 2006; Rosowski *et al.*, 2004) are also difficult to distinguish from normal responses. Despite the changes in middle-ear mechanics resultant from a SC opening, it does not appear, at this point, that the mechanical changes alone can be utilized effectively for the diagnosis of SCD syndrome. However, since the description of the changes in middle-ear motion is opposite that of the many other forms of conductive hearing loss, umbo velocity measurements may be useful in the diagnosis of SCD syndrome in the presence of conductive hearing loss (Chen *et al.* 2006).

As we mentioned in the Introduction, SCD syndrome is associated with both auditory and vestibular symptoms. The changes in Y_{ME} and *Vs* observed in this study indicate that the SCD changes the impedance of the inner ear in response to auditory sounds due to the addition of a shunt pathway in parallel with the cochlea. The introduction of such a shunt via the semicircular canal could lead to pathological stimulation of vestibular end-organs in the frequency range where shunting is active, i.e., where changes in Y_{ME} and *Vs* were observed.

V. CONCLUSIONS

The third window hypothesis predicts that changes in inner ear mechanics due to experimental manipulations of the SC will alter middle ear function. Our data support the predictions of the third window hypothesis demonstrating increases in both Y_{ME} and *Vs* in response to SC opening that are reversible when the SC hole is patched. These results indicate that pathologies of the vestibular system, such as SCD syndrome, can impact the mechanics of hearing.

ACKNOWLEDGMENTS

This work has been supported by an NSF graduate fellowship and NIH Grant Nos. T32 DC-00038 and R01 DC-00194. Saumil Merchant, Bill Peake, and Mike Ravicz provided insights and suggestions. Melissa Wood assisted with data collection, animal surgery and figure preparation.

[1]Those requiring additional reading on the admittance of single acoustic elements are directed to the text of Fletcher (1992).

Barany, E. (**1938**). "A contribution to the physiology of bone conduction," Acta Oto-Laryngol., Suppl. **26**, 1–223.

Bekesy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).

Brantberg, K., Bergenius, J., Mendel, L., Witt, H., Tribukait, A., and Ygge, J. (**2000**). "Symptoms, findings and treatment in patients with dehiscence of the superior semicircular canal," Acta Oto-Laryngol. **121**, 68–75.

Carey, J., Minor, L., and Nager, G. (**2000**). "Dehiscence or thinning of bone overlying the superior semicircular canal in temporal bone survey," Arch. Otolaryngol. Head Neck Surg. **126**, 137–147.

Chen, W., Ravicz, M., Rosowski, J., and Merchant, S. (**2006**). "Measurements of human middle- and inner-ear mechanics with dehiscence of the superior semicircular canal," Otol. Neurotol. (submitted).

Cremer, P., Minor, L., Carey, J., and Santina, C. (**2000**). "Eye movements in patients with superior canal dehiscence syndrome align with the abnormal canal," Neurology **55**, 1833–1841.

Dallos, P. (**1970**). "Low-frequency auditory characteristics: Species dependence," J. Acoust. Soc. Am. **48**, 489–499.

Fletcher, N. (**1992**). *Acoustic Systems in Biology* (Oxford U.P., Oxford).

Guinan, J., and Peake, W. (**1967**). "Middle-ear characteristics of anesthetized cats," J. Acoust. Soc. Am. **41**, 1237–1261.

Hirvonen, T., Carey, J., Liang, C., and Minor, L. (**2001**). "Superior canal dehiscence: Mechanisms of pressure sensitivity in a chinchilla model," Arch. Otolaryngol. Head Neck Surg. **127**, 1331–1336.

Huang, G., Rosowski, J., Puria, S., and Peake, W. (**2000**). "A noninvasive method for estimating acoustic admittance at the tympanic membrane," J. Acoust. Soc. Am. **108**, 1128–1146.

Kim, D., Siegel, J., and Molnar, C. (**1980**). "Postmortem effects and species differences for acoustic input characteristics at the eardrum of the chinchilla and cat," in *Society of Neuroscience Abstracts*, Vol. **6**, p. 41.

Lynch, T., Peake, W., and Rosowski, J. (**1994**). "Measurements of the acoustic input impedance of cat ears: 10 Hz to 20 kHz," J. Acoust. Soc. Am. **96**, 2184–2209.

Mikulec, A., McKenna, M., Ramsey, M., Rosowski, J., Herrmann, B., Rauch, S., Curtin, H., and Merchant, S. (**2004**). "Superior semicircular canal dehiscence presenting as conductive hearing loss without vertigo," Otol. Neurotol. **25**, 121–129.

Miller, J. (**1970**). "Audibility curve of the chinchilla," J. Acoust. Soc. Am. **48**, 513–523.

Minor, L. (**2000**). "Superior canal dehiscence syndrome," Am. J. Otol. **21**, 9–19.

Minor, L., Solomon, D., Zinreich, J., and Zee, D. (**1998**). "Sound- and/or pressure-induced vertigo due to bone dehiscence of the superior semicircular canal," Arch. Otolaryngol. Head Neck Surg. **124**, 249–258.

Minor, L., Carey, J., Cremer, P., Lustig, L., and Streubel, S. (**2003**). "Dehiscence of bone overlying the superior canal as a cause of apparent conductive hearing loss," Otol. Neurotol. **24**, 270–278.

Moller, A. (**1965**). "An experimental study of the acoustic impedance of the middle ear and its transmission properties," Acta Oto-Laryngol. **60**, 129–149.

Puria, S., and Allen, J. (**1998**). "Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay," J. Acoust. Soc. Am. **104**, 3463–3481.

Ravicz, M., Rosowski, J., and Voigt, H. (**1992**). "Sound-power collection by the auditory periphery of the mongolian gerbil meriones unguiculatus: I middle-ear input impedance," J. Acoust. Soc. Am. **92**, 157–177.

Rosowski, J., and Ravicz, M. (**2001**). "The middle-ear input admittance in chinchilla: Effect of middle-ear cavities and some low-frequency peculiarities," *Abstracts of the Twenty-Fourth ARO*, p. 62.

Rosowski, J., Ravicz, M., and Songer, J. (**2006**). "The acoustic admittance of the middle ear of the chinchilla: Structures that contribute to middle-ear frequency dependence at frequencies less than 10 kHz," J. Comp. Physiol. A (submitted).

Rosowski, J., Songer, J., Nakajima, H., Brinsko, K., and Merchant, S. (**2004**). "Investigations of the effect of superior semicircular canal dehiscence on hearing mechanisms," Otol. Neurotol. **25**, 323–332.

Ruggero, M., Rich, N., and Shivapuja, B. (**1990**). "Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea," J. Acoust. Soc. Am. **87**, 1612–1629.

Ruggero, M., Rich, N., Shivapuja, B., and Temchin, A. (**1996**). "Auditory nerve responses to low-frequency tones: Intensity dependence," Aud. Neurosci. **2**, 159–185.

Songer, J., and Rosowski, J. (**2005**). "The effect of superior canal dehiscence on cochlear potential in response to air-conducted stimuli in chinchilla," Hear. Res. **210**, 53–62.

Songer, J., Brinsko, K., and Rosowski, J. (**2004**). "Superior semicircular canal dehiscence and bone conduction in chinchilla," in *The Proceedings of the Third International Symposium on Middle Ear Research and Oto-Surgery*, edited by K. Gyo, H. Wada, N. Hato, and T. Koike, (World Scientific, Singapore), pp. 234–241.

Songer, J., Wood, M., and Rosowski, J. (**2005**). "Superior semicircular canal dehiscence decreases sensitivity to air conduction in chinchillas," *Abstracts of the Twenty-Eighth ARO*, p. 77.

Tonndorf, J. (**1972**). *Foundations of Modern Auditory Theory* (Academic, New York), Vol. **II**, chapter Bone Conduction.

Vrettakos, P., Dear, S., and Saunders, J. (**1988**). "Middle ear structure in the chinchilla: A quantitative study," Am. J. Otolaryngol. **9**, 58–67.

Distortion product otoacoustic emission fine structure analysis of 50 normal-hearing humans

Karen Reuter[a] and Dorte Hammershøi
Department of Acoustics, Aalborg University, Denmark

(Received 15 November 2005; revised 21 April 2006; accepted 24 April 2006)

When distortion product otoacoustic emissions (DPOAEs) are measured with a high-frequency resolution, the DPOAE shows quasi-periodic variations across frequency, called DPOAE fine structure. In this study the DPOAE fine structure is determined for 50 normal-hearing humans using fixed primary levels of $L_1/L_2=65/45$ dB. An algorithm is developed, which characterizes the fine structure ripples in terms of three parameters: ripple spacing, ripple height, and ripple prevalence. The characteristic patterns of fine structure can be found in the DPOAE of all subjects, though the DPOAE fine structure characteristics are individual and vary from subject to subject. On average the ripple spacing decreases with increasing frequency from $\frac{1}{8}$ oct at 1 kHz to $\frac{3}{32}$ oct at 5 kHz. The ripple prevalence is two to three ripples per $\frac{1}{3}$ oct, and ripple heights of up to 32 dB could be detected. The 50 normal-hearing subjects were divided into two groups, the subjects of group A having slightly better hearing levels than subjects of group B. The subjects of group A have significantly higher DPOAE levels. The overall prevalence of fine structure ripples do not differ between the two groups, but are higher and narrower for subjects of group B than for group A.
 [DOI: 10.1121/1.2205130]

PACS number(s): 43.64.Jb [BLM] Pages: 270–279

I. INTRODUCTION

Otoacoustic emissions (OAEs) are sounds generated by the inner ear as part of the normal hearing process. They can occur spontaneously and can be evoked by stimulating the ear acoustically. OAEs are caused by the cochlear amplifier involving active mechanical feedback from the outer hair cells (OHCs) to the basilar membrane. This feedback enhances the vibration of narrow regions on the basilar membrane and improves low-level sensitivity and sharpness of tuning. OAEs are assumed to be a by-product of this active process. The OHCs are the part of the hearing that is most sensitive to overexposure. Since the OHC function is monitored by OAEs, it seems reasonable that OAEs might be at least as sensitive to detect early hearing losses as other audiometric methods. In the literature it has been suggested that there is a correlation between the presence of OAEs and normal hearing threshold and the absence of OAEs and hearing loss. This suggestion holds for large datasets but not for individual diagnoses. Many researchers have tried to find a relation between OAE level and hearing threshold, but an unequivocal correlation is not established (Avan *et al.*, 1991; Bonfils and Avan, 1992; Gaskill and Brown, 1990, 1993; Gorga *et al.*, 1993). This may, however, be partly due to the fact that the OAE and hearing threshold inherently reflect two different processes. The hearing threshold also reflects the state of the inner hair cells (IHCs) and depends on the further neuronal processing, including cognitive detection.

Distortion product otoacoustic emission (DPOAE) is the response of the inner ear to two pure-tone stimuli (the primaries f_1 and f_2). Because of nonlinear interaction of the two tones in the cochlea, the primaries evoke a series of combination tones, the most prominent being at the frequency $2f_1-f_2$. For the measurement of DPOAE over a particular frequency range the frequencies of the primaries are varied simultaneously while keeping their frequency ratio constant. The origin of DPOAE is not completely understood. It is generally agreed that $2f_1-f_2$ DPOAE is generated by two sources, (1) the distortion component generated at the region of primary overlap near f_2 and (2) the reflection component from the distortion product frequency at $2f_1-f_2$ (Kalluri and Shera, 2001; Knight and Kemp, 2000; Konrad-Martin *et al.*, 2002; Mauermann *et al.*, 1999a; Talmadge *et al.*, 1998). The two sources at two distinct cochlear locations are generated by two different mechanisms: nonlinear distortion induced by the traveling wave and linear coherent reflection off preexisting micromechanical impedance perturbations (Kalluri and Shera, 2001). Recent discussions contradict this theory: Siegel *et al.* (2005) found that the group delays of stimulus-frequency OAEs were shorter than basilar-membrane group delays for frequencies <4 kHz. Ren (2004) detected forward-traveling waves on the basilar membrane and found that the stapes vibrate earlier than the basilar membrane, suggesting that OAE are emitted through the cochlear fluids as compression waves rather than along the basilar membrane as backward-traveling waves.

According to the backward-traveling wave theory the distortion component, which has a frequency of $2f_1-f_2$, is generated at the region of overlap near f_2. The energy of the distortion component travels bi-directionally, basally toward the ear canal and apically toward the distortion product frequency location $2f_1-f_2$, also called the characteristic frequency. At the characteristic frequency the energy undergoes linear coherent reflection. The reflection component travels basally toward the ear canal. At the stapes some distortion

[a]Electronic mail: kr@acoustics.aau.dk

 0001-4966/2006/120(1)/270/10/$22.50

product energy passes on to the middle ear, while some energy is reflected due to the impedance mismatch at the stapes and causes multiple internal reflections (Dhar *et al.*, 2002; Zweig and Shera, 1995). According to this theory the $2f_1-f_2$ DPOAE is a vector sum of the distortion component, the reflection component, and multiple internal reflections. The resulting interference pattern leads to variation in the sound pressure level and phase of the composite DPOAE. This variation in the sound pressure level of the DPOAE is quasi-periodic with frequency and is known as fine structure.

The DPOAE fine structure is characterized by consistent maxima and minima in dependence of frequency with depth of the notches up to 20 dB (Gaskill and Brown, 1990; He and Schmiedt, 1993; Heitmann *et al.*, 1996) and a periodicity of $\frac{3}{32}$ oct (He and Schmiedt, 1993; Mauermann *et al.*, 1997b). A similar periodicity structure can be found in all other types of OAE, in the hearing threshold (Kemp, 1979a, b; Long, 1984; Schloth, 1983; Zwicker and Schloth, 1984), and, as Mauermann *et al.* (2004) recently have shown, in the low-level equal loudness contours. No direct correlations could be established between frequencies of DPOAE maxima and minima and the threshold fine structure in humans (Mauermann *et al.*, 1997a; Talmadge *et al.*, 1998), which might be due to the fact that the DPOAE does not simply reflect properties of cochlear status near f_2.

Mauermann *et al.* (2004) suggested that the fine structure might serve for the identification of early hearing loss: "The high sensitivity of fine structure [obtained from psychoacoustic experiment or on OAE measurements] to cochlear damage may offer the opportunity to further categorize the group of the "normal-hearing" subjects and to find methods for early diagnosis of incipient cochlear damage." In the literature there are some indications that the presence of cochlear fine structure might be a property of the healthy ear. DPOAE fine structure reappears at a very late stage of recovery after a sudden hearing loss (Mauermann *et al.*, 1999b). A reduction of DPOAE fine structure after aspirin consumption was observed by Rao *et al.* (1996). Overexposure experiments, which cause a temporary change in the auditory system, have shown to flatten the DPOAE fine structure (Engdahl and Kemp, 1996). Mauermann *et al.* (1999b) determined DPOAE fine structures for subjects with mild to moderate cochlear hearing losses with certain shapes of hearing loss. The subjects for that study were chosen based on the results of computer modeling (Mauermann *et al.*, 1999a). When the primaries were located in the region of normal or near normal hearing, but DP frequencies were located in a region of impairment, the distortion product $2f_1-f_2$ was still observable, but the DPOAE fine structure disappeared. When the DP frequencies fell into a region of normal hearing, fine structure was preserved as long as DPOAEs could be recorded. In the simulations (Mauermann *et al.*, 1999a) the fine structure disappeared, when the roughness in the modeling was removed only around the characteristic frequency. The roughness in the modeling reflects random inhomogeneity in the placement and behavior of cells along the cochlea, especially the outer hair cells. When the roughness is removed from the primary region, no effect is seen on the fine structure. This means that a disappearance or flattening of fine structure can be expected if the reflection source is affected by cochlear damage while the distortion source falls in a relative healthy region, as it was shown in Mauermann *et al.* (1999b). For subjects with mild increasing high-frequency hearing loss the reflection source is less affected than the distortion source. In these cases the vector components of the two sources are still of similar size or even adjusted to be more similar. According to the modeling, the fine structure is in these cases assumed to be preserved or even increased.

In summary, there are indications in the literature that the DPOAE fine structure might contain more information about the state of hearing than the DPOAE level alone. It is not known whether there is a systematic change of DPOAE fine structure—either an increase or a decrease—with the state of hearing. In the present study the DPOAE fine structure was determined in 50 normal-hearing subjects over a frequency range of three octaves. The purpose of the measurements was to study the prevalence of fine structure in a group of normal-hearing subjects and to develop a classification algorithm for the characterization of fine structure ripples.

II. MATERIALS AND METHODS

In the experiment all 50 subjects had their pure-tone thresholds on both ears tested. DPOAE were measured in one ear, which was chosen randomly. During the entire test the subjects were seated in a double-walled, sound-isolated audiometry chamber, complying with ISO 8253-1:1989. For eight subjects the DPOAE measurement was repeated after 4 weeks.

A. Subjects

The subjects were aged between 20 and 29 (mean =23.6), 24 male and 26 female, and reported no known incidents of excessive exposures or known hearing losses. All subjects had hearing levels below 25 dB in the measured frequency range.

B. Pure-tone audiometry

For 40 subjects the pure-tone audiometry was performed with a custom-built audiometer (Lydolf, 1999), using Sennheiser HDA 200 headphones and the ascending method that complies with the norms for automatic audiometries (ISO-8253-1:1989). The system was calibrated using the B&K type 4153 artificial ear according to IEC-60318-1:1992 and IEC-60318-2:1992. Hearing thresholds were measured in $\frac{1}{2}$ oct from 250 Hz to 4 kHz. For ten subjects the hearing thresholds were measured using the Madsen Orbiter 922 in a frequency range from 250 Hz to 8 kHz and a frequency resolution of one octave.

C. DPOAE measurement

The $2f_1-f_2$ DPOAEs were measured using the ILO96 Research system from Otodynamics. DPOAEs were measured in the frequency range of 903 Hz$<f_2<$6201 Hz with $f_2/f_1=1.22$ and fixed primary levels of $L_1/L_2=65/45$ dB.

The choice of primary levels is based on several considerations. The DPOAE fine structure can be measured with both equal and unequal primary levels. Measurements with varying equal primary levels ($L_1=L_2=40$ to 70 dB) have shown a flattening of the fine structure for some subjects at high levels (He and Schmiedt, 1993; Heitmann *et al.*, 1996; Mauermann *et al.*, 1997b). The pattern of DPOAE fine structure gets shifted along the frequency axis when the primary levels are increased (He and Schmiedt, 1993; Mauermann *et al.*, 1997b). Low level primaries with $L_1>L_2$ have been shown to be most sensitive to overexposure effects (Sutton *et al.*, 1994). Therefore low-level primaries are likely to detect fine structure ripples and to be more sensitive to small permanent changes in the auditory system, i.e., to an early hearing loss. The level combination of 65/45 dB was chosen based on results from Whitehead *et al.* (1995), who measured DPOAEs for various primary level combinations at different frequencies (1.39, 2.79, and 5.57 kHz). The primary level combination 65/45 dB showed relatively high level DPOAEs for the tested frequencies. It was assumed to be the best compromise between measurable DPOAEs, presence of fine structure, and high sensitivity to detect small changes.

The DPOAE fine structure was measured using the frequency resolution "micro." It presents 17 primary tones within 200-Hz intervals for $f_2<3$ kHz and within intervals of 400 Hz for $f_2>3$ kHz. Each pair of primary tones was averaged in the time domain (32 subaverages), corresponding to approximately 3-s measurement time for one pair of frequencies. The system uses a constant average time, independent off the signal-to-noise ratio (SNR). To cover the full frequency range, 22 measurements in different frequency ranges were measured. The probe was not removed between these measurements, unless the measurement system indicated that the probe fit was altered. The measurement of a DPOAE covering the entire frequency range lasted approximately 30 min, including the breaks where data were saved.

Prior to each measurement a checkfit procedure is performed, where two broadband click stimuli are alternately delivered by the two output transducers. The checkfit result is stored in an array and used during data collection to balance and normalize the two stimuli levels. All spectrum analyses are performed by the system. A fast Fourier transform (FFT) with a frequency resolution of 12.2 Hz is performed. The noise is estimated from the ten Fourier components nearest to but not including the $2f_1-f_2$ frequency. The noise is represented as all levels within two standard deviations of the background noise, i.e., the limits of the 95% confidence region. All measurements are saved as spreadsheet files and further analyzed. The 22 measurements of the different frequency regions are concatenated to one DPOAE measurement covering the measured frequency range.

D. Classification algorithm

The DPOAE fine structure is analyzed by an automatic classification algorithm, which is described in detail in the following. The DPOAE fine structure is characterized by consistent maxima and minima, the frequency locations of this ripple structure being very individual. The mean DPOAE over all subjects would smooth out any fine structure. Therefore, the fine structure of all subjects is analyzed individually. For the fine structure analysis the ripple maxima and minima are detected. Two reasons complicate this ripple structure analysis: (1) low-level variations of the DPOAE and (2) DPOAE levels that are below the noise floor.

Narrow-band level variations exist, some presumably due to measurement errors. With the frequency resolution used here, this would lead to the identification of an excessive amount of narrow ripples, impeding the identification of true fine structure ripples. The first determination of maxima and minima is therefore based on the average of every set of five neighbor frequencies (frequency smoothing). Characteristic levels are determined from the raw data at the minima and maxima frequencies. Any mishap from the smoothing is at the same time repaired: If the real extreme value is at either one of the neighbor frequencies, then the frequency and levels of the real extreme is used.

A ripple is rejected, whenever the maximum is less than 3 dB above the noise floor. In the following this is referred to as the $\mathrm{SNR}_{\mathrm{max}}$ criterion, where the noise floor is estimated by the average noise at the DP frequency $2f_1-f_2$ and its four nearest neighbors. Ripples are also rejected, when they do not fulfill a certain ripple height criterion, i.e., ripples have to have a certain height in order to be detected as real ripples. The fine structure is characterized by the following parameters:

(i) ripple center frequency: the frequency centered between two minima (on a logarithmic axis)
(ii) ripple spacing: the frequency distance between two minima,
(iii) ripple height: the level difference between the maximum and the mean of the two minima, and
(iv) ripple prevalence: number of ripples are counted in $\frac{1}{3}$-oct frequency bands (in the present case with center frequencies of $f_2=1260$, 1590, 2000, 2520, 3170, 4000, and 5040 Hz).

A "schematic" DPOAE fine structure can be derived from the first three parameters: ripple spacing, ripple height, and ripple center frequency. Also these may favorably be analyzed in given frequency ranges, e.g., for comparison across subjects and populations in $\frac{1}{8}$-oct frequency bands (in the present case with center frequencies of $f_2=1091$, 1189, 1297, 1414, 1542, 1682, 1834, 2000, 2181, 2378, 2594, 2828, 3084, 3364, 3668, 4000, 4362, 4757, and 5187 Hz).

Two examples for the derivation of the schematic fine structure are illustrated in Fig. 1. The top figures illustrate fine structure measurement examples for two subjects, zoomed into a narrow frequency range. The top figures also show the noise floor, the smoothed DPOAE, and the detected maxima and minima. The bottom figures illustrate the derived schematic fine structure for both subjects. The subject shown on the left panel has a high-level DPOAE and a pronounced fine structure with ripple heights of up to 32 dB, whereas the subject shown on the right panel has lower-level DPOAE and little fine structure. For the analysis a ripple

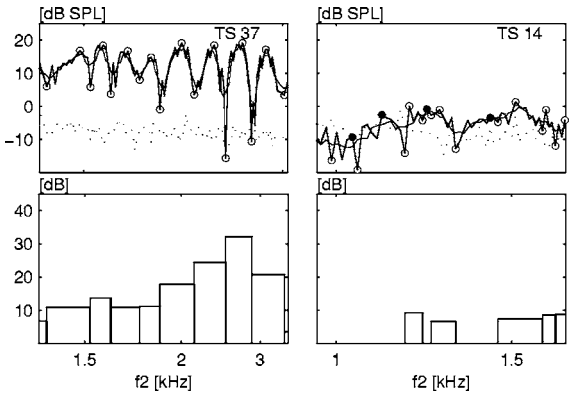


FIG. 1. Derivation of schematic DPOAE fine structure, examples for two subjects. Top panel: DPOAE fine structure (black line), noise floor (dotted gray line), smoothed DPOAE fine structure (gray line), maxima and minima (white circles), and maxima of rejected ripples (black circles). Lower panel: schematic representation of the fine structure ripples.

height criterion of 3 dB and a SNR_{max} criterion of 3 dB were chosen. For the data shown on the right panel many of the fine structure ripples fall below the 3-dB ripple height criterion and are therefore rejected. The fine structure of this subject also has DPOAE levels very close to the noise floor. Since many of the maxima are not 3 dB above the noise floor, these ripples are also rejected and not detected as "real" maxima.

III. RESULTS

A. Repeatability of DPOAE fine structure

For 8 of the 50 subjects repeated DPOAE measurements were taken. Figure 2 illustrates the repeated measurements for three subjects. In total three DPOAE measurements were taken for each of the eight subjects. In the first session the DPOAE was measured twice, without removing the probe between the two measurements. After 4 weeks the measurement of one DPOAE was repeated. The average measurement time was twice as long for the third measurement (~3 s) as for the first and second measurement (~1.5 s), therefore the noise floor of the third measurement is higher. The repeated measurements show that the fine structure pattern remains stable over time also for the repeated measurement after 4 weeks.

On the bottom of Fig. 2 the schematic fine structures, derived from the ripple parameters, are shown for the first two measurements. The repeatability of the schematic fine structure is not perfect, i.e., the DPOAE measurement and/or the fine structure analysis is not perfectly robust against measurement uncertainties. In few cases the classification algorithm detects one broadband ripple, whereas in the repeated measurement this ripple is detected as two narrow-band ripples or vice versa (e.g., Fig. 2, top panel at 1 and 4 kHz). In these cases all parameters can vary significantly. Otherwise the parameters' ripple width and ripple prevalence seem to be stable. The parameter ripple height can show some variations, especially when the DPOAE levels are close to the noise floor (e.g., Fig. 2, second panel at 1–2 kHz).

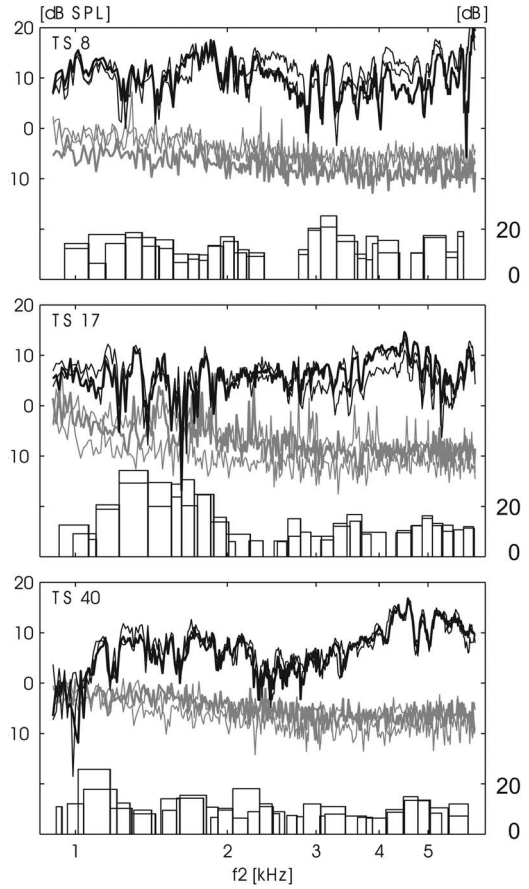


FIG. 2. Repeated DPOAE fine structure measurements for three subjects: two successive measurements (thin lines), one repetition after 4 weeks (thick line), the noise floor of the measurement (gray). Schematic presentation of DPOAE for two measurements without probe reinsertion (right ordinate).

B. Optimization of classification algorithm

The classification algorithm is based on several parameters, e.g., the number of frequencies included in smoothing, the SNR_{max} criterion, and the ripple height criterion. These parameters were optimized based on the repeated measurements. The ripple parameters ripple spacing, ripple height, and ripple prevalence were determined for each of the three repeated measurements. Standard deviations of ripple parameters for these three measurements were calculated for each subject and then averaged over all eight subjects. For the fine structure analysis different parameters were tested, i.e., the number of frequencies included in smoothing (three, five, seven, and nine values), the ripple height criterion (3, 5, and 7 dB), and the SNR_{max} criterion (0, 3, 5, and 10 dB or no rejection). In this way it could be tested, under which conditions ripple parameters could be detected with the lowest spread. Figures 3–5 show the results of the analysis. When different rejection criteria are used in the analysis (Figs. 3 and 4), similar results are obtained for the repeatability of the analysis method. It was decided to use a ripple height criterion of 3 dB and a SNR_{max} criterion of 3 dB for the automatic classification algorithm. The number of frequencies included in smoothing (Fig. 5) has a higher influence on the repeatability of the algorithm than the rejection criteria, i.e.,

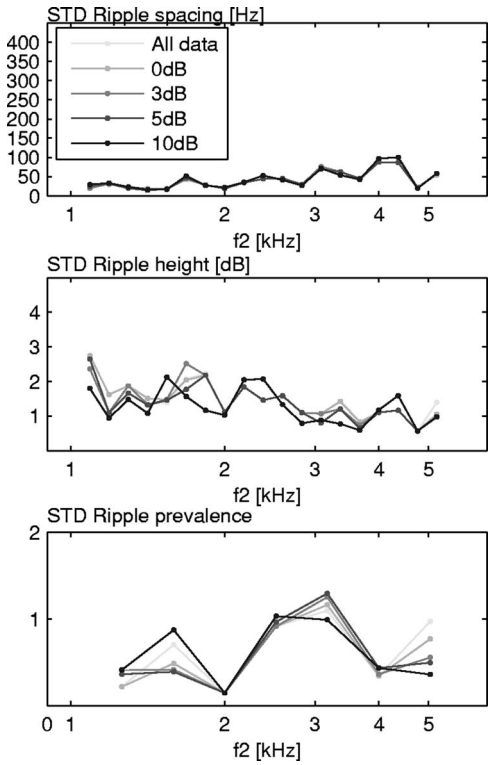


FIG. 3. Standard deviations (STD) of three repeated measurements (averaged over eight subjects) for derived parameters ripple spacing (top panel), ripple height (mid panel), and ripple prevalence (lower panel). The SNR_{max} criterion is varied (no rejection=all data, 0, 3, 5, and 10 dB) for a fixed frequency smoothing of 5 and a ripple height criterion of 3 dB.

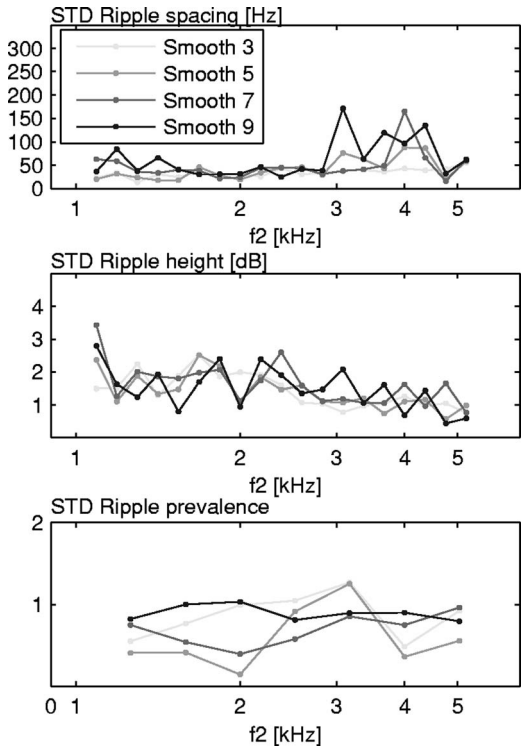


FIG. 5. Standard deviations (STD) of three repeated measurements (averaged over eight subjects) for derived parameters ripple spacing (top panel), ripple height (mid panel), and ripple prevalence (lower panel). The frequency smoothing is varied (three, five, seven, or nine frequencies) for a fixed SNR_{max} criterion of 3 dB and a fixed ripple height criterion of 3 dB.

the standard deviations show variations from each other depending on how many frequencies are included in data smoothing. The smoothing of five frequencies results in a relatively low standard deviation of all parameters and was therefore chosen for further analysis.

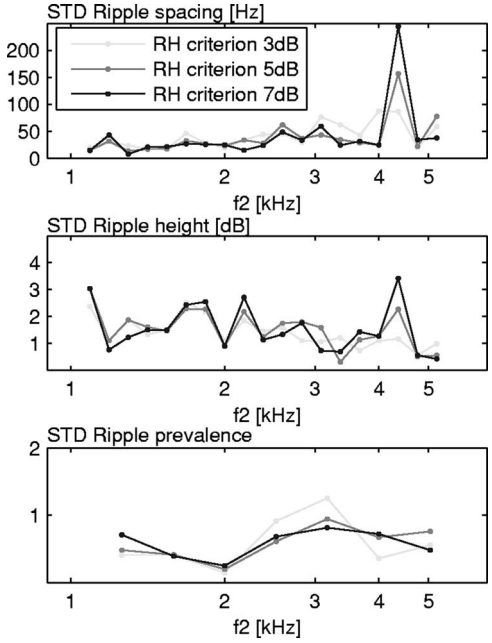


FIG. 4. Standard deviations (STD) of three repeated measurements (averaged over eight subjects) for derived parameters ripple spacing (top panel), ripple height (mid panel) and ripple prevalence (lower panel). The **ripple height** criterion is varied (3, 5, and 7 dB) for a fixed frequency smoothing of 5 and a SNR_{max} criterion of 3 dB.

DPOAE recordings can have values very close to or below the noise floor. For the determination of the ripple height the level difference between maxima and minima are taken. The minima often lie below the estimated noise floor of the measurement, therefore the levels at minima are very sensitive to noise, and the repeatability of the DPOAE level at minima is rather low. It can be discussed whether measurements below the noise floor should be taken into account in the analysis. In many cases there are ripples with a maximum value far above the noise floor and minima with values below the noise floor. In most cases these ripples are repeatable and are therefore considered to be true ripples (e.g., Fig. 2, second panel at 1–2 kHz). Therefore it was decided not to reject these ripples and accept them, if the level at the maximum is 3 dB above the noise floor (SNR_{max} criterion =3 dB). Since the levels at minima are rather uncertain, the ripple height is not highly repeatable, which has to be kept in mind in the further analysis.

C. Presentation of individual data

Figure 6 shows the individual data for 15 randomly selected subjects. The figures show the DPOAE fine structure and the schematic fine structure, which is derived from the calculation of ripple spacing, height, and center frequency of the ripples, as described in Sec. II D. It can be seen that the characteristics of DPOAE fine structure exist for all subjects. The fine structure is not evenly pronounced over the measured frequency range, i.e., for most subjects there are narrow frequency ranges, in which the fine structure is more

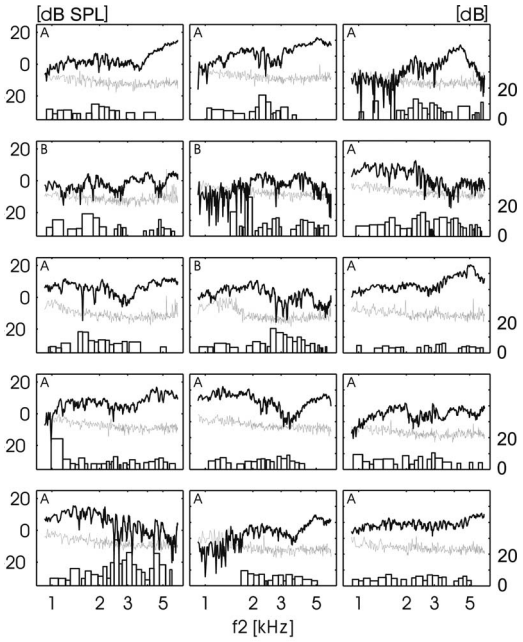


FIG. 6. DPOAE fine structures (black line, left ordinate) with noise floor (gray line, left ordinate) and schematic DPOAE (right ordinate) of 15 randomly selected subjects. The upper left corner indicates whether subjects belong to group A or group B; subjects belonging to group A having better hearing thresholds than subjects of group B.

pronounced than in other frequency regions. Some subjects have fine structures with ripple heights up to 32 dB, whereas the ripple heights of other subjects is very low over the measured frequency range. The letter A or B in the upper left corner of each figure indicates whether a subject belongs to group A or group B. Subjects of group A have slightly better hearing thresholds than subjects of group B. A detailed analysis of these two groups is described in Sec. III E.

D. Analysis of DPOAE fine structure

The fine structure parameters ripple spacing, ripple height, and ripple prevalence were determined for each subject over the measured three octaves and averaged over all 50 subjects. In Figs. 7–9 the results of the fine structure analysis are illustrated.

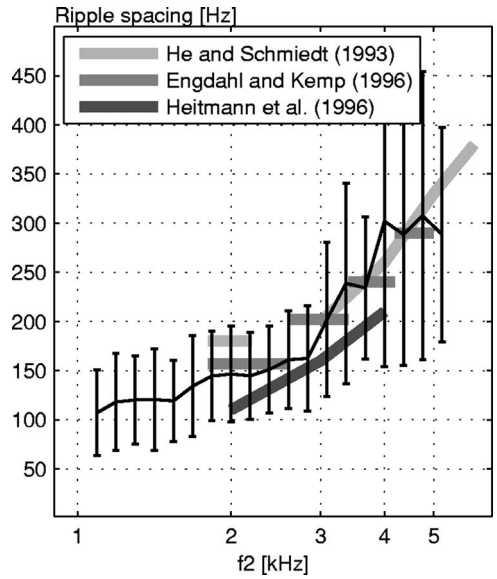


FIG. 7. Mean ripple spacing of 50 subjects in $\frac{1}{8}$-oct bands. Errorbars are standard deviations between subjects. Gray bars: reference data from previous studies.

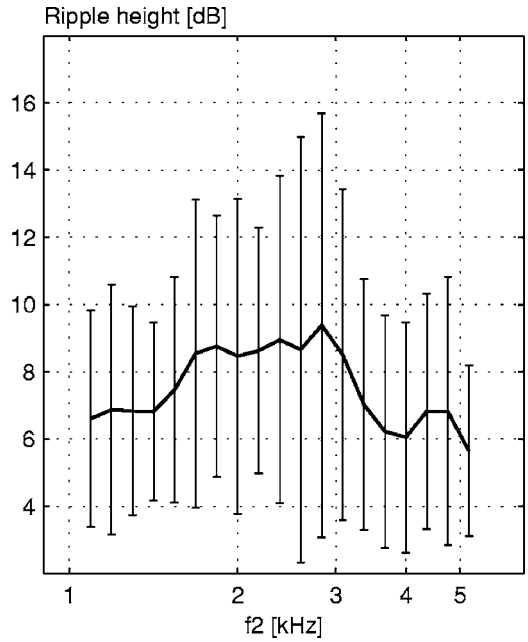


FIG. 8. Mean ripple height of 50 subjects in $\frac{1}{8}$-oct bands. Errorbars are standard deviations between subjects.

The ripple spacing is plotted in Fig. 7 as a function of the primary frequency f_2. Means and standard deviations over the 50 subjects are plotted in $\frac{1}{8}$-oct bands. The figure also contains data from literature. The ripple spacing found in this study increases from 100 Hz at the lowest frequency to 300 Hz at the highest frequencies. This corresponds to a decrease from $\frac{1}{8}$ to $\frac{3}{32}$ oct. The ripple spacings found in this study agree very well with the ripple spacings found in the

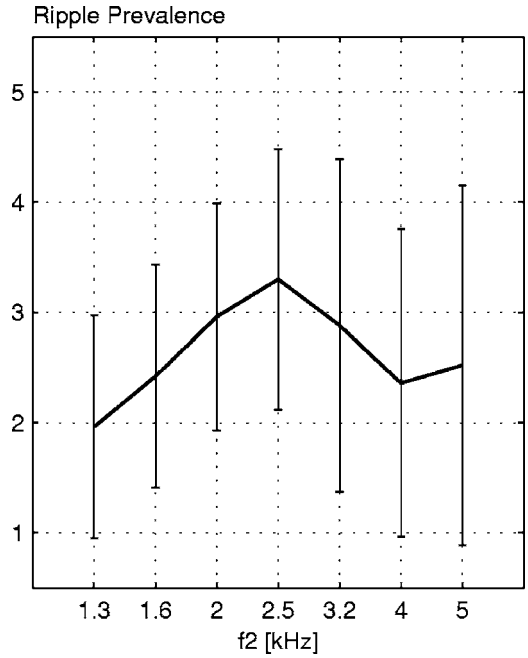


FIG. 9. Mean ripple prevalence of 50 subjects in $\frac{1}{3}$-oct bands. Errorbars are standard deviations between subjects.

literature. Engdahl and Kemp (1996) measured DPOAE fine structures with a ripple spacing of 157 Hz at 1.8–2.6 kHz, 202 Hz at 2.6–3.4 kHz, and 240 Hz at 3.4–4.2 kHz. He and Schmiedt (1993) measured frequency spacings of $\frac{3}{32}$ oct at 3–6 kHz and $\frac{1}{8}$ oct at 2 kHz, and Heitmann *et al.* (1996) found ripple spacings of 0.078 octaves corresponding to 180 Hz at 2–4 kHz.

Figure 8 shows the ripple height as a function of f_2. The spread of ripple heights between subjects is high; ripples with heights between 0 dB and up to 32 dB are found. On average the height of the ripples is between 6 and 9 dB. In the mid frequency range (1.5–3 kHz) the ripples are slightly higher.

The prevalence of ripple fine structure is plotted in Fig. 9. On average the 50 subjects have two to three ripples per $\frac{1}{3}$-oct band. Ripples are most prevalent in the mid frequency range (at 2.5 kHz).

E. Group differences

The 50 subjects had all normal-hearing thresholds with hearing levels below 25 dB. If the DPOAE reflects early stages of hearing loss, it might be possible to further categorize the group of subjects. The subjects were therefore divided into two groups. Group A consists of 39 subjects and group B of 11 subjects, where the subjects of group B have hearing levels exceeding 10 dB for at least two frequencies. The hearing levels for the subjects of each group can be seen in Fig. 10. The mean hearing levels for subjects of group B are slightly raised in the measured frequency range compared to the hearing levels for subjects of group A.

The overall DPOAE levels for subjects of group A and group B are plotted in Fig. 11. Subjects of group B have significantly lower DPOAE levels than subjects of group A over the measured frequency range. The ripple parameters ripple spacing, height and prevalence were calculated for the subjects of the two groups. Subjects of group A have broader fine structure ripples than subjects of group B (Fig. 12). Group B subjects have higher ripples (Fig. 13) and a higher prevalence of DPOAE fine structure (Fig. 14). The latter difference is, however, not significant (see ANOVA Table I).

IV. DISCUSSION

The fine structure characteristics found in this study substantially agree with the characteristics found by other authors (Engdahl and Kemp, 1996; He and Schmiedt, 1993; Heitmann *et al.*, 1996). Fine structures with ripple heights of up to 32 dB and ripple spacings around $\frac{1}{8}$ to $\frac{3}{32}$ oct were found. Small differences between results might be caused by the choice of DPOAE measurement parameters, the frequency resolution of DPOAE measurements, and the way the fine structure characteristics are calculated.

Different methods can be chosen to measure the ripple spacing. Some authors determined the ripple width from the maxima-to-maxima distance of the ripples (He and Schmiedt, 1993). Another possibility could have been to determine the equivalent rectangular bandwidth (ERB), as described by Moore (2003). The ERB uses all information on the ripple to describe it, whereas, e.g., ripple spacing between maxima or minima is affected by the accuracy of the determination of a few points in the measurements. In this work the ripple spacing was calculated from the minimum-to-minimum distance. The repeated measurements have shown that the frequency locations of minima are highly

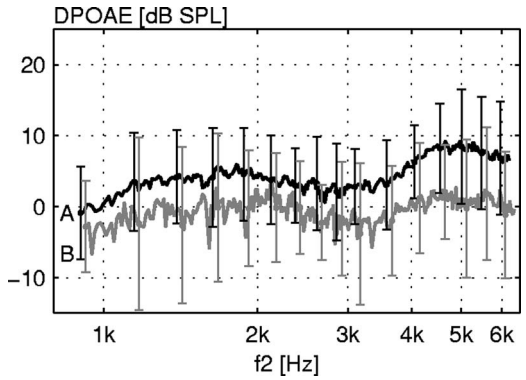


FIG. 11. Mean DPOAE levels and standard deviations for subjects of group A and group B. Subjects belonging to group A have better hearing thresholds than subjects of group B.

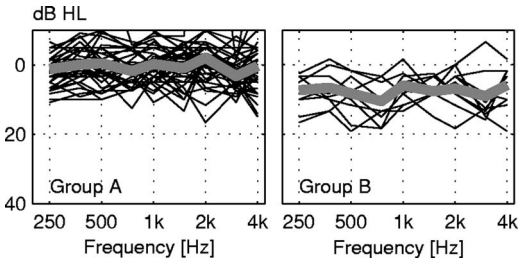


FIG. 10. Mean (gray line) and individual (black lines) hearing levels and standard deviations for subjects of group A (left panel) and group B (right panel). Subjects belonging to group A have better hearing thresholds than subjects of group B.

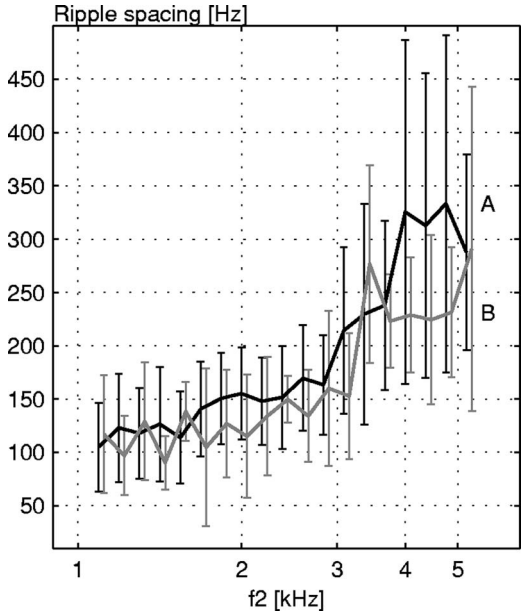


FIG. 12. Ripple spacing in $\frac{1}{8}$-oct bands for subjects for group A (black line) and group B (gray line). Subjects belonging to group A have better hearing thresholds than subjects of group B. Errorbars are standard deviations between the subjects.

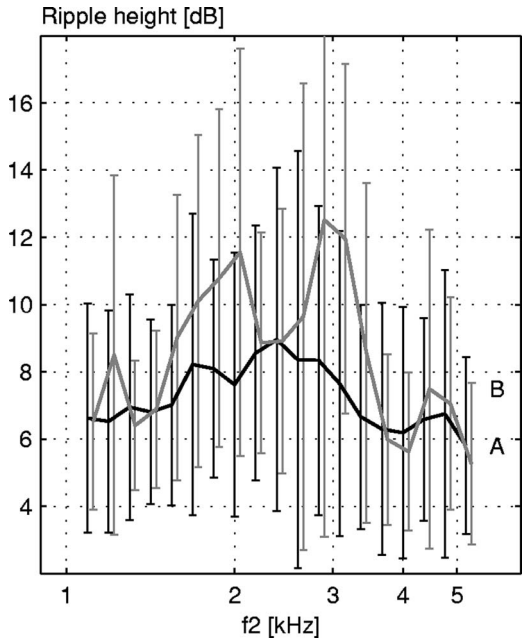


FIG. 13. Ripple height in $\frac{1}{8}$-oct bands for subjects of group A (black line) and group B (gray line). Subjects belonging to group A have better hearing thresholds than subjects of group B. Errorbars are standard deviations between the subjects.

repeatable. Therefore a relatively accurate determination of ripple spacing is possible by taking the distance between neighboring minima (when a sufficiently high-frequency resolution of the DPOAE measurement is used to find the right maxima and minima locations).

For the determination of ripple height the level difference between the maximum and the mean of the minima levels is determined. The minima often have levels below the noise floor, which means that their levels are not very reliable, and therefore the accuracy of the determination of ripple height may depend on the level of the noise floor. For the classification algorithm this calculation was evaluated to be the best approximation of ripple height. For a better determination of ripple height the DPOAE measurements should be performed with a longer average time, i.e., a better SNR.

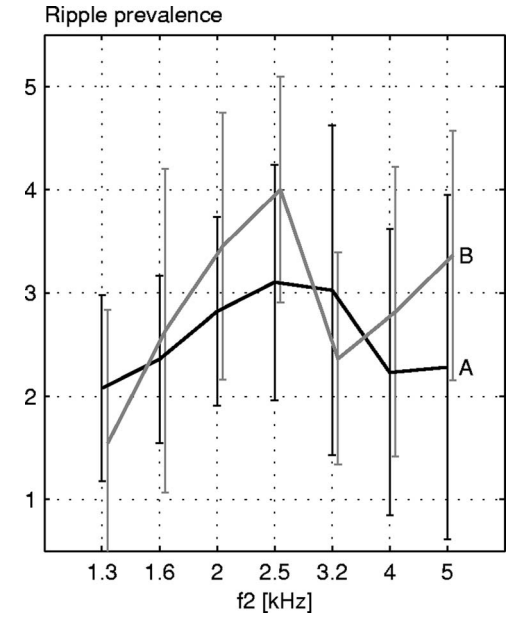


FIG. 14. Number of ripples per $\frac{1}{3}$-oct bands for subjects of group A (black line) and group B (gray line). Subjects belonging to group A have better hearing thresholds than subjects of group B. Errorbars are standard deviations between the subjects.

TABLE I. Unbalanced two-way-ANOVA p-values. Tested parameters: DPOAE, ripple spacing (RS), ripple height (RH), and ripple prevalence (RP).

Source	DPOAE level	RS	RH	RP
Frequency	0[b]	0[b]	0[b]	0[b]
Group	0[b]	0.0003[b]	0.0014[b]	0.0444[a]
Frequency * Group	1	0.029[a]	0.4484	0.0209[a]

[a] $p<0.05$.
[b] $p<0.01$.

Measurements of DPOAE fine structure with different primary levels have shown that the prevalence of fine structure depends on the choice of primary levels. High primary levels give less pronounced fine structure than lower level primaries for some subjects, i.e., the fine structure flattens out with increasing primary levels (He and Schmiedt, 1993; Mauermann *et al.*, 1997b). The calibration of the primary levels in most measurement systems is performed in the ear canal. Depending on the calibration and the transmission characteristics of the middle ear the same primary levels might result in different stimulation strengths of the basilar membrane for different subjects, i.e., for subjects with very good middle ear transmission the primary levels of $L_1/L_2 = 65/45$ dB might be chosen relatively high with the risk of flattening possible fine structures, whereas for subjects with lower middle ear transmission the same primary levels might result in less stimulation strength of the basilar membrane. Therefore it might be necessary to optimize the primary levels for each subject individually.

The results of this study show that subjects with normal hearing thresholds can have different prevalence of DPOAE fine structure. Subjects with good hearing thresholds might have little fine structure, whereas subjects with raised hearing thresholds might have a more pronounced fine structure. All subjects have hearing thresholds in the range of normal hearing, but a "raised" threshold according to the authors' definition, when the hearing level exceeds 10 dB for at least two tested frequencies. The hearing levels are below 20 dB for most of the subjects. There are always some uncertainties in the hearing threshold measurements, i.e., it is not certain that group A subjects have better hearing abilities than subjects of group B. A comparison of overall DPOAE levels of the two groups (Fig. 11) shows that the DPOAE levels are significantly lower for group B subjects than for group A subjects. This DPOAE level difference supports the idea that there is a difference in the state of the hearing, although it is not certain.

By dividing the normal-hearing subjects into two groups, it could be investigated whether there is a systematic change in the DPOAE fine structure—either an increase or decrease of the fine structure—as it is suggested in the lit-

erature. A difference between the two groups of subjects was found for the ripple spacing of DPOAE fine structure: group A subjects have significantly broader ripples than subjects of group B. Group B subjects have generally lower DPOAE levels than group A subjects and therefore a lower SNR and a lower repeatability of DPOAE measurements. Low-level DPOAE recordings might have small amplitude variations that are caused by measurement uncertainty. These small level variations might be interpreted as ripples by the automatic classification algorithm. When in the analysis all ripples with spacing less than 100 Hz are filtered out for all subjects, then there is (not surprisingly) no significant difference in ripple spacing between the two groups, but the difference in ripple height is still significant. It might be argued whether these narrow ripples are "true" ripples or caused by measurement uncertainty. The difference in ripple spacing between the two groups of subjects still exist over the entire measured frequency range, also at high frequencies, where the SNR of the DPOAE measurement is relatively good. Thus it seems unlikely that the differences observed are due to measurement uncertainties.

V. CONCLUSION

DPOAE fine structures with a fixed-level paradigm have been determined in 50 normal-hearing subjects. A classification algorithm has been developed that finds individual ripples in the DPOAE. The fine structures are individual and stable over time, i.e., the locations of maxima and minima are highly repeatable. Fine structure ripples with a height of up to 32 dB were determined. On average the ripple spacing decreases with increasing frequency from $\frac{1}{8}$ oct at 1 kHz to $\frac{3}{32}$ oct at 5 kHz. A prevalence of two to three ripples per $\frac{1}{3}$ oct band was found. In this study the 50 normal-hearing subjects were divided in two groups, the subjects of group A having slightly better hearing thresholds than the subjects of group B. Group A subjects show on average significantly higher DPOAE levels than subjects of group B. The subjects of group B have generally higher, but more narrow, ripples.

ACKNOWLEDGMENTS

This work was financed by the William Demant Foundation (Oticon) and The Danish Council for Technology and Production Sciences. The authors would like to thank all subjects for participating in this experiment. We would like to thank Rodrigo Ordoñez for committed discussion on data analysis and Miguel Angel Aranda de Toro for participation in data collection. We are also grateful for the comments and suggestions on the manuscript by two reviewers.

Avan, P., Bonfils, P., Loth, D., Narcy, P., and Trotoux, J. (**1991**). "Quantitative assessment of human cochlear function by evoked otoacoustic emissions," Hear. Res. **52**, 99–112.

Bonfils, P., and Avan, P. (**1992**). "Distortion-product otoacoustic emissions—Values for clinical use," Arch. Otolaryngol. Head Neck Surg. **118**, 1069–1076.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**, 2882–2897.

Engdahl, B., and Kemp, D. T. (**1996**). "The effect of noise exposure on the details of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **99**, 1573–1587.

Gaskill, S. A., and Brown, A. M. (**1993**). "Comparing the level of the acoustic distortion product $2f_1-f_2$ with behavioural threshold audiograms from normal-hearing and hearing-impaired ears," Br. J. Audiol. **27**, 397–407.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

Gorga, M. P., Neely, S. T., Bergman, B., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). "Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses," J. Acoust. Soc. Am. **93**, 2050–2060.

He, N., and Schmiedt, R. A. (**1993**). "Fine structure of the $2f_1-f_2$ acoustic distortion product: Changes with primary level," J. Acoust. Soc. Am. **94**, 2659–2669.

Heitmann, J., Waldmann, B., and Plinkert, P. K. (**1996**). "Limitations in the use of distortion product otoacoustic emissions in objective audiometry as the result of fine structure," Eur. Arch. Otorhinolaryngol. **253**, 167–171.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Kemp, D. T. (**1979a**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Oto-Rhino-Laryngol. **224**, 37–45.

Kemp, D. T. (**1979b**). "The evoked cochlear mechanical response and the auditory microstructure—Evidence for a new element in cochlear mechanics," Scand. Audiol. Suppl. **9**, 35–47.

Knight, R. D., and Kemp, D. T. (**2000**). "Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1, f_2 area study," J. Acoust. Soc. Am. **107**, 457–473.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**). "Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears," J. Acoust. Soc. Am. **111**, 1800–1809.

Long, G. R. (**1984**). "The microstructure of quiet and masked thresholds," Hear. Res. **15**, 73–87.

Lydolf, M. (**1999**). "The thresholds of hearing and contours of equal loudness," Ph.D. dissertation, Aalborg University.

Mauermann, M., Long, G. R., and Kollmeier, B. (**2004**). "Fine structure of hearing threshold and loudness perception," J. Acoust. Soc. Am. **116**, 1066–1080.

Mauermann, M., Uppenkamp, S., and Kollmeier, B. (**1997a**). "Zusammenhang zwischen unterschiedlichen otoakustischen Emissionen und deren Relation zur Ruhehörschwelle" ("Correlation between different types of otoacoustic emissions and their relation to the hearing threshold"), in *Fortschritte der Akustik—DAGA 97*, edited by P. Wille (DEGA e.V., Oldenburg), pp. 242–243.

Mauermann, M., Uppenkamp, S., and Kollmeier, B. (**1997b**). "Periodizität und Pegelabhängigkeit der spektralen Feinstruktur von Verzerrungsprodukt-Emissionen" ("Periodicity and dependence on level of the distortion product otoacoustic emission spectral fine-structure"), Audiol. Akustik **36**(2), 92–104.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999a**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1," J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999b**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Moore, B. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, USA).

Rao, A., Long, G. R., Narayan, S., and Dhar, S. (**1996**). "Changes in the temporal characteristics of TEOAE and the fine structure of DPOAEs with aspirin consumption," 19th ARO Midwinter Research Meeting, Abstract, p. 27.

Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**(4), 333–334.

Schloth, E. (**1983**). "Relation between spectral composition of spontaneous otoacoustic emissions and fine structure of threshold in quiet," Acustica **53**, 250–256.

Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P.,

and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.

Sutton, L. A., Lonsbury-Martin, B. L., Martin, G. K., and Whitehead, M. L. (**1994**). "Sensitivity of distortion-product otoacoustic emissions in humans to tonal overexposure: time course of recovery and effects of lowering L_2," Hear. Res. **75**, 161–174.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L_1, L_2 space," J. Acoust. Soc. Am. **97**, 2359–2377.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different oto-acoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

Low-level otoacoustic emissions may predict susceptibility to noise-induced hearing loss[a)]

Judi A. Lapsley Miller,[b)] Lynne Marshall, Laurie M. Heller,[c)] and Linda M. Hughes
Naval Submarine Medical Research Laboratory, Groton, Connecticut 06349-5900

(Received 6 October 2005; revised 11 April 2006; accepted 19 April 2006)

In a longitudinal study with 338 volunteers, audiometric thresholds and otoacoustic emissions were measured before and after 6 months of noise exposure on an aircraft carrier. While the average amplitudes of the otoacoustic emissions decreased significantly, the average audiometric thresholds did not change. Furthermore, there were no significant correlations between changes in audiometric thresholds and changes in otoacoustic emissions. Changes in transient-evoked otoacoustic emissions and distortion-product otoacoustic emissions were moderately correlated. Eighteen ears acquired permanent audiometric threshold shifts. Only one-third of those ears showed significant otoacoustic emission shifts that mirrored their permanent threshold shifts. A Bayesian analysis indicated that permanent threshold shift status following a deployment was predicted by baseline low-level or absent otoacoustic emissions. The best predictor was transient-evoked otoacoustic emission amplitude in the 4-kHz half-octave frequency band, with risk increasing more than sixfold from approximately 3% to 20% as the emission amplitude decreased. It is possible that the otoacoustic emissions indicated noise-induced changes in the inner ear, undetected by audiometric tests. Otoacoustic emissions may therefore be a diagnostic predictor for noise-induced-hearing-loss risk. [DOI: 10.1121/1.2204437]

PACS number(s): 43.64.Jb, 43.64.Wn [BLM] Pages: 280–296

I. INTRODUCTION

Evoked otoacoustic emissions (OAEs) are sounds produced by the inner ear in response to acoustic stimulation (Kemp, 1978). These sounds can be measured in the ear canal with a low-noise microphone. OAEs are thought to be generated by the outer hair cells (OHCs), which are susceptible to noise damage (e.g., Liberman *et al.*, 1986; Nordmann *et al.*, 2000; Rask-Andersen *et al.*, 2000). Diminished OAE amplitudes may be an early warning sign of incipient noise-induced hearing loss (NIHL), and therefore they may have a role to play in hearing-conservation programs.

Cross-sectional studies have shown that OAEs are sensitive indicators of permanent noise-induced damage to the inner ear in groups of noise-exposed people, with lower OAE levels associated with higher audiometric thresholds (e.g., LePage and Murray, 1993; LePage *et al.*, 1993; LePage and Murray, 1998; Desai *et al.*, 1999; Mansfied *et al.*, 1999; Attias *et al.*, 2001). Furthermore, noise-exposed people tend to have lower OAEs than people with similar audiometric thresholds but no noise exposure (Bicciolo *et al.*, 1993; LePage and Murray, 1993; LePage *et al.*, 1993; Murray and LePage 1993; Attias *et al.*, 1995, 1998; Xu *et al.*, 1998; Desai *et al.*, 1999; Attias *et al.*, 2001).[1] This finding has led to the hypothesis that, in individuals, OAEs may decrease prior to changes in audiometric thresholds. However, a cross-sectional design means the purported progression of changes in OAEs due to noise exposure and the relationship to changes in audiometric thresholds cannot easily be demonstrated in individual ears.

Longitudinal studies have shown that permanent changes in OAEs and permanent changes in audiometric thresholds do not necessarily occur together, both for groups of noise-exposed people and for noise-exposed individuals (Engdahl *et al.*, 1996; Murray *et al.*, 1998; Murray and LePage, 2002; Lapsley Miller *et al.*, 2004; Konopka *et al.*, 2005; Seixas *et al.*, 2005a, b). Typically, group changes in OAEs are seen, but often there are no group changes in audiometric thresholds. Studies that have considered changes in individual ears have also found that permanent threshold shifts (PTSs) do not necessarily correlate with changes in OAEs (Murray *et al.*, 1998; Murray and LePage, 2002; Lapsley Miller *et al.*, 2004). The actual progression of OAE changes and hearing loss in individuals has not been documented to date, and the existing data are ambiguous and inconsistent, partly for methodological reasons. These reasons include (a) noise exposures that were not severe enough to permanently elevate audiometric thresholds; (b) study durations that were not long enough to measure slowly progressing hearing loss; (c) OAE stimulus levels that were too high to optimally detect OAE changes (Sutton *et al.*, 1994; Marshall and Heller, 1998; Marshall *et al.*, 2001); (d) difficulty getting volunteers who had not been recently exposed to noise [i.e., baseline measurements were contaminated by temporary threshold shifts (TTS)]; (e) difficulty getting an appropriate age-matched and sex-matched control group; (f) getting volunteers without previous noise exposure; (g) separating out the effects of aging and NIHL; and (h) achieving

[a)]Preliminary results have been reported at the Association for Research in Otolaryngology Midwinter Meeting, St. Petersburg Beach, FL, January 2002, and the International Military Noise Conference, Baltimore, MD, April 2001.

[b)]Electronic mail: judi@psychophysics.org
[c)]Now at the Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI 02912.

sufficiently low test-retest variability (especially in field settings) to enable small changes in audiometric thresholds and OAEs to be detected.

It is unclear whether diminished OAEs are predictive of eventual NIHL, especially within an individual. The ideal study would follow a large number of volunteers, many of whom would eventually get PTS, over a period of years. Many subjects are needed because the incidence of NIHL is low in any one year, even in severely noise-exposed populations. To date, no one has amassed enough PTS cases to identify the best predictors. This question can be addressed in a more limited way by testing two points in time (before and after a particular noise exposure) to determine whether those with low amplitude OAEs on the preexposure test are more at risk for NIHL as measured postexposure. While this is not as desirable as a long-term multi-measurement longitudinal study because it does not provide information about why the OAE is at a low level, such a study can provide some information about which OAE parameters and properties seem to be the most predictive of PTS. In most populations, NIHL is a gradual process, and age can be a confounding factor, so studying this issue is more easily accomplished in a young population exposed to high levels of noise. One such population is the crew of an aircraft carrier.

An aircraft carrier, especially during flight operations, is one of the noisiest working environments known (Yankaskas, 1999; Yankaskas and Shaw, 1999). This environment puts sailors at risk for NIHL because even when using hearing protection as recommended, noise dosages still can exceed risk limits.[2] Naval hearing-conservation regulations mandate single hearing protection when noise levels exceed 84 dBA or impulse noise exceeds 140 dB pSPL, and double hearing protection (earplugs plus muffs or cranial helmets) when levels exceed 104 dBA (Navy Occupational Health and Safety Program, 1999). Double hearing protection ideally can provide attenuation up to 30 dB, but cannot provide sufficient attenuation to remove the risk of NIHL in the extreme noise levels present on an aircraft carrier. Furthermore, unlike many noise-hazardous industrial environments, there may be no truly quiet time for ears to recover from these shipboard exposures. This also implies that damage-risk criteria, which assume a daily quiet recovery time (Passchier-Vermeer, 1993), may not apply to this population. Poor hearing protection usage (Bjorn *et al.*, 2005), coupled with very high noise levels and with little quiet time for recovery, means sailors on aircraft carriers are at high risk for noise-induced hearing loss.

The main aim of the present study was to assess changes in audiometric thresholds and OAEs in sailors after 6 months of hazardous noise exposure on an aircraft carrier. A modified test battery was used, based on earlier studies (Sutton *et al.*, 1994; Kummer *et al.*, 1998; Marshall and Heller, 1998; Lapsley Miller *et al.*, 2004), which indicated that lower-level OAE stimuli were more sensitive to NIHL. The hypotheses were that (a) group average audiometric thresholds would increase (worsen), and group average OAE amplitudes would decrease (worsen); (b) individual cases of noise-induced PTS would be associated with significant emission shifts (SESs), but there would be more sailors with SESs than PTSs, and (c) ears with low-level or absent OAEs at predeployment testing would be more likely to show PTS at postdeployment testing.

II. MATERIALS AND METHODS

A. Volunteers

Audiometric thresholds and OAEs were measured in 338 sailors (35 women, median age 22 years, range 18 to 46 years; 303 men, median age 22 years, range 18 to 41 years) before and after 9 months on a Nimitz-class aircraft carrier, including 6 months at sea. Approximately 47% of the sailors were from the Air Department, who worked around aircraft and their launch and recovery mechanisms on or below the flight deck, as well as in the hangar bays; 19% were from the Engineering Department, who worked in various locations below deck; 32% were from the Reactor Department, who worked in the machinery spaces; and 2% were from other departments. Additionally, a control group of 28 volunteers (sailors and research staff; 8 women, median age 31 years, range 20 to 53 years; 20 men, median age 26 years, range 20 to 47 years) completed an identical protocol with no intervening noise exposure between pre- and posttesting. The posttest for the control group occurred 20 min to 2 days after the predeployment testing. A suitable age- and sex-matched control group that could be noise-free over 9 months was not available.

B. Stimuli and equipment

Pure-tone audiograms were obtained at frequencies 0.5, 1, 2, 3, 4, and 6 kHz using a modified Hughson-Westlake procedure (with the usual 10-dB descending and 5-dB ascending steps). Four microprocessor-controlled audiometers were used (three Tremetrics RA400 and one RA500), all with TDH 39 earphones and MX-41/AR cushions, and one Beltone 120 manual audiometer, with TDH 50P earphones and MX-41/AR cushions. For the most part, the Tremetrics audiometers were used in automated mode. Middle-ear pressures were estimated from the peak of an immittance tympanogram with a 226-Hz tone using a Grason Stadler GSI 33 version 2 analyzer at a sweep speed of 12.5 daPa/s to minimize hysteresis.

OAEs were measured with the ILO292 Echoport system (Otodynamics Ltd., England), using the distortion-product OAE (DPOAE) probe. It was covered by an acoustic-immittance probe tip, which had been enlarged using a grinding tool, to allow better placement and manipulation in the ear canal.

C. OAE test battery

Transient-evoked OAEs (TEOAEs) evoked with a 74 dB pSPL click (abbreviated herein to $TEOAE_{74}$) were measured in nonlinear mode, where responses to three clicks at one polarity and one click 9.5 dB higher with opposite polarity were added together to reduce linear artifact from the stimulus (Bray, 1989). TEOAEs were collected and averaged until 260 low-noise averages were obtained.[3] The results were windowed, filtered, and analyzed into half-octave

bands [which is optimal according to Marshall and Heller (1996)]. At predeployment testing, every attempt was made to get a flat stimulus spectrum during calibration. At postdeployment testing, every attempt was made to get the same stimulus pattern during calibration as in predeployment testing.

In order of presentation, DPOAEs were measured with stimulus levels $L_1/L_2=57/45$, 59/50, 61/55, and 65/45 dB SPL (abbreviated herein to $DP_{57/45}$, $DP_{59/50}$, $DP_{61/55}$, and $DP_{65/45}$). The first three levels specified a DPOAE I/O function (Kummer *et al.*, 1998); the fourth level is sensitive to TTS (Marshall *et al.*, 2001). For all stimulus levels, the f_2/f_1 ratio was 1.22, with $f_2=1.8$, 2.0, 2.2, 2.5, 2.8, 3.2, 3.6, 4.0, and 4.5 kHz.[4]

Individual in-the-ear calibration was used for both TEOAEs and DPOAEs.

D. Procedure

All OAE and audiometric testing occurred in single-walled sound-attenuating booths. OAE testing was done pierside, near the ship, in a mobile test van. Most audiometric testing was done in the medical department on the ship, which was docked at the pier, but some testing was done in the mobile van to expedite testing as many volunteers as possible. The left ear was tested first. At predeployment testing, volunteers were screened for clear ear canals (cerumen was removed if present), audiometric thresholds of ≤25 dB HL from 0.5 to 3 kHz and ≤30 dB HL at 4 kHz, and peak immittance within the range of ±50 daPa atmospheric pressure, with grossly normal amplitude, slope, and smoothness of the tympanogram. 85% of the ears had normal audiometric thresholds, using a strict criterion of ≤15 dB HL at 1 to 4 kHz. If the definition of normal is relaxed to include thresholds at 20 dB HL (which is often used for hearing screening using OAEs, e.g., Gorga *et al.*, 1993), 98% of the ears had normal thresholds. There was a greater incidence of slight hearing losses at higher frequencies. At 1 kHz, 98% had thresholds ≤15 dB HL and 2% had 20 dB HL thresholds. At 2 kHz, 97% had thresholds ≤15 dB HL and 2% had 20 dB HL thresholds. At 3 kHz, 94% had thresholds ≤15 dB HL, 5% had 20 dB HL thresholds, and 1% had 25 dB HL thresholds. At 4 kHz, 90% had thresholds ≤15 dB HL, 7% had 20 dB HL thresholds, 1% had 25 dB HL thresholds, and 1% had 30 dB HL thresholds. These percentages were similar for the group that got PTS during the deployment and the group that did not. Volunteers who did not meet screening criteria did not continue in the study.

At postdeployment testing, volunteers were asked to complete a detailed noise history covering the previous 9 months. They then underwent the same testing as for predeployment testing. At that time, they were screened only for peak immittance within ±50 daPa atmospheric pressure, and in all cases the tympanometric peak was within this range shortly before OAE testing.

During postdeployment data collection, the Navy hearing-conservation significant-threshold-shift (STS) criteria at that time of the study (a shift of at least 15 dB at 1, 2, 3, or 4 kHz, or an average shift of at least 10 dB at 2, 3, and 4 kHz) were used to detect changes in audiometric thresholds in individuals (Navy Occupational Health and Safety Program, 1999). STSs were confirmed with manual audiometry, immediately if possible, or as soon as possible thereafter (up to 9 days). If the volunteer had been noise-free and the STS was confirmed, it was considered a PTS. If the volunteer had recently been exposed to noise, they were asked to return for a 14-h noise-free follow-up to see if their STS was permanent or temporary.

E. Data definitions, cleaning, and reduction

Because the testing was conducted in a military working environment, some of the data were unavoidably affected by background ambient and electrical noise, despite testing in a sound-attenuating booth calibrated to ANSI standards (ANSI, 1991) and using a power-line conditioner. Data-collection logistics meant that much of the OAE testing was done using a hook-up to the naval base's mains power supply rather than batteries. Once the problem was identified, the equipment was run on battery as much as possible. The short testing time available for each volunteer meant that it was not always possible to obtain clean data. Data points and/or test conditions contaminated with extreme stimulus levels, bad calibrations, high noise levels, large differences in noise level between tests, or many unexplained outliers were removed from the data set in an objective fashion, using the same elimination rules across the entire dataset of all volunteers.[5]

An OAE was considered present if, for $TEOAE_{74}$, the amplitude was greater than 0 dB SNR above the noise level, and, for DPOAEs, the amplitude was greater than the noise level, which was defined as two standard deviations above the noise floor.

For the remaining "good" frequencies and levels, the percentage of measurable OAEs was calculated (i.e., those OAE amplitudes with good SNR) and any frequencies where less than 70% of OAEs were measurable were dropped ($TEOAE_{74}$ at 0.7 and 5.7 kHz).

Although the actual criteria used were liberal at each stage of screening, a large amount of data was rendered unusable. Losing $DP_{61/55}$ and the two DPOAE frequencies (see footnote 5) meant that planned analyses involving DPOAE input-output functions and half-octave analyzed DPOAEs had to be dropped. The remaining test conditions were $TEOAE_{74}$, which was analyzed into half-octave bands centered at 1.0, 1.4, 2.0, 2.8, and 4.0 kHz, and $DP_{65/45}$, $DP_{59/50}$, and $DP_{57/45}$ at 1.8, 2.0, 2.5, 2.8, 3.2, 3.6, and 4.0 kHz.

For some cases, a predeployment OAE was measurable, but the postdeployment OAE was below the noise level. These postdeployment OAE amplitudes with bad SNR were substituted with the noise level in some circumstances [similarly to Lapsley Miller *et al.* (2004)]. This occurred only if the noise level was below the predeployment OAE amplitude (otherwise, a high noise level may masquerade as an increase in OAE amplitude). This enabled the use of more data, such as the important cases where a normal OAE at predeployment testing disappeared below the noise level by postde-

TABLE I. The number of ears in each group that contributed to each analysis, listed by the section number. The number in parentheses is the total number of volunteers in the group. The numbers varied at each test frequency, OAE level, and OAE type because all good data were used. The exception was for the ANOVAs where volunteers were required to have complete OAE data sets for both ears at 2, 3, and 4 kHz.

Analyses	No. of ears (volunteers)	Group	Notes
III A ANOVA	150 (75)	Noise	Ear was a factor in ANOVAs.
III B Correlations among changes in audiometric thresholds and changes in OAEs	169–338 (338)	Noise	The left and right ears were in separate analyses.
III C 1 Forming STS and SES criteria	33–56 (28)	Control	Ears were pooled across volunteers.
III C 2 Applying STS and SES criteria to noise-exposed and control groups	Noise: 473–675 (338) Control: 33–56 (28)	Noise and Control	Noise and control groups were analyzed separately. Within groups, ears were pooled across volunteers.
III C 3–4 Identifying and describing PTS cases	18 (15)	Noise	Both men and women were analyzed but no woman got PTS.
III C 4 Correlations between SES status and PTS status	PTS: 10–17 non-PTS: 473–572	Noise	PTS and non-PTS volunteers were in the same analysis. Ears were pooled across volunteers.
III C 4 Correlations among SESs for PTS ears	15 (12)	Noise	Three ears were not included due to missing data.
III D Susceptibility	PTS: 16–18 non-PTS: 524–559	Noise	Only data from the male volunteers were used. Ears were pooled.

ployment testing, with the caveat that true decreases in OAE amplitude may have been underestimated. For $TEOAE_{74}$, 6% of postdeployment measurements were replaced with the noise level. For DPOAEs, 5% of postdeployment measurements were replaced with the noise level.

For the susceptibility analyses, it was of interest to know if low or absent OAEs at predeployment increased the chance of PTS at postdeployment. Some of the analyses required estimating amplitudes for missing predeployment OAEs. To do this, the noise level was substituted for missing OAEs, providing the noise floor was not high. A noise level was considered acceptably low if it was within the tenth percentile of the corresponding OAE amplitude (not the noise floor) based on the group. Again, this process is conservative because it overestimates the actual amplitude of the OAE.

This research was conducted in compliance with all applicable federal regulations governing the protection of human subjects in research.

III. RESULTS

Table I provides a breakdown of the number of ears contributing to each analysis, and whether the ears were noise exposed or controls. The numbers varied at each test frequency, OAE level, and OAE type because all good data were used. The exception was for the ANOVAs, where data from 75 volunteers with complete data sets for both ears were used.

A. Group OAE and audiometric thresholds before and after noise exposure

The primary interest was to see if there were any changes in audiometric thresholds or OAEs between pre- and postdeployment tests. Of secondary interest was whether these changes differed across frequency, stimulus level (for DPOAEs), or ears. There were not enough female volunteers to group by sex.

Separate, repeated-measures ANOVAs were conducted on audiometric threshold, TEOAE, and DPOAE data for the

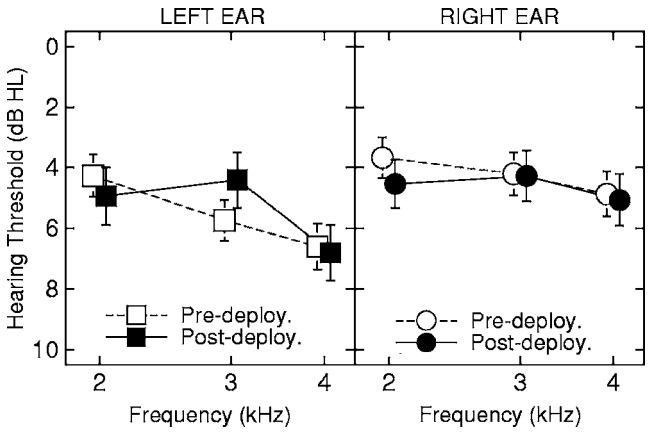


FIG. 1. Average group audiometric thresholds for left and right ears and pre- and postdeployment tests for the subgroup of 75 noise-exposed sailors with complete data sets used in the ANOVA. Error bars indicate one standard error of the mean. Frequency is plotted on a $\log_2$ scale. Data points are offset either side of the labeled frequency to aid interpretation.

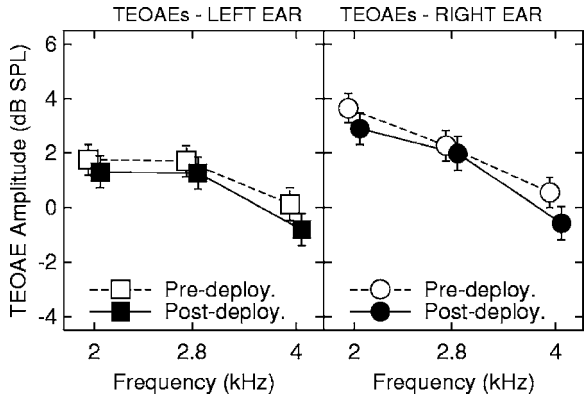


FIG. 2. Between pre- and postdeployment, average group TEOAE amplitudes significantly decreased by 1 dB at 4 kHz (combined over ears) for the 75 noise-exposed sailors with complete data sets used in the ANOVA. Left panel shows average group $TEOAE_{74}$ amplitudes for left ears; right panel shows average group $TEOAE_{74}$ amplitudes for right ears. Error bars indicate one standard error of the mean. Frequency is plotted on a $\log_2$ scale. Data points are offset either side of the labeled frequency to aid interpretation.

subgroup of 75 volunteers with "complete" data sets (8 women, median age 21 years; 67 men, median age 22 years). To maximize the number of volunteers with complete data, only the frequencies 2, 2.8 (or 3 for audiograms), and 4 kHz were included (i.e., there could be missing data at other frequencies). By selecting volunteers with complete data, a bias may have been introduced, because those volunteers with more missing data may have more noise-induced damage. However, by using complete data sets, comparisons across OAE stimulus types and frequencies could be more fairly made.

Figure 1 shows group average audiometric thresholds for left and right ears and pre- and postdeployment tests. A three-way, repeated-measures ANOVA was conducted for audiometric thresholds (test: pre- versus postdeployment; ear: left versus right; and frequency: 2, 3, and 4 kHz). There was no significant change in audiometric thresholds (main effect) between pre- and postdeployment tests ($F_{1,74}=0.05$, ns). There were, however, significant differences between ears ($F_{1,74}=5.82, p<0.05$) and across frequency ($F_{2,148}=3.68, p<0.05$). There was also a two-way interaction for test-by-frequency ($F_{2,148}=3.32, p<0.05$). Bonferroni *post hoc t*-test comparisons were used to establish which frequencies contributed to the test-by-frequency, two-way interaction. The familywise significance level was $p<0.05$, so, for three comparisons, $p<0.017$ was used. None were significant.

Figure 2 shows the group average TEOAE amplitudes for left and right ears and pre- and postdeployment tests. A three-way, repeated-measures ANOVA was conducted for $TEOAE_{74}$ amplitude (test: pre- versus postdeployment; ear: left versus right; and frequency: 2, 2.8, and 4 kHz). All three factors showed significant main effects. Particularly, there was a 0.66-dB decrease in $TEOAE_{74}$ amplitude between pre- and postdeployment testing ($F_{1,74}=12.3, p<0.05$). Ears also differed ($F_{1,74}=8.6, p<0.05$) as did frequency ($F_{2,148}=4.4, p<0.05$). There were two significant two-way interactions: test-by-frequency ($F_{2,148}=19.5, p<0.05$) and ear-by-frequency ($F_{2,148}=3.2, p<0.05$). Bonferroni *post hoc t*-test comparisons were used to establish which frequencies contributed to the test-by-frequency, two-way interaction. The familywise significance level was $p<0.05$, so, for three comparisons, $p<0.017$ was used. There was a significant 1.0-dB decrement in $TEOAE_{74}$ amplitude at 4 kHz.

Figure 3 shows the group average DPOAE amplitudes for each level, and pre- and postdeployment tests (ears combined). A four-way, repeated-measures ANOVA was conducted for DPOAE amplitude (test: pre- versus postdeployment; ear: left versus right, level: stimulus levels of 65/45, 59/50, and 57/45 dB SPL; and frequency: 2, 2.8, and 4 kHz). There was a 1.28-dB decrement in DPOAE amplitude between pre- and postdeployment testing ($F_{1,74}=27.4, p<0.05$). There were also main effects for level ($F_{2,148}=190.8, p<0.05$) and frequency ($F_{2,148}=24.2, p<0.05$) but not for ear ($F_{1,74}=0.08$,ns). There were three significant two-way interactions: test-by-level ($F_{2,148}=9.1, p<0.05$), ear-by-level ($F_{2,148}=10.1, p<0.05$), and level-by-frequency ($F_{4,296}=28.4, p<0.05$). Bonferroni *post hoc t*-test comparisons were used to establish which of the three levels contributed to the test-by-level, two-way interaction. The familywise significance level was $p<0.05$, so, for three comparisons, $p<0.017$ was used. Postdeployment DPOAE amplitudes for $DP_{59/50}$ and$DP_{57/45}$ were significantly lower than predeployment amplitudes (by 1.5 dB). None of the three- or four-way interactions were significant.

B. Changes in OAEs and audiometric thresholds: Correlations

The relationship between *changes* in OAEs and *changes* in audiometric thresholds was assessed using Pearson corre-

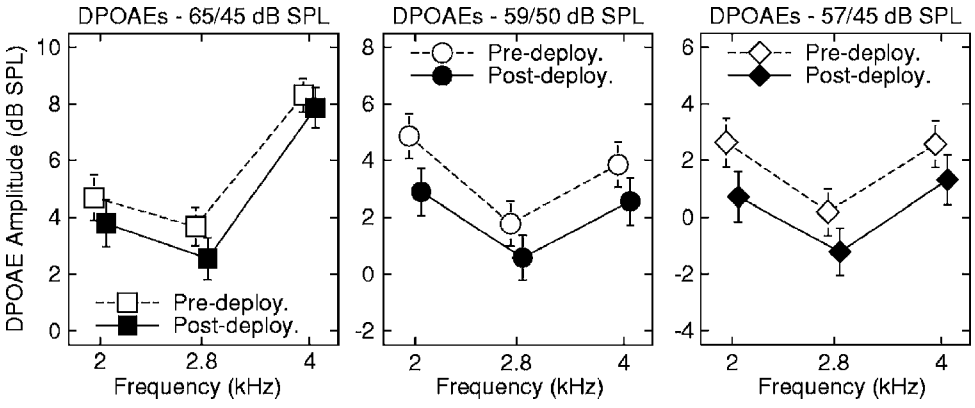


FIG. 3. Between pre- and postdeployment, average group DPOAE amplitudes for $DP_{59/50}$ and $DP_{57/45}$ significantly decreased by 1.5 dB for the 75 noise-exposed sailors with complete data sets used in the ANOVA. Left panel shows average group $DP_{65/45}$, middle panel shows average group $DP_{59/50}$, and right panel shows average group $DP_{57/45}$ amplitudes. Plots are averaged over ears. Error bars indicate one standard error of the mean. Frequency is plotted on a $\log_2$ scale. Data points are offset either side of the labeled frequency to aid interpretation.

lation coefficients for the arithmetic difference between pre- and postdeployment OAE amplitudes and audiometric thresholds, for all the valid data at every test, level, and frequency. The number of volunteers contributing to each correlation ranged from 169 to 338. Right and left ears were considered separately.

1. Correlations between changes in audiometric thresholds and changes in TEOAEs

There was no correlation greater than 0.22 at any frequency, and most were not statistically significant at $p < 0.05$.

2. Correlations between changes in audiometric thresholds and changes in DPOAEs

There was no correlation greater than 0.19 at any frequency or stimulus level, and most were not statistically significant at $p < 0.05$.

3. Correlations between changes in DPOAEs and changes in TEOAEs

Generally, the strongest correlations were from 0.5 to 0.6 (statistically significant at $p < 0.05$), which occurred for same-frequency combinations and for $TEOAE_{74}$ frequencies 0.5 to 1 octave lower than the DPOAE frequency. This may reflect separate correlations with the two DPOAE components, with the reflection source originating from the $2f_1$-f_2 place and the distortion-source originating from near the f_2 place (both with a frequency of $2f_1$-f_2). Sometimes the correlation was highest when the $2f_1$-f_2 frequency was in the same $TEOAE_{74}$ half-octave and sometimes when the f_2 frequency was in the same $TEOAE_{74}$ half-octave, but the differences were not great, and many cases showed similar correlations across two or three $TEOAE_{74}$ half-octaves. Correlations between $TEOAE_{74}$ and DPOAEs showed no consistent pattern across DPOAE stimulus level. In general, $DP_{57/45}$ and $DP_{59/50}$ showed stronger correlations with $TEOAE_{74}$ than $DP_{65/45}$. The strongest correlations tended to be for $TEOAE_{74}$ at 1 or 1.4 kHz with DPOAEs at 1.8 kHz (the lowest DPOAE frequency), regardless of stimulus level.

There was no evidence that changes in OAEs were correlated with changes in audiometric thresholds. Although DPOAEs and TEOAEs did tend to shift together, the correlation was only moderate. Furthermore, there was little-to-no correlation between left and right ears for either audiometric thresholds or OAEs, even for the same test type and frequency. The lack of correlation between changes in audiometric threshold and changes in OAEs may be due to the small number of ears that actually had *significant* changes in audiometric threshold or OAEs—the larger number of non-PTS ears, where changes are just due to test-retest variability, may have swamped any effect. This was investigated further by considering the OAEs of the PTS ears.

C. Association between changes in audiometric threshold and changes in OAEs in individuals

Note that a +STS and a −STS indicate a worsening of audiometric thresholds and OAE amplitude, respectively, whereas a −STS and a +STS indicate an improvement of audiometric thresholds and OAE amplitude, respectively. The plus or minus sign comes from subtracting the predeployment test result from the postdeployment test result.

TABLE II. Significant threshold shift (STS) criteria calculated from the control group (56 ears). Shown for each audiometric frequency and some averaged frequency combinations are the group average threshold shifts (postdeployment-predeployment), the standard error of measurement (SE_{MEAS}), the resulting STS criteria, and the Navy STS criteria, which was used to diagnose PTS onsite.

Frequency (kHz)	Average shift (dB)	ΔSE_{MEAS} (dB)	STS (dB)	Navy STS (dB)
1	−1.3	2.8	15	15
2	−0.2	2.1	15	15
3	−1.6	3.4	15	15
4	−0.2	3.8	20	15
6	−1.1	5.5	25	…
Mean 2 and 3	−0.9	2.4	10	…
Mean 3 and 4	−0.5	2.8	12.5	…
Mean 2, 3 and 4	−0.5	2.2	8.3	10

1. Standard error of measurement used to define individual significant shifts

Criteria based on the SE_{MEAS} were used to detect significant audiometric thresholds and OAE shifts between pre- and postdeployment tests (similarly to Lapsley Miller and Marshall, 2001; Marshall *et al.* 2001, 2002; Lapsley Miller *et al.*, 2004),[6] for each frequency of interest from the group of 28 control volunteers, who had received no intervening noise exposure (see Table I; 56 ears were included for audiometric thresholds and between 29 to 43 ears for OAEs, because only OAEs with good SNR at both pre- and postdeployment were used).[7]

Tables II and III summarize SE_{MEAS} and the resulting STS and SES criteria, respectively, for each frequency of interest.

2. STS and SES cases identified using derived criteria

Table IV shows the percentage of STSs detected and Table V shows the percentage of SESs detected when applying the derived criteria to the data set of 338 volunteers. The percentages are relative to the amount of good data (i.e., after removing the cases with poor calibrations, etc., as described earlier). Virtually no STSs were detected in the control group, but more were detected in the noise-exposed group. Nearly as many −STSs (improvement of audiometric thresholds) were seen as +STSs (deterioration of audiometric thresholds), except for the averaged frequencies of 2 and 3 kHz, and 2, 3, and 4 kHz. This indicated that the test-retest variability was too great to reliably see noise-induced audiometric-threshold shifts at single frequencies, and it was only when a wider frequency band was examined that significant noise-induced changes were apparent. The STS criterion for 6 kHz was deemed too large (at 25 dB) to reliably detect shifts at this frequency and was therefore not used. For subsequent analyses, data for ears with STS were used only

TABLE III. $DP_{57/45}$, $DP_{59/50}$, $DP_{65/45}$, and $TEOAE_{74}$ significant emission shift (SES) criteria. Shown for each single DPOAE frequency and half-octave TEOAE band are the number of control-group ears going into the calculation, SE_{MEAS}, and the resulting SES criterion.

OAE type	Frequency (kHz)	Ears	SE_{MEAS} (dB)	SES (dB)
$DP_{57/45}$	1.8	41	2.3	6.9
	2.0	33	2.7	8.0
	2.5	39	1.9	5.7
	2.8	34	2.6	7.8
	3.2	37	2.1	6.3
	3.6	39	2.7	8.2
	4.0	38	1.7	5.2
$DP_{59/50}$	1.8	40	2.2	6.5
	2.0	39	2.9	8.5
	2.5	38	2.3	7.0
	2.8	38	2.6	7.9
	3.2	36	2.5	7.5
	3.6	39	2.2	6.5
	4.0	35	2.0	6.1
$DP_{65/45}$	1.8	42	2.0	6.1
	2.0	42	1.9	5.6
	2.5	39	1.5	4.6
	2.8	40	1.7	5.1
	3.2	43	1.8	5.4
	3.6	43	1.6	4.7
	4.0	42	1.9	5.7
$TEOAE_{74}$	1.0	40	2.5	7.5
	1.4	46	2.0	6.1
	2.0	40	1.1	3.2
	2.8	39	1.3	3.8
	4.0	35	1.2	3.7

if the STS was confirmed to be PTS (with a repeat audiogram showing the STS was maintained, noise-free for at least 14 h prior to testing, and a noise history consistent with hazardous noise exposure). The data sets for ears with no STS were used for comparison with the PTS ears.

TABLE IV. Percentage of significant threshold shifts (STSs) detected with the derived criteria based on the SE_{MEAS}, for the noise-exposed group (n =338) and for the control group (n=28). Shown are the percentages of good data, the percentages of +STSs (deterioration in audiometric thresholds), relative to the good data, and the percentage of −STSs (improvement in audiometric thresholds), relative to the good data.

	Noise-exposed ears			Control ears		
Frequency (kHz)	Good data (%)	+STS (%)	−STS (%)	Good data (%)	+STS (%)	−STS (%)
1.0	100	1	1	100	0	0
2.0	100	2	0	100	0	0
3.0	100	1	1	100	0	0
4.0	100	1	1	100	2	0
6.0	100	1	1	100	2	0
Mean 2 and 3	100	3	1	100	0	0
Mean 3 and 4	100	2	2	100	0	0
Mean 2, 3, and 4	100	5	1	100	0	0

TABLE V. Percentage of significant emission shifts (SESs) for each OAE type detected with the derived criteria based on the SE_{MEAS}, for the noise-exposed group (n=338), and for the control group (n=28). Shown are the percentages of good data, the percentages of −SESs (decrease in OAE amplitude), relative to the good data, and the percentage of +SESs (increase in OAE amplitude), relative to the good data.

		Noise-exposed ears			Control ears		
Test	Frequency (kHz)	Good data (%)	−SES (%)	+SES (%)	Good data (%)	−SES (%)	+SES (%)
$DP_{57/45}$	1.8	75	11	2	72	5	2
	2.0	72	6	2	64	3	3
	2.5	70	12	3	71	2	0
	2.8	73	5	2	62	3	6
	3.2	74	8	3	69	5	0
	3.6	76	3	2	72	2	0
	4.0	80	13	6	69	0	8
$DP_{59/50}$	1.8	79	10	3	72	5	0
	2.0	79	4	2	74	0	2
	2.5	75	6	2	69	5	3
	2.8	77	5	1	69	0	3
	3.2	79	6	2	67	0	5
	3.6	80	7	4	69	5	3
	4.0	83	9	4	62	3	3
$DP_{65/45}$	1.8	83	7	2	74	0	5
	2.0	83	8	4	74	2	0
	2.5	82	12	4	72	7	0
	2.8	83	11	3	69	0	3
	3.2	86	7	3	74	2	0
	3.6	86	10	6	76	2	0
	4.0	88	6	4	74	5	2
$TEOAE_{74}$	1.0	85	5	1	75	5	2
	1.4	88	5	2	82	2	4
	2.0	81	12	7	75	5	0
	2.8	75	8	4	70	0	3
	4.0	71	12	3	68	0	5

OAEs, on the other hand, showed more evidence of noise-induced changes. Table V summarizes the percentage of SESs found for each OAE type, level, and frequency, for both the noise-exposed group and for the control group. Shown is the percentage of good data relative to all data (i.e., the percentage of pre- and postdeployment measurement pairs that could be used to calculate differences), the percentage of −SESs (deterioration of OAEs), and the percentage of +SESs (improvement of OAEs), both relative to the amount of good data. There was little difference between the percentages of +SESs for noise-exposed and control groups, indicating that many increments were just due to variability (even though it is theoretically possible for OAEs to increase in amplitude with noise damage). There were, however, more −SESs for the noise-exposed group, compared to the control group, indicating the effects of noise exposure on the OAEs. There does not appear to be any indication that higher frequencies showed more OAE changes, as might have been expected.

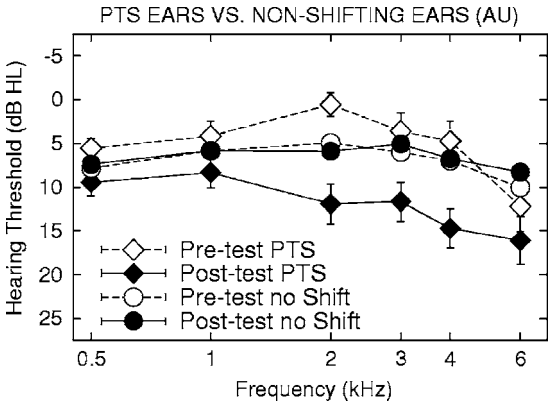


FIG. 4. Average audiometric thresholds for PTS ears ($n=18$) and non-PTS ears ($n=559$) for the volunteers with no shifts in either ear. Audiometric thresholds for PTS ears were, as expected, worse at postdeployment, with the biggest increases in threshold at 2 to 4 kHz. Error bars indicate one standard error of the mean and are smaller than the symbols for the non-PTS ears.

3. Permanent threshold shifts

Fifteen sailors (18 ears) were diagnosed with PTS (3 bilateral, 10 left ears, 2 right ears); this is 4.4% of sailors who were tested postdeployment. Their median age was 24 years (compared with 22 years for the group of sailors with no shifts), median length of service was 3.5 years (compared with 3 years for the group of sailors with no shifts), and all were male. Figure 4 shows the average audiograms for pre- and postdeployment for the 18 PTS ears and for the group of sailors with no shifts in either ear (599 ears). The PTS ears had slightly better audiometric thresholds at predeployment, but increased thresholds at postdeployment, especially in the 2- to 4-kHz range. Two volunteers with PTS at 4 kHz were diagnosed with an STS using the Navy criterion of 15 dB, which is less strict than the derived criterion of 20 dB. Both, however, had follow-up audiograms confirming the shift, so they were included in the PTS group. Most PTSs were on the order of 15 dB, with the largest being 25 dB.

From the noise histories of the volunteers with PTS, there were no particular commonalities among the noise exposures, other than there being proportionally more sailors with PTS from the reactor department (11%), compared with the air (4%) and engineering (3%) departments.

Thirteen of the 15 sailors with PTS had incomplete data sets, so only the data for two sailors with PTS contributed to the subset of 75 sailors used for the ANOVA reported in the previous section. The data sets tended to be incomplete due to the absence of OAEs rather than measurement problems. This tendency for absent OAEs among the PTS cases is examined later.

4. Pattern of SESs for the 18 PTS ears

For $TEOAE_{74}$, 11 ears showed no changes (though there were some missing data for seven ears), five ears showed changes in one half-octave band (though only one ear had no missing data), one ear showed changes across more than one band and also had some missing data, and one ear had essentially no measurable OAEs. For DPOAEs, generalizing over level, seven ears had no OAE changes at any level, four ears had OAE changes at all levels (though not necessarily at the same frequency), three ears had some OAE changes but not at all levels, one ear had essentially no OAEs, and three ears had no usable data at any frequency. There were many missing data (due to bad calibrations, etc.), so potentially some OAE shifts were not detected. Excluding the cases where all data were missing due to measurement problems or where OAEs were absent at most frequencies at predeployment, 31% of PTS ears showed at least one SES in TEOAEs, and 50% of PTS ears showed at least one SES in DPOAEs (across levels). However, many of these SESs were improvements in OAE amplitude, many were not in the same frequency band as the PTS, and there was only some consistency between changes in TEOAEs and changes in DPOAEs. Consistency among changes within DPOAE levels was also not high.

The nonparametric phi coefficient (Siegel, 1956) was used as a measure of association for the 2×2 cross-tabulated tables of PTS ears versus non-PTS ears (at any frequency) versus ears with and without SESs (at any frequency) to determine whether PTSs and SESs tended to occur together (PTS ears: $n=10$ to 17; non-PTS ears: $n=473$ to 572). The phi coefficient can be interpreted similarly to a correlation coefficient and can be used for small data sets. Because of the small number of PTS cases and the large amount of missing data, an ear was considered to have an SES if there was an SES at any frequency within an OAE type and level. Ears with no measurable SESs (due to missing data) were excluded. There was no correlation between PTS status and SES status for any OAE type.

To investigate if there was an association between TEOAE SESs and DPOAE SESs in the PTS ears, each PTS ear (excluding the three cases with extensive missing data) was flagged as having either (a) no SESs at any frequency or (b) at least one SES at any frequency, for the conditions $TEOAE_{74}$, $DP_{57/45}$, $DP_{59/50}$, and $DP_{65/45}$. The phi coefficient was again used as a measure of association for the resulting six 2×2 cross-tabulated tables. Phi was 0.58 for $TEOAE_{74}$ versus $DP_{57/45}$; 0.70 for $TEOAE_{74}$ versus $DP_{59/50}$; and 0.87 for $TEOAE_{74}$ versus $DP_{65/45}$. It is fair to say that when there was an SES for one OAE type, then there was often an SES for the other OAE type. Similarly, among the DPOAE levels, the association between levels $DP_{57/45}$ and $DP_{59/50}$ was 0.80, between $DP_{57/45}$ and $DP_{65/45}$ was 0.87, and between $DP_{59/50}$ and $DP_{65/45}$ was 0.93. DPOAE SESs and TEOAE SESs were associated with each other in the PTS ears, indicating that the SESs for the PTS ears were unlikely to be due to random fluctuations. However, this does not indicate that these SESs are related to the PTSs; it merely reinforces the finding from Sec. III B 3 that SESs across OAE type are related.

5. Summary of changes in audiometric thresholds and concomitant changes in OAEs

Although there is no compelling relationship between changes in audiometric thresholds and changes in OAEs, there is an association among changes in OAEs across OAE types, levels, and frequencies.

The number of PTS cases with low-level or absent OAEs was notable. If an OAE is already low level, it is

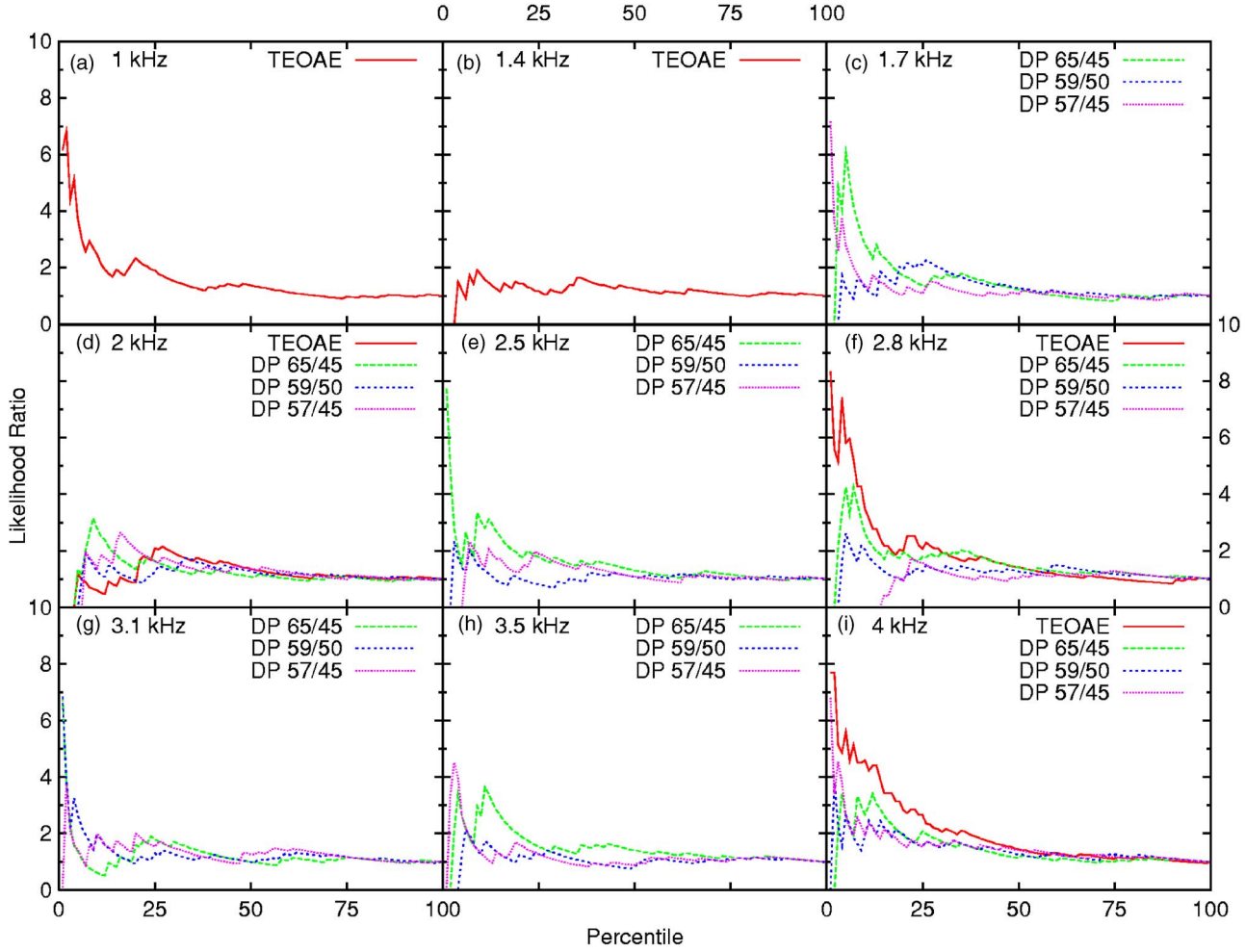


FIG. 5. (Color online) The likelihood ratio by percentile for each OAE test and frequency. For many OAE conditions, the likelihood of a PTS ear having an OAE level below the percentile criterion, compared with having an OAE level above the percentile criterion, increased as the percentile criterion (based on OAE amplitude) decreased. Not all OAE tests had measurements made at the same frequency; note that TEOAEs are half-octave band and DPOAEs are single-frequency measurements.

unlikely that further changes will be detected. A possible explanation is that noise damage prior to this study left many of these ears with subclinical damage, which makes these ears more likely to acquire hearing loss with further noise exposure. To investigate this theory, low-level and absent OAEs at predeployment were examined to see if they were predictive of PTS.

D. Predictors of susceptibility to PTS

Earlier observations suggested that low-level or absent OAEs were more likely among the PTS ears. PTS risk (defined by likelihood ratios and positive predictive values) was estimated as a function of predeployment OAE level, with OAE levels converted into percentiles to enable comparison across OAE types.[8] By considering all possible percentiles, a prediction of PTS risk for any OAE amplitude is possible. Because no female sailors got PTS, this analysis was restricted to data from the male sailors. The greatest number of ears with good data was used for each condition, so the number of ears varied across conditions. As the total number of ears and the number of PTS ears were not constant across conditions, some caution must be taken in interpreting the results, particularly in comparing the advantage of various stimuli in predicting susceptibility.

In medical diagnostics, the likelihood ratio is a ratio of two probabilities: the probability of a particular test result among patients with a condition to the probability of that particular test result among patients without the condition (Zhou *et al.*, 2002). Here, the likelihood ratio indicates the relative probability that a predeployment OAE amplitude was below a given percentile in the group of ears that subsequently got a PTS, relative to the same result in the group of ears that did not subsequently get a PTS. For instance, a likelihood ratio of 1 would indicate that a particular predeployment test result was equally likely to occur for ears that subsequently got PTS and ears that did not get PTS. A likelihood ratio of 4 for a particular test result indicates the result was four times more likely among ears that got PTS than ears that did not get PTS. The likelihood ratio does not take the base rate (*a priori* probability) of PTS into account.

A cutoff defined by an OAE percentile can be applied as a diagnostic criterion for PTS risk. This cutoff is independent of the actual condition (presence or absence of PTS), and it is

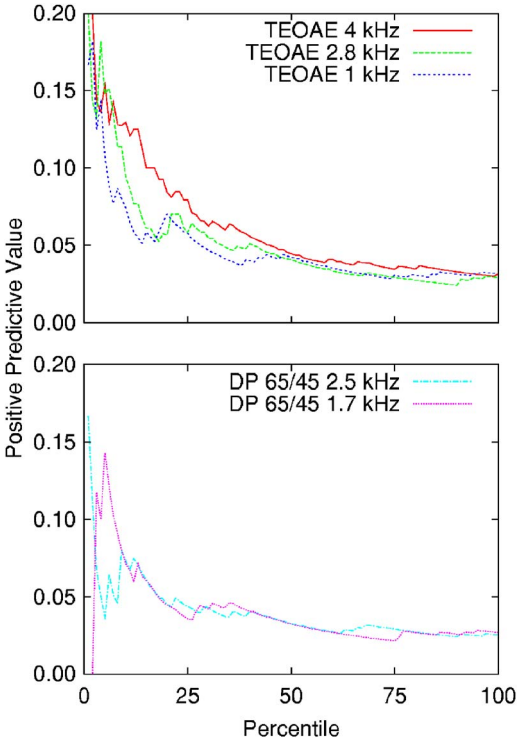


FIG. 6. (Color online) The positive predictive value (PPV) of a low-level OAE amplitude increased as the percentile criterion defining a low-level OAE amplitude decreased. PPV is the probability that a predeployment OAE amplitude was predictive of subsequent PTS. Shown are the PPVs, as a function of OAE percentile criterion, for the five OAE tests and frequencies that were most predictive of subsequent PTS. TEOAEs are shown in the top panel and DPOAEs in the bottom panel. These data should be viewed as indicative only; the jaggedness is caused by the small number of PTS cases contributing to the analysis. However, the general tendency is for TEOAEs to provide more predictive power than DPOAEs.

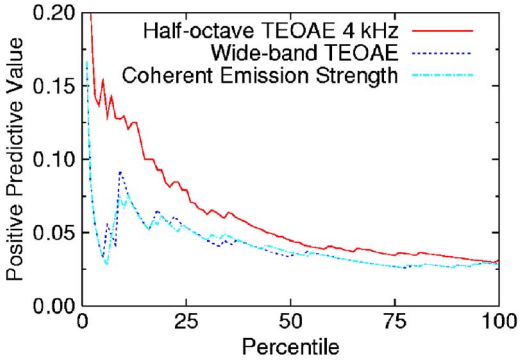


FIG. 7. (Color online) Positive predictive values (PPVs) as a function of percentile criterion for TEOAEs: half-octave 4 kHz, coherent emission strength, and the wideband TEOAE response. There is essentially no difference between CES and the wideband TEOAE, whereas the half-octave TEOAE at 4 kHz is a substantially better predictor of subsequent PTS.

of interest to find an optimal cutoff value. For any specific percentile cutoff value, the likelihood ratio is defined as the ratio of the probability that a test result was below the cutoff given there was a PTS to the probability that a test result was below a percentile cutoff given there was not a PTS. For readers familiar with the theory of signal detectability and ROC analysis, this is equivalent to the ratio of the hit rate and false-alarm rate, though the data are transformed so that the PTS group is the "signal" and the non-PTS group is the "noise."

Figure 5 shows likelihood ratio as a function of percentile cutoff for each OAE and test frequency. For TEOAEs, there were 16 to 18 PTS ears and 524 to 559 non-PTS ears included in the analysis. For DPOAEs, there were 16 PTS ears and 546 to 550 non-PTS ears included in the analysis. For many cases, there was a clear trend of increasing risk with decreasing percentile cutoff. $TEOAE_{74}$ at 4, 2.8, and 1 kHz are the clearest cases—each shows that low-level TEOAE amplitudes were more likely among the ears that subsequently developed PTS in this population and noise environment. The risk starts to increase as the TEOAE amplitude moves below the 25th percentile. DPOAEs show a similar trend to TEOAEs, but they are not as consistent, nor do they reach as high a likelihood ratio.

The positive predictive value (PPV) (Zhou *et al.*, 2002), on the other hand, is the conditional probability of an ear from this population *getting* PTS within 9 months in this particular noise environment, *given* a test result of a low-level OAE. The PPV is also known as the *a posteriori* conditional probability: $P(\text{PTS}\,|\,\text{OAE} \leq \text{cutoff})$, and it takes the base rate of PTS into account. PPVs are more useful than likelihood ratios for diagnosticians, because they can be used to estimate the probability of getting a PTS for a given population, timeframe, and noise environment.[9]

Figure 6 shows the best three $TEOAE_{74}$ frequencies and the best two DPOAE frequencies from Fig. 5, replotted as PPV. The base rate for an ear incurring PTS is approximately 3%.[10] For ears with results in the low percentiles, the risk of PTS increases to between 17% and 20% for the best TEOAE conditions and to between 14% and 17% for the best DPOAE conditions, depending on the percentile cutoff chosen.

LePage and Murray have also considered low-level TEOAEs as a predictor for hearing loss, based on their cross-sectional data set. They used an empirically derived TEOAE measure: coherent emission strength (CES, dB SPL), which represents the noise-free part of the TEOAE (LePage and Murray, 1993; LePage *et al.*, 1993; LePage, 1998). CES is a reweighting of the average TEOAE wide-band response by the square of the reproducibility (when linearly rescaled and transformed from $[-1:1]$ to $[0:1]$). For comparison, PPVs were calculated for the TEOAE data using CES and the TEOAE wide-band response, and compared to the best performing 4-kHz half-octave band (see Fig. 7). There were 563 non-PTS ears and 17 PTS ears contributing to the analysis. Data for eight ears were removed (including one PTS case) because the noise level was too high. The noise level was substituted for wide-band TEOAEs that were less than 0 dB SNR. No substitutions were made for CES, because this method allows for the use of TEOAEs with SNR less than 0 dB. Figure 7 shows PPVs for CES, wide-band TEOAEs, and 4-kHz half-octave TEOAEs. CES and wide-band TEOAEs were almost identical in their ability to predict PTS, and both were substantially worse than 4 kHz TEOAEs. The performance of CES and wide-band TEOAEs were similar to the best DPOAEs shown in Fig. 6. Focusing on the area most likely to be damaged by noise (4 kHz and above) increases the predictability for TEOAEs. The predict-

ability for DPOAEs at 4 kHz, on the other hand, is slightly worse than CES and wide-band TEOAEs, especially at the lowest percentiles.

These predictors are based on a small number of PTS cases, and they should be treated as indicative only. The predictors are very much dependent on the specific population, elapsed time (only 9 months in this study), and noise environment studied. However, these data show promising signs that OAEs may be used as predictors for susceptibility to PTS.

E. Summary of main findings

(1) Average audiometric thresholds did not change between pre- and postdeployment testing, but both average TEOAE and DPOAE amplitudes decreased for the group of 75 sailors with relatively complete data sets.
(2) There was no correlation between changes in audiometric thresholds and changes in OAEs for the entire group of noise-exposed sailors. There were, however, significant correlations between changes in OAEs across OAE types.
(3) Fifteen sailors (18 ears) were diagnosed with PTS. It was expected that significant changes in audiometric thresholds would be mirrored with significant changes in OAEs, but this was not the case in the majority of ears. Instead, the main observation was that PTS ears had many low-level or absent OAEs.
(4) There was no correlation between ears with PTS and without PTS and ears with SES or without SES. There was, however, a correlation among SESs across OAE types for the PTS ears, indicating that the SESs were probably not random fluctuations.
(5) Low-level and absent predeployment OAEs were predictive of postdeployment PTS.
(6) The best predictor of postdeployment PTS was predeployment TEOAE amplitude at 4 kHz, with lower amplitudes indicating increased risk.

IV. DISCUSSION

A. Why did PTS occur?

PTS occurred due to a combination of high noise levels and imperfect hearing protection usage. Fifteen sailors had a documented PTS in at least one ear after 6 months' deployment on an aircraft carrier, despite an active hearing-conservation program and the use of hearing protection. As reported earlier, the noise levels on aircraft carriers often exceed the maximum noise reduction ability of hearing protection. Furthermore, from the self-reported noise-exposure histories in the current study, the majority of sailors with PTS used hearing protection inconsistently. However, many sailors without PTS were also poor users of hearing protection. It was common for sailors to report using only single hearing protection in situations where double hearing protection was required. Sometimes no hearing protection was used when single hearing protection was required. It is also likely that many were not fitting hearing protection correctly. In a study across multiple platforms (aircraft carriers and amphibious assault ships), the vast majority of sailors omitted their earplugs some or all of the time, and did not insert them to a proper depth (Bjorn *et al.*, 2005). It is likely that these results generalize to the population of the current study, and it is therefore of no surprise that there was documented PTS.

B. Derived STS and SES criteria

It was important to develop site-specific STS criteria in the current study, because the Navy STS criteria used at the time were not based on test-retest reliability measurements. The STS criteria were identical to the Navy criteria at 1, 2, and 3 kHz, but were larger at 4 kHz and smaller at the average of 2, 3, and 4 kHz. In comparison to previously published studies, the STS criteria were larger at 4 and 6 kHz than the criteria developed in Lapsley Miller *et al.* (2004). The TEOAE criteria were smaller than in this earlier study, possibly because the testers were more experienced, and possibly because every attempt was made to match the postdeployment stimulus waveform and spectrum to the predeployment waveform and spectrum by manipulating the angle and depth of the probe in the volunteer's ear. DPOAEs were not comparable because here they were based on measurements at individual frequencies, rather than averaged within half-octave bands. However, they were comparable to those reported elsewhere (Franklin *et al.*, 1992; Beattie and Bleech, 2000; Seixas *et al.*, 2005b). OAE reliability depends on the measurement paradigm and equipment, therefore the values from the current study may not generalize to other settings.

C. Relationships between PTS and SES

As summarized in the introduction, cross-sectional human studies and longitudinal animal studies all indicate, for various reasons, that there should be a relationship between noise-induced PTS and SES. However, longitudinal human studies have yet to offer a clear-cut picture. Differences in noise exposures during the study, differences in prior noise exposures, hearing-protection usage, age, sex, and individual susceptibility all mean that each person is at a different stage in developing noise-induced inner-ear changes and noise-induced hearing loss. In the current study, there were sailors with no changes in OAEs or audiometric thresholds, changes in OAEs but not audiometric thresholds, changes in audiometric thresholds but not OAEs, and changes in OAEs and audiometric thresholds. Can all these scenarios be accounted for by current theories, or is it a sign that there is no relationship between OAEs, audiometric thresholds, and noise-induced change?

1. No changes in OAEs or audiometric thresholds

Out of the 338 noise-exposed sailors, only 12 (3.6%) had no measurable changes in audiometric thresholds or in any of their OAEs in either ear. A further 44 (13%) had no measurable changes, but also had some missing OAE data. All the other sailors had at least one significant shift in either audiometric thresholds or OAEs, though many of these are likely to be false positives. Why did some sailors have no changes? They may have been better users of hearing protection. They may have been lucky to not be as severely

noise exposed, and so avoided any noise-induced damage. They may have had particularly tough ears (Cody and Robertson, 1983) or toughened ears [in laboratory rodents, intermittent noise exposure may increase resistance to noise damage (Henderson *et al.*, 1993)]. There are likely to be some undetected changes (false negatives), in part due to missing OAE data (from bad calibrations, high noise, etc.) and also due to test-retest variability. However, there were proportionally more ears in the control group with no changes across any of the measures, suggesting that the noise-exposed group did indeed have more changes due to noise exposure.

2. Changes in OAEs but not audiometric thresholds

More ears showed significant OAE shifts than permanent threshold shifts in the noise-exposed group. Furthermore, the ANOVAs indicated no group changes in audiometric thresholds, but small, significant, decreases in group OAE levels for the 75 volunteers with complete data sets.[11] This is mostly consistent with the other longitudinal studies in the literature where small group decreases in OAEs are often seen, but concomitant changes in group audiometric thresholds are not (Engdahl *et al.*, 1996; Seixas *et al.*, 2005a). Konopka *et al.* (2005) found an approximately 2-dB decrease in TEOAEs, but the only significant changes in audiometric thresholds were at 10 and 12 kHz (frequencies not normally measured). Lapsley Miller *et al.* (2004) also showed changes in audiometric thresholds along with changes in OAEs, with a standard audiometric-frequency range, but this result was not as clear-cut when considering individual PTS cases. In comparison, no consistent changes were found among an orchestra group over 5 and 9 years (Murray *et al.*, 1998; Murray and LePage, 2002), but there were issues with recent noise exposure and possible TTS at baseline, which would reduce the magnitude of any audiometric threshold shift.

There are at least four explanations for why there were more SESs than PTSs: sensitivity of the audiogram, high-frequency hearing loss, outer-hair-cell redundancy, and age-related changes. The parsimonious explanation is that OAE measurements have smaller test-retest variability than audiometry (as measured using a standard clinical protocol and audiometer), so smaller noise-induced changes to the inner ear can be detected. Although the audiometric reliability in the present study was not as low as possible (because of the requirement that the Navy's nonoptimal shipboard audiometers were used), even in the best of circumstances audiometric reliability is worse than OAE reliability when comparing the same frequency band (e.g., Lapsley Miller *et al.*, 2004).[12] Audiometric resolution is similar to OAE resolution if multiple audiometric test frequencies are combined, but at the price of a decrease in frequency specificity (11 out of 18 PTS cases had PTS over an average of two or three frequency bands). However, it would be expected that if audiometric resolution was much lower than OAE resolution, the PTS cases that were identified should show consistent SESs. This was not the case—only one-third of PTS cases also showed SESs (and these SESs were not necessarily consistent with the PTS). Differences in resolution cannot explain all the findings.

Only changes in audiometric thresholds up to 4 kHz were considered, because of the high test-retest variability at 6 kHz. It is possible that undetected high-frequency hearing loss at 6 kHz and above affected lower frequency OAEs. The mechanism by which this occurs is still being debated, but it could be due to intermodulation distortion of the OAE components (Avan *et al.*, 1995; Yates and Withnell, 1999; Withnell *et al.*, 2000). Recently, Konopka *et al.* (2005) reported group decreases in TEOAEs between 2 and 4 kHz, concomitant with group increases in high-frequency audiometric thresholds at 10 and 12 kHz, but no significant change in audiometric thresholds at the TEOAE frequencies.

An alternative theory is that there is outer-hair-cell (OHC) redundancy such that only some are required for normal hearing. Animal studies have shown that there can be extensive OHC loss without changes in hearing thresholds (e.g., Bohne and Clark, 1982; Hamernik *et al.*, 1989; Altschuler, 1992), and that OAEs can be more sensitive to the effects of noise damage to the inner ear than pure-tone thresholds (Hamernik *et al.*, 1996). LePage and Murray (1993) argue that because OAEs are a more direct measurement of OHC activity, the loss of some OHCs is more likely to show up as diminished OAEs levels rather than as hearing loss. OAEs would therefore show noise-induced changes prior to hearing loss (i.e., even if audiometric thresholds could be measured more sensitively, there would not be any increase in the amount of noise-induced PTS). This implies that OAE changes can indicate subclinical NIHL. LePage (1992) proposes that the ear is able to remap the cochlear place-to-frequency conversion to avoid gaps in frequency detection because of OHC loss, and hearing loss occurs only when this ability to remap is exceeded. In their large cross-sectional study, Le Page and Murray found that coherent emission strength (CES) may decrease by 80% before there is a change in audiometric thresholds (LePage *et al.*, 1993). They concluded that the "pure tone audiogram may not be a direct measure of cochlear damage so much as a measure of how much the cochlea can maintain normal performance despite ongoing damage" (LePage *et al.*, 1993). This is supported by the current findings that (a) although there were changes in audiometric thresholds and changes in OAEs, the two were not related contemporaneously, and (b) ears with low OAEs have less resistance to hazardous noise and are more likely to get PTS with continued noise exposure.

Another possibility is age-related OAE changes. Some studies have shown that OAEs decrease with age, even when audiometric thresholds are controlled (Dorn *et al.*, 1998; Cilento *et al.*, 2003, for women but not for men). Murray and LePage (1993) hypothesize that the OHC loss that occurs from birth throughout life causes the aging effect seen with OAEs, and that noise exposure accelerates the loss of OHCs. However, the present study was only 9 months long, and the volunteers were relatively young, so aging is unlikely to be a major factor.

These four possibilities—lower audiometric sensitivity, high-frequency hearing loss, OHC redundancy, and age-related changes—are not mutually exclusive, nor is it pos-

sible to easily tease out which, if any, is operating here. However, none of the explanations are contradictory—all can explain these findings to some extent.

3. Changes in audiometric threshold, but not OAEs

About two-thirds of the PTS ears did not have significant OAE changes consistent with their PTS. In most cases, this was because OAEs were already at low levels or were absent at predeployment testing. It is possible that OAEs had already decreased from earlier noise exposures (the large majority of volunteers had considerable military noise exposure prior to the study), and some sailors may also have had low-level OAEs due to genetics, illness, or environmental factors such as chemical exposure. Regardless of the cause, having low-level or absent OAEs was predictive of subsequent noise-induced hearing loss.

Massive, traumatic noise exposure can simultaneously affect both audiometric thresholds and OAEs, but here the largest PTS was only 25 dB, and most were only 15 dB. It is more likely that changes in OAEs preceded the changes in audiometric thresholds. By the time the audiometric thresholds were affected, the OAEs may have been sufficiently low or even absent. It is difficult to measure a change in a low-level OAE because measurements near or below the noise floor are not reliable. Even in the ears where PTSs and SESs occurred together, it is conceivable that the OAEs may have diminished before the audiometric thresholds, but both may have changed within the 9 months, so only the final outcome was observed.

Other explanations as to why some PTS ears showed no changes in OAEs include interaction of DPOAE sources and differences in intrasubject variability, though these factors are probably less influential than missing data. The DPOAE measured at frequency $2f_1$-f_2 actually consists of two frequency components—both with frequency $2f_1$-f_2, but with different magnitudes and phases. These two components—the reflection component and the distortion component—are thought to be generated from different parts of the cochlea (Shera and Guinan, 1999). If a DPOAE was dominated by the distortion component, then changes in the cochlea at the source of the reflection component (thought by Shera to be the more likely indicator of damage) may not show up (Shera, 2004). In the current study, the definition of a significant OAE shift was based on the group SE_{MEAS}. Some people exhibit little variation in OAEs over time—some show a great deal (Marshall and Heller, 1996). The criteria used in the current study may have been too strict for some people with very stable OAEs, therefore missing some SESs.

Finally, damage to structures other than OHCs may have caused the threshold shift. For example, stria vascularis may be affected by long-duration exposures (e.g., Bohne and Clark, 1982), and inner-hair cells may also be damaged by noise exposure, but usually not until greater amounts of hearing loss are observed (e.g., Hamernik *et al.*, 1989).

4. Both audiometric thresholds and OAEs change together

Only one-third of the PTS ears showed SESs; however, there was no strong or consistent pattern of SESs across ears, or across OAE types and frequencies. The ANOVA and correlational analyses gave scant evidence of audiometric thresholds and OAEs changing concomitantly. Sometimes OAEs improved when audiometric thresholds diminished or stayed the same. Although there were some positive SESs, they were at about the same rate as for the control group, so they most likely reflected random test-retest variability [although inner-ear damage can in some cases produce an increase in OAE amplitude (Withnell *et al.*, 2000)].

D. Susceptibility

Ears with low-level or absent OAEs were more likely to get PTS. Low-level or absent OAEs may be a sign of genetic variability, or, more likely, a sign that these ears had already experienced subclinical damage from previous hazardous noise exposures. Although most volunteers had normal hearing at predeployment (some had a slight high-frequency hearing loss), most, if not all, had received considerable noise exposure prior to this study. The PTS ears had slightly better audiometric thresholds at predeployment compared with the non-PTS ears (see Fig. 4), so the initial audiometric thresholds themselves cannot explain the difference in OAE levels.

In the current study, TEOAEs were better PTS predictors than were DPOAEs. This may be because normal DPOAE microstructure has spectral nulls due to the interaction of the two DPOAE sources (e.g., Kalluri and Shera, 2001; Mauermann and Kollmeier, 2004). Measurements at or near nulls could easily show low-level DPOAE amplitudes. For instance, Shaffer *et al.* (2003) quite clearly showed how points in the DPOAE microstructure for a normal-hearing person can fall below norms. This would inflate the number of low-level DPOAEs seen in healthy ears. Such cases would not necessarily be predictive of hearing loss. Half-octave TEOAEs are less affected by fluctuations in microstructure because they represent the average level over a range of frequencies. A DPOAE design where frequencies were clustered closer together and the resulting amplitudes averaged may reduce the problem of nulls.

Dancer stated "There would be great interest in finding a test which predicts individual susceptibility to permanent threshold shift" (Dancer, 2000, p. 5-1). It looks promising that low-level or absent OAEs may indeed form the basis for such a test. Here, problems inherent with many previous susceptibility tests were bypassed in that the measurements were made on ears that developed PTS, rather than using TTS as a substitute (Ward, 1965; Humes, 1977). This is important since the mechanisms underlying TTS and PTS are different (e.g., Saunders *et al.*, 1985; Slepecky, 1986; Nordmann *et al.*, 2000).

Using low-level OAEs to measure PTS susceptibility is not like using a gene marker, which is stable and innate. Instead, an OAE reflects the current state of the inner ear which is likely a result of both genetic susceptibility and

acquired susceptibility. Susceptibility to NIHL probably varies over time depending on factors such as noise exposure, illness, chemical exposure, and age. It will not be enough to take just one measurement at one point in time to determine susceptibility. Instead, personnel will require regular monitoring to see if their susceptibility is changing as they continue to accumulate OHC damage. With more data—especially longitudinal data over a longer timeframe and information about the outcome costs—optimal criteria for risk detection can be developed.

In the future, OAEs may also be used in other ways to gauge susceptibility to NIHL. For instance, the reflex strength of the auditory efferent medial olivocochlear system, whose fibers synapse primarily on the outer hair cells, may indicate NIHL susceptibility. One of the suggested physiological functions of the efferent system has been protection from acoustic overexposure (Cody and Johnstone, 1982; Reiter and Liberman, 1995; Maison and Liberman, 2000; Luebke and Foster, 2002). Maison and Liberman (2000) predicted susceptibility to NIHL from the strength of the auditory efferent reflex, as measured with ipsilateral DPOAE adaptation. Guinea pigs with a high reflex strength exhibited only small or no PTS whereas guinea pigs with a low reflex strength exhibited PTS. Muller *et al.* (2005) showed that measuring DPOAE amplitude changes in humans with a contralateral AER elicitor (which is easier to measure in humans) is also a suitable measure for determining AER strength. Others have suggested using TEOAEs (e.g., Berlin *et al.*, 1995) or stimulus-frequency otoacoustic emissions (e.g., Guinan *et al.*, 2003). No matter which OAE type is used, the challenge is to develop a clinical test that shows a large range of auditory efferent reflex strength across people, relative to the intrasubject test-retest reliability, to be able to validly distinguish people with large and small efferent reflex strength. Such a test must also be fast for use in clinical and field settings.

In the future, measuring both auditory efferent reflex strength and absolute OAE amplitude in normal-hearing ears may provide powerful indications of individual NIHL risk before significant hearing loss has occurred.

E. Conclusions

When sailors are exposed to hazardous levels of shipboard noise, OAEs show the accumulated damage to the inner ear before hearing loss shows up in an audiogram. Moreover, diminished OAEs are predictive of subsequent hearing loss if the sailor remains in the noise-hazardous environment.

ACKNOWLEDGMENTS

Thanks to Linda Westhusin, Michael McFadden, Joe Bertoline, Dan Pawluk, Doug Clingan, and John Correia for their assistance with study coordination, volunteer recruitment, and data collection. A special thanks to the crew of the *USS Dwight D. Eisenhower*, Bob Rogers, and the staff from Norfolk Naval Hospital Occupational Health Audiology Department. Thanks to the Commander in Chief, US Atlantic Fleet; the Commander, Naval Air Force, US Atlantic Fleet; and the Commander, Naval Sea Systems Command. Thanks to Vit Drga for comments and suggestions on data analyses; Wayne Horn, Jeff Gertner, Michael Qin, and Loring Crepeau for their comments on an earlier version of the manuscript; Kurt Yankaskas for comments on noise exposure levels; and many useful comments and suggestions from the anonymous reviewers.

This research was supported by grants from the Office of Naval Research (N0001400WX20195, N0001400WX20244, and 0601153N4508) and the US Army Medical Research and Materiel Command (DAMD17-01-IA-0679). The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the United States Government.

[1]These studies tend not to consider whether noise-exposed people with normal OAEs have similar audiometric thresholds to non-noise-exposed people with normal OAEs. Therefore, it is not entirely clear that OAEs are providing an advantage in detecting the early stages of NIHL.

[2]On a Nimitz-class aircraft carrier, flight-deck noise produced from aircraft launches ranges from 126 to 148 dBA peak depending on the proximity to the aircraft. The arresting gear and water brakes, as well as tools such as needle guns, grinders, and hydro-blasters, generate noise levels around 94 dBA peak in work and berthing areas. Sound levels in the hangar bay under the flight deck can exceed 120 dBA peak during flight operations. Other noisy areas include the main propulsion machinery space, machinery rooms, (work) shops, and the laundry, which is located above the propellers. Many of the berthing spaces are directly below the flight deck—some sailors even wear hearing protection while sleeping. Dosimetry data from a Nimitz-class aircraft carrier, reported by Rovig *et al.* (2004), showed 8-h time-weighted averages (using an 85 dBA damage-risk criterion with a 3 dB exchange rate) of 109 dBA (ranging from 96 to 120 dBA) for flight deck crew and 92 dBA (ranging from 79 to 98 dBA) for engineering crew. The average workday was 11.5 h. Unweighted peak noise levels were regularly clipped by the dosimeter's 150 dB SPL ceiling, so these noise levels are underestimated. Rovig *et al.* (2004) also found that in sailors with 4 or more years of service, 30% of flight deck crew and 37% of engineering crew had audiometric thresholds greater than 25 dB HL (at 1, 2, 3, or 4 kHz), compared with 5% of administrative crew. Many sailors reported not wearing double hearing protection because they felt it jeopardized speech communication.

[3]The noise-rejection level was set at 4 mPa by default, but was usually adjusted by the tester to approximately one standard deviation above the mean of the noise-level histogram, which was usually lower than 4 mPa.

[4]Each frequency was measured three times with 15 subaverages and then averaged. The frequencies where DPOAEs had a signal-to-noise ratio (SNR) less than 3 dB were automatically retested until the test time (50 s) expired. The noise-rejection level was set at 5 mPa.

[5]$TEOAE_{74}$ stimulus levels were considered on-target if they were within ±4 dB of 74 dB pSPL (no data points were eliminated; 99.9% of data points were within ±3 dB of the target). DPOAE stimulus levels were considered on target if they were within ±6 dB of the target level, for both L_1 and L_2 (1.2% of data were eliminated across all levels and frequencies). Sometimes the ILO program could not obtain a good DPOAE calibration (usually due to standing waves). In these cases it automatically used an estimated level instead. The resulting DPOAEs produced many outliers—either unusually high or low, so all cases where levels were estimated were dropped from the DPOAE analyses (2.5% of all DPOAE data). Consideration of outliers showed many more outliers for $DP_{61/55}$, compared to the other levels. The reason for this discrepancy could not be traced, so all data at this level were dropped. Furthermore, all DPOAE data at f_2=2.2 and 4.5 kHz were dropped because they were contaminated with a large, intermittent harmonic artifact at these frequencies. Sometimes the artifact elevated the noise level and sometimes it appeared to elevate the DPOAE amplitude, so it was not possible to identify the affected cases by just looking for high noise levels. When considering changes in OAEs, some outliers appeared to be due to large differences in the noise floor between pre- and postdeployment testing. Therefore, the cases where the absolute average difference between the pre- and postdeployment noise levels was larger than 3.5 dB (when averaged across 2.5 to 3.6 kHz for DPOAEs, and

when averaged across the half-octave bands centered at 2, 2.8, and 4 kHz for $TEOAE_{74}$) were removed. This affected 2.8% of $TEOAE_{74}$ measurements and 6.5% of DPOAE measurements.

[6]The SE_{MEAS} can be used to specify the magnitude of a statistically significant change within an individual (Ghiselli, 1964) and is defined as $SE_{MEAS}=\sqrt{\frac{1}{2}(s_1^2+s_2^2)(1-r)}$ where s_1^2 and s_2^2 are the pre- and postdeployment variances, and r is the correlation between pre- and postdeployment tests. Because the focus here is on the difference between pre- and postdeployment testing, ΔSE_{MEAS} is defined as $\sqrt{2}SE_{MEAS}$ (Beattie, 2003; Beattie *et al.*, 2003). Multiplying ΔSE_{MEAS} by an appropriate multiplier then gives the desired confidence interval. Here a multiplier of 2.12 is used, which gives a 98% confidence interval.

[7]The STS criteria were adjusted by adding the group average audiometric thresholds to account for a slight mean shift (most likely a practice effect). For individual ears, the resolution of the audiogram was 5 dB for single frequencies, and, due to averaging, 2.5 dB for the average of 2 and 3 kHz and the average of 3 and 4 kHz, and 1.66 dB for the average of 2, 3, and 4 kHz. In other words, the smallest change that can be detected is defined by the resolution. Each STS criterion was adjusted accordingly by rounding up to the next largest step. Since STS criteria are usually specified as inclusive (i.e., $\geqslant X$ dB, rather than $>X$ dB), another resolution step was added to give the STS criteria reported in Table II. For example, the ΔSE_{MEAS} for the average of 2, 3, and 4 kHz was 2.16 dB and the minimum resolution is 1.66 dB. Converting into an STS criterion for the 98% confidence interval gives $2.12\sqrt{2}\Delta\ SE_{MEAS}=6.48$ dB; rounding up to the next resolution step gives 6.66 dB, and then another step to 8.33 dB gives the inclusive criterion. Resolution was not an issue for SESs because the SE_{MEAS} were orders of magnitude larger than the measurement resolution.

[8]The percentiles were calculated from the predeployment OAE data for the entire group of noise-exposed male sailors (606 ears, 303 volunteers), including the PTS ears and ears with absent OAEs, where noise levels were substituted to quantify absent OAEs, providing the noise level was low (as described earlier). Any ears with absent OAEs with noise levels higher than the cutoff were not included in these analyses. OAE amplitudes were converted to percentiles for each TEOAE and DPOAE level and frequency. Percentiles were calculated for left and right ears separately and for all ears combined. For the same percentile, OAE amplitude differed by up to 2.4 dB between the left and right ears.

[9]The PPV can be related to the likelihood ratio by using odds ratios where the *a posteriori* odds are equal to the *a priori* odds multiplied by the likelihood ratio (Zhou *et al.* 2002). In the current scenario, $PPV/(1-PPV)=\text{likelihood ratio}\times P(\text{PTS})/P(\text{no PTS})$. Other formulations and relationships may be derived using Bayes' theorem for conditional probabilities.

[10]This PTS rate is the percentage of ears with PTS and without high noise floors. It differs from the earlier PTS rate, which was the percentage of sailors with PTS.

[11]Because only 75 volunteers had complete data sets across 2 to 4 kHz, a further ANOVA was conducted for the group of 206 volunteers with complete data sets at just 4 kHz for TEOAEs and audiometric thresholds, to see if the larger group showed any changes between pre- and postdeployment. The two-way, repeated-measures ANOVA for TEOAE amplitude (test: pre- versus postdeployment; ear left versus right) showed a 0.95-dB decrement between pre- and postdeployment testing ($F_{1,205}=49.2, p<0.05$), but no significant difference between ears, and no significant interaction. The two-way, repeated-measures ANOVA for audiometric thresholds (test: pre- versus postdeployment; ear left versus right) showed no significant effect for test, but a significant difference between ears ($F_{1,205}=24.8, p<0.05$). The interaction was not significant. Even with the increased numbers, there was no significant change in audiometric thresholds.

[12]The ability to detect a PTS or SES is dependent on the test-retest reliability. Poorer reliability results in a larger criterion to decide that a significant shift has occurred. For example, compare the OAE and hearing-threshold criteria for 4 kHz from Tables II and III. To make this a fair comparison between OAEs and audiometric thresholds, the SES criteria are multiplied by 2.5 to convert them into "equivalent dB HLs" (Marshall and Heller, 1998) to give criteria of 9.3 dB for $TEOAE_{74}$, 14.0 dB for $DP_{57/45}$, 15.2 dB for $DP_{59/50}$, and 14.2 dB for $DP_{65/45}$. Both the TEOAE and DPOAEs are superior to single-frequency audiometric thresholds in their ability to detect a shift. However, when averaging the audiometric threshold over 2, 3, and 4 kHz, the criterion decreases to be comparable with TEOAEs at 4 kHz. A closer relationship between changes in OAEs and audiometric thresholds is found in the laboratory in animal studies and in human TTS studies (e.g., Marshall and Heller, 1998) when there is more precise information about both the audiometric changes and the noise that produced them. Differences in results between TTS studies and PTS studies could be attributed to the fact that the mechanisms underlying TTS and PTS differ (e.g., Saunders *et al.*, 1985; Slepecky, 1986; Nordmann *et al.*, 2000). However, this conclusion may not be warranted because the precision of the audiometric and noise measurements is higher for TTS studies.

Altschuler, R. A. (**1992**). "Acoustic stimulation and overstimulation in the cochlea: A comparison between basal and apical turns of the cochlea," in *Noise-Induced Hearing Loss*, edited by A. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik (Mosby-Year Book, St. Louis), pp. 60–72.

ANSI (1991). "Maximum permissible ambient noise levels for audiometric test rooms (ANSI S3.1)" (American National Standards Institute, New York).

Attias, J., Horovitz, G., El-Hatib, N., and Nageris, B. (**2001**). "Detection and clinical diagnosis of noise-induced hearing loss by otoacoustic emissions," Noise Health **3**, 19–31.

Attias, J., Bresloff, I., Reshef, I., Horowitz, G., and Furman, V. (**1998**). "Evaluating noise induced hearing loss with distortion product otoacoustic emissions," Br. J. Audiol. **32**, 39–46.

Attias, J., Furst, M., Furman, V., Reshef, I., Horowitz, G., and Bresloff, I. (**1995**). "Noise-induced otoacoustic emission loss with or without hearing loss," Ear Hear. **16**, 612–618.

Avan, P., Bonfils, P., Loth, D., Elbez, M., and Erminy, M. (**1995**). "Transient-evoked otoacoustic emissions and high-frequency acoustic trauma in the guinea pig," J. Acoust. Soc. Am. **97**, 3012–3020.

Beattie, R. C. (**2003**). "Distortion product otoacoustic emissions: comparison of sequential versus simultaneous presentation of primary-tone pairs," J. Am. Acad. Audiol **14**, 471–484.

Beattie, R. C., and Bleech, J. (**2000**). "Effects of sample size on the reliability of noise floor and DPOAE," Br. J. Audiol. **34**, 305–309.

Beattie, R. C., Kenworthy, O. T., and Luna, C. A. (**2003**). "Immediate and short-term reliability of distortion-product otoacoustic emissions," Int. J. Audiol. **42**, 348–354.

Berlin, C. I., Hood, L. J., Hurley, A. E., Wen, H., and Kemp, D. T. (**1995**). "Binaural noise suppresses linear click-evoked otoacoustic emissions more than ipsilateral or contralateral noise," Hear. Res. **87**, 96–103.

Bicciolo, G., Ruscito, P., Rizzo, S., and Frenguelli, A. (**1993**). "Evoked otoacoustic emissions in noise-induced hearing loss," Acta Otorhinolaryngol. Ital. **13**, 505–515.

Bjorn, V. S., Albery, C. B., Shilling, R., and McKinley, R. L. (**2005**). "Navy Flight Deck Hearing Protection Use Trends: Survey Results," in *NATO Human Factors and Medicine Panel Symposium, New Directions for Improving Audio Effectiveness* (Amersfoort, The Netherlands).

Bohne, B., and Clark, W. W. (**1982**). "Growth of hearing loss and cochlear lesion with increasing duration of noise exposure," in *New Perspectives on Noise-induced Hearing Loss*, edited by R. P. Hamernik, D. Henderson, and R. Salvi (Raven, New York), pp. 283–301.

Bray, P. J. (1989). "Click evoked otoacoustic emissions and the development of a clinical otoacoustic hearing test instrument," unpublished Ph.D. thesis, Institute of Laryngology and Otology, Univ. College and Middlesex School of Medicine, London.

Cilento, B. W., Norton, S. J., and Gates, G. A. (**2003**). "The effects of aging and hearing loss on distortion product otoacoustic emissions," Otolaryngol.-Head Neck Surg. **129**, 382–389.

Cody, A. R., and Johnstone, B. M. (**1982**). "Temporary threshold shift modified by binaural acoustic stimulation," Hear. Res. **6**, 199–205.

Cody, A. R., and Robertson, D. (**1983**). "Variability of noise-induced damage in the guinea pig cochlea: electrophysiological and morphological correlates after strictly controlled exposures," Hear. Res. **9**, 55–70.

Dancer, A. (2000). "Individual susceptibility to NIHL and new perspective in treatment of acute noise trauma," in *RTO HFM Lecture Series on Damage Risk from Impulse Noise*, held in Maryland, USA, 5–6 June 2000 and Meppen, Germany, 15–16 June 2000, and published in RTO EN-11.

Desai, A., Reed, D., Cheyne, A., Richards, S., and Prasher, D. (**1999**). "Absence of otoacoustic emissions in subjects with normal audiometric thresholds implies exposure to noise," Noise Health **2**, 58–65.

Dorn, P. A., Piskorski, P., Keefe, D. H., Neely, S. T., and Gorga, M. P. (**1998**). "On the existence of an age/threshold/frequency interaction in distortion product otoacoustic emissions," J. Acoust. Soc. Am. **104**, 964–971.

Engdahl, B., Woxen, O., Arnesen, A. R., and Mair, I. W. (**1996**). "Transient evoked otoacoustic emissions as screening for hearing losses at the school

for military training," Scand. Audiol. **25**, 71–78.
Franklin, D. J., McCoy, M. J., Martin, G. K., and Lonsbury-Martin, B. L. (**1992**). "Test/retest reliability of distortion-product and transiently evoked otoacoustic emissions," Ear Hear. **13**, 417–429.
Ghiselli, E. E. (**1964**). *Theory of Psychological Measurement* (McGraw-Hill, New York).
Gorga, M. P., Neely, S. T., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (**1993**). "A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **94**, 2639–2648.
Guinan, J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). "Medial olivocochlear efferent reflex in humans: otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," J. Assoc. Res. Otolaryngol. **4**, 521–540.
Hamernik, R. P., Ahroon, W. A., and Lei, S. F. (**1996**). "The cubic distortion product otoacoustic emissions from the normal and noise-damaged chinchilla cochlea," J. Acoust. Soc. Am. **100**, 1003–1012.
Hamernik, R. P., Patterson, J. H., Turrentine, G. A., and Ahroon, W. A. (**1989**). "The quantitative relation between sensory cell loss and hearing thresholds," Hear. Res. **38**, 199–211.
Henderson, D., Subramaniam, M., and Boettcher, F. A. (**1993**). "Individual susceptibility to noise-induced hearing loss: an old topic revisited," Ear Hear. **14**, 152–168.
Humes, L. E. (**1977**). "Review of four new indices of susceptibility to noise-induced hearing loss," J. Occup. Med. **19**, 116–118.
Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: a test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.
Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.
Konopka, W., Pawlaczyk-Luszczynska, M., Sliwinska-Kowalska, M., Grzanka, A., and Zalewski, P. (**2005**). "Effects of impulse noise on transiently evoked otoacoustic emission in soldiers," Int. J. Audiol. **44**, 3–7.
Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of the 2 f1-f2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.
Lapsley Miller, J. A., and Marshall, L. (**2001**). "Monitoring the effects of noise with otoacoustic emissions," Semin. Hear. **22**, 393–403.
Lapsley Miller, J. A., Marshall, L., and Heller, L. M. (**2004**). "A longitudinal study of changes in evoked otoacoustic emissions and pure-tone thresholds as measured in a hearing conservation program," Int. J. Audiol. **43**, 307–322.
LePage, E. L. (**1992**). "Hysteresis in cochlear mechanics and a model for variability in noise-induced hearing loss," in *Noise-Induced Hearing Loss*, edited by A. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik (Mosby-Year Book, St. Louis), pp. 106–115.
LePage, E. L. (**1998**). "Occupational noise-induced hearing loss: Origin, characterisation and prevention," Acoust. Aust. **26**, 57–61.
LePage, E. L., and Murray, N. M. (**1993**). "Click-evoked otoacoustic emissions: Comparing emission strengths with pure tone audiometric thresholds," Aust. J. Audiol. **15**, 9–22.
LePage, E. L., and Murray, N. M. (**1998**). "Latent cochlear damage in personal stereo users: a study based on click-evoked otoacoustic emissions," Med. J. Aust. **169**, 588–592.
LePage, E. L., Murray, N. M., Tran, K., and Harrap, M. J. (**1993**). "The ear as an acoustical generator: Otoacoustic emissions and their diagnostic potential," Acoust. Aust. **21**, 86–90.
Liberman, M. C., Dodds, L. W., and Learson, D. A. (**1986**). "Structure-function correlation in noise-damaged ears: A light and electron-microscopic study," in *Basic and Applied Aspects of Noise-induced Hearing Loss*, edited by R. J. Salvi, D. Henderson, R. P. Hamernik, and V. Colletti (Plenum, New York), pp. 163–177.
Luebke, A. E., and Foster, P. K. (**2002**). "Variation in inter-animal susceptibility to noise damage is associated with alpha 9 acetylcholine receptor subunit expression level," J. Neurosci. **22**, 4241–4247.
Maison, S. F., and Liberman, M. C. (**2000**). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," J. Neurosci. **20**, 4701–4707.
Mansfield, J. D., Baghurst, P. A., and Newton, V. E. (**1999**). "Otoacoustic emissions in 28 young adults exposed to amplified music," Br. J. Audiol. **33**, 211–222.
Marshall, L., and Heller, L. M. (**1996**). "Reliability of transient-evoked otoacoustic emissions," Ear Hear. **17**, 237–254.
Marshall, L., and Heller, L. M. (**1998**). "Transient-evoked otoacoustic emissions as a measure of noise-induced threshold shift," J. Speech Lang. Hear. Res. **41**, 1319–1334.
Marshall, L., Lapsley Miller, J. A., and Heller, L. M. (**2001**). "Distortion-product otoacoustic emissions as a screening tool for noise-induced hearing loss," in *Noise Induced Hearing Loss Basic Mechanisms, Prevention and Control*, edited by D. Henderson, D. Prasher, R. D. Kopke, R. Salvi, and R. P. Hamernik (Noise Research Network, London), pp. 453–470.
Marshall, L., Lapsley Miller, J. A., Hughes, L. M., and Westhusin, L. J. (**2002**). "Changes in evoked otoacoustic emissions and hearing thresholds after a six-month deployment on an aircraft carrier," Assoc. Res. Otolaryngol. Abs., **25**, 203.
Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.
Muller, J., Janssen, T., Heppelmann, G., and Wagner, W. (**2005**). "Evidence for a bipolar change in distortion product otoacoustic emissions during contralateral acoustic stimulation in humans," J. Acoust. Soc. Am. **118**, 3747–3756.
Murray, N. M., and LePage, E. L. (**1993**). "Age dependence of otoacoustic emissions and apparent rates of ageing of the inner ear in an Australian population," Aust. J. Audiol. **15**, 59–70.
Murray, N. M., and LePage, E. L. (2002). "A nine-year longitudinal study of the hearing of orchestral musicians," in *International Auditory Congress*, Melbourne, Australia.
Murray, N. M., LePage, E. L., and Mikl, N. (**1998**). "Inner ear damage in an opera theatre orchestra as detected by otoacoustic emissions, pure tone audiometry and sound levels," Aust. J. Audiol. **20**, 67–78.
Navy Occupational Health and Safety Program (1999). OPNAVINST 5100.23E: Hearing conservation and noise abatement (Chief of Naval Operations, Washington DC).
Nordmann, A. S., Bohne, B. A., and Harding, G. W. (**2000**). "Histopathological differences between temporary and permanent threshold shift," Hear. Res. **139**, 13–30.
Passchier-Vermeer, W. (1993). Noise and Health (Publication No. A93/02E) (Health Council of The Netherlands, The Hague).
Rask-Andersen, H., Ekvall, L., Scholtz, A., and Schrott-Fischer, A. (**2000**). "Structural/audiometric correlations in a human inner ear with noise-induced hearing loss," Hear. Res. **141**, 129–139.
Reiter, E. R., and Liberman, M. C. (**1995**). "Efferent-mediated protection from acoustic overexposure: relation to slow effects of olivocochlear stimulation," J. Neurophysiol. **73**, 506–514.
Rovig, G. W., Bohnker, B. K., and Page, J. C. (**2004**). "Hearing health risk in a population of aircraft carrier flight deck personnel," Mil. Med. **169**, 429–432.
Saunders, J. C., Dear, S. P., and Schneider, M. E. (**1985**). "The anatomical consequences of acoustic injury: A review and tutorial," J. Acoust. Soc. Am. **78**, 833–860.
Seixas, N. S., Goldman, B., Sheppard, L., Neitzel, R., Norton, S. J., and Kujawa, S. G. (**2005a**). "Prospective noise induced changes to hearing among construction industry apprentices," Occup. Environ. Med. **62**, 309–317.
Seixas, N. S., Neitzel, R., Brower, S., Goldman, B., Somers, S., Sheppard, L., Kujawa, S. G., and Norton, S. (**2005b**). "Noise-related changes in hearing: a prospective study among construction workers," in 30th Annual NHCA National Hearing Conservation Conference, Tucson, AZ.
Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**). "Sources and mechanisms of DPOAE generation: implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.
Shera, C. A. (**2004**). "Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions," Ear Hear. **25**, 86–97.
Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Siegel, S. (**1956**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, New York).
Slepecky, N. (**1986**). "Overview of mechanical damage to the inner ear: noise as a tool to probe cochlear function," Hear. Res. **22**, 307–321.
Sutton, L. A., Lonsbury-Martin, B. L., Martin, G. K., and Whitehead, M. L. (**1994**). "Sensitivity of distortion-product otoacoustic emissions in humans

to tonal over-exposure: time course of recovery and effects of lowering L_2," Hear. Res. **75**, 161–174.
Ward, W. D. (**1965**). "The concept of susceptibility to hearing loss," J. Occup. Med. **7**, 595–607.
Withnell, R. H., Yates, G. K., and Kirk, D. L. (**2000**). "Changes to low-frequency components of the TEOAE following acoustic trauma to the base of the cochlea," Hear. Res. **139**, 1–12.
Xu, Z. M., Van Cauwenberge, P., Vinck, B., and De Vel, E. (**1998**). "Sensitive detection of noise-induced damage in human subjects using transiently evoked otoacoustic emissions," Acta Otorhinolaryngol. Belg. **52**, 19–24.
Yankaskas, K. D. (1999). "Hearing conservation: The engineering part of the equation," Navy Med. Sept.-Oct., 21–25.
Yankaskas, K. D., and Shaw, M. F. (1999). "Landing on the roof: CVN noise," Nav. Eng. J. July, 23–34.
Yates, G. K., and Withnell, R. H. (**1999**). "The role of intermodulation distortion in transient-evoked otoacoustic emissions," Hear. Res. **136**, 49–64.
Zhou, X.-H., Obuchowski, N. A., and McClish, D. K. (**2002**). *Statistical Methods in Diagnostic Medicine* (Wiley-Interscience, New York).

Semirealistic models of the cochlea

Norman Sieroka[a)]
Sektion Biomagnetismus, Neurologische Klinik, Universität Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany and ETH Zürich, RAC G 16, 8092 Zürich, Switzerland

Hans Günter Dosch
Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany

André Rupp
Sektion Biomagnetismus, Neurologische Klinik, Universität Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany

(Received 12 July 2005; revised 19 April 2006; accepted 20 April 2006)

The aim of this paper is the introduction and comparison of consistent albeit passive mechanical models for the whole cochlea. A widely used transmission line filterbank, which hydrodynamically speaking is a long wave approximation (L model), suffers from a well-known inconsistency: its main modeling assumption is not valid within the resonance region, where most of the overall excitation takes place. In the present paper two approaches to overcome this inconsistency are discussed. One model is the average pressure (AP) model by Duifhuis, the other is obtained by a combination of a long and a short wave approximation (LS model). Considerable differences between the L and the LS model are observed. All models are compared by inserting them into the full integral equation obtained from the hydrodynamic equations and the boundary conditions. Here the LS model fares better than the AP model for small damping, whereas the opposite is true for higher damping. As expected, the L model fails badly in the resonance region.
 [DOI: 10.1121/1.2204438]

PACS number(s): 43.64.Kc, 43.64.Bt [AK] Pages: 297–304

I. INTRODUCTION

Auditory models of the cochlea provide useful tools to understand the responses of the first station of the pathway, namely the auditory nerve, to the excitations of the basilar membrane and allow the investigation of psychoacoustic phenomena that are based on properties of basilar membrane mechanics. Widely used practical approaches implemented to serve as filterbanks are phenomenological ones as given by the gammatone auditory filters (Aertsen and Johannesma, 1980) and recently developed compressive gammachirp filters that provide level-dependent modeling in cochlear filtering (Irino and Patterson, 2001). A second type of approach is the one-dimensional transmission line model of the cochlea (Strube, 1985; Giguère and Woodland, 1994; Giguère *et al.*, 1997). The advantage of this type of model is that it has a physiological basis and that it can include nonlinear effects. It suffers, however, from an inconsistency: the conditions used for its derivation from a simplified model for the cochlea are not fulfilled by its solutions.

In terms of hydrodynamics the one-dimensional transmission line model is a long wave approximation. The application of both long and short wave models (L model and S model) to cochlea dynamics goes back to the work of Ranke (1950) and Zwislocki (1950, 1953). The discussion was later revived by Siebert (1974), who allied with Ranke and the L model, and by Schroeder (1975), who like Zwislocki favored the S model. To our knowledge no one has suggested a combination of the two models. Either the authors strictly favored one type of model over the other or saw them as incompatible heuristic tools. The first who claimed to have overcome the debate between S and L models are Viergever and Kalker (1975). Their model is based on an average pressure analysis (AP model) and was further elaborated in particular by Duifhuis (1988). A recent application of the AP model on questions of scattering and otoacoustic emissions is given by Shera *et al.* (2005).

In this paper we investigate approximate solutions to a three-dimensional mechanical model, based on the average pressure analysis of Duifhuis (1988) (AP model) and on a solution pasted together from a shallow water (long wave length) and a deep water (short wave length) solution (LS model). These models avoid the inconsistency of the one-dimensional long wave models (L model). A comparison of these three-dimensional models with the widely used one-dimensional ones allows us to isolate the specific effects of the more realistic approach.

This paper is organized as follows: In Sec. II we briefly discuss some basic mechanical features of the inner ear and of a simplified rectangular model of the cochlea. Next, in Sec. III, the approximate solutions to the relevant hydrodynamic equations and further specifications of the model are given. We discuss the reliability of specific features of the models in Sec. IV.

[a)]Electronic mail: sieroka@phil.gess.ethz.ch

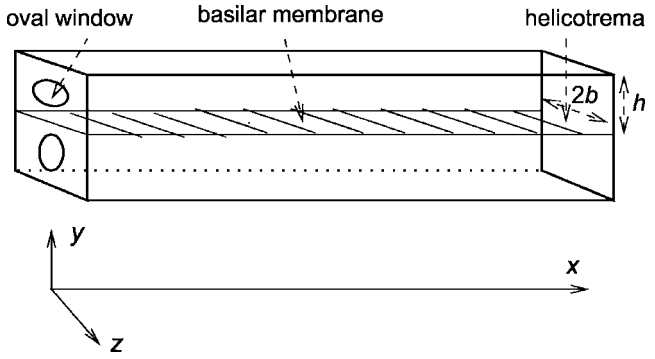


FIG. 1. Idealized model of the cochlea in side view.

II. MODEL OF THE INNER EAR

A. Geometry of the cochlea

The inner ear allows the transformation of acoustic into neuronal signals. The acoustically relevant part is the cochlea with its geometrical and mechanical features. Figure 1 shows the idealized model of the cochlea which forms the basis of our theoretical considerations. Here we shall give only a very short and rough sketch of the main mechanical features and refer for details to the excellent reviews of de Boer (1980, 1984, 1996) and the literature quoted there.

B. Dynamics

Generally the following approximations (cf. de Boer, 1980) are made:

(i) The friction in the lymph and its compressibility are neglected; the latter is justified for frequencies for which the sound wavelength in water is large compared to the linear dimension of the cochlea, that is for $\nu \ll 40\,000$ Hz.

(ii) Nonlinear contributions in the hydrodynamic Euler equations for an ideal liquid are not taken into account, thus excluding vortex solutions. This is justified by the smallness of the involved velocities and displacements.

For simplicity we consider only harmonic time dependence of the signal; in a linear model the general time dependence can be constructed by Fourier transform.

Let $p(\mathbf{x},\omega)$ be the hydrodynamic pressure at position $\mathbf{x}$ and the angular frequency $\omega=2\pi\nu$,

$$p(\mathbf{x},t) = p(\mathbf{x},\omega)e^{i\omega t}. \tag{1}$$

Under the assumptions discussed above the velocity is given by the gradient of the pressure:

$$\mathbf{v}(\mathbf{x},\omega) = \frac{i}{\omega\rho}\boldsymbol{\partial} p(\mathbf{x},\omega), \tag{2}$$

and the Laplace equation holds:

$$\boldsymbol{\partial}^2 p(\mathbf{x},\omega) = 0. \tag{3}$$

In the following we do not indicate the ω dependence of the pressure and velocity explicitly and denote by $p(x,y,z)$ the difference between the pressure in the upper and lower chamber at the fixed frequency ω:

$$p(x,y,z) = p_u(x,y,z,\omega) - p_l(x,-y,z,\omega). \tag{4}$$

The solution to the Laplace equation is specified by the boundary conditions, namely that the normal components of the velocity at walls of the cochlea are zero and that the normal component of the velocity at the basilar membrane is determined by the properties of the latter. Usually it is also assumed that the pressure difference is zero at the helicotrema. We replace this by the condition that only solutions corresponding to waves traveling to the right are admitted.

C. Model for the cochlea

In the rectangular model of Fig. 1 the following expression (de Boer, 1984) fulfills the boundary conditions at the walls of the cochlea, namely $v_y=0$ for $y=h$ and $v_z=0$ at $z=\pm b$:

$$p(\mathbf{x},\omega) = \sum_n \int_0^\infty \frac{dk}{2\pi} e^{-ikx} p_0(k) \left(\frac{\cosh[m_0(h-y)]}{\cosh[m_0 h]} + \eta \frac{m_0 \tanh[m_0 h]\cosh[m_1(h-y)]}{m_1 \tanh[m_1 h]\cosh[m_1 h]} \cos\left[\frac{\pi z n}{b}\right]\right). \tag{5}$$

In order to facilitate comparison with two-dimensional models we have introduced in (5) the constant η. For a three-dimensional model it is 1 and we have chosen this value in all applications of this paper. The corresponding two-dimensional model is obtained by putting $\eta=0$.

The basilar membrane is hinged at the borders of the cochlea, that is $v_y(x,0,\pm b)=0$. In the following we shall only consider the principal mode of excitations in the z direction, i.e., we confine ourselves to $n=1$.

The Laplace equation (3) yields

$$m_0 = \sqrt{k^2} = k; \quad m_1 = \sqrt{k^2 + \pi^2/b^2} \geqslant \max\left(k, \frac{\pi}{b}\right). \tag{6}$$

If the velocity at the middle of the basilar membrane can be expressed by an impedance ξ, one obtains

$$\partial_y p(x,0,0) = -i\omega\rho v_y(x,0,0) = \frac{2i\omega\rho}{\xi(x,\omega)} p(x,0,0). \tag{7}$$

Equations (5) and (7) yield the integral equation

$$\int_0^\infty dk\, e^{-ikx} Q(k)\hat{p}(k) = \frac{-i\omega\rho}{\xi(x,\omega)} \int_0^\infty dk\, e^{-ikx}\hat{p}(k), \tag{8}$$

with the kernel

$$Q(k) \equiv \frac{1+\eta}{2} \frac{k \tanh(kh) m_1 \tanh(m_1 h)}{\eta k \tanh(kh) + m_1 \tanh(m_1 h)}. \tag{9}$$

The expression $\hat{p}(k)$ is the Fourier transform of $p(x,0,0)$:

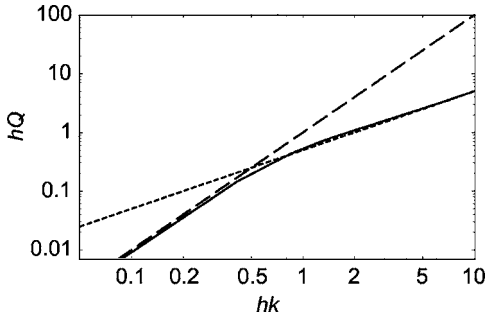


FIG. 2. Comparison of the exact kernel $Q(k)$, see Eq. (9), with its approximations given in (11); solid line: $hQ(k)$; long dashed line: h^2k^2 (long wave approximation); short dashed line: $hk/2$ (short wave approximation). The intersection point of the two dashed lines at $hk=\frac{1}{2}$ marks a sensible value for changing from a long to a short wave approximation.

$$\hat{p}(k) \equiv p_0(k)\left(1+\eta\frac{k\tanh(kh)}{m_1\tanh(m_1h)}\right)=\int_{\infty}^{\infty} dx\, e^{ikx}p(x,0,0). \tag{10}$$

The kernel can be simplified to monomial form in the two limiting cases of long and short waves:

$$Q\approx\begin{cases}\dfrac{1+\eta}{2}hk^2 & \text{for } kh\ll 1 \text{ long waves},\\ |k|/2 & \text{for } kh\gg 1 \text{ short waves}.\end{cases} \tag{11}$$

In these limiting cases it is more convenient to work with differential equations for $p(x,0,0)$ than with the integral equation (8). This is achieved by the replacement $\hat{p}(k)\rightarrow p(x)$, $k\rightarrow i\partial_x$, $k^2\rightarrow-\partial_x^2$. Here we have assumed that the waves are right moving so that we can replace $|k|$ by k. We come back to this point in Sec. IV (Fig. 11).

The local wave vector is defined as

$$k(x)\equiv i\partial_x\log[p(x,0,0)]. \tag{12}$$

The value of the local wave vector indicates which of the two conditions of (11) is more appropriate.

We follow the modeling of the impedance for the basilar membrane discussed extensively in de Boer (1980) and assume it to be of the following form:

$$2p=-\xi\partial_t Y=-\xi v_y,$$

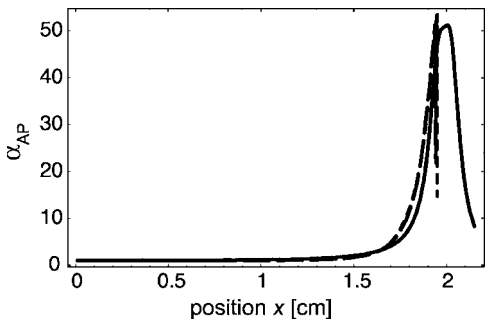


FIG. 3. The function $(\alpha_{AP})_n(x)$, see (34), for the first (solid), second (long dashes), and third (short dashes) iteration.

$$\xi=\frac{1}{i\omega}(S-\omega^2m+i\omega R), \tag{13}$$

with a transverse stiffness $S(x)$, a resistance $R(x)$, and an effective specific mass m:

$$S(x)=C_0e^{-\alpha x}-a;\quad R(x)=R_0e^{-\alpha x/2},\quad R_0\geqslant 0; \tag{14}$$

all other quantities are taken to be independent of x. The constant a in the stiffness $S(x)$ has been introduced in order to get into better agreement with physiological and psychoacoustic data for the resonance position (Greenwood, 1990). The numerical values of the parameters are given in Sec. III C.

III. ANALYTICAL SOLUTIONS AND APPROXIMATIONS

A. Long-, short- and combined-wave approximation (L, S, and LS models)

For $kh\ll 1$ the kernel in (8) is given by $Q\approx hk^2$ [see (11)]. The expression

$$p_L(x,0,0)=\sqrt{\frac{G_L(x)}{g(x)}}H_0^{(2)}(G_L(x)) \tag{15}$$

with

$$g(x,\omega)=\sqrt{\frac{2}{1+\eta}}\frac{\omega\sqrt{\rho}}{\sqrt{h(C_0e^{-\alpha x}-a-m\omega^2+i\omega R_0e^{-\alpha x/2})}}, \tag{16}$$

$$G_L(x,\omega)=\int_0^x dx'\, g(x')+\frac{2}{\alpha}g(0) \tag{17}$$

is a very good approximate solution for the resulting differential equation

$$\partial_x^2p(x,0,0)=\frac{2}{1+\eta}\frac{i\rho\omega}{h\xi(x,\omega)}p(x,0,0)=\frac{2}{1+\eta}\frac{-\omega^2\rho}{h(S(x)-\omega^2m+i\omega R(x))}p(x,0,0). \tag{18}$$

An analytical expression for $G_L(x)$ is given in the Appendix.

The local wave vector is given approximately by

$$k_L(x)=\sqrt{\frac{2}{1+\eta}}\omega\sqrt{\frac{\rho}{h(S(\omega)-\omega^2m+i\omega R(x))}}. \tag{19}$$

Note that $\eta=1$ in the three-dimensional case.

Near the resonance region $S(x)\approx mw^2$ the real part of the wave vector becomes very large and hence the long wave solution is no longer adequate.

For $kh\gg 1$, the kernel $Q\approx k/2$. The expression

$$p_S(x,0,0)=\exp[-G_S(x,\omega)] \tag{20}$$

with

$$G_S(x,\omega)=i\int_0^x dx'\,\frac{2\omega^2\rho}{S(x')-\omega^2m+i\omega R(x')}. \tag{21}$$

is a right-moving traveling wave solution to

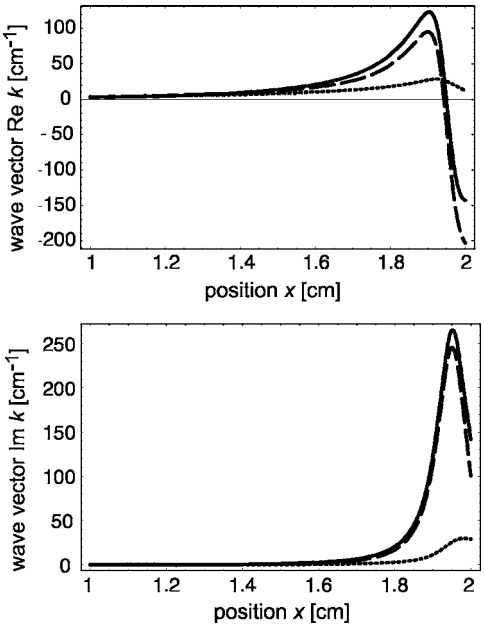


FIG. 4. Real part (top) and imaginary part (bottom) of the local wave vector for different models; the combined long and short wave (LS model, solid), the average pressure (AP model, long dashes), and the long wave (L model, dots) model; in this scale of the figure the curve for the short wave approximation (S model) cannot be distinguished from the LS model; $\nu=1200$ Hz, $\delta=0.15$.

$$\partial_x p(x,0,0) = \frac{-2i\omega^2\rho}{S(x)-\omega^2 m + i\omega R(x)} p(x,0,0). \tag{22}$$

In this case the local wave vector is

$$k_S(x) = \frac{2\omega^2\rho}{S(x)-\omega^2 m + i\omega R(x)}. \tag{23}$$

In Fig. 2 we show the quantity $hQ(k)$ (solid line), displayed as a function of kh (with $b=2h$). Also shown are the approximations k^2h^2 (long dashes) and $kh/2$ (short dashes). Since the transition between the two regimes at $kh=1/2$ is quite sharp, a promising strategy is to construct a continuous solution which satisfies the long wave equation (18) below this point and the short wave equation (22) above this point. If we approximate k by $k(x)$, the local wave vector (19) and (23), the transition point x_{LS} from long to short waves is determined by the condition

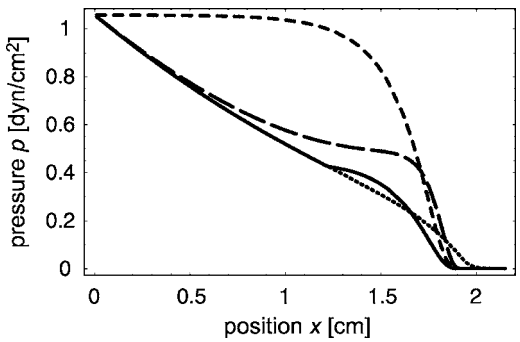


FIG. 5. Modulus of the pressure $p(x,0,0)$ for different models, solid LS, long dashes AP, short dashes S, and dots L model; $\nu=1200$ Hz, $\delta=0.15$, the pressure values are normalized to the same value at $x=0$.

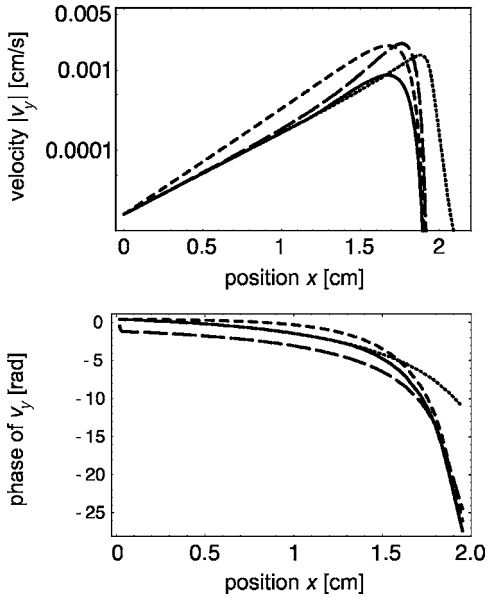


FIG. 6. Modulus (top) and phase (bottom) of the velocity of the basilar membrane for different models at high damping ($\delta=0.15$), solid LS, long dashes AP, short dashes S, and dots L model; $\nu=1200$ Hz.

$$\mathrm{Re}\left(\frac{1}{k_S(x_{LS})}\right) = \mathrm{Re}\left(\frac{1}{k_L(x_{LS})}\right), \tag{24}$$

that is

$$x_{LS} = \frac{1}{\alpha}\log\left(\frac{C_0}{2(1+\eta)\omega^2\rho h + a + m\omega^2}\right). \tag{25}$$

A continuous approximation to the solution in the whole cochlea is given by

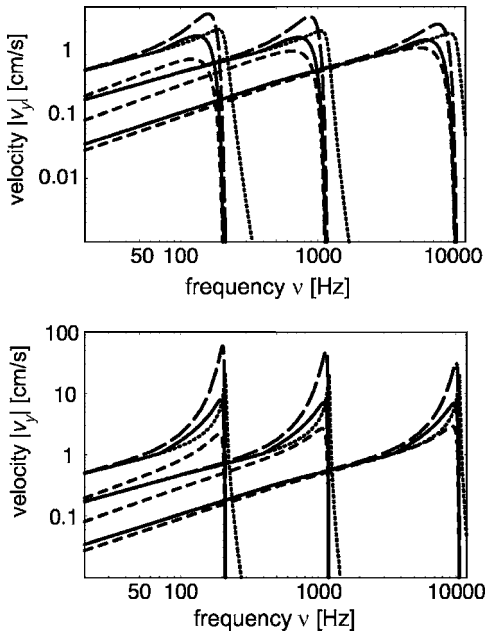


FIG. 7. The modulus $|v_y(x)|$ for the different models as a function of the frequency ν for three different positions on the basilar membrane, $x=0.5$, 1.95, and 3 cm; solid LS, long dashes AP, short dashes S, and dots L model. The pressure at the oval window is adjusted to give approximately equal maxima for each position. Top: damping parameter $\delta=0.15$; bottom: damping parameter $\delta=0.015$.

$$p_{LS}(x,0,0) = \begin{cases} p_L(x,0,0) & \text{for } x < x_{LS}, \\ p_L(x_{LS},0,0)e^{G_S(x)-G_S(x_{LS})} & \text{for } x_{LS} \leqslant x. \end{cases} \tag{26}$$

$p_L(x,0,0)$ is given in Eq. (15); the analytical expression for $G_S(x)-G_S(x_{LS})$ is given in the Appendix.

B. The averaged pressure approach by Duifhuis

The Laplace equation and the boundary conditions can be unified in a single equation if one considers the pressure integrated over the cross section $\mathcal{F}$ of the scala vestibuli,

$$\bar{p}(x) = \int_{\mathcal{F}} dy\, dz p(x,y,z). \tag{27}$$

Performing this integration in the Laplace equation (3) one obtains using Gauss's theorem and the fact that the normal velocity at the walls of the cochlea is zero

$$\partial_x^2 \bar{p}(x) = -\int_{-b}^{b} dz \partial_y p(x,0,z) = \frac{1}{\beta}\partial_y p(x,0,0). \tag{28}$$

From expression (5) follows $\beta=(1+\eta)/2b$.
Introducing the function

$$\alpha_{AP}(x) = \frac{2bhp(x,0,0)}{\bar{p}(x)}, \tag{29}$$

one obtains the differential equation for the integrated pressure:

$$\partial_x^2 \bar{p}(x) = \frac{2i\omega\rho}{\xi}\frac{\alpha_{AP}(x)}{(1+\eta)h}\bar{p}(x) = -\alpha_{AP}(x)(k_L(x,\omega))^2\bar{p}(x) \tag{30}$$

where k_L is given by (19).

This exact equation contains the unknown function $\alpha(x)$. If one approximates the Fourier variable k in Eq. (5) locally through the local wave vector $k_{AP}(x)$ of the solution of the differential equation (30), one obtains (Duifhuis, 1988)

$$\alpha_{AP}(x) = k_{AP}(x)h\left(\frac{1}{\tanh[k_{AP}(x)h]} + \eta\frac{k_{AP}(x)}{\sqrt{k_{AP}(x)^2+\pi^2/b^2}\tanh[\sqrt{k_{AP}(x)^2+\pi^2/b^2}h]}\right). \tag{31}$$

From (30) follows

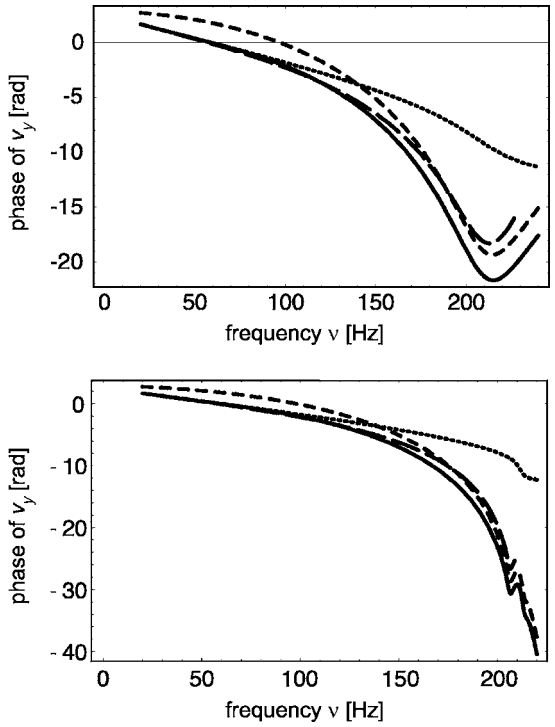


FIG. 8. The phase of $v_y(x)$ for the different models as a function of the frequency ν for the position $x=3$ cm on the basilar membrane; solid LS, long dashes AP, short dashes S, and dots L model. Top: damping parameter $\delta=0.15$; bottom: damping parameter $\delta=0.015$.

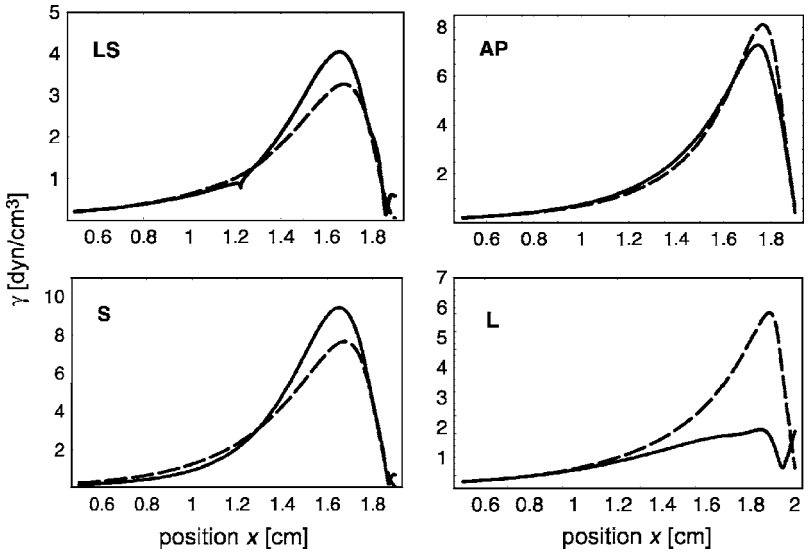


FIG. 9. Left-hand side $\gamma_l(x)$ (solid) and right-hand side $\gamma_r(x)$ (dashed) of the full integral equation (8) for different models as indicated. $\nu=1200$ Hz, $\delta=0.15$.

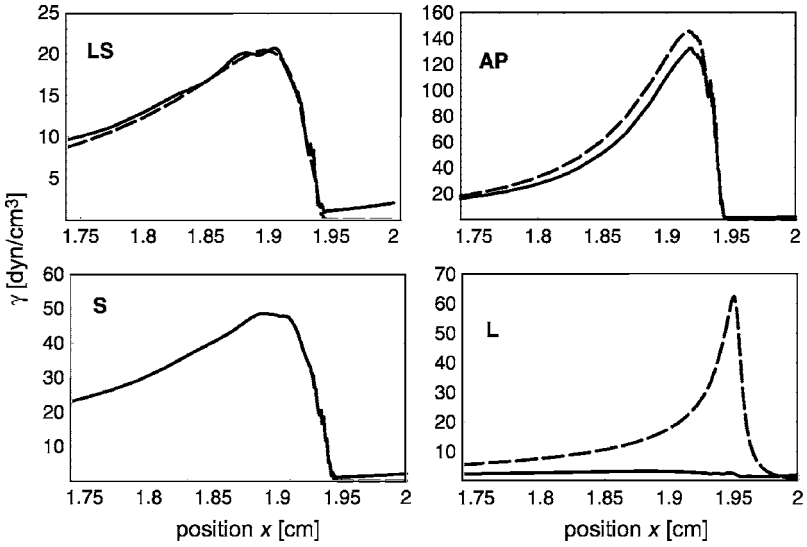


FIG. 10. Left-hand side $\gamma_l(x)$ (solid) and right-hand side $\gamma_r(x)$ (dashed) of the full integral equation (8) for different models as indicated. $\nu=1200$ Hz, $\delta=0.015$.

$$k_{AP}(x) \approx k_L(x)\sqrt{\alpha_{AP}(x)}, \tag{32}$$

and from that the transcendental equation (Duifhuis, 1988):

$$\sqrt{\alpha_{AP}} = k_L h\left(\frac{1}{\tanh[k_L\sqrt{\alpha_{AP}}h]} + \eta\frac{k_L\sqrt{\alpha_{AP}}}{\sqrt{k_L^2\alpha_{AP}+\pi^2/b^2}\tanh[\sqrt{k_L^2\alpha_{AP}+\pi^2/b^2}h]}\right) \tag{33}$$

For values of x below the resonance position (see Fig. 3) this equation can be approximately solved by iteration:

$$\sqrt{(\alpha_{AP})_n} = k_L h\left(\frac{1}{\tanh[k_L\sqrt{(\alpha_{AP})_{n-1}}h]} + \eta\frac{k_L\sqrt{(\alpha_{AP})_{n-1}}}{\sqrt{k_L^2(\alpha_{AP})_{n-1}+\pi^2/b^2}\tanh[\sqrt{k_L^2(\alpha_{AP})_{n-1}+\pi^2/b^2}h]}\right) \tag{34}$$

with $(\alpha_{AP})_0=1$.

C. Comparison of the models

We shall now compare the different models discussed above:

(i) **LS**, the combined long-short wave model (26),
(ii) **AP**, the average pressure model of Sec. III B,
(iii) **L**, the pure long-wave model (15), and
(iv) **S**, the pure short-wave model (20).

The following parameter values were taken from de Boer (1980): $\rho=1.0\text{g/cm}^3$, $h=0.1$ cm, $m=0.05\text{g/cm}^2$, $C_0=10^9\text{g/(s}^2\text{cm}^2)$, $\alpha=3.0\text{ cm}^{-1}$, and for the additional constant $a=35\,000\text{g/(s}^2\text{cm}^2)$. The dimensionless constant $\delta=R_0/(\sqrt{C_0 m})$, specifying the damping of the membrane, was normally set to 0.15. This value was determined heuristically by adjusting the duration of poststimulus ringing in the LS model to that given by the transmission line filterbank adequate for a SPL of about 65 dB.

In Fig. 3 different iterations for $(\alpha_{AP})_n$ for $\nu=1200$ Hz are displayed. For $n\geq 2$, $(\alpha_{AP})_n$ becomes unstable beyond the resonance position of the impedance function, for our parameters and $\nu=1200$Hz beyond $x=1.95$ cm. This is a consequence of the multiple roots of (33); for a more detailed discussion see Duifhuis (1988). They may play some role for reflections and emissions (Shera *et al.*, 2005). In the following we shall use the well-behaved first iteration $(\alpha_{AP})_1$.

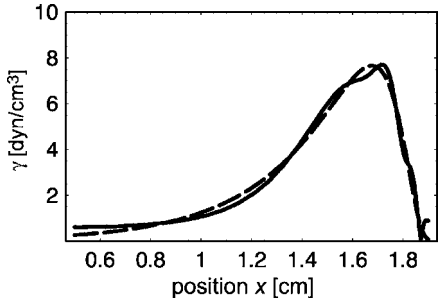


FIG. 11. Left-hand side $\gamma_l(x)$ (solid) and right-hand side $\gamma_r(x)$ (dashed) of the integral equation (8) with $Q=|k/2|$ for the S model, for $\nu=1200$ Hz, $\delta=0.15$.

The local wave vectors for the different models are displayed in Fig. 4; we note the similarity between the LS model and the AP model. The typical feature of these two models as compared to the widely used L model is a sharp increase of the imaginary part well below the resonance position of the impedance. This leads to a sharp decrease in the pressure before the resonance point as can be clearly seen in Fig. 5. Since this happens in a region of strong variation of the impedance, the influence on the velocity of the basilar membrane is even more marked, as can be seen in Fig. 6. At smaller values of the damping constant the difference between the models becomes even more apparent.

In Figs. 7 and 8 the modulus and the phase of the transverse velocity v_y is displayed as function of the frequency ν, for different positions on the basilar membrane, for the different models, and for different damping parameters.

IV. DISCUSSION

The marked differences of the solutions, especially for small damping, necessitate a discussion of the validity of the different models. We expect the S model to be inadequate far below the resonance point and the L model to be inadequate near the resonance point. In the transition region the AP model is certainly more reliable than the LS model, but the determination of $\alpha_{AP}(x)$, as discussed in Sec. III B, might become questionable very near the resonance region, where the local wave vector depends strongly on x; see (33) and (34), and Fig. 3.

In order to quantify these results we insert the solutions of the different models in the integral equation (8) with the full kernel (9). We define the following internal quantities from (8): the left-hand side

$$\gamma_l(x) \equiv \int_0^\infty dk\, e^{-ikx} Q(k) \hat{p}(k)$$

and right-hand side

$$\gamma_r(x) \equiv \frac{-i\omega\rho}{\xi(x,\omega)} \int_0^\infty dk\, e^{-ikx} \hat{p}(k),$$

which should be equal for an exact solution. In Fig. 9 we display the left-hand side $\gamma_l(x)$ (solid) and the right-hand side $\gamma_r(x)$ (dashed) for the different models, with strong damping (δ=0.15). As expected the L model fails badly in the resonance region. The agreement between $\gamma_l(x)$ and $\gamma_r(x)$ is slightly better for the AP than for the LS model, especially below the peak. For the latter the artefact due to the joining of different solutions is clearly visible at x=1.22 cm. Above the peak the LS model fares slightly better than the AP model. The difference between the right-and left-hand sides has opposite sign for the AP and the LS models.

For small damping (δ=0.015, Fig. 10) the agreement between $\gamma_l(x)$ and $\gamma_r(x)$ in the resonance region is much better for the LS than for the AP model. This is in agreement with the expectation that the estimate of $\alpha_{AP}(x)$ from the local wave vector is worse for rapidly varying local wave vectors than for slowly varying ones; see (33) and (34), and Fig. 3.

The characteristic differences between the LS and AP models on the one side and the L model on the other side are due to the fact that the imaginary part of the wave vector increases much faster for the realistic models than for the L model. This leads to a considerable broadening of the peak in the v_y amplitude and a shift of its position to the left. Part of that can be compensated by a change of the parameters of the basilar membrane, but the different behavior of the phase (Fig. 8), e.g., is very stable against a change of parameters. Furthermore, if one has a hope to relate the model parameters with physical properties of the cochlea one needs to use a realistic and consistent model.

This simple S model with its exact solution to the linear differential equation (22) describes quite well the qualitative features, especially for small damping. The differential equation (22) corresponds to a replacement of $|k|$ by k in the approximate short-wave kernel (11). This has no large influence, as can be seen from Fig. 11. There the solution of the differential equation (22) has been inserted into (8) with the kernel $Q=|k|/2$. The agreement between both sides is still satisfactory. Therefore the S model may be useful as a simple orientation, especially for small damping.

A realistic treatment of the cochlea has to take into account nonlinear effects due to energy input through the outer hair cells. It is a convenient feature of the LS model that nonlinearities, as discussed, e.g., in Duke and Jülicher (2003), can be incorporated into it as easily as into the L model.

ACKNOWLEDGMENT

This project was supported by the Deutsche Forschungsgemeinschaft (Ru 652/1-3).

APPENDIX: ANALYTICAL EXPRESSIONS

Analytical expressions for the function G_L [(17)] and G_S [(21) and (26)]:

$$G_L(x,\omega) = \frac{2g(0,\omega)}{\alpha} - \frac{2\sqrt{2}i\omega}{\alpha\sqrt{h(\omega^2 m + a)}} \times \log((-2ie^{\alpha/2x}(m\omega^2 + a) - \omega R_0 + 2\sqrt{m\omega^2 + a}\sqrt{C_0 - (m\omega^2 + a)e^{\alpha x} + i\omega R_0 e^{\alpha x/2}}) \times ($$
$$-2i(m\omega^2 + a) - \omega R_0 + 2\sqrt{m\omega^2 + a}\sqrt{C_0 - (m\omega^2 + a) + i\omega R_0})^{-1});$$

$$
\begin{aligned}
G_S(x,\omega) - G_S(x_{LS},\omega) = \frac{1}{\alpha(a+m\omega^2)}\omega^2\Bigg(& 2\alpha x_{LS} - 2\alpha x - 2i\arctan\left(\frac{e^{\alpha x_{LS}/2}\omega R}{-C_0 + e^{\alpha x_{LS}}(a+m\omega^2)}\right) \\
& + 2i\arctan\left(\frac{e^{\alpha x/2}\omega R}{-C_0 + e^{\alpha x}(a+m\omega^2)}\right) \\
& - \frac{4i\omega R\sqrt{-4aC_0+\omega^2(-4C_0m+R^2)}\,\arctan(2C_0e^{-\alpha x_{LS}/2} + i\omega R/\sqrt{-4aC_0+\omega^2(-4C_0m+R^2)})}{4aC_0+\omega^2(4C_0m-R^2)} \\
& - \frac{4i\omega R\,\arctan(2C_0e^{-\alpha x/2} + i\omega R/\sqrt{-4aC_0+\omega^2(-4C_0m+R^2)})}{\sqrt{-4aC_0+\omega^2(-4C_0m+R^2)}} + \log\left(a^2 + \frac{C_0^2}{e^{2\alpha x_{LS}}} - \frac{2aC_0}{e^{\alpha x_{LS}}} + 2am\omega^2\right. \\
& \left. - \frac{2C_0m\omega^2}{e^{\alpha x_{LS}}} + m^2\omega^4 + \frac{\omega^2R^2}{e^{\alpha x_{LS}}}\right) - \log\left(a^2 + \frac{C_0^2}{e^{2\alpha x}} - \frac{2aC_0}{e^{\alpha x}} + 2am\omega^2 - \frac{2C_0m\omega^2}{e^{\alpha x}} + m^2\omega^4 + \frac{\omega^2R^2}{e^{\alpha x}}\right) \\
& + i\pi\,\mathrm{Sign}(-C_0 + e^{\alpha x_{LS}}(a+m\omega^2)) - i\pi\,\mathrm{Sign}(-C_0 + e^{\alpha x}(a+m\omega^2))\Bigg).
\end{aligned}
$$

Aertsen, A. M. H. J., and Johannesma, P. I. M. (**1980**). "Spectro-temporal receptive fields of auditory neurons in the grassfrog," Biol. Cybern. **38**, 223–234.

de Boer, E. (**1980**). "Auditory physics. Physical principles in hearing theory. I," Phys. Rep. **62**, 87–174.

de Boer, E. (**1984**). "Auditory physics. Physical principles in hearing theory. II," Phys. Rep. **105**, 141–226.

de Boer, E. (**1996**). "Mechanics of the cochlea: Modeling efforts, " in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 258–317.

Duifhuis, H. (**1988**). "Cochlear macromechanics," in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Covan (Wiley, New York), pp. 189–211.

Duke, T., and Jülicher, F. (**2003**). "Active traveling wave in the cochlea," Phys. Rev. Lett. **90**, 15801.

Giguère, C., and Woodland, P. C. (**1994**). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," J. Acoust. Soc. Am. **95**, 331–342.

Giguère, C. Smoorenburg, G. F. and Kunov, H. (**1997**). "The generation of psychoacoustic combination tones in relation to two-tone suppression effects in a computational model," J. Acoust. Soc. Am. **102**, 2821–2830.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Irino, T., and Patterson, R. D. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. **109**, 2008–2022.

Ranke, O. F. (**1950**). "Theory of operation of the cochlea: a contribution to the hydrodynamics of the cochlea," J. Acoust. Soc. Am. **22**, 772–777.

Schroeder, M. R. (**1975**). "Models of hearing," Proc. IEEE **63**, 1332–1350.

Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflexion in a two-dimensional cochlea: short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.

Siebert, W. M. (**1974**). "Ranke revisited—a simple short-wave cochlear model," J. Acoust. Soc. Am. **56**, 594–600.

Strube, H. W. (**1985**). "A computionally efficient basilar-membrane model," Acustica **58**, 207–214.

Viergever, M. A., and Kalker, J. J. (**1975**). "A two-dimensional model for the cochlea," J. Eng. Math. **9**, 353–365.

Zwislocki, J. (**1950**). "Theory of the acoustical action of the cochlea," J. Acoust. Soc. Am. **22**, 778–784.

Zwislocki, J. (**1953**). "Review of the recent mathematical theories of cochlear dynamics," J. Acoust. Soc. Am. **25**, 743–751.

A potential carry-over effect in the measurement of induced loudness reduction

Michael Epstein[a)]
Institute for Hearing, Speech, and Language, Communication Research Laboratory, Auditory Modeling and Processing Laboratory, Department of Speech-Language Pathology and Audiology (106A FR), and Communications and Digital Signal Processing Center, ECE Department (440 DA), Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Elizabeth Gifford
Institute for Hearing, Speech and Language, Communication Research Laboratory and Department of Speech-Language Pathology and Audiology (106A FR), Northwestern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 30 June 2005; revised 9 April 2006; accepted 11 April 2006)

The majority of studies on induced loudness reduction (ILR) use an experimental paradigm that results in an underestimation of the amount of ILR. Most of those studies utilize loudness matches between tones of two different frequencies (a test tone and a comparison tone) with (experimental condition) and without (baseline condition) an inducer tone at the test frequency. The change in level of the comparison tone between the baseline and experimental conditions is the amount of ILR. In those experiments, the level of the comparison tone in the baseline condition tends to be substantially higher (often about 10 dB) than in the experimental condition. Because of this level difference, exposure to the baseline condition immediately prior to the experimental condition causes unintended ILR for the comparison tone. In this study, the delay between the baseline and experimental conditions was varied and it was determined that the amount of ILR is underestimated by about 30% and the variability is increased when the experimental condition is run immediately after the baseline condition. A second experiment using a Békésy-tracking procedure showed that ILR maximizes rapidly upon exposure to an inducer and decays over the course of several minutes after the inducer is removed. [DOI: 10.1121/1.2202867]

PACS number(s): 43.66.Cb [JHG] Pages: 305–309

I. INTRODUCTION

Induced loudness reduction (ILR), previously called loudness recalibration (Marks, 1994; Arieh and Marks, 2001; Arieh and Marks, 2003), is a phenomenon by which the loudness of a sound is reduced when it is preceded by a higher-intensity sound (inducer) at a close frequency with an equal or longer duration (Marks and Warner, 1991; Nieder *et al.*, 2003). The majority of studies on this topic have used a procedure to determine the amount of ILR that involves the presentation of a test tone and an inducer tone at one frequency, and a comparison tone at another frequency under two conditions. In the first condition, called the "baseline," the test and comparison tones are presented without the inducer in order to find the point of subjective equality for the loudness of the two tones by having the subject adjust the level of the comparison tone. Then, an "experimental" condition immediately follows with the regular presentation of the inducer (usually every trial) and a second measurement of the point of subjective equality for the loudness of the test and comparison tones. In the experimental condition, the loudness of the test tone is reduced by the inducer and therefore, the comparison tone is typically set to a lower level than in the baseline condition to achieve equal loudness. The amount of ILR is defined as the decrease in the level of the comparison tone from the baseline to the experimental condition. The amount of ILR under this experimental paradigm has been reported to be around 10 dB (e.g., Nieder *et al.*, 2003). However, some studies have shown that there is some effect of frequency on amount of ILR, and that the effect of ILR may be greater at lower frequencies (Marks and Warner, 1991; Mapes-Riordon and Yost, 1999), though others do not show this effect (Wagner and Scharf, 2006).

It is also known that ILR lasts for at least several minutes (Mapes-Riordon and Yost, 1999; Arieh *et al.*, 2005; Wagner and Scharf, 2006). The above described experimental paradigm assumes that ILR only affects the loudness of the test tone (Marks, 1994; Scharf *et al.*, 2002). However, because the baseline condition results in the comparison tone being set to a higher level than in the experimental condition, it is possible that the repeated presentation of the comparison tone in the baseline condition could result in ILR on the comparison tone in the experimental condition. For example, in a baseline condition a listener may set the comparison tone to roughly the same level as the test tone; 70 dB SPL perhaps. In the experimental condition, the loudness of the 70 dB SPL test tone would be reduced making it likely that the comparison tone might be set to 60 dB SPL to achieve equal loudness. In fact, the amount of ILR (the difference between the comparison tones in the two conditions) is often around 10 dB. A difference of 10 dB between the inducer

[a)]Electronic mail: m.epstein@neu.edu

and the test tone results in nearly maximal ILR under the usual test conditions (Mapes-Riordon and Yost, 1999; Nieder *et al.*, 2003). In this scenario, the comparison tone, which is not intended to undergo the effects of ILR, is on average about 70 dB SPL in the baseline condition and about 60 dB SPL in the experimental case. This sequence is highly likely to cause ILR for the comparison tone.

The purpose of this paper is primarily to determine whether the procedure typically used for determining the amount of ILR is susceptible to unexpected contextual effects on the comparison tone. Second, if the commonly used procedure does have unexpected contextual effects, then it is necessary to determine how that procedure must be altered so that it does not have those undesired effects. In order to do so, the exact time course of ILR must be understood. Some studies have shown portions of the time course of ILR (Arieh and Marks, 2001; Arieh *et al.*, 2005), but none has closely examined the full course in a single session. In addition, these studies have always used the paradigm that may be susceptible to unexpected contextual effects.

II. METHOD

In Experiment 1, the amount of ILR was measured as a function of the delay between the baseline and the experimental condition using the typical experimental paradigm at three different delays: directly after the baseline condition was run, 15 min after the baseline condition was run, and several hours after the baseline condition was run in order to determine whether the baseline condition itself causes ILR.

In Experiment 2, the amount of ILR and recovery from ILR were measured as a function of time using a Békésy-tracking procedure. An alternating test-comparison tone pattern was presented continuously without waiting for a listener response. In the experimental condition, an inducer preceded these tones. The comparison tone changed level with each presentation and the direction of the level change was controlled by the subject responses. In this experiment, the measurements began with the experimental condition and after 6 min, the inducer was removed and recovery back to baseline was observed for 9 min.

A. Stimuli

The test tone and inducer were 2500 Hz and the comparison tone was 500 Hz. The level of the inducer was 80 dB SPL and the level of the test tone was 70 dB SPL. In Experiment 1, the comparison had an initial level of 60 dB SPL in order to start below the likely level match in the baseline condition and near the level match in the experimental condition. In Experiment 2, the comparison tone started at a level near the expected level of the comparison tone during the experimental condition determined from Experiment 1. All tones were presented monaurally and had equivalent rectangular durations of 200 ms.

B. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, recorded the listeners' responses, and executed the procedure. The sample rate was 48 kHz. The output of the 16 bit D/A converter (TDT DD1) was attenuated (TDT PA4), low-pass filtered (TDT FT5, f_c = 20 kHz, 135 dB/oct), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed one earphone of a Sony MDR-V6 headset. Listeners sat in a sound-attenuating booth (Acoustic Systems), and the stimuli were presented monaurally to the preferred ear. For routine calibration, the output of the headphone amplifier was led to a 16 bit A/D converter (TDT DD1) such that the computer could sample the wave form, calculate its spectrum and rms voltage, and display the results before each block of trials. The SPLs reported in the following assume a frequency-independent output at the earphone of 116 dB SPL for an input of 1 V rms.

C. Procedure

1. Experiment 1

Listeners participated in two conditions: a baseline and an experimental condition. In the baseline condition, only the test tone and the comparison tone were presented with a 500 ms silent period in between. The listener's task was to indicate which sound was louder by pressing a key on a small computer terminal. The response initiated the next trial after a 1000 ms delay. No feedback was provided. In the experimental condition, an inducer preceded the other tones by 800 ms.

The level of the comparison tone was adjusted according to a simple up-down method (Jesteadt, 1980). If the listener indicated that the comparison tone was louder, its level was reduced; otherwise it was increased. The step size was 5 dB until the second reversal after which it was reduced to 2 dB. A track ended after nine reversals. This procedure converges at the level corresponding to the 50% point on the psychometric function (Levitt, 1971). The equal-loudness level for each track was calculated as the average of the last four reversals.

To reduce biases that may occur when only a single fixed sound is presented in a series of trials, two interleaved tracks were used to obtain concurrent loudness matches. This procedure was run twice and data from all four tracks were averaged to obtain the baseline result. The experimental condition was similar to the baseline condition except that it began with 12 presentations of the inducer separated by 200 ms and included an inducer preceding the test tone and variable comparison tone by 800 ms on each trial. The listener was told to ignore the inducer.

The baseline condition was run twice directly followed by the experimental condition twice. The two baseline and two experimental blocks lasted about 10 min total. Then, after at least 2 h had passed, the experimental condition was run twice again. The same baseline value for the comparison tone was assumed. On a separate day, the baseline condition was run twice followed by a 15 min break after which the experimental condition was run twice.

The whole of Experiment 1 was run twice and all eight tracks from the corresponding experimental conditions were averaged to determine the amount of ILR.

2. Experiment 2

The second experiment was conducted using a modified Békésy procedure. The trials were presented continuously without waiting for a listener response. For the first 90 trials, an inducer was presented 1250 ms before the test and variable tone, which were separated by 550 ms and followed by 1600 ms of silence for a total trial time of 4 s. For the remaining 150 trials, the inducer was replaced with silence. The listener's task was to ignore the inducer when present and to indicate whether the test tone or the comparison tone was louder by pressing a key on a small computer terminal.

On each trial, if the last response indicated that the comparison tone was louder then the level was increased by 2 dB, otherwise it was reduced by 2 dB. If the listener did not respond to a particular trial, the level continued to move in the same direction as from the previous trial. The experiment was run twice and the two runs were averaged. The amount of ILR for each listener was determined by subtracting the average of 20 stable trials near the end of the trials with the inducer (the experimental level) from 20 stable trials near the end (the baseline level). Because listeners did not always stabilize their judgments at the same time during the experiment, it was difficult to select a group of trials that was suitable for ILR estimation for all listeners.

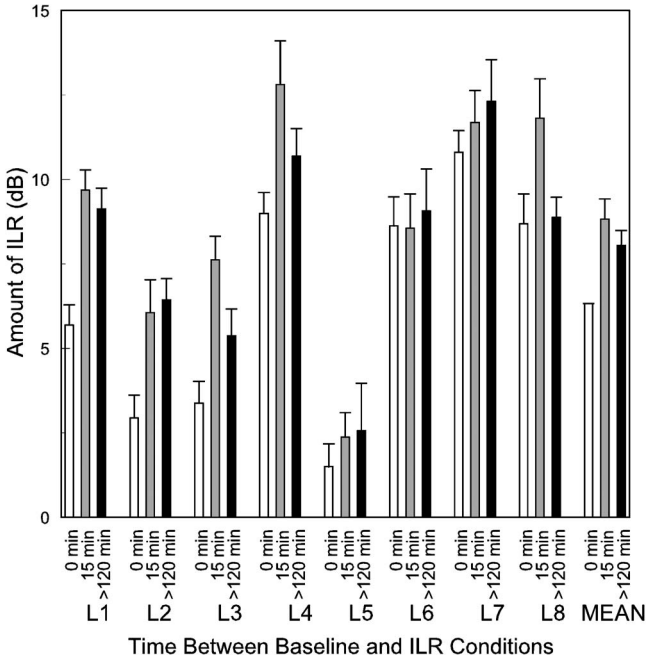


FIG. 1. The amount of ILR for eight listeners after each of three time delays between the baseline and experimental conditions. Error bars indicate the standard errors. The error bars on the mean show a standard error for which all individual baseline conditions were normalized to the overall mean for the 0 min condition in order to eliminate intersubject variance.

D. Listeners

The same eight listeners, four male and four female, participated in both experiments. None had a history of hearing loss and all had pure-tone thresholds at or below 10 dB HL at 500 and 2500 Hz (ANSI, 1996). They ranged in age from 20 to 31 years. All participants were members of the laboratory and volunteered participation.

III. RESULTS AND DISCUSSION

A. Experiment 1

Figure 1 shows the amount of ILR measured under each of the three delay conditions. The error bars show standard errors. The error bars on the mean show a standard error for which all individual baseline conditions were normalized to the overall mean in order to eliminate intersubject variance.

The overall amount of ILR is less than seen in several similar studies, but is within the overall range seen in the literature (Marks, 1994; Mapes-Riordon and Yost, 1999; Nieder *et al.*, 2003). Although there is some variability, all listeners showed more ILR in the experimental conditions when the measurements were not made directly after the baseline condition, except for the 15 min condition for listener L6. It is possible that recovery rates differ in individuals and that some listeners are not affected as profoundly by the baseline condition.

A univariate analysis of variance (SPSS 11.5) indicates that time delay had a significant effect on the amount of ILR ($p<0.001$; $df=2168$; $F=17.552$). Scheffe post-hoc analysis showed that significant differences existed between the amounts of ILR measured when the experimental condition was run with no delay and with delays of 15 and at least 120 min ($p<0.001$ for both). However, the difference in the amount of ILR between the two delayed conditions was not found to be significant ($p=0.204$).

Because two experimental blocks of trials were run after

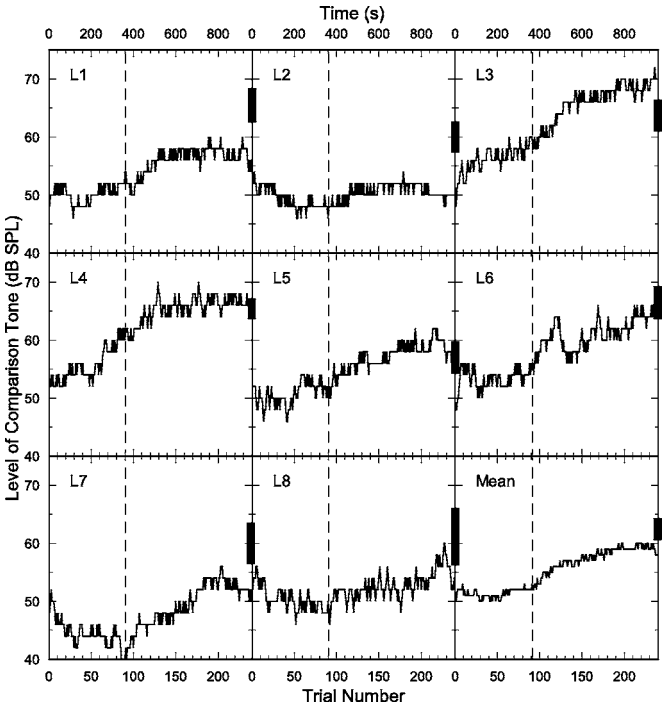


FIG. 2. The level of the comparison tone is shown as a function of trial number and time for eight listeners and the mean. The inducer was present until trial 90, which is marked with a dashed line. The black bar on the right of each graph indicates the baseline measured in Experiment 1± one standard deviation.

TABLE I. Summary of the amount of ILR found for all eight listeners and the mean and standard error of the mean in Experiments 1 and 2.

	L1	L2	L3	L4	L5	L6	L7	L8	Mean	S.E
After 0 min	5.69	2.94	3.38	9.00	1.50	8.63	10.81	8.69	6.33	1.21
After 15 min	9.69	6.06	7.63	12.81	2.38	8.56	11.69	11.81	8.83	1.23
After >120 min	9.13	6.44	5.38	10.69	2.56	9.06	12.31	8.88	8.05	1.10
Békésy	6.29	3.62	11.71	12.48	7.05	11.24	9.33	3.62	8.17	1.25

two baseline blocks of trials, an additional ANOVA was performed to determine whether there was a significant effect of time in even the short span between the start of the two blocks (approximately 2.5 min). Although all subjects did show slightly more ILR for the second block (the mean was 0.9 dB more with a standard error of 0.25), the statistical difference between the two trials was not significant ($p=0.088; df=148; F=3.035$). In contrast, in the two delayed experimental conditions, there was slightly less ILR in the second block than the first block (the mean was 0.3 dB less with a standard error of 0.47). However, two listeners showed more ILR in the second block. An ANOVA showed no effect of trial ($p=0.50; df=1112; F=0.452$). The contrast in these results between the delayed and undelayed conditions makes it seem possible, though not certain, that some small recovery effect might be present, even over the course of just the time it takes to run each experimental block. This also indicates that it is likely that ILR accumulates and reaches a maximum relatively rapidly and recovery at least begins quickly after the inducer is removed, consistent with other work (Arieh *et al.*, 2005; Wagner and Scharf, 2006).

Overall, these statistics indicate that it is likely that the baseline condition does cause ILR on the comparison tone in the experimental condition. Because there was no statistical difference between the amounts of ILR in the delayed conditions, a delay of 15 min is likely to be sufficient to recover from the ILR that results from the baseline condition.

B. Experiment 2

Figure 2 shows the level of the comparison tone as a function of trial number for each of the eight listeners and for the mean of all listeners. At trial 90, when the inducer tone was shut off, the listeners started recovery and generally stabilized near the end of the experiment. As seen in the results of Experiment 1, Fig. 2 also confirms that generally, recovery occurs in less than 15 min. However, the amount of time needed for recovery varied from listener to listener. The ending value of the comparison tone should be equivalent to the value of the comparison tone in the baseline condition of Experiment 1. The mean baseline value was 59.25 dB SPL with a standard error of 2.33 for Experiment 2 and 62.50 dB SPL with a standard error of 1.25 for Experiment 1. For some individuals, the baseline value found in Experiment 1 differs markedly from the baseline value found in Experiment 2, but no overall pattern of difference was evident. It is possible that a central-tendency bias had some role in reducing the final level (the baseline) in Experiment 2 for some of the listeners. When listeners are given the opportunity to directly control the level of a stimulus, some tend to adjust more toward moderate levels than high or low levels (Stevens and Greenbaum, 1966).

Because the baseline condition was not run prior to the experimental condition and the starting level generally remained below or slightly above the final stabilized matching level, the comparison tone should not be affected by ILR in Experiment 2. The amount of ILR for both experiments is summarized in Table I.

The mean amounts of ILR determined in the delayed condition in Experiment 1 and the Békésy method used in Experiment 2 show nearly identical amounts of ILR and all conditions had nearly identical standard errors. Some listeners showed very different amounts of ILR in the two experiments. The Békésy method demonstrates that it is necessary to wait as much as 10–15 min for the undesired ILR on the comparison tone to dissipate in order to obtain an accurate account of the amount of ILR. This method also exposes the rapidity of the onset of ILR when inducer exposure begins and the slow course of recovery when the inducer is removed.

IV. CONCLUSIONS

The method typically used for measurements of the amount of ILR is susceptible to unintentional effects resulting from the baseline condition. The comparison tone is set to a higher level in the baseline condition than in the experimental condition in order to match the level of the test tone. If the experimental condition is run immediately after the baseline condition, then the level difference is sufficient to produce ILR in the comparison tone. This effectively reduces the amount of ILR seen in the test tone and results in an underestimation of the quantity of ILR. A delay of 15 min between the baseline and experimental conditions was shown to be sufficient to avoid this effect. Additionally, a Békésy-style tracking method showed the specific time course of recovery and provided evidence that ILR onset is rapid and the rate of recovery from ILR is relatively slow.

ACKNOWLEDGMENTS

We wish to thank Eva Wagner, Jeremy Marozeau, Bertram Scharf, reviewer Yoav Arieh, and an anonymous reviewer for helpful comments on an earlier draft of this manuscript. This research was supported by NIH/NIDCD Grant No. R01DC02241.

ANSI (**1996**). "American National Standard Specification for Audiometers," Journal ANSI S3.6-1996.

Arieh, Y., Kelly, K., and Marks, L. E. (**2005**). "Tracking the time to recovery

after induced loudness reduction (L)," J. Acoust. Soc. Am. **117**, 3381–3384.
Arieh, Y., and Marks, L. E. (**2001**). "Recalibration of loudness: Sensory vs. decisional processes," in *Fechner Day 2001*, edited by Sommerfeld, E., Kompass, R. and Lachmann, T. (Pabst, Berlin).
Arieh, Y., and Marks, L. E. (**2003**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.
Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.
Marks, L. E. (**1994**). "'Recalibrating' the auditory system: The perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.
Marks, L. E., and Warner, E. (**1991**). "Slippery context effect and critical bands," J. Exp. Psychol. Hum. Percept. Perform. **17**, 986–996.
Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test- and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846–2855.
Scharf, B., Buus, S., and Nieder, B. (**2002**). "Loudness enhancement: Induced loudness reduction in disguise? (L)," J. Acoust. Soc. Am. **112**, 807–810.
Stevens, S. S., and Greenbaum, H. B. (**1966**). "Regression effect in psychophysical judgement," Percept. Psychophys. **1**, 439–446.
Wagner, E., and Scharf, B. (**2006**). "Induced loudness reduction as a function of exposure time and signal frequency," J. Acoust. Soc. Am. **112**, 1012–1020

Spectral and threshold effects on recognition of speech at higher-than-normal levels

Judy R. Dubno,[a] Amy R. Horwitz, and Jayne B. Ahlstrom
Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Avenue, P.O. Box 250550, Charleston, South Carolina 29425

(Received 30 May 2005; revised 26 April 2006; accepted 28 April 2006)

To examine spectral and threshold effects for speech and noise at high levels, recognition of nonsense syllables was assessed for low-pass-filtered speech and speech-shaped maskers and high-pass-filtered speech and speech-shaped maskers at three speech levels, with signal-to-noise ratio held constant. Subjects were younger adults with normal hearing and older adults with normal hearing but significantly higher average quiet thresholds. A broadband masker was always present to minimize audibility differences between subject groups and across presentation levels. For subjects with lower thresholds, the declines in recognition of low-frequency syllables in low-frequency maskers were attributed to nonlinear growth of masking which reduced "effective" signal-to-noise ratio at high levels, whereas the decline for subjects with higher thresholds was not fully explained by nonlinear masking growth. For all subjects, masking growth did not entirely account for declines in recognition of high-frequency syllables in high-frequency maskers at high levels. Relative to younger subjects with normal hearing and lower quiet thresholds, older subjects with normal hearing and higher quiet thresholds had poorer consonant recognition in noise, especially for high-frequency speech in high-frequency maskers. Age-related effects on thresholds and task proficiency may be determining factors in the recognition of speech in noise at high levels. [DOI: 10.1121/1.2206508]

PACS number(s): 43.66.Dc, 43.71.Es, 43.66.Sr [JHG] Pages: 310–320

I. INTRODUCTION

Studebaker *et al.* (1999) observed declines in word recognition at higher-than-normal speech and noise levels for listeners with and without hearing loss. To explore these effects, recognition of monosyllabic words in high- and low-context sentences from the Speech Perception in Noise test (SPIN; Kalikow *et al.*, 1977) was measured over a wide range of speech and noise levels with "effective" signal-to-noise ratio for each subject held constant (Dubno *et al.*, 2000). Subjects were younger and older adults with normal hearing. The results suggested that over the range of speech and noise levels used, key word recognition in sentences generally remained constant or decreased only slightly.

Based on these findings and others, several questions remained unanswered regarding the detrimental effects of high speech and noise levels on speech recognition. First, the mechanism responsible for the decline was not known, in part because masked thresholds at frequencies within the spectrum of the speech were rarely measured. Second, the influence of experimental variables such as speech material, masker type, and spectral content of the speech and/or masker was unclear. Studebaker *et al.* (1999) concluded that the largest effects were observed for nonsense syllables and monosyllabic words and when speech was presented in a masker with a spectrum that matched the speech spectrum. Molis and Summers (2003) observed that recognition of key words in sentences in quiet declined at high levels more for high-pass-filtered sentences than for low-pass-filtered sentences. Analysis of patterns of consonant confusions by Hornsby *et al.* (2005) also suggested a differential effect of spectral content in that the duration and place-of-articulation features (e.g., higher frequency cues) were more affected by high levels than the voicing feature (e.g., lower frequency cues), which was least affected. However, the influence of spectral content was not straightforward, as other features with lower frequency cues, such as nasality, were susceptible to level effects.

Third, results of studies that assessed effects of subject variables such as hearing loss or elevated thresholds were equivocal. Studebaker *et al.* (1999) observed similar declines in recognition of bandpass-filtered words in noise for subjects with normal and elevated thresholds, when scores were corrected for differences in audibility among subjects. Summers and Cord (2005) found larger declines in sentence recognition for high-frequency speech than for broadband or low-frequency speech for subjects with normal hearing, but similar declines across spectral conditions for subjects with hearing loss. In contrast to conclusions of Studebaker *et al.* (1999), these differences were not attributed to audibility differences among subjects. However, given that masked thresholds were not measured in either study, it is not known if declines in speech recognition for subjects with higher quiet thresholds may be attributed to differences in nonlinear growth of masking and if differential effects of spectral content remain once differences in effective signal-to-noise ratio due to masking growth are taken into account.

These questions were addressed in a series of three experiments. In the first two experiments, subjects were young adults with normal hearing; in the third experiment (reported

[a]Electronic mail: dubnojr@musc.edu

here), subjects were younger adults with normal hearing and older adults with normal hearing but significantly higher average quiet thresholds. The purpose of the first experiment (Dubno *et al.*, 2005a) was to test the hypothesis that reduced speech audibility at high levels contributed to the decline in speech recognition in noise. Given the variance among studies, stimuli were NU#6 monosyllabic words and NU#6-talker-shaped maskers, selected to maximize the likelihood of observing a decline in speech recognition at high levels and to use stimuli that were identical to those in Studebaker *et al.* (1999). Word recognition was measured in nine conditions, corresponding to all combinations of three signal-to-noise ratios (+8, +3, −2 dB) and three speech-shaped masker levels (70, 77, 84 dB SPL). Pure-tone thresholds were measured in quiet and in all maskers. If word recognition was determined entirely by signal-to-noise ratio and was independent of overall speech and masker levels, scores at a given signal-to-noise ratio should remain constant with increasing level. However, consistent with results of Studebaker *et al.* (1999), word recognition declined significantly with increasing speech level. Using audibility estimates based on the Articulation Index (AI), the deterioration in word recognition was attributed to nonlinear growth of masking in the speech-shaped masker, which resulted in reduced effective signal-to-noise ratio with increasing signal level.

Spectral effects on word recognition for speech and noise at high levels were assessed in the second experiment (Dubno *et al.*, 2005b). Recognition of NU#6 words was measured for low-pass-filtered speech and speech-shaped maskers and for high-pass-filtered speech and speech-shaped maskers at three speech levels in each of three masker levels. With the exception of filtering of speech and maskers, methods were the same as in Dubno *et al.* (2005a). For both low- and high-frequency speech in low- and high-frequency maskers, recognition declined significantly with increasing speech level while signal-to-noise ratio was held constant. The decline in recognition of low-frequency words in low-frequency maskers at high levels was attributed to nonlinear growth of masking in the speech-shaped masker which resulted in reduced effective signal-to-noise ratio at high levels, similar to results for broadband speech and speech-shaped maskers (Dubno *et al.*, 2005a). However, an unexpected finding was that masking growth accounted for some but not all of the decline in recognition of high-frequency speech in high-frequency maskers at high levels.

These spectral effects on word recognition in noise at high levels may relate to level-dependent differences in processing of low- and high-frequency speech information. Alternatively, the results could be a function of the experimental stimuli and listening conditions, or the range of subjects' quiet thresholds. Therefore, to confirm and extend the results to other subjects, stimuli, and conditions, a final experiment was conducted and is reported here whose purpose was to explore spectral and threshold effects on consonant recognition at high levels for subjects with normal hearing and a range of quiet thresholds. A broadband "threshold-matching noise" (TMN) was always present to equalize audibility between subject groups and across presentation levels.

In this experiment, the term "threshold effects" was used as an alternative to "hearing-loss effects" because subjects (even those with higher thresholds) were not typically described as having "hearing loss." The purpose of the TMN was not to eliminate these threshold effects, but to minimize differences in audibility that result from differences in quiet thresholds among subjects. Deficits in auditory function (and cognitive/proficiency deficits, if any) that may accompany threshold elevation remain and their effects on speech recognition can be assessed once any confounding audibility differences are minimized.

Examining declines in speech recognition in noise at high levels in individuals with relatively normal hearing is important for at least three reasons. First, results of these studies provide a better understanding of the mechanisms underlying perception of complex sounds in the normal auditory system. Second, speech recognition for subjects with relatively normal hearing is less likely to be confounded by large differences in speech audibility among subjects, simplifying interpretation of the results. Third, results of studies with younger and older subjects who have relatively normal hearing provide information needed to explain speech recognition for subjects with hearing loss, many of whom are older. In particular, understanding differential effects of spectral content is important because listening to high-frequency speech at high levels is a common occurrence for individuals with high-frequency hearing loss who wear hearing aids. Under some circumstances, providing high-frequency amplification may not improve and may even diminish speech recognition, which may relate to reported declines in performance for high-frequency speech observed even for subjects with relatively normal hearing (see Dubno *et al.*, 2005a, b for additional discussion).

II. METHODS

A. Subjects

Twelve younger subjects were initially recruited ("younger" was defined as 18–30 years), followed by a group of 12 older subjects ("older" was defined as ⩾60 years). Older subjects were recruited with the goal of increasing the range of quiet thresholds and to assign subjects to two groups defined by average quiet thresholds. The hearing criterion for all subjects was audiometric thresholds in the test ear ⩽25 dB HL (ANSI, 1996) at octave frequencies from 0.25 to 4.0 kHz and normal immittance measures. Of the 24 subjects, 23 were subsequently organized into a "lower threshold" group and a "higher threshold" group, which also separated subjects by age. The lower threshold group included 12 subjects ranging in age from 19 to 26 years (mean age: 22.2 years). The higher threshold group included 11 subjects ranging in age from 63 to 76 years (mean age: 67.9 years).[1] To determine group assignment, weighted-average quiet thresholds for pure tones from 0.2 to 6.3 kHz were computed using weights from the frequency importance function for the nonsense syllables used in this experiment (Dirks *et al.*, 1990). Weighted-average thresholds are discussed further in Sec. III A.

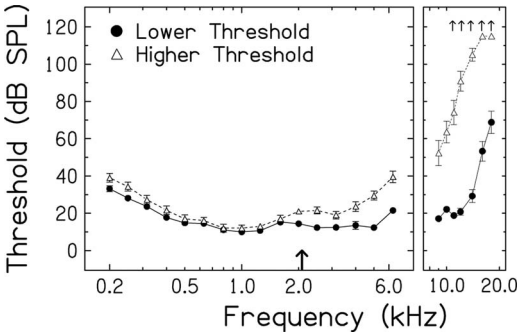


FIG. 1. Mean thresholds (±1 SE) for lower threshold subjects (filled) and higher threshold subjects (open) for frequencies from 0.2 to 6.3 kHz (left) and extended high frequencies (right). Arrows in the right panel denote frequencies in which at least one subject had no response at the maximum output of the audiometer. The arrow along the abscissa is plotted at the cutoff frequency for the speech and speech-shaped masker.

Mean pure-tone thresholds (±1 standard error, SE) for the two subject groups are shown in Fig. 1. Thresholds for frequencies from 0.2 to 6.3 kHz are in the left panel and thresholds for extended high frequencies are in the right panel (see below for apparatus and procedures). The mean weighted-average quiet threshold was 14.5 dB SPL for the lower threshold group and 21.7 dB SPL for the higher threshold group ($t_{21}=6.35$, $p<0.001$). Across frequency, extended high-frequency thresholds averaged 55.1 dB higher for subjects in the higher threshold group than the lower threshold group (this difference may be underestimated because more thresholds were beyond the limits of the audiometer for the higher than lower threshold group).

B. Apparatus and stimuli

1. Tonal and speech signals

Tonal signals were the same as in Dubno *et al.* (2005a, b). Thresholds for extended high frequencies were measured with a Madsen audiometer and Sennheiser HDA 200 earphones. Speech signals were 57 consonant-vowel and 54 vowel-consonant syllables formed by combining the consonants /b,tʃ,d,f,g,k,l,m,n,ŋ,p,r,s,ʃ,t,θ,v,w,j,z/ with the vowels /ɑ,i,u/ spoken by one male and one female talker without a carrier phrase (a total of 222 syllables).[2] Descriptions of the speech stimuli are in Dubno and Schaefer (1992) and Dubno *et al.* (2003).

For low-frequency speech, each syllable was filtered from 0.16 to 2.08 kHz (2 cascaded TDT PF1s; 101 dB/oct); for high-frequency speech, each syllable was filtered from 2.08 to 7.40 kHz (106 dB/oct). The 2.08-kHz cutoff frequency was selected to achieve nearly equal scores for low- and high-frequency speech in speech-shaped maskers, as determined by the AI. The cutoff frequency for equally intelligible bands (also referred to as the "crossover frequency") was higher in this experiment that used nonsense syllables (2.08 kHz) than for the previous two experiments that used NU#6 monosyllabic words (1.41 kHz).

Due to the shape of the speech spectrum, after filtering, the overall sound-pressure level of the low-frequency nonsense syllables was 15 dB higher than the overall sound-pressure level of the high-frequency nonsense syllables. In two earlier studies with monosyllabic words that compared recognition among high-level broadband, low-frequency, and high-frequency speech (Dubno *et al.*, 2005a, b), the design called for low-frequency and high-frequency speech to be maintained at the same relative levels as in the broadband condition (i.e., low-frequency words were 16 dB higher than high-frequency words). As reviewed earlier, subsequent results were somewhat different for high-frequency words than for low-frequency words. We hypothesized that the divergent results could be due, in part, to their 16-dB level difference. Thus, although our original rationale for maintaining relative levels was sound, the level difference may also have been a confounding factor in comparing recognition of low- and high-frequency speech at high levels. To address this question and simplify score interpretation in the current experiment, the level of the high-frequency speech was increased by 15 dB, equating it to the level of the low-frequency speech; there was no broadband condition. Thus, low-frequency and high-frequency nonsense syllables were each presented at 74, 84, and 94 dB SPL.

2. Maskers

For masking of speech and pure tones, the masker was a noise whose one-third-octave band spectrum matched the long-term rms levels of the nonsense syllables ("speech-shaped masker"). Using the same cutoff frequencies as for filtering speech, the low-frequency speech-shaped masker was filtered from 0.16 to 2.08 kHz and the high-frequency speech-shaped masker was filtered from 2.08 to 7.40 kHz. These maskers were presented at 62, 72, and 82 dB SPL. With speech levels of 74, 84, and 94 dB SPL, a constant +12-dB signal-to-noise ratio was maintained for the three low-frequency conditions and the three high-frequency conditions. The signal-to-noise ratio and speech levels were chosen based on previous results with NU#6 words where the largest effects were observed for the most advantageous signal-to-noise ratio, to avoid floor and ceiling effects for the more-difficult nonsense syllables, and to provide a wide range of signal levels.

As noted earlier, to minimize the influence of differences in quiet thresholds among subjects, a second masker (TMN) was always present. A broadband noise was digitally generated and its spectrum then adjusted at one-third-octave intervals to produce equivalent masked thresholds for all subjects. Band levels of the TMN were set to achieve masked thresholds of 20–25 dB HL from 0.2 to 3.15 kHz, 30 dB HL at 4.0 and 5.0 kHz, and 40 dB HL at 6.3 kHz. The overall level of the TMN was 62 dB SPL. For conditions in which the speech-shaped masker was presented at 72 and 82 dB SPL, the TMN was also presented at 72 and 82 dB SPL, respectively, to maintain constant audibility across conditions. Maskers were always present during the measurement of pure-tone thresholds and speech recognition. See Dubno *et al.* (2005a, b) for additional details on generation and presentation of speech-shaped maskers and TMN.

C. Procedures

Procedures were similar to those in Dubno *et al.* (2005a, b) and are briefly reviewed here. For each subject, thresholds for pure tones from 0.2 to 6.3 kHz were measured in quiet, in three levels of TMN alone, and in three levels of the low-frequency and high-frequency speech-shaped masker plus the TMN. Pure-tone thresholds were obtained using a single-interval (yes-no) maximum-likelihood psychophysical procedure (Green, 1993; Leek *et al.*, 2000). Signal level was varied adaptively.

Following measurement of pure-tone thresholds, recognition of low-frequency and high-frequency nonsense syllables in noise was measured at three speech levels in three masker levels, thus maintaining a fixed signal-to-noise ratio of +12 dB. The set of possible consonants was presented on a computer monitor and subjects responded by pointing with a mouse. After an initial familiarization period during which the experimenter verbally reinforced subjects' responses, no feedback was provided. During data collection, presentation order was randomized for the three speech levels. The order of filter condition (low-frequency speech or high-frequency speech) was counterbalanced. To reduce variance and increase statistical power, the six conditions were repeated so that the final scores for each condition were the percentage of correct responses to 444 syllables. AI values and predicted scores were computed using procedures similar to ANSI (1969) and ANSI (1997) and the frequency importance function and AI-recognition transfer function developed for these materials (Dirks *et al.*, 1990).[3] Using low-frequency and high-frequency rau-transformed scores (Studebaker, 1985), differences due to speech level and subject group were assessed by repeated-measures ANOVAs.

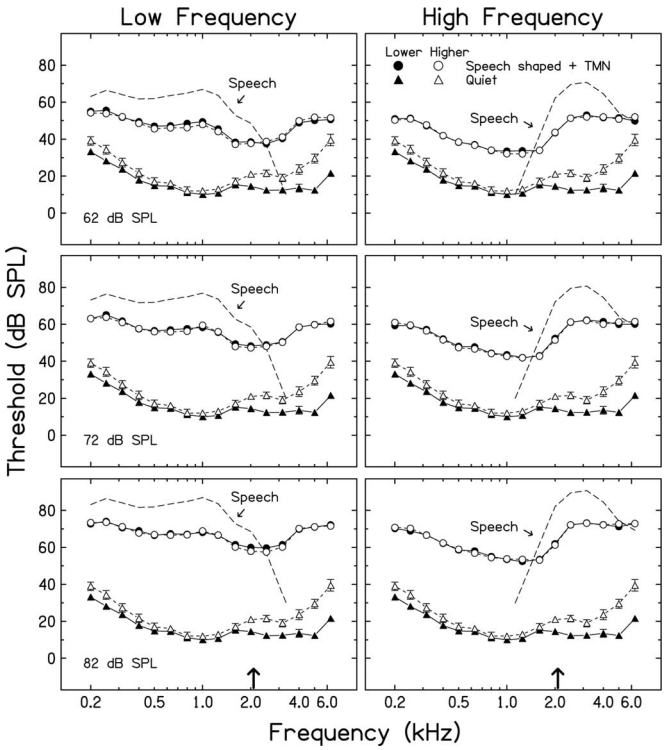


FIG. 2. Mean thresholds (±1 SE) for pure tones from 0.2 to 6.3 kHz measured in three levels of the low-frequency speech-shaped masker plus threshold-matching noise (TMN) (circles, left panels) and three levels of the high-frequency speech-shaped masker plus TMN (circles, right panels), for lower threshold subjects (filled) and higher threshold subjects (open). Each panel also contains mean thresholds measured in quiet (triangles). Masker levels are indicated in the left panels. Standard error ranges exceed the size of the data points for some quiet thresholds only. The dashed lines are the one-third-octave rms spectrum of the low-frequency nonsense syllables (left) or the high-frequency nonsense syllables (right) plotted at the speech level used with each masker level. The arrow along the abscissa is plotted at the cutoff frequency for the speech and speech-shaped masker.

III. RESULTS AND DISCUSSION

A. Masked thresholds

Figure 2 shows mean thresholds (circles) for pure tones from 0.2 to 6.3 kHz measured in low-frequency speech-shaped maskers (left column) and high-frequency speech-shaped maskers (right column), at each of three levels. Given that TMN was also present, these masked thresholds are for the combined speech-shaped and TMN maskers. Also shown are quiet thresholds, replotted from Fig. 1 (triangles). Only very small differences in mean masked thresholds between subject groups were observed in low- and high-frequency maskers across frequency and at each masker level.

Growth of masking for low-frequency and high-frequency maskers can have a differential effect on effective signal-to-noise ratios in low-frequency and high-frequency speech regions and may explain differences between declines in recognition of low-frequency vs high-frequency speech, or between lower threshold and higher threshold subjects. To assess these effects, weighted-average thresholds for pure tones measured in low- and high-frequency maskers were computed using weights from the nonsense-syllable frequency importance function. For the low-frequency and high-frequency maskers, weighted averages included the range of frequencies where there was audible low-frequency or high-frequency speech. Computed in this way, weighted-average thresholds take into account the relative importance of these frequencies to consonant recognition. Thresholds were analyzed using a repeated-measures ANOVA, with subject group as the between-subject variable (lower threshold and higher threshold); repeated measures were masker filter (low frequency and high frequency) and masker level (62, 72, 82 dB SPL). Results revealed that weighted-average masked thresholds did not differ significantly between the two subject groups $[F(1,21)=1.00, p=0.329]$. A *posthoc* test of the interaction of group and masker level revealed that increases in masked thresholds for the two groups with increases in masker level did not differ significantly $[F(2,42)=0.18, p=0.834]$. Averaged across the two subject groups, with the low-frequency masker increasing by 10 dB from 62 to 72 dB SPL and from 72 to 82 dB SPL, weighted-average masked thresholds increased by 10.3 dB (0.3 dB SE) and 10.5 dB (0.3 dB SE), respectively, providing some evidence of nonlinear growth of masking (>1.0 dB/dB) at frequencies where there was audible low-frequency speech. With the high-frequency masker increasing in 10-dB steps, weighted-average masked thresholds increased by 9.3 dB (0.3 dB SE) and 11.1 dB (0.3 dB SE), respectively, suggesting that nonlinear growth at frequencies where there was

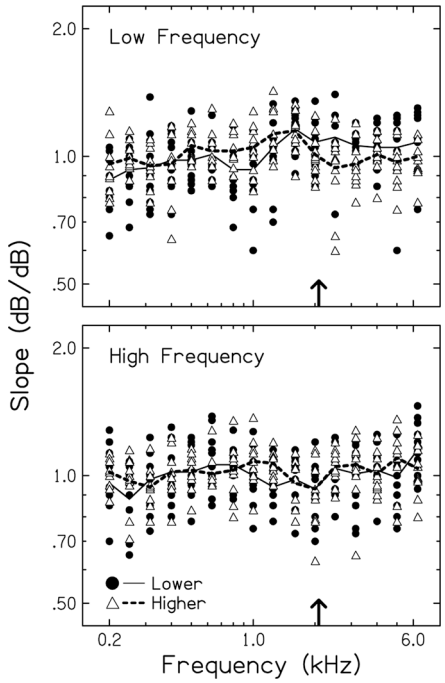


FIG. 3. Slopes of growth-of-masking functions for pure tones from 0.2 to 6.3 kHz for subjects in the lower threshold group (filled) and the higher threshold group (open) measured in three levels of the low-frequency speech-shaped masker plus threshold-matching noise (top) and high-frequency speech-shaped masker plus threshold-matching noise (bottom), computed using linear regression. Mean slopes are shown by solid and dashed lines. The arrow along the abscissa in each panel is plotted at the cutoff for the speech-shaped masker.

audible high-frequency speech occurred only between the two higher masker levels.

To view growth of masking with more frequency specificity, Fig. 3 displays individual slopes (symbols) and mean slopes (lines) of growth-of-masking functions for pure tones from 0.2 to 6.3 kHz, measured in three levels of the low-frequency (top) and high-frequency (bottom) speech-shaped maskers plus TMN. Slopes at each frequency were derived using linear regression, and so were dependent only on thresholds at the lowest and highest masker levels. For low-frequency speech-shaped maskers, linear growth of masking was observed at low frequencies and slightly steeper growth-of-masking slopes at middle frequencies, consistent with nonlinear growth of upward spread of masking attributed to the low-to-midfrequency peaks in the speech-shaped masker. An ANOVA with subject group as a between-subjects variable and signal frequency as a repeated measure revealed that frequency had a significant effect on growth-of-masking slope $[F(15,315)=3.98, p<0.001]$, but slopes did not differ significantly for the two subject groups $[F(1,21)=0.004, p=0.951]$. *Posthoc* tests showed a second-order trend, which was significant for slopes peaking at 1.6 kHz, that is, slopes increasing from 0.2 to 1.6 kHz and decreasing from 1.6 to 6.3 kHz, for the lower threshold group $[F(1,21)=8.14, p=0.010]$ and the higher threshold group $[F(1,21)=6.77, p=0.017]$. Results were generally the same when the analysis was restricted to slopes at frequencies where there was audible low-frequency speech (0.2 to 2.5 kHz), except that *posthoc* tests showed a significant linear trend $[F(1,21)=18.06, p=0.0004]$. This suggested that slopes within this frequency range increased linearly with increasing frequency for both groups.

Growth-of-masking slopes across frequency for low-frequency maskers were generally closer to linear than was observed in the previous experiment with low-frequency NU#6-shaped maskers (Dubno *et al.*, 2005b), despite differences in experimental conditions that were expected to increase slopes. For example, masker levels spanned a 20-dB range in the current experiment, but only a 14-dB range in the previous experiment, although levels were similar (current: 62–82 dB SPL; previous: 70–84 dB SPL). However, the band levels of the TMN used in the current experiment were higher to accommodate the quiet thresholds of the higher threshold group. Also, in contrast to results for the NU#6 words, where the steepest growth-of-masking slopes coincided with the peak in the NU#6 frequency importance function, the steepest slope here (1.6 kHz) was lower in frequency than the frequency-importance maxima for the nonsense syllables, which range from 2.5 to 4.0 kHz. This suggests that nonlinear growth of masking may have a smaller effect on weighted speech audibility for these low-frequency nonsense syllables than for the low-frequency NU#6 words.

Thresholds measured in the high-frequency speech-shaped masker at three levels generally revealed linear growth of masking across frequency, with slopes varying around 1.0 at frequencies within the passband of the high-frequency masker. Slopes differed significantly as a function of frequency $[F(15,315)=3.09, p<0.001]$ but did not differ significantly between the two subject groups $[F(1,21)=0.39, p=0.539]$. *Posthoc* tests showed a significant linear trend $[F(1,21)=7.13, p=0.014]$, suggesting that slopes increased with increasing frequency. Results remained the same when the analysis was restricted to slopes at frequencies where there was audible high-frequency speech (1.6 to 6.3 kHz). Linear growth of masking for the high-frequency masker was also similar to that observed in the previous experiment, despite high-frequency maskers in the current experiment that were higher by 15 dB, because they were equalized to the low-frequency maskers, and were narrower in bandwidth because of a higher cutoff frequency (2.08 vs 1.41 kHz).

B. Recognition of speech in noise

1. Effects of spectral content and speech level

a. Nonsense syllables. The top panels of Fig. 4 contain mean observed consonant-recognition scores plotted as a function of speech level for low-frequency (left) and high-frequency (right) speech and maskers for lower threshold subjects (filled) and higher threshold subjects (open). Scores (in rau) were analyzed with a repeated-measures ANOVA, with subject group as the between-subjects variable; repeated measures were speech level (74, 84, 94 dB SPL) and spectral content (low frequency and high frequency). Results revealed a significant main effect of filter $[F(1,21)=36.80, p<0.0001]$ and group $[F(1,21)=31.79, p<0.0001]$ and a significant interaction between group and filter $[F(1,21)=11.73, p=0.0025]$. That is, although scores for high-

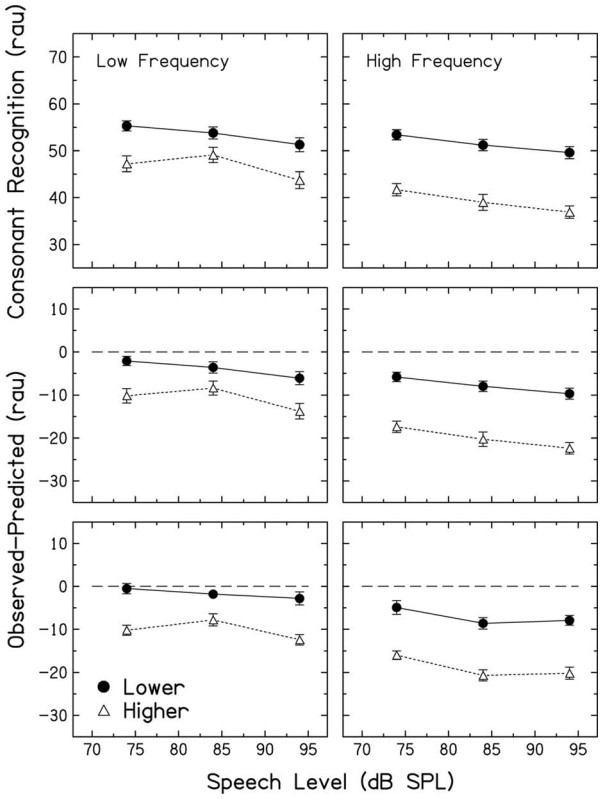


FIG. 4. Top row: Mean consonant-recognition scores (±1 SE) plotted as a function of speech level for lower threshold subjects (filled) and higher threshold subjects (open). Scores for low-frequency syllables obtained in low-frequency speech-shaped maskers plus threshold-matching noise are in the left panel and scores for high-frequency syllables obtained in high-frequency speech-shaped maskers plus threshold-matching noise are in the right panel. Middle row: Mean differences (±1 SE) between observed scores and AI-predicted scores computed using each subject's quiet thresholds and the levels and spectra of the speech and speech-shaped plus threshold-matching noises, plotted as a function of speech level. A dashed line is drawn at an observed-predicted difference of 0 rau. Bottom row: Same as middle panel, but scores were predicted using the speech spectrum and each subjects' thresholds measured in the speech-shaped maskers plus threshold-matching noise.

frequency syllables and maskers were significantly poorer than for low-frequency syllables and maskers, a *posthoc* test of the filter × group interaction attributed this result to significantly poorer scores for high-frequency syllables than low-frequency syllables for the higher threshold group only $[F(1,21)=43.16, p<0.0001]$. Thus, although equivalent AI values for low-frequency and high-frequency speech and maskers predicted equivalent scores for all subjects, recognition of high-frequency speech for subjects with higher thresholds was significantly poorer than recognition of low-frequency speech.

The ANOVA also revealed a significant main effect of speech level $[F(2,42)=49.19, p<0.0001]$, a significant interaction of speech level and spectral content $[F(2,42)=4.15, p=0.023]$, and a nonsignificant interaction of speech level and group $[F(2,42)=2.28, p=0.115]$. The significant decline in scores confirmed that consonant recognition for low- and high-frequency syllables in low- and high-frequency speech-shaped maskers decreased at high levels when signal-to-noise ratio was held constant, and decreased for subjects with lower and higher thresholds. Although the declines in scores with increases in speech level differed significantly with spectral content, *posthoc* tests attributed this result to the change in scores between the 74- and 84-dB low-frequency speech levels for the higher threshold subject group.

The middle panels of Fig. 4 show differences between observed and predicted consonant-recognition scores for low-frequency syllables and maskers (left) and high-frequency syllables and maskers (right) plotted as a function of speech level. In this case, predicted scores were estimated from AI values computed using each subject's quiet thresholds and the levels and spectra of the nonsense syllables and the speech-shaped masker plus TMN. Predicted scores (not shown) were identical for subjects in the lower and higher threshold groups because audibility was limited by speech-shaped plus TMN maskers rather than by the (lower) quiet thresholds. Predicted scores remained constant with increasing speech level because, as low-frequency and high-frequency speech level increased by 10 dB, masker level also increased by 10 dB, maintaining a constant signal-to-noise ratio for low- and high-frequency syllables. When low- and high-frequency consonant-recognition scores were predicted without taking into account effects of nonlinear growth of masking, observed-predicted differences declined with increasing level in the same manner as the observed scores, and group differences remained the same (note the similarity in the pattern of results between the top and middle rows of Fig. 4). These results are similar to findings from the earlier studies using broadband words and speech-shaped maskers (Dubno *et al.*, 2005a) and low- and high-frequency words and maskers (Dubno *et al.*, 2005b).

The bottom panels of Fig. 4 plot differences between observed scores and scores predicted from AI values computed using the speech spectrum and each subject's thresholds measured in the low-frequency or high-frequency speech-shaped masker plus TMN (similar to the weighted-average masked thresholds discussed in the second paragraph of Sec. III A). To the extent that speech audibility was reduced due to nonlinear growth of masking, predicted scores determined using masked thresholds should decline with increasing speech level. To the extent that declines in observed scores were also due to this reduction in audibility, observed-predicted differences should remain constant with increasing speech level. Observed-predicted differences were analyzed using a repeated-measures ANOVA (same design as described above for observed scores) and results revealed a significant interaction of speech level and spectral content $[F(2,42)=7.31, p=0.0019]$. Therefore, changes in observed-predicted differences with increasing speech level will be described separately for low-frequency and high-frequency speech and maskers.

For low-frequency syllables and maskers, predicted scores based on AI values computed using low-frequency speech spectra and low-frequency masked thresholds declined with increasing speech level for both subject groups (not shown), reflecting the nonlinear increase in low-frequency masked thresholds. To examine changes in observed-predicted differences with speech level, *posthoc*

analyses were conducted comparing the lowest and highest low-frequency speech levels (lower left panel of Fig. 4). Observed-predicted values did not differ significantly for either the lower threshold group [$F(1,21)=2.81, p=0.108$] or the higher threshold group [$F(1,21)=2.24, p=0.150$]. For other level comparisons and for both subject groups, where there were significant declines in observed scores, observed-predicted values did not differ significantly as speech level increased, with one exception (see below). Taken together, these results suggest that the decrease in recognition of low-frequency nonsense syllables at higher levels may be attributed to nonlinear growth of masked thresholds (and reduced effective signal-to-noise ratio) in the low-frequency speech-shaped masker. It is notable that these results are similar to recognition of low-frequency words (Dubno *et al.*, 2005b), despite less nonlinear masking growth for low-frequency maskers presented with syllables than words. The one exception was a significant decline in scores between the two higher speech levels for the higher threshold group, where observed-predicted differences also declined significantly [$F(1,21)=14.20, p=0.001$]. Thus, factors other than masking growth may contribute to declines in low-frequency consonant recognition at higher levels for these subjects.

The pattern of results for the high-frequency syllables and maskers (right) was somewhat different than that for the low-frequency syllables and maskers, and the findings did not differ between the two subject groups. Here, when consonant-recognition scores were predicted using high-frequency speech spectra and high-frequency masked thresholds, predicted scores declined only slightly between the two highest speech levels, consistent with nonlinear masking growth between these levels, discussed earlier (see Sec. III A). With observed scores declining significantly in the high-frequency masker, observed-predicted values also declined significantly between the lowest and highest speech level [$F(1,21)=15.74, p=0.0007$] and between the lowest and middle speech level [$F(1,21)=22.45, p=0.0001$], but did not differ significantly between the middle and highest speech level [$F(1,21)=0.50, p=0.487$]. Thus, declines in recognition of high-frequency syllables at high levels could not entirely be attributed to nonlinear growth of masking, similar to previous results for recognition of high-frequency words (Dubno *et al.*, 2005b). Where nonlinear growth of masking was observed, the subsequent reduction in effective signal-to-noise ratios in the higher frequencies accounted for declines in recognition of high-frequency syllables at high levels.

b. Comparing nonsense syllables and monosyllabic words. Recall that a purpose of this experiment was to confirm and extend results for low-frequency and high-frequency NU#6 monosyllabic words and NU#6-shaped maskers using different stimuli and conditions. In general, results were similar across studies for both low- and high-frequency speech (i.e., comparing the lower threshold group to the previous study). Consistent findings across low-frequency conditions were expected because speech and masker spectra for low-frequency nonsense syllables were generally similar to those for low-frequency words, with the exception of location of peaks, perhaps due to differences in talker gender, differences in bandwidth, and the restricted vowel set for the nonsense syllables. In contrast, findings for high-frequency conditions were less predictable because speech and masker spectra for high-frequency nonsense syllables were substantially different than those for high-frequency words. Relative to the spectra for words and maskers, the spectra for nonsense syllables and maskers were higher in level by 15 dB, contained a prominent peak rather than a uniform shape, and were narrower in bandwidth.

Across subjects, comparing scores at the lowest and highest speech levels, there was a nonsignificant trend for the decline in consonant-recognition scores to be larger for high-frequency than low-frequency syllables and maskers, comparable to the pattern observed previously for monosyllabic words. In the current experiment, larger declines in scores were expected for both low- and high-frequency conditions, due to the wider range of syllable and masker levels (20 vs 14 dB). Still larger declines for high-frequency nonsense syllables were expected because the high-frequency syllables and maskers had been increased by 15 dB. The smaller decline that resulted may relate to the narrower bandwidth for high-frequency nonsense syllables and maskers as compared to high-frequency words and maskers, due to the higher cutoff frequency (2.08 kHz for nonsense syllables vs 1.41 kHz for words). However, this explanation is not supported by results of Molis and Summers (2003), who reported the largest decline in recognition of high-frequency sentences with increasing level for the subject with the narrowest high-frequency bandwidth.

Generally consistent findings for words and nonsense syllables, notwithstanding spectral characteristics unique to these stimulus sets, suggest that differences in susceptibility of cues available in low- and high-frequency speech to deterioration at high levels may be common across speech materials and may also relate to level-dependent differences in processing of low- and high-frequency speech information. For example, higher frequency cues, such as for place information, may require more fine spectral resolution than some lower frequency cues, such as for voicing information or to distinguish among manner classes, which may be conveyed by periodicity or envelope cues (consistent with patterns of consonant confusions for high syllable levels reported by Hornsby *et al.*, 2005). Thus, the normally broadened auditory filters at higher levels and at higher frequencies may have a greater detrimental effect on high-frequency speech cues than on low-frequency speech cues. In addition, high-level, lower frequency speech information may be dominated by vowel segments, which contribute relatively little to intelligibility (as noted by Kates and Arehart, 2005). Thus, deterioration of these cues at high levels would have a relatively smaller effect on recognition of speech. Results from studies of auditory-nerve responses also suggest frequency-dependent effects for processing of high-level speech. For example, a loss of $F2$ synchrony, which could relate to auditory-nerve fibers' increasing response to $F1$, occurs at a lower level for high-frequency fibers than for low-frequency fibers (Wong *et al.*, 1998).

2. Subject effects

In addition to assessing effects of spectral content on consonant recognition at high levels, an additional goal of the current experiment was to explore effects of subjects' quiet thresholds. As noted earlier, subjects had normal hearing but differed significantly in their average quiet thresholds. Consonant-recognition scores for the higher threshold group were significantly poorer than scores for the lower threshold group, and scores for high-frequency speech were significantly poorer than for low-frequency speech for the higher threshold group only. Moreover, although declines in recognition with increasing speech level did not differ significantly for the two groups, declines in recognition of low-frequency speech for the higher threshold group were not entirely accounted for by changes in effective signal-to-noise ratio resulting from nonlinear masking growth. These results were unexpected due to the subjects' generally good hearing, the use of a threshold-matching noise to minimize audibility differences among subjects, and the similarity between groups in low-frequency and high-frequency masked thresholds and masking growth. Given that subjects in the lower and higher threshold groups also differed in age, the generally poorer scores across conditions for older subjects may be attributed, at least in part, to age-related differences in task proficiency and/or cognitive abilities among subjects.

Although age-related proficiency/cognitive effects may have contributed to an overall reduction in scores for older subjects, these factors do not provide reasonable explanations for the differential effects of spectral content for older subjects, that is, significantly poorer scores for high- than low-frequency syllables. In addition to being older, these subjects had significantly higher average quiet thresholds. Recent evidence suggests that elevations in quiet thresholds are consistent with reductions in nonlinearities in the basilar-membrane response (e.g., Plack *et al.*, 2004). Given that threshold differences between groups were largest in the higher frequencies (see Fig. 1), it is possible that changes in nonlinearities that were greater in higher frequencies could have a differential effect on recognition of high-frequency syllables for higher threshold subjects.

To examine this question further, associations between consonant recognition and quiet thresholds were assessed for individual subjects. Specifically, Pearson correlation coefficients were computed between high-frequency consonant-recognition scores at each speech level and weighted-average high-frequency quiet thresholds (i.e., including thresholds at frequencies where there was audible high-frequency speech); correlations were also computed for comparable pairs of low-frequency scores and thresholds. Statistically significant negative correlations were observed for all comparisons, ranging from −0.76 to −0.77 for high-frequency scores and thresholds and from −0.54 to −0.65 for low-frequency scores and thresholds, as shown in Fig. 5. Within the higher threshold (older) group, there were no significant correlations between weighted-average quiet thresholds and age, suggesting a negligible influence of age on these associations. Furthermore, in this experiment, a dependence of consonant recognition on quiet thresholds does not reflect differences in speech audibility among subjects because audibility estimates were determined by speech-shaped plus TMN maskers and not by quiet thresholds.

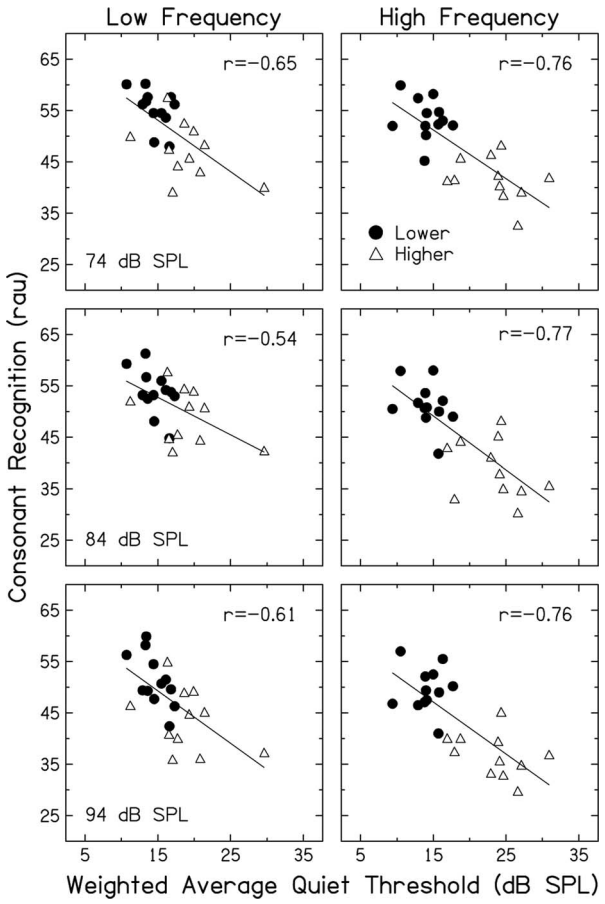


FIG. 5. Left column: Consonant-recognition scores for low-frequency syllables in low-frequency maskers for each of three speech levels plotted against weighted-average low-frequency quiet thresholds for subjects in the lower threshold group (filled) and subjects in the higher threshold group (open). Right column: Same as left, but for high-frequency syllables and weighted-average high-frequency quiet thresholds. Pearson correlation coefficients and linear regression functions are included in each panel.

Associations between consonant recognition and quiet thresholds are consistent with the assumption that threshold elevations (and concomitant reductions in nonlinearities) may have contributed to poorer performance for higher threshold subjects. Stronger negative correlations between high-frequency than low-frequency scores and thresholds may also imply that recognition of high-frequency syllables may be more susceptible to elevated high-frequency thresholds and reductions in nonlinearities. Molis and Summers (2003) and Summers and Cord (2005) suggested that reductions in active cochlear processing at high levels and at high frequencies contribute to *declines* in speech recognition at high levels. This hypothesis would predict greater declines in recognition of high- than low-frequency syllables and greater declines in recognition for subjects with lower than higher thresholds. However, neither of these results was observed in the current study.

An additional subject-related issue with regard to recognition of speech at high levels is the role of the acoustic reflex. Given that middle-ear muscle contractions attenuate

sounds below 2.0 kHz, with the largest changes occurring below 1.0 kHz, potential effects for the current study would be limited to low-frequency speech and maskers. One could speculate that the acoustic reflex provides an overall attenuation of lower frequency speech and maskers that may reduce declines in recognition with increasing level, because an equivalent effective signal-to-noise ratio was maintained. However, the effect of the reflex is complex in its attenuation pattern and temporal course. The largest attenuation is at the lowest frequencies which contribute less to intelligibility of nonsense syllables than higher frequencies. With regard to temporal effects, in the current experiment the low-frequency speech-shaped masker plus TMN were on continuously, whereas the syllables were presented at random intervals, controlled by the timing of the subjects' responses. Therefore, it is not known if muscle contractions were elicited initially by the maskers and adapted over time, but then were also elicited by each high-level syllable, or only by some syllables. These complex factors make it difficult to speculate on how contraction of the middle-ear muscles may contribute to the overall interpretation of the results. Nevertheless, it is possible that effects of the acoustic reflex may add to the between-subject variance in some findings, such as individual differences in the slopes of the growth-of-masking functions (see Fig. 3). However, to the extent that nonlinear growth of masking explained declines in scores, the role of the acoustic reflex should be relatively small. Finally, given evidence that acoustic reflex thresholds do not differ significantly as a function of age (e.g., Thompson *et al.*, 1980; Gates *et al.*, 1990), it is unlikely that this factor provides an explanation for age-related differences in consonant recognition.

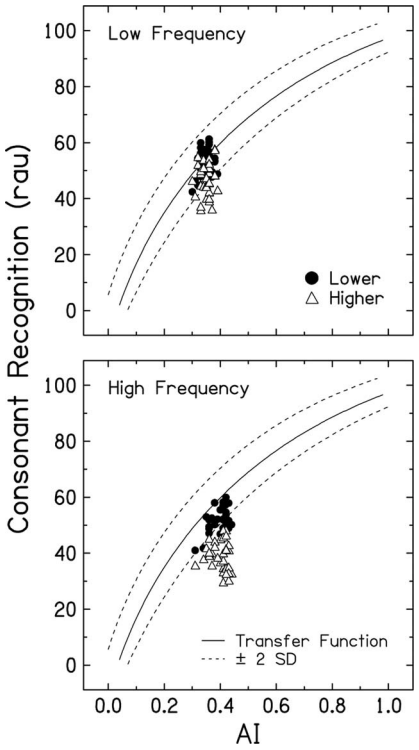


FIG. 6. Top: Low-frequency consonant-recognition scores plotted against AI for subjects in the lower threshold group (filled) and subjects in the higher threshold group (open). AI values were computed using the low-frequency speech spectrum and thresholds measured in the low-frequency masker plus threshold-matching noise. The solid line is the AI-recognition transfer function for the nonsense syllables used in this experiment; the dashed lines encompass the 95% confidence limits. Bottom: Same as top, but for high-frequency consonant-recognition scores, with AI values computed using the high-frequency speech spectrum and thresholds measured in the high-frequency masker plus threshold-matching noise.

3. Differences in observed and predicted recognition scores for low-frequency and high-frequency syllables and maskers at high levels

The relationship between observed and predicted scores is also shown in Fig. 6, wherein consonant-recognition scores for low-frequency speech and maskers (top) and high-frequency speech and maskers (bottom) are plotted against AI, with AI values computed using speech spectra and thresholds measured in the speech-shaped maskers plus TMN. The solid line is the transfer function relating the AI to consonant recognition established for the speech materials used in this experiment and represents the predicted scores; the dashed lines encompass the 95% confidence limits. As noted earlier, observed scores for high-frequency speech and maskers were significantly poorer than observed scores for low-frequency speech and maskers, but only for the higher threshold group. Observed-predicted differences were also larger (i.e., more negative) for high-frequency than low-frequency speech and maskers for both subject groups [main effect of spectral content: $F(1,21)=60.85$, $p<0.0001$; interaction of group and spectral content: $F(1,21)=3.47$, $p=0.077$]. Consistent with these results, Fig. 6 shows scores for individuals from both groups for high-frequency speech that were substantially poorer than predicted, more so than scores for low-frequency speech (see also Fig. 4, bottom panels). For the higher threshold group, observed scores for higher frequency speech were poorer than the 95% confidence interval of the predicted scores.

In regard to these results, an additional question left unresolved following the previous experiment with filtered words concerned differences in observed scores and observed-predicted scores for low-frequency and high-frequency speech and maskers. Although bandwidths were selected to achieve nearly equal scores for low-frequency and high-frequency words and maskers, word-recognition scores averaged 10.4 rau higher for high-frequency words than for low-frequency words. In addition, observed scores for low-frequency words and maskers were generally worse than predicted, and observed scores for high-frequency words and maskers were generally better than predicted. We speculated that these differences resulted from the AI-computation method, wherein speech peaks were fixed at 15 dB across frequency (see Dubno *et al.*, 2005b). Because the physical speech peaks (Sherbecoe *et al.*, 1993) and the effective speech peaks (Studebaker and Sherbecoe, 2002) increase with increasing frequency, actual signal-to-noise ratio may have been larger in the higher frequencies and smaller in the lower frequencies than estimated with speech peaks fixed across frequency. In the current experiment, AI values computed with the actual speech peaks for these nonsense syllables were expected to yield better estimates of speech

recognition in low- and high-frequency regions and the cutoff frequency for equally intelligible bands. Indeed, in the current experiment, averaged across speech level, the difference in consonant-recognition scores between low-frequency and high-frequency syllables was reduced to 2.1 rau (for the lower threshold group, the appropriate comparison to the previous study). For low-frequency syllables and maskers, observed scores remained worse than predicted, but observed-predicted differences now averaged only 1.7 rau. For high-frequency syllables and maskers, observed scores were now also worse than predicted, consistent with expectations using actual speech peaks.

IV. SUMMARY AND CONCLUSIONS

Recognition of low- and high-frequency nonsense syllables in low- and high-frequency nonsense-syllable-shaped maskers was measured at three speech levels, with signal-to-noise ratio held constant. Pure-tone thresholds were measured in each masker and in quiet. Subjects were younger and older adults with normal hearing who were organized into two groups according to their average quiet thresholds. Results may be summarized as follows.

(1) Pure-tone thresholds measured in a low-frequency speech-shaped masker plus TMN increased linearly with increasing masker level at lower frequencies. For midfrequency signals, steeper growth-of-masking slopes were observed, consistent with nonlinear growth of upward spread of masking attributed to the low-to-midfrequency peak in the speech-shaped masker. Masked thresholds measured in a high-frequency speech-shaped masker plus TMN generally revealed linear growth of masking with increasing masker level, with some increase in slope at higher frequencies. No significant differences were observed in growth of masking between subjects with lower and higher quiet thresholds.

(2) Recognition of low-frequency nonsense syllables in low-frequency maskers and high-frequency nonsense syllables in high-frequency maskers generally declined with increasing speech level when signal-to-noise ratio was held constant, for subjects with lower and higher quiet thresholds.

(3) For subjects with lower thresholds, declines in recognition of low-frequency nonsense syllables in low-frequency maskers at high levels were attributed to nonlinear growth of masking in the speech-shaped masker plus TMN, which reduced the effective signal-to-noise ratio at high levels. In contrast, the decline in scores for subjects with higher thresholds could not be explained by masking growth. For both subject groups, masking growth did not entirely account for declines in recognition of high-frequency nonsense syllables in high-frequency maskers. These results were similar to those observed previously for broadband words and speech-shaped maskers and for low- and high-frequency words in maskers.

(4) Generally consistent results for words and nonsense syllables suggested that differences in susceptibility of cues available in low- and high-frequency speech to deterioration at high levels may be common across speech materials and may also relate to level-dependent differences in processing of low- and high-frequency speech information.

(5) Relative to younger subjects with normal hearing and lower quiet thresholds, older subjects with normal hearing and higher quiet thresholds demonstrated poorer consonant recognition in noise, especially for high-frequency speech in high-frequency maskers. Generally poorer consonant recognition may be attributed to age-related differences in task proficiency and/or cognitive abilities, but differential effects of spectral content may be a function of threshold-related differences in cochlear function.

ACKNOWLEDGMENTS

This work was supported (in part) by Grants R01 DC00184 and P50 DC00422 from NIH/NIDCD, the James E. and Pamela Knowles Foundation, and the MUSC General Clinical Research Center (M01 RR01070). This investigation was conducted in a facility constructed with support from Research Facilities Improvement Program Grant Number C06 RR14516 from the National Center for Research Resources, National Institutes of Health. The authors thank Chris Ahlstrom for computer and signal-processing support, Fu-Shing Lee for advice on data analysis, Rebecca McDonald and Jillanne Schulte for assistance with data collection, and Associate Editor John Grose and two anonymous reviewers for helpful suggestions on earlier versions of this manuscript.

[1]An examination following measurement of quiet thresholds for pure tones from 0.2 to 6.3 kHz revealed that only one older subject would have been placed in the lower threshold group. To simplify the analysis and separate subjects by both age and average quiet thresholds, we elected not to include the results for this one older subject. Inspection of this subject's masked thresholds and consonant-recognition scores suggested that conclusions would not have changed had these data been included in the analyses.

[2]Given that the focus of the study was on spectral effects on speech recognition, it was necessary to consider whether it was appropriate to include syllables spoken by both male and female talkers. The rationale for including both male and female talkers was as follows. First, only very small differences in the long-term speech spectra for the male and female talkers who recorded these speech materials were observed, consistent with the literature. For example, Byrne *et al.* (1994) state that spectra for male and female talkers were "virtually identical over the frequency range from 250 to 5000 Hz" (p. 2115). According to these authors, gender differences were larger at and below 160 Hz, where importance weights for the nonsense syllables are 0.0. Most studies conclude that an average spectrum including both males and females can appropriately represent the spectrum of either gender (e.g., Cox and Moore, 1988; Olsen *et al.*, 1987). Second, the current study was not focused on revealing fine spectral differences (e.g., differences in fundamental and formant frequencies between male and female talkers) but rather on more general higher and lower frequency cues (e.g., place vs voicing).

[3]According to Pavlovic (personal communication, 2005), the key difference between the standards describing the AI (ANSI, 1969) and the Speech Intelligibility Index (SII; ANSI, 1997) is that the SII allows for the use of frequency importance functions for specific speech materials. Thus, given that the frequency importance function for the specific set of nonsense syllables was used in the current study, our procedures are consistent with the standard that describes the SII. In the implementation of the AI used in this experiment, no speech or masker level corrections were included.

ANSI (1969). ANSI S3.5-1969, "American National Standard Methods for the Calculation of the Articulation Index" (American National Standards Institute, New York).
ANSI (1996). ANSI S3.6-1996, "American National Standard Specification for Audiometers" (American National Standards Institute, New York).
ANSI (1997). ANSI S3.5-1997, "American National Standard Methods for Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).
Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, 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., and Frolenkov, G. I. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.
Cox, R. M., and Moore, J. N. (**1988**). "Composite speech spectrum for hearing aid gain prescriptions," J. Speech Hear. Res. **31**, 102–107.
Dirks, D. D., Dubno, J. R., Ahlstrom, J. B., and Schaefer, A. B. (**1990**). "Articulation index importance and transfer functions for several speech materials," Asha **32**, 91.
Dubno, J. R., and Schaefer, A. B. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110–2121.
Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (**2000**). "Use of context by younger and older adults with normal hearing," J. Acoust. Soc. Am. **107**, 538–546.
Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing," J. Acoust. Soc. Am. **113**, 2084–2094.
Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2005a**). "Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking," J. Acoust. Soc. Am. **118**, 914–922.
Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2005b**). "Recognition of filtered words in noise at higher-than-normal levels: Decreases in scores with and without increases in masking," J. Acoust. Soc. Am. **118**, 923–933.
Gates, G. A., Cooper, J. C., Kannel, W. B., and Miller, N. J. (**1990**). "Hearing in the elderly: The Framingham cohort, 1983–1985. Part 1. Basic audiometric test results," Ear Hear. **11**, 247–256.
Green, D. M. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes-no task," J. Acoust. Soc. Am. **93**, 2096–2105.
Hornsby, B. W.Y., Trine, T. D., and Ohde, R. N. (**2005**). "The effects of high presentation levels on consonant feature transmission," J. Acoust. Soc. Am. **118**, 1719–1729.
Kalikow, D., Stevens, K., and Elliott, L. (**1977**). "Development of a test of speech intelligibility in noise using test material with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.
Kates, J. M., and Arehart, K. H. (**2005**). "Coherence and the speech intelligibility index," J. Acoust. Soc. Am. **117**, 2224–2237.
Leek, M. R., Dubno, J. R., He, N.-j., and Ahlstrom, J. B. (**2000**). "Experience with a yes-no single-interval maximum-likelihood procedure," J. Acoust. Soc. Am. **107**, 2674–2684.
Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low- and high-frequency speech," ARLO **4**, 124–128.
Olsen, W. O., Hawkins, D. B., and Van Tasell, D. J. (**1987**). "Representations of the long-term spectra of speech," Ear Hear. **8**, 100S–108S.
Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.
Sherbecoe, R. L., Studebaker, G. A., and Crawford, M. R. (**1993**). "Speech spectra for six recorded monosyllabic word tests," Ear Hear. **14**, 104–111.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Studebaker, G. A., and Sherbecoe, R. L. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422–1436.
Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.
Summers, V., and Cord, M. T. (**2005**). "Effects of high intensity on recognition of low- and high-frequency speech in noise," J. Acoust. Soc. Am. **117**, S2606.
Thompson, D. J., Sills, J. A., Recke, K. S., and Bui, D. M. (**1980**). "Acoustic reflex growth in the aging adult," J. Speech Hear. Res. **23**, 405–418.
Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., and Young, E. D. (**1998**). "Effects of high sound levels on responses to the vowel /ɛ/ in cat auditory nerve," Hear. Res. **123**, 61–77.

Auditory filter shapes of CBA/CaJ mice: Behavioral assessments

Bradford J. May[a)]
The Center for Hearing and Balance, Department of Otolarygnology-HNS, Johns Hopkins University, Baltimore, Maryland 21205

Sarah Kimar and Cynthia A. Prosen
Department of Psychology, Northern Michigan University, Marquette, Michigan 49855

(Received 24 January 2006; revised 11 April 2006; accepted 14 April 2006)

Auditory filter shape and frequency tuning may be derived by measuring changes in pure tone thresholds as a function of the bandwidth of notched-noise maskers. When these psychophysical methods were applied to CBA/CaJ mice, the resulting filter shapes were well fit by $roex(p,r)$ functions originally developed for human subjects. The equivalent rectangular bandwidths (ERBs) of the filter shapes ranged from 16 to 19% of test frequencies between 8 to 16 kHz. These ERBs correspond well to the performance of humans at high frequencies and the limited number of mammalian species that have been characterized with notched-noise procedures. Frequency tuning was maintained throughout most of the adult lifespan and then showed a selective high-frequency loss at ages beyond 2 years. These results suggest that auditory filtering effects in adult CBA/CaJ mice are similar to normal processes in other mammalian species and provide an excellent model of human presbycusis when they begin to degrade in aging individuals.
 [DOI: 10.1121/1.2203593]

PACS number(s): 43.66.Gf, 43.66.Dc, 43.66.Sr [AJO] Pages: 321–330

I. INTRODUCTION

Hearing loss is one of the most pervasive public health issues in the United States. Over 20 million Americans report difficulty hearing or understanding speech (Ries, 1994; Mitchell, 2006). By age 65, less than 10% of our population demonstrates auditory abilities that would be considered normal in terms of audiological standards (Mosicki *et al.*, 1985; Gates and Cooper, 1991). Age-related hearing loss (AHL), or presbycusis, is usually a sensorineural disorder involving cochlear hair cells (Nadol, 1979; Wright *et al.*, 1987). The onset and rate of the progressive loss ranges widely between individuals and may be accelerated by hereditary, medical, and environmental factors (Johnsson *et al.*, 1990; Schuknecht and Gacek, 1993). These sources of variation in human populations have made mouse models with more predictable patterns of hearing loss increasingly popular over the last decade (Erway *et al.*, 2001; Francis *et al.*, 2003).

The increasing sophistication of genetic approaches has led to the generation of inbred and transgenic mice with selectively altered auditory function (Mullen and Ryan, 2001). These potentially subtle changes require phenotyping methods that extend beyond basic threshold measures (Parham, 1997; Barsz *et al.*, 2002; Prosen *et al.*, 2003). From a psychophysical perspective, valuable insights into perception have been gained by conceptualizing the early stages of auditory processing as an array of bandpass filters that separate complex or competing sounds into their constituent frequency components (Fletcher, 1940; Patterson, 1974). This filtering effect is determined by the active electromechanical tuning of the inner ear (Moore *et al.*, 1999), and deteriorates with sensorineural neural hearing loss (Patterson *et al.*, 1982; Tyler *et al.*, 1984). Consequently, the perceptual manifestations of hearing loss are multidimensional with impaired listeners reporting problems not only hearing sounds but also understanding speech (Turner and Robb, 1987), tolerating abnormal loudness effects (Beattie and Warren, 1982), and listening in noise (Lyregaard, 1982).

The frequency resolution of the mouse auditory system has been previously characterized in terms of the critical band (Ehret, 1976, 1979). This approach measures changes in pure tone thresholds as a function of the bandwidth of bandpass masking noise (Scharf, 1961). The critical band is reached when further widening of the noise band has no effect on threshold, presumably because the additional energy falls outside the hypothetical auditory filter. In contrast to values that typically remain within 10–30 % of the test frequency for a broad sampling of mammalian species (Fay, 1988), critical bands ranged from 40 to 280 % in mice.

This study revisited the question of auditory frequency tuning in mice using notched-noise masking procedures (Patterson, 1974; Glasberg and Moore, 1990). The notched-noise method is used in our study to complement previous critical band measures (Ehret, 1976) and to provide a more complete description of filter shape and bandwidth. Psychophysical assessments were performed on adult CBA/CaJ mice because this strain does not show early onset hearing loss and therefore may provide stable normal baselines during the initial stages of training and testing. Our estimates of the auditory filter in younger mice conformed well to human performance at high frequencies (Moore and Glasberg, 1983; Glasberg and Moore, 1990). Older mice displayed a selective loss of high-frequency tuning that was reminiscent of human presbycusis (Margolis and Goldberg, 1980; Patterson *et al.*, 1982).

[a)]Author to whom all correspondence should be addressed. Electronic mail: bmay@jhu.edu

II. METHODS

All of the following procedures were approved by the Institutional Animal Care and Use Committee of Northern Michigan University.

A. Subjects

Experiments were performed on 13 female CBA/CaJ mice that were procured from Jackson Labs at an approximate age of 4–6 weeks. The subjects were housed in laboratory facilities on a reverse day-night cycle with free feeding. When trained, mice received most of their daily water in the testing environment. A brief watering period was provided in the evening to ensure that all mice maintained normal weights throughout the course of experiments. The mice remained healthy under these housing conditions and continued producing behavioral thresholds at ages approaching 2.5 years.

B. The tone detection task

The effects of notched noise on the detection of pure tones were measured with a positive reinforcement operant procedure using water rewards (Prosen *et al.*, 2000). The contingencies of the tone detection task are shown in Fig. 1. Mice initiated a testing cycle by entering the observation compartment of the cage. A variable time interval (3–9 sec) began when the interruption of photocells indicated an observing response. The cycle advanced to a trial state if the subject remained in the observing compartment for the duration of the waiting period.

The onset of tone bursts signaled the transition to a detection trial. A speaker above the observation compartment pulsed the auditory stimuli with 250-msec on and off periods. The subject gained access to a water dipper by crossing the cage partition to the detection compartment before the termination of the 6-sec trial window. The dipper remained active for 3 sec and then retracted beneath the floor of the detection compartment.

Mice generated time out intervals by entering the detection compartment during wait intervals, or by failing to enter the compartment during detection trials. Mice learned to avoid these response errors because the 6-sec interruption of the testing cycle delayed the next period of water access.

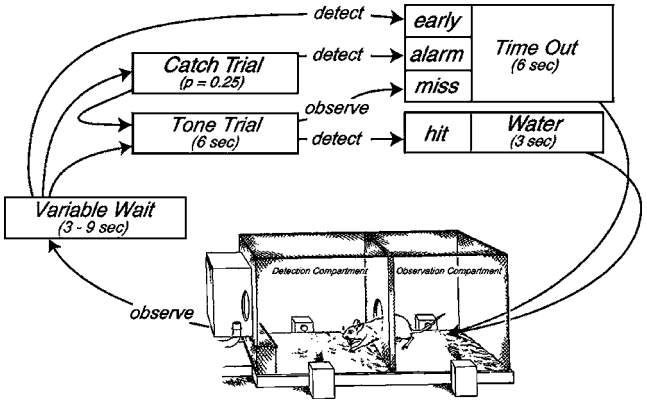


FIG. 1. The behavioral testing cycle. Mice initiated tone presentations by waiting in the observation compartment of the test cage, then entered the response compartment to activate a water dipper.

Calculation of the tone detection threshold required the presentation of detectable and undetectable stimulus levels. Consequently, a portion of correct detection responses reflect chance responding and not actual tone detection. Catch trials were presented on 25% of all trials to monitor the probability of false positive responses. Catch trials shared all of the timing properties of detection trials but did not include tone bursts. The subject advanced immediately to a trial state by remaining in the observation area until the end of the 6-sec catch trial interval. As in other error responses, entering the detection compartment during the catch trial (false alarm) produced a 6-sec time out.

The level of the tone bursts was fixed within a trial but varied at random across trials. The range of levels included stimuli around threshold, as well as values clearly above threshold. High-level tone presentations did not contribute to threshold determinations, but improved the stability of behavioral performance by increasing the probability of reinforced trials.

Threshold was based on the signal detection criterion $d' = 1$ (Green and Swets, 1974). This statistic was calculated from the response probabilities for detection and catch trials, as shown in Eq. (1)

$$d' = z(P_{\text{hit}}) - z(P_{\text{false alarm}}), \tag{1}$$

where $z(P_{\text{hit}})$ is the standardized probability (z score) of correct responses at each stimulus level and $z(P_{\text{false alarm}})$ is the standardized probability (z score) of false alarms. When psychometric functions relating d' scores to tone level did not include stimulus values corresponding to threshold, the predicted level was derived by linear interpolation between the two values bracketing threshold. The same frequency was tested over multiple sessions until daily thresholds converged on stable performance.

C. The notched-noise method

The power spectrum model of masking assumes that threshold represents a constant ratio of signal and noise energy (Fletcher, 1940). The integration of noise energy across frequency is determined by the shape of a hypothetical auditory filter that is centered on the frequency of the tone. Consequently, filter shape may be derived by selectively removing noise energy with a spectral notch and measuring the change in threshold.

The first stages of our notched-noise paradigm began by establishing the tone detection threshold under quiet conditions. When these tests were completed, the detection task was modified to include a low-level of continuous broadband noise. Gaussian noise with a nominal bandwidth of 100 kHz was generated with a digital-to-analog converter (model RP2, Tucker-Davis Technologies; Alachua, FL) and transduced with a free-field electrostatic speaker (model ES1, Tucker-Davis Technologies). Acoustic calibrations indicated that the effective bandwidth of the speaker ranged from 2–48 kHz with less than ±10 dB variation. The level of the masker was increased over several sessions until thresholds

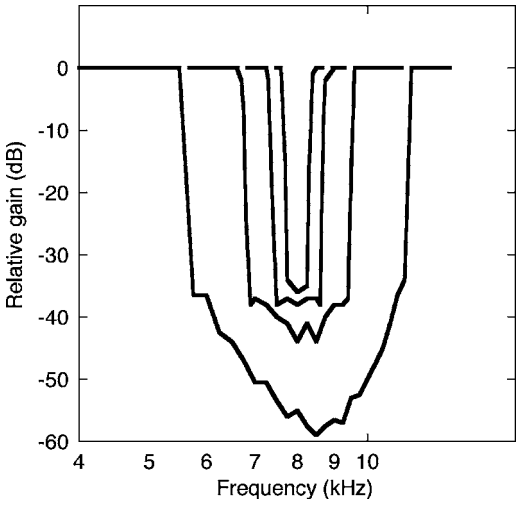


FIG. 2. Calibrations for notched-noise backgrounds with a center frequency of 8 kHz.

in noise elevated 40(±4.0) dB relative to thresholds in quiet. Noise spectrum levels of 13.4(±4.5) dB/Hz produced the requisite threshold shift.

Subsequent experiments were conducted by adding spectral notches to the noise background. The notches were logarithmically centered on the signal frequency with octave bandwidths of 0.125, 0.25, 0.5, 1.0, and 2.0. Only one bandwidth was presented in each session and the same bandwidth was repeated until performance reached stable values. Filter coefficients for generating notched noise were computed using MATLAB's implementation of the finite duration impulse response constrained by least-squares (function FIRCLS). The average slope of the spectral edges was ±26 dB/100 Hz. Sample calibrations with an analog spectrum analyzer (General Radio Company type 1900-A) are shown in Fig. 2.

D. Derivation of filter shape

The power spectrum model of masking is mathematically summarized by Eq. (2) (Glasberg and Moore, 1990):

$$P_s = K\int_{-\infty}^{\infty} N(f)W(f)df, \tag{2}$$

where P_s is the masked threshold, $N(f)$ is the power spectrum of the noise, and $W(f)$ is the weighting function of the auditory filter. The constant K is a scaling adjustment. Given this relationship, $W(f)$ may be estimated by subtracting frequency regions from the power spectrum of the noise with spectral notches and measuring the resulting masked thresholds.

Previous psychoacoustic studies (Patterson *et al.*, 1982) have found that the general shape of the auditory filter may be approximated by the exponential shown in Eq. (3)

$$W(g) = (1-r)(1+pg)\exp(-pg) + r. \tag{3}$$

The parameters p and r describe the filter skirts near the center frequency and at remote frequencies, respectively. The parameter g is the frequency deviation of the edge of the notch normalized by its center frequency $(\Delta f/f_c)$. The frequency deviation of broadband noise is designated 0.

Normal baselines and aging effects in CBA/CaJ mice were quantified by applying $roex(p,r)$ fits to the notched-noise data that were collected with the tone detection task. These software tools were distributed for general use by Glasberg and Moore (1990), where the analysis is described in detail. The calculations involve restating Eq. (2) in terms of normalized frequency deviation, substituting the $roex(p,r)$ approximation of W, then solving the equation analytically.

In human psychoacoustic studies, notches are generally described in linear frequency so that deviations of the upper and lower edges are symmetrical. Notches described in logarithmic frequency, such as those used in the present study, produce asymmetric frequency deviations. The $roex(p,r)$ function accepts the lower and upper deviation of each notch as input and calculates fitting parameters for both linear and log symmetrical notches. Masked thresholds and filter magnitudes are plotted in terms of the deviation at the near (lower) edge, following the convention of Glasberg and Moore (1990).

E. Equivalent rectangular bandwidth

The equivalent rectangular bandwidth (ERB) is a convenient method for comparing the selectivity of auditory frequency tuning across species and studies (Moore and Glasberg, 1983). The ERB transforms the $roex(p,r)$ filter into a rectangular shape with the same peak and total transmission. The sharpness of frequency tuning is indicated by the bandwidth of the rectangle.

After the parameters p and r were obtained by application of the $roex(p,r)$ fitting procedure of Glasberg and Moore (1990), the ERB of the auditory filter shape was calculated according to Eq. (4).

$$\mathrm{ERB} = 2f_c\{(1-r)p^{-1}[2-(2+p)\exp(-p)] + r\}. \tag{4}$$

This analysis also was automatically performed by the $roex(p,r)$ software of Glasberg and Moore.

III. RESULTS

Initial behavioral testing was conducted with 8-kHz tones. To sample a wider range of frequencies before the onset of AHL, the mice were then divided into two groups and tested at either 11.2 or 16 kHz. The following results begin with a description of normal baselines in adult mice and end with aging effects that were revealed by tracking the same subjects over 2.5 years.

A. Normal baselines

Thresholds were derived from psychometric functions that combined a minimum of five days stable performance under the same stimulus conditions. Representative functions are shown in Fig. 3. These data were obtained with 8-kHz tones under quiet conditions, in broadband noise, and with three bandwidths of notched noise. The subject's percentage of correct responses at a given tone level and false alarm rates for catch trials are plotted in Fig. 3(a). Masking effects are indicated by the rightward shift of the tone-in-noise functions relative to responses in quiet. The largest shift was

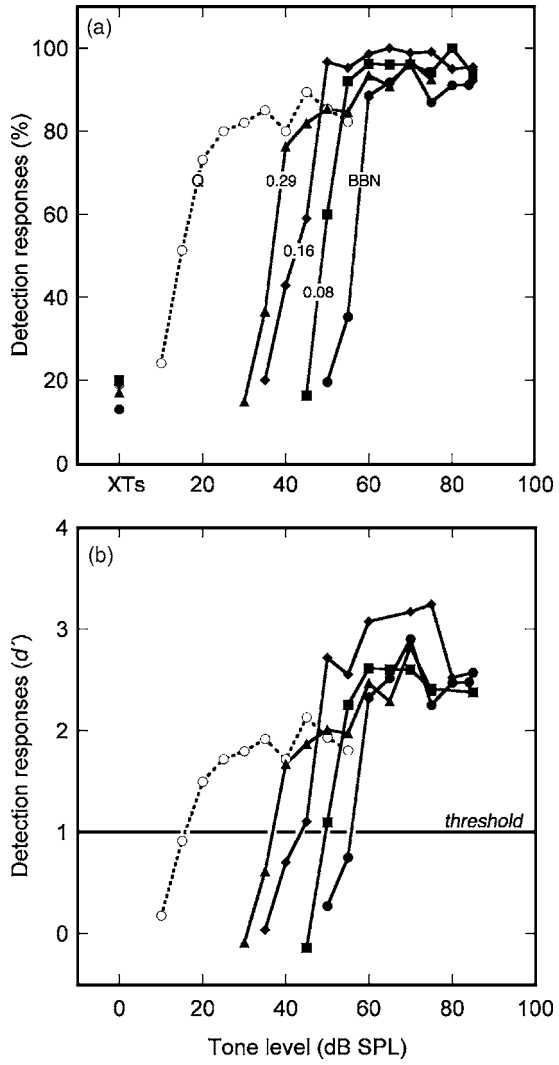


FIG. 3. The summary psychometric functions of mouse CBA03sa. Detection scores are plotted as the percentage of correct responses (a) and the signal detection statistic d' (b) The data were obtained for 8-kHz tones in quiet (Q), broadband noise (BBN), and three bandwidths of notched noise (0.08, 0.16, and 0.29 deviation). The threshold criterion $d'=1$ is indicated by the. horizontal line. Responses to catch trials are plotted to the left of the psychometric functions (XTs).

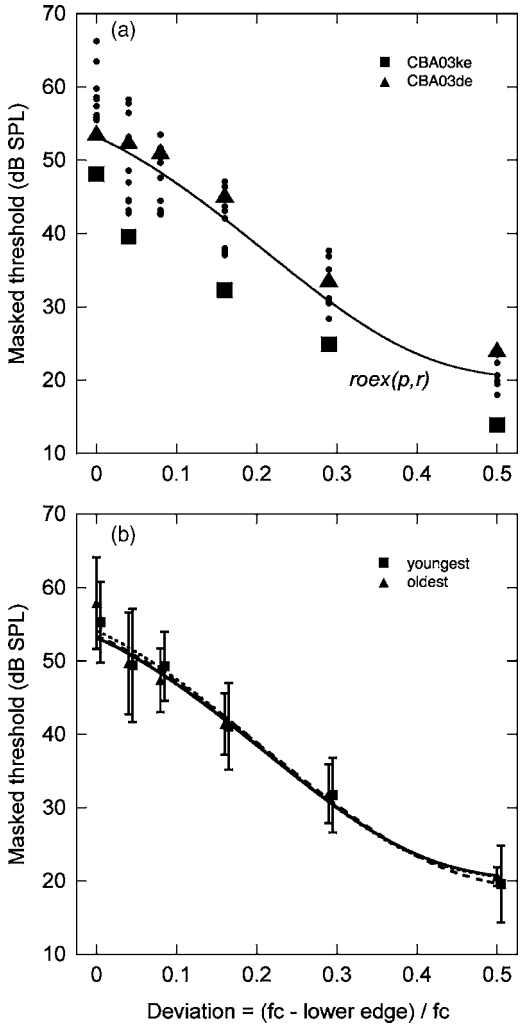


FIG. 4. Effects of notch bandwidth on the detection of 8-kHz tones. (a) Symbols indicate the thresholds of 13 mice. Data from mice CBA03ke and CBA03de are highlighted with unique symbols to illustrate consistent performance differences between subjects. The $roex(p,r)$ function of Glasberg and Moore (1990) has been fit to the sample averages of each deviation, (b) Threshold averages (±1 standard deviation) when the same data are divided into the five youngest versus the five oldest subjects. $Roex(p,r)$ functions have been fit to each dataset (youngest, dashed line; oldest, dotted line). For comparison, the fit to the original dataset is shown (solid line).

observed for broadband noise. Removing noise energy by addition of a spectral notch reduced the magnitude of the shift in proportion to the deviation of the notch from center frequency.

The percentage of responses to tones and catch trials were substituted into Eq. (1) to produce the d' values in Fig. 3(b). Thresholds are indicated by the intersection of each psychometric function with the criterion $d'=1$. Threshold changes in the four masking conditions mirror the shifts of the psychometric functions in Fig. 3(a).

Results of the $roex(p,r)$ analysis are not strongly influenced by the selection of the d' threshold criterion. Because the psychometric functions in Fig. 3(b) are quite steep, halving or doubling the criterion changes absolute thresholds by less than 10 dB. Because the transitional phases of the psychometric functions are essentially parallel across noise conditions, relative threshold differences remain the same. The auditory filter shape is derived from these relative differences.

Under quiet conditions, 8-kHz thresholds ranged from 8–21 dB SPL for the 13 mice and produced an average threshold of 16.1 dB SPL. The standard deviation of the sample was 3.6 dB. Similar thresholds were subsequently observed in quiet at frequencies of 11.2 and 16 kHz, which produced average thresholds of 12.5(±3.0) and 14.2(±1.5) dB SPL. These results compare favorably with several published audiograms of laboratory mice (Fay, 1988).

Masked thresholds were averaged to derive P_s values for input to the $roex(p,r)$ function. Individual thresholds for 8-kHz tones are compared in Fig. 4(a). The variation in thresholds between subjects was often systematic. Data from two mice are highlighted in the figure to illustrate a subject whose thresholds were consistently better (CBA03ke) or worse (CBA03de). The general shape of the threshold function was usually maintained across subjects regardless of these differences.

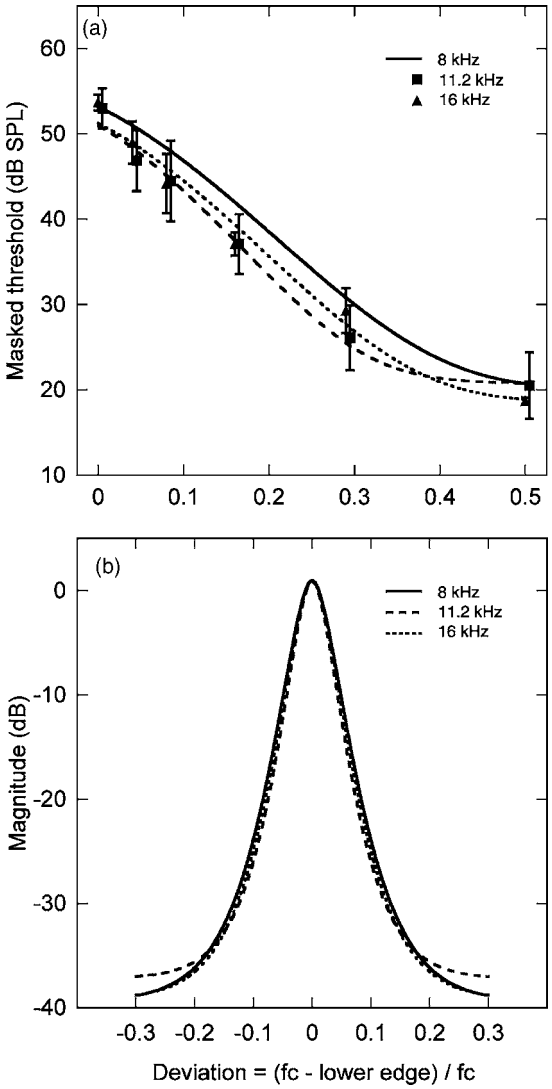


FIG. 5. Effects of frequency on auditory filter shape. Average notched-noise thresholds at 8, 11.2, and 16 kHz were fit with the *roex*(p,r) model (a) to derive filter weighting functions (b). Error bars indicate ±1 standard deviation.

Individual mice proceeded through the series of threshold measures at their own rates. Although initial threshold values were completed at ages between 19 and 22 weeks, slower rates of acquisition in some subjects and the introduction of intermediate bandwidth conditions in others delayed the completion of all thresholds to ages approaching one year. Relative to other strains, the potential confounds of AHL are somewhat lessened in CBA/CaJ mice. In Fig. 4(b), the 8-kHz threshold data have been separated in two groups based on age of completion. This analysis yielded essentially identical fitting parameters for the youngest and oldest mice.

The effects of frequency on auditory filter shapes were investigated by repeating threshold measures with center frequencies of 11.2 and 16 kHz. The fits for all three test frequencies are shown in Fig. 5(a). The parameters p and r from these fits were substituted into Eq. (3) to derive the filter shapes in Fig. 5(b). Because notch bandwidth is specified in terms of deviation, which is normalized by center frequency,

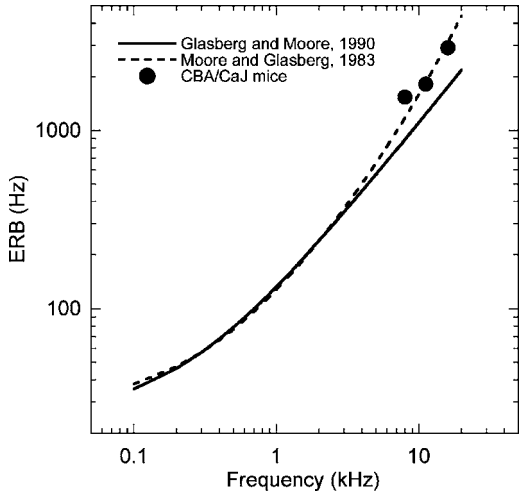


FIG. 6. Comparison of mouse and human ERBs. Estimates of human frequency resolution are based on equations derived from the notched-noise studies of Moore and Glasberg (1983) and Shailer *et al.* (1990).

the filters remain constant across frequency. This stability suggests that normal baselines were obtained at all three frequencies prior to significant AHL.

The three filter shapes in Fig. 5(b) were converted to ERBs by substituting their p and r values into Eq. (4). Center frequencies of 8, 11.2, and 16 kHz produced ERBs of 1540, 1820, and 2914 Hz. These filter widths fall within 16 to 19 % of the center frequency and reproduce results at similar frequencies from studies of human hearing (Shailer *et al.*, 1990).

The ERBs of CBA/CaJ mice are compared with model fits from human notched-noise studies in Fig. 6. Normal baseline measures in mice corresponded well with psychoacoustic performance at the upper limits of human hearing. These results are further evidence that threshold measures reported in Fig. 5 were completed before significant AHL.

B. Aging effects

Threshold measures were continued after the completion of normal baselines to track age-related changes in auditory filter shapes at 11.2 and 16 kHz. These data were obtained across ages that ranged from 106–127 weeks. When testing was conducted under quiet conditions, average thresholds at 11.2 kHz (±1 standard deviation) increased from normal baselines of 12.5(±3.0) to final values of 23.3(±3.5) dB SPL. Thresholds at 16 kHz showed a similar increase from 14.2(±1.5) to 30.2(±4.2) dB SPL. Although threshold elevations were small in magnitude, they were universally observed in the two sample groups. The loss, therefore, was statistically significant (paired t test, $p<0.005$).

The average notched-noise thresholds (±1 standard deviation) of aged mice are compared with normal baselines in Fig. 7. As shown in Fig. 7(a), the final threshold replications at 11.2 kHz indicate a uniform 9-dB threshold elevation relative to baseline performance. The filter shapes in Fig. 7(b) were derived from the fitting parameters of the *roex*(p,r) functions. Because the parallel threshold curves can be fit with the same values of p and r, there was no significant change in filter shape despite the modest loss of sensitivity.

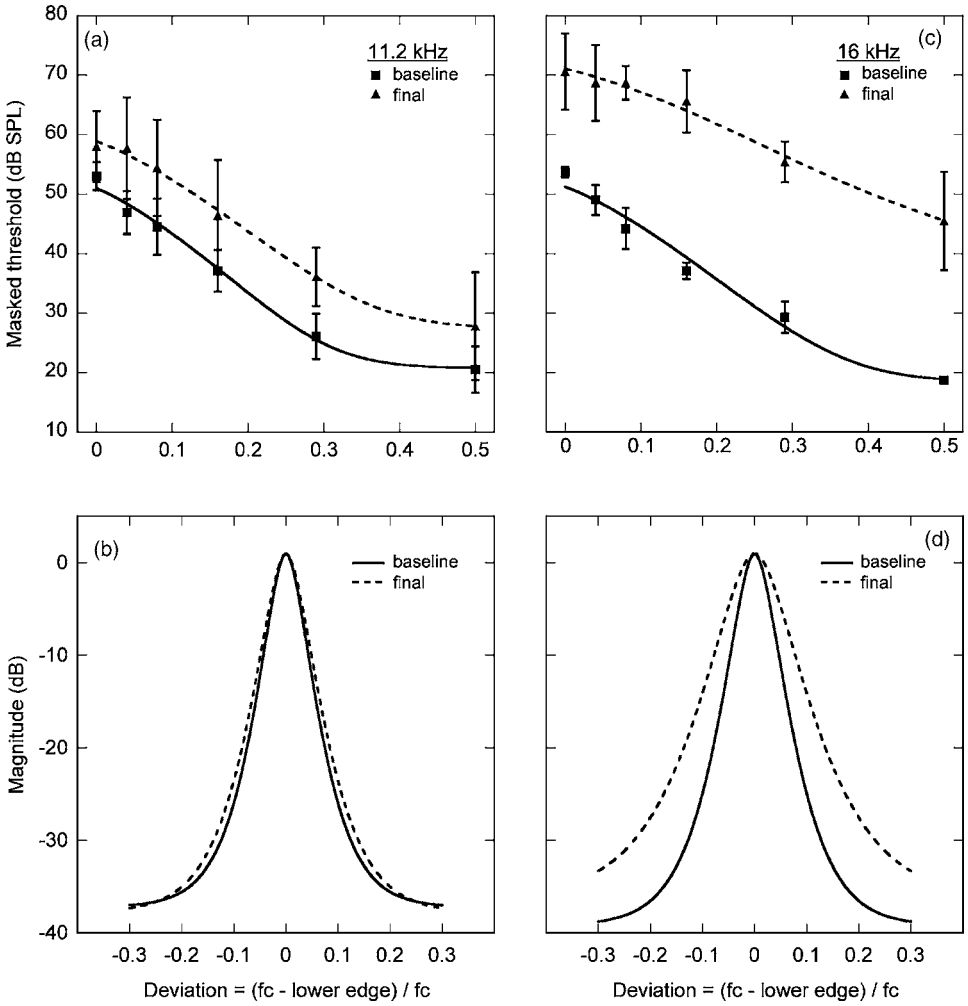


FIG. 7. Effects of aging on auditory filter shape. (a) Comparison of the *roex*(p,r) fits for initial baseline measures and final threshold replications at 11.2 kHz. (b) Filter shapes derived from the *roex*(p,r) fits, (c) (d) Results obtained at 16 kHz from a second group of subjects.

The final notched-noise replications at 16 kHz reveal more pronounced effects of age on threshold and frequency tuning. As shown in Fig. 7(c), the threshold elevations of older mice ranged from 20 dB at small deviations to 29 dB at large deviations. The increased masking effectiveness of noise bands with wider spectral notches indicates a broadening of the auditory filter, as shown in Fig. 7(d). Relative to the initial baseline ERB of 2914 Hz, the filter expanded to 4615 Hz.

The specification of age-related hearing deficits in Fig. 7(d) is problematic because mice of the same age may exhibit different patterns of hearing loss. This variability is described in Fig. 8, which compares the masked thresholds and corresponding auditory filter shapes for two subjects at three time periods. Mouse CBA03ch maintained a normal ERB of 3112 Hz during these tests, whereas mouse CBA03bi showed substantial loss of frequency tuning.

The *roex*(p,r) fits of the two mice and their corresponding filter shapes are compared in Figs. 8(a) and 8(c). The initial baseline measures from both subjects showed steeply sloped threshold functions that produced ERBs of 2948 and 2690 Hz, respectively, for mice CBA03ch and CBA03bi. These results fall close to the baseline ERB of 2914 Hz that was derived from the threshold averages of all subjects.

Intermediate threshold replications were conducted at ages ranging from 92–108 weeks. The threshold functions of both mice show a largely parallel shift relative to baseline measures. The fit for mouse CBA03bi has a shallower slope because the rate of hearing loss was higher at wider notch deviations. As a result, this subject's ERB increased to 3595 Hz.

Final threshold replications were obtained at ages ranging from 107–124 weeks. The thresholds of mouse CBA03ch elevated approximately 25 dB during this time period. Nevertheless, the mouse continued to produce a steeply sloped threshold function and an ERB within 1% of normal baselines. Mouse CBA03bi showed a progressive flattening of the threshold function which expanded the auditory filter to 7302 Hz.

The effects of age on masking conditions are summarized in Fig. 8(b). Selected detection thresholds from the fits in Fig. 8(a) have been replotted in relation to the subject's age at test completion. The parallel threshold curves of mouse CBA03ch indicate a uniform change in sensitivity and therefore preservation of auditory filter shape. By contrast, the thresholds of mouse CBA03bi are marked by an upward compression that accelerated during the final weeks of behavioral testing. This increased masking was prominent at the widest notch deviations and contributed to a lifting of the filter skirts in Fig. 8(c).

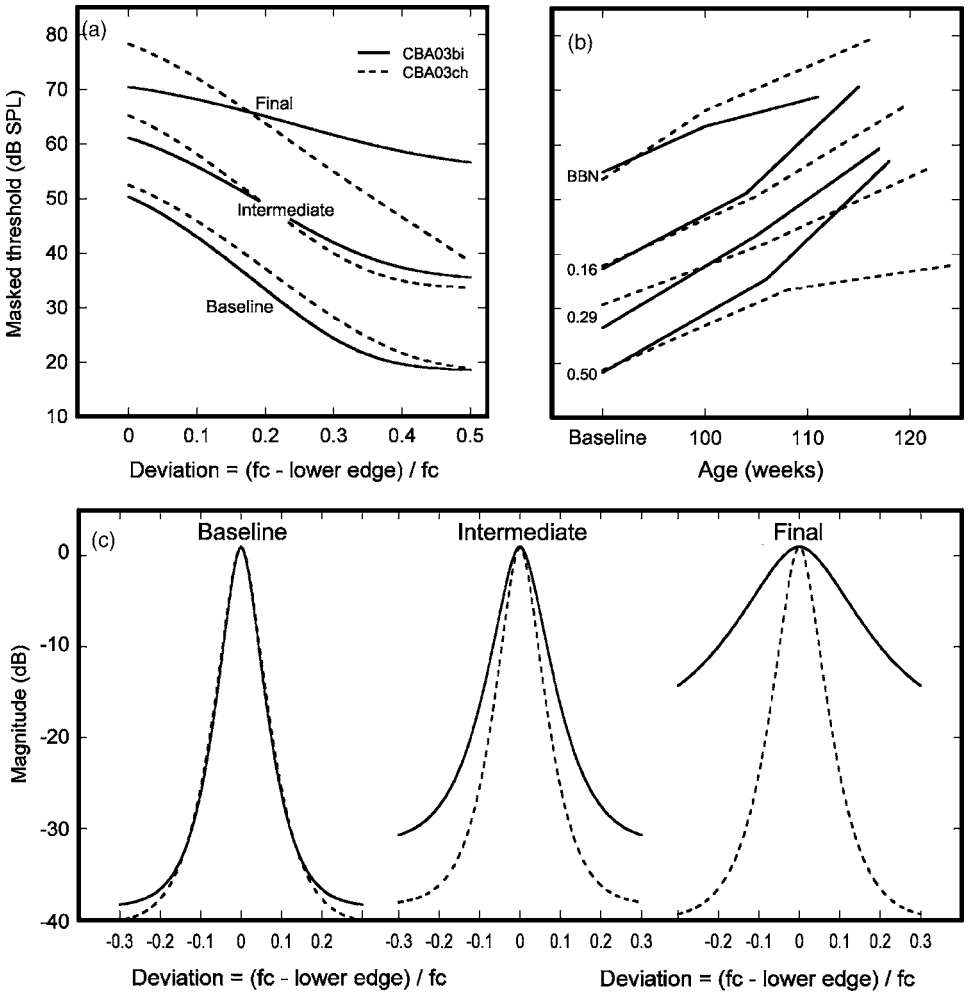


FIG. 8. Intersubject variation in aging effects. (a) *Roex*(p,r) fits for mice CBA03bi and CBA03ch at three age periods. The thresholds were obtained with 16-kHz tones. (b) The same data replotted to highlight the importance of masking condition on the magnitude of threshold shifts. (c) Filter shapes derived from the *roex*(p,r) fits.

IV. DISCUSSION

The major finding of this study was that CBA/CaJ mice demonstrate auditory filtering effects that are equivalent in shape and bandwidth to other mammalian species. Long-term behavioral subjects maintained normal frequency tuning throughout most of their adult life. When aged mice showed a loss of frequency tuning, the deficits followed the high-to-low frequency progression that characterizes human presbycusis.

A. Normal baselines

The critical bands (CBs) and auditory filter characteristics of laboratory mice have been described in one previous psychophysical study (Ehret, 1976). Those experiments were conducted in NMRI mice, which is an outbred albino strain with no obvious hearing deficiencies. The evoked potential thresholds (Muller *et al.*, 1997) and neural tuning curves (Egorova *et al.*, 2001; Egorova *et al.*, 2002) of NMRI mice are comparable to those of the CBA/CaJ strain (Ma *et al.*, 2006).

Behavioral results from the NMRI strain suggest exceptionally poor frequency tuning in laboratory mice. At frequencies from 20 to 40 kHz, where mice display their most sensitive hearing, the CB varied from 40 to 52 % of filter frequencies. By contrast, the CB remains near 16 to 25 % of filter frequencies across most of the bandwidth of hearing in humans (Zwicker *et al.*, 1957; Scharf, 1961) and common animal models such as macaque monkeys (Gourevitch, 1970), domestic cats (Pickles, 1975; Nienhuys and Clark, 1979), and chinchilla (Seaton and Trahiotis, 1975; Niemiec *et al.*, 1992). In the present study, the ERBs of CBA/CaJ mice ranged, from 16 to 19 % of filter frequencies. These tuning characteristics represent a substantial departure from earlier behavioral estimates in mice but correspond well to other mammalian species.

Testing method is one potential source of the discrepancy between our present results in CBA/CaJ mice and previous assessments of the NMRI strain. In situations where filter estimates from band widening and notched-noise paradigms can be compared in the same species, notched-noise data indicate consistently better frequency resolution. The procedure-based differences may be of considerable magnitude in animal psychophysical studies. For example, the tuning bandwidth of the Atlantic bottlenoise dolphin is 17 kHz at a center frequency of 30 kHz when measured with band-widening procedures (Au and Moore, 1990); but only 5 kHz when it is derived from notched-noise data (Finneran *et al.*, 2002). Direct within-study comparisons of band-widening versus notched-noise paradigms have noted a similar three-fold decrease in the frequency tuning of the chinchilla (Niemiec *et al.*, 1992).

Animal studies of the CB and neural frequency tuning

suggest that the fundamental properties of auditory frequency selectivity are established at the earliest stages of sound processing. The structural elements that contribute to the active mechanical properties of the cochlea are similar in mice and humans (Pujol, 1985; Cheatham *et al.*, 2004). The physiological tuning of the mouse cochlea, as reflected in the sound-driven discharge rates of auditory-nerve fibers, displays the same filter characteristics as other mammalian species (Taberner and Liberman, 2005). Our results indicate that the perceptual behaviors of the CBA/CaJ strain also conform to the basic patterns of mammalian frequency selectivity that are assumed to arise from cochlear tuning.

B. Preservation of function

The goal of the present experimental design was to determine optimal baseline measures and subsequent aging effects by repeated testing at the same frequencies. The time interval between threshold replications was minimized by focusing on a small number of signal and masker combinations in each subject. Test frequencies were chosen to fall within the most audible region of the mouse audiogram but below frequencies that show early age-related hearing loss. In addition, frequencies from 8 to 16 kHz are audible to humans with normal high-frequency hearing and therefore allow direct comparisons to human perception. A limitation of this focused approach is that the more generalized effects of frequency on the shape and bandwidth of the auditory filter remain unknown.

Extensive training and testing at each masking condition delayed the collection of baseline measures of frequency selectivity until mice were aged at least 4–5 months. The early manifestation of age-related hearing loss is known to affect the high-frequency hearing of some mouse strains within this time period. CBA/CaJ mice, however, are one of the most resistant strains to aging effects (Hunter and Willott, 1987; Egorova *et al.*, 1993; Henry, 2004). They maintain a full complement of inner hair cells throughout their adult lifespan, and do not show the onset of outer hair cell loss until 18 months of age (Spongr *et al.*, 1997).

Our initial measures of auditory filter shape were obtained at a center frequency of 8 kHz. As expected by the delayed onset of anatomical changes in the CBA/CaJ cochlea, tests at this apical transduction site indicated normal function. Thresholds in quiet conformed to previously published hearing assessments (Fay, 1988; Zheng *et al.*, 1999), auditory filter shapes were well fit by standard $roex(p,r)$ procedures, and the resulting ERBs were in good agreement with psychoacoustic data from mammalian species with long lifespans. Subsequent testing verified the preservation of filter shape and bandwidth at two more basally located frequencies. These results remained stable during the first two replications of the threshold series.

C. Aging effects

Repeated testing in aging mice eventually demonstrated threshold changes that were small in magnitude but followed the orderly progression of human presbycusis. The deficits were most pronounced at 16 kHz, which was the highest frequency in our stimulus set. Average masked thresholds at 16 kHz remained within 2 dB of initial baselines until 95 weeks of age and then elevated by 10 and 23 dB when replications were completed at ages of 102 and 117 weeks. Thresholds at 11.2 kHz were within 7 dB of normal baselines at ages up to 112 weeks, then increased by 13 dB when the final replication was completed at 123 weeks.

The magnitude of AHL showed only minor variation between age-matched subjects. The final notched-noise replications at 16 kHz revealed threshold shifts ranging from a minimum of 17 dB in mouse CBA03lo to a maximum of 32 dB in mouse CBA03ch. Under quiet conditions, the two mice showed shifts of 15 and 20 dB.

As in humans (Florentine *et al.*, 1980; Sommers and Humes, 1993), deficits in frequency selectivity were correlated with age-related hearing loss. Changes in filter shape, however, were not intractably linked to the magnitude of hearing loss. When threshold shifts were uniform across masker conditions, the shape and bandwidth of the auditory filter were preserved. When shifts were restricted to a limited range of deviations, the bandwidth of tuning was substantially altered. For example, mouse CBA03bi experienced a 270% increase in the ERB during the final replication of 16-kHz thresholds although its average threshold shifts were typical of other mice.

The threshold shifts in our behavioral subjects were slow to develop and modest in magnitude when compared with the rapid profound deafness that is observed in strains such as C57BL/6J (Parham, 1997; Spongr *et al.*, 1997; Prosen *et al.*, 2003). Our present behavioral approach may not be suitable for characterizing the accelerated patterns of AHL that are observed in these specialized strains. It is conceivable, however, that similar masking effects may be captured with electrophysiological wave forms such as the auditory brainstem response or compound action potential (Dallos and Cheatham, 1976; Walton *et al.*, 1995). The refinement of efficient nonbehavioral screening procedures represents an opportunity to extend functional measures of auditory frequency selectivity to a broader range of subject ages, stimulus conditions, and mouse strains. Although CBA/CaJ mice may be regarded as a relatively uninteresting model of AHL, the preservation of normal function throughout the adult lifespan of this strain makes it an ideal preparation for the psychoacoustic validation of alternative electrophysiological approaches.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. R01 DC04841 (B.J.M.) and No. R15 DC04405 (C.A.P.). Undergraduate students at Northern Michigan University conducted the behavioral experiments. $Roex(p,r)$ analysis software was generously provided by Brian C. J. Moore. The authors thank Andrew Oxenham and two anonymous reviewers for their comments on an earlier version of this manuscript.

Au, W. W., and Moore, P. W. (**1990**). "Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **88**, 1635–1638.
Barsz, K., Ison, J. R., Snell, K. B., and Walton, J. P. (**2002**). "Behavioral and

neural measures of auditory temporal acuity in aging humans and mice," Neurobiol. Aging **23**, 565–578.
Beattie, R. C., and Warren, V. G. (**1982**). "Relationships among speech threshold, loudness discomfort, comfortable loudness, and PB max in the elderly hearing impaired," Am. J. Otol. **3**, 353–358.
Cheatham, M. A., Huynh, K. H., Gao, J., Zuo, J., and Dallos, P. (**2004**). "Cochlear function in Prestin knockout mice," J. Physiol. (Paris) **560**, 821–830.
Dallos, P., and Cheatham, M. A. (**1976**). "Compound action potential (AP) tuning curves," J. Acoust. Soc. Am. **59**, 591–597.
Egorova, M., Ehret, G., Vartanian, I., and Esser, K. H. (**2001**). "Frequency response areas of neurons in the mouse inferior colliculus. I. Threshold and tuning characteristics," Exp. Brain Res. **140**, 145–161.
Egorova, M. A., Vartanyan, I. A., and Ehret, G. (**2002**). "Neural critical bands and inhibition in the auditory midbrain of the house mouse (Mus domesticus)," Dokl. Biol. Sci. **382**, 5–7.
Ehret, G. (**1976**). "Critical bands and filter characteristics of the ear of the house mouse (Mus musculus)," Biol. Cybern. **24**, 35–42.
Ehret, G. (**1979**). "Correlations between cochlear hair cell loss and shifts of masked and absolute behavioral auditory thresholds in the house mouse," Acta Oto-Laryngol. **87**, 28–38.
Erway, L. C, Willott, J. F., Archer, J. R., and Harrison, D. E. (**1993**). "Genetics of age-related hearing loss in mice: I. Inbred and Fl hybrid strains," Hear. Res. **65**, 125–132.
Erway, L. C., Zheng, Q. Y., and Johnson, K. R. (**2001**). "Inbred strains of mice for genetics of hearing in mammals: Searching for genes for hearing loss," in *Handbook of Mouse Auditory Research: From Behavior to Molecular Biology*, edited by J. F. Willott (CRC Press, Boca Raton, Florida), pp 429–439.
Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysical Databook*. Winnetka (Hill-Illinois Fay Associates).
Finneran, J. J., Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002**). "Auditory filter shapes for the bottlenose dolphin (Tursiops truncatus) and the white whale (Delphinapterus leucas) derived with notched noise," J. Acoust. Soc. Am. **112**, 322–328.
Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–61.
Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (**1980**). "Frequency selectivity in normally-hearing and hearing-impaired observers," J. Speech Hear. Res. **23**, 646–669.
Francis, H. W., Ryugo, D. K., Gorelikow, M. J., Prosen, C. A., and May, B. J. (**2003**). "The functional age of hearing loss in a mouse model of presbycusis. II. Neuroanatomical correlates," Hear. Res. **183**, 29–36.
Gates, G. A., and Cooper, J. C. (**1991**). "Incidence of hearing decline in the elderly," Acta Oto-Laryngol. **111**, 240–248.
Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Gourevitch, G. (**1970**). "Detectability of tones in quiet and in noise by rats and monkeys," in *Animal Psychophysics: The Design and Conduct of Sensory Experiments*, edited by W. C. Stebbins (Appleton Century Crofts, New York), pp 67–97.
Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Robert E. Krieger, New York).
Henry, K. R. (**2004**). "Males lose hearing earlier in mouse models of late-onset age-related hearing loss; females lose hearing earlier in mouse models of early-onset hearing loss," Hear. Res. **190**, 141–148.
Hunter, K. P., and Willott, J. F. (**1987**). "Aging and the auditory brainstem response in mice with severe or minimal presbycusis," Hear. Res. **30**, 207–218.
Johnsson, L. G., Felix, H., Gleeson, M., and Pollak, A. (**1990**). "Observations on the pattern of sensorineural degeneration in the human cochlea," Acta Oto-Laryngol., Suppl. **470**, 88–95; discussion 95-86.
Lyregaard, P. E. (**1982**). "Frequency selectivity and speech intelligibility in noise," Scand. Audiol. **15**, 113–122.
Ma, W. L., Hidaka, H., and May, B. J. (**2006**). "Spontaneous activity in the inferior colliculus of CBA/J mice after manipulations that induce tinnitus," Hear. Res. **212**, 9–21.
Margolis, R., and Goldberg, S. M. (**1980**). "Auditory frequency selectivity in normal and presbycusic subjects," J. Speech Hear. Res. **23**, 603–613.
Mitchell, R. E. (**2006**). "How many deaf people are there in the United States? Estimates from the survey of income and program participation," J. Deaf Stud. Deaf Educ. **11**, 112–119.
Moore, B. C., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Moore, B. C., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.
Mosicki, J. K., Elkins, E. F., Baum, H. M., and Mcnamara, P. M. (**1985**). "Hearing loss in the elderly: an epidemiologic study of the framingham heart study cohort.," Ear Hear. **6**, 184–190.
Mullen, L. M., and Ryan, A. F. (**2001**). "Transgenic mice: Genome manipulation and induced mutations," in *Handbook of Mouse Auditory Research: From Behavior to Molecular Biology*, edited by J. F. Willott (CRC Press, Boca Raton, Florida), pp 457–474.
Muller, M., Smolders, J. W., Meyer Zum Gottesberge, A. M., Reuter, A., Zwacka, R. M., Weiher, H., and Klinke, R. (**1997**). "Loss of auditory function in transgenic Mpvl7-deficient mice," Hear. Res. **114**, 259–263.
Nadol, J. B., Jr. (**1979**). "Electron microscopic findings in presbycusic degeneration of the basal turn of the human cochlea," Otolaryngol.-Head Neck Surg. **87**, 818–836.
Niemiec, A. J., Yost, W. A., and Shomer, W. P. (**1992**). "Behavioral measures of frequency selectivity in the chinchilla," J. Acoust. Soc. Am. **92**, 2636–2649.
Nienhuys, T. G., and Clark, G. M. (**1979**). "Critical bands following the selective destruction of cochlear inner and outer hair cells," Acta Oto-Laryngol. **88**, 350–358.
Parham, K. (**1997**). "Distortion product otoacoustic emissions in the C57BL/6J mouse model of age-related hearing loss," Hear. Res. **112**, 216–234.
Patterson, R. D. (**1974**). "Auditory filter shape," J. Acoust. Soc. Am. **55**, 802–809.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Pickles, J. O. (**1975**). "Normal critical bands in the cat," Acta Oto-Laryngol. **80**, 245–254.
Prosen, C. A., Dore, D. J., and May, B. J. (**2003**). "The functional age of hearing loss in a mouse model of presbycusis. I. Behavioral assessments," Hear. Res. **183**, 44–56.
Prosen, C. A., Bath, K. G., Vetter, D. E., and May, B. J. (**2000**). "Behavioral assessments of auditory sensitivity in transgenic mice," J. Neurosci. Methods **97**, 59–67.
Pujol, R. (**1985**). "Morphology, synaptology and electrophysiology of the developing cochlea," Acta Oto-Laryngol., Suppl. **421**, 5–9.
Ries, P. W. (**1994**). "Prevalence and characteristics of persons with hearing trouble: United States, 1990-91," Vital Health Stat. 10 **188**, 1–75.
Scharf, B. (**1961**). "Complex sounds and critical bands," Psychol. Bull. **58**, 205–217.
Schuknecht, H. F., and Gacek, M. R. (**1993**). "Cochlear pathology in presbycusis," Ann. Otol. Rhinol. Laryngol. **102**, 1–16.
Seaton, W. H., and Trahiotis, C. (**1975**). "Comparison of critical ratios and critical bands in the monaural chinchilla," J. Acoust. Soc. Am. **57**, 193–199.
Shailer, M. J, Moore, B. C., Glasberg, B. R., Watson, N., and Harris, S. (**1990**). "Auditory filter shapes at 8 and 10 kHz," J. Acoust. Soc. Am. **88**, 141–148.
Sommers, M. S., and Humes, L. E. (**1993**). "Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners," J. Acoust. Soc. Am. **93**, 2903–2914.
Spongr, V. P., Flood, D. G., Frisina, R. D., and Salvi, R. J. (**1997**). "Quantitative measures of hair cell loss in CBA and C57BL/6 mice throughout their life spans," J. Acoust. Soc. Am. **101**, 3546–3553.
Taberner, A. M., and Liberman, M. C. (**2005**). "Response properties of single auditory nerve fibers in the mouse," J. Neurophysiol. **93**, 557–569.
Turner, C. W., and Robb, M. P. (**1987**). "Audibility and recognition of stop consonants in normal and hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 1566–1573.
Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. G, and Patterson, R. D. (**1984**). "Auditory filter asymmetry in the hearing impaired," J. Acoust. Soc. Am. **76**, 1363–1368.
Walton, J. P., Frisina, R. D., and Meierhans, L. R. (**1995**). "Sensorineural hearing loss alters recovery from short-term adaptation in the C57BL/6 mouse," Hear. Res. **88**, 19–26.

Wright, A., Davis, A., Bredberg, G., Ulehlova, L., and Spencer, H. (**1987**). "Hair cell distributions in the normal human cochlea," Acta Oto-Laryngol., Suppl. **444**, 1–48.
Zheng, Q. Y., Johnson, K. R., and Erway, L. C. (**1999**). "Assessment of hearing in 80 inbred strains of mice by ABR threshold analyses," Hear. Res. **130**, 94–107.
Zwicker, E., Flottorp, G., and Stevens, S. (**1957**). "Critical bandwidths in loudness summation," J. Acoust. Soc. Am. **29**, 548–557.

Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners

Rainer Beutelmann[a] and Thomas Brand[b]
Medizinische Physik, Fakultät V, Carl-von-Ossietzky-Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 18 March 2005; revised 7 April 2006; accepted 12 April 2006)

Binaural speech intelligibility of individual listeners under realistic conditions was predicted using a model consisting of a gammatone filter bank, an independent equalization-cancellation (EC) process in each frequency band, a gammatone resynthesis, and the speech intelligibility index (SII). Hearing loss was simulated by adding uncorrelated masking noises (according to the pure-tone audiogram) to the ear channels. Speech intelligibility measurements were carried out with 8 normal-hearing and 15 hearing-impaired listeners, collecting speech reception threshold (SRT) data for three different room acoustic conditions (anechoic, office room, cafeteria hall) and eight directions of a single noise source (speech in front). Artificial EC processing errors derived from binaural masking level difference data using pure tones were incorporated into the model. Except for an adjustment of the SII-to-intelligibility mapping function, no model parameter was fitted to the SRT data of this study. The overall correlation coefficient between predicted and observed SRTs was 0.95. The dependence of the SRT of an individual listener on the noise direction and on room acoustics was predicted with a median correlation coefficient of 0.91. The effect of individual hearing impairment was predicted with a median correlation coefficient of 0.95. However, for mild hearing losses the release from masking was overestimated.
 [DOI: 10.1121/1.2202888]

PACS number(s): 43.66Pn, 43.71An, 43.55Hy, 43.71Ky [AK] Pages: 331–342

I. INTRODUCTION

A binaural model, capable of predicting speech intelligibility under the influence of noise, reverberation, and hearing loss, may help in understanding the underlying mechanisms of binaural hearing and may assist in the development and fitting of hearing aids. In this study, a binaural model of speech intelligibility based on an approach by vom Hövel (1984) is presented and the model predictions are compared to measurement data. It combines two established models, the binaural equalization-cancellation (EC) processing (Durlach, 1963) with the monaural speech intelligibility index (SII, ANSI, 1997).

A number of studies are concerned with measuring the effects of spatial unmasking of speech. A detailed overview can be found in a review by Bronkhorst (2000). Research has focused on the influence of synthetic and natural spatial cues on speech intelligibility (Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997; Platte and vom Hövel, 1980; Plomp and Mimpen, 1981), on the influence of reverberation (Haas, 1972; Moncur and Dirks, 1967; Nábělek and Pickett, 1974) and hearing loss (Bronkhorst and Plomp, 1989; Duquesnoy and Plomp, 1983; Festen and Plomp, 1986; Irwin and McAuley, 1987).

Spatial unmasking of speech is based on spatial differences between target talker and interfering sources and can cause a benefit of speech reception threshold (SRT) of up to 12 dB (Bronkhorst, 2000). The basic cues for binaural processing are interaural time differences (ITD) due to the distance between the ears and interaural level differences (ILD) mainly due to the head shadowing effect. There are also spectral cues, mainly caused by the geometry of the pinna, but they play a less important role in spatial unmasking of speech (Mesgarani *et al.*, 2003).

A number of standardized methods of monaural speech intelligibility prediction exist in the literature, for instance the articulation index (AI, ANSI, 1969; Fletcher and Galt, 1950) and the speech intelligibility index (SII, ANSI, 1997), which was derived from the AI. A recent development by Müsch and Buus (2001a, b, 2004), the speech recognition sensitivity model, incorporates interactions between frequency bands which were neglected by the AI and SII. In this study, the standardized SII (ANSI, 1997) was used. However, the binaural part of the model is independent of the method for speech intelligibility prediction. Consequently, other methods can be used as well.

Models of binaural interaction in psychoacoustics, such as the models by Jefress (1948), Osman (1971), Colburn (1977a), and Lindemann (1986), provide a basis for some binaural speech intelligibility models. Zerbs (2000) and Breebaart *et al.* (2001a) each described a binaural signal detection model that uses peripheral preprocessing (modeled outer/middle ear, basilar membrane, and haircells) which converts the signals arriving at the ears into an internal representation. The binaural processing is done by an EC type of operation according to the theory by Durlach (1972). Both models differ in details, mainly in the way the internal inaccuracies are handled. The model presented here also makes use of the EC theory, but is kept simpler by omitting the peripheral preprocessing and working directly on the signals.

[a]Electronic mail: rainer.beutelmann@uni-oldenburg.de
[b]Electronic mail: thomas.brand@uni-oldenburg.de

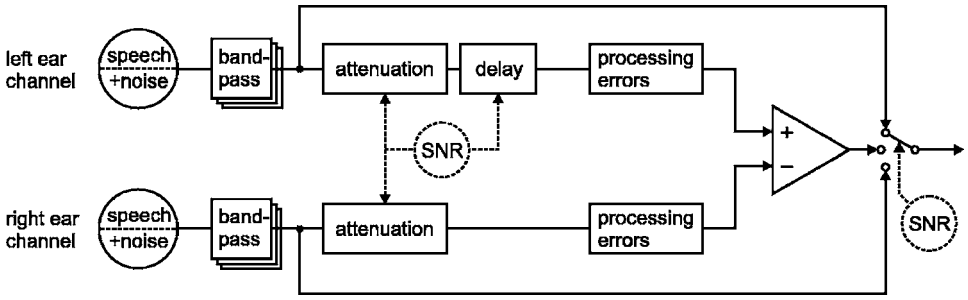


FIG. 1. Binaural processing using the modified, multifrequency channel EC model according to vom Hövel (1984). The speech and noise signals are processed identically, but separately for exact SNR calculation. The noise signal part includes the internal masking noise. Attenuation is only applied to one of the channels, depending on which of them contains more noise energy compared to the other.

The model of Culling and Summerfield (1995) in some way spans the gap between rather psychoacoustic binaural models and models related to binaural speech perception. It has been used to predict the release of masking for vowel intelligibility, but only qualitatively in the form of processed vowel spectra, where certain features could be identified or not. It incorporates most of the elements which were also used in this study, namely wave forms as input signals, a peripheral filter bank, and an EC type mechanism. Particularly, it features independent delays in each frequency band. There was no need for level equalization, because the stimuli contained only binaural time or phase differences.

Existing models of binaural speech intelligibility (Levitt and Rabiner, 1967; vom Hövel, 1984; Zurek, 1990) have certain common elements. They act as a preprocessing unit for monaural speech intelligibility models like the AI (Levitt and Rabiner, 1967; Zurek, 1990) or a modified version of the AI (vom Hövel, 1984). The benefit due to binaural interaction is expressed as a reduction of masking noise level after binaural processing. The models differ in the way they calculate the release of masking. Levitt and Rabiner (1967) used frequency dependent binaural masking level differences (BMLD) for interaurally phase reversed tones in diotic noise, taken from Durlach (1963), and subtracted these from the masking noise level. Zurek (1990) calculated the release from masking with the help of an equation from Colburn (1977b), using measured interaural level differences and an analytical expression for interaural phase differences. Vom Hövel (1984) derived an expression for the increase in signal-to-noise ratio based on EC theory. He used interaural parameters from actual transfer functions and incorporated a coarse estimate of the influence of reverberation.

The model presented in this study processes signal wave forms. Two uncorrelated internal masking noises accounting for the individual hearing thresholds of the two ears are added prior to dividing the binaural input signals into frequency bands and further processing. Independent EC stages in each band with artificial errors, which simulate human inaccuracy, calculate residual monaural signals consisting of speech and noise with the best possible signal-to-noise ratio. These signals are resynthesized into one broadband signal and with the aid of the SII a speech reception threshold is computed. Speech and noise have to be available as separate signals.

The goal of the present work was to determine the ability of such a straightforward functional model to predict binaural speech intelligibility under realistic conditions such as spatial sound source configuration, reverberation, and hearing loss. Model predictions were compared to observed SRTs for various combinations of noise source azimuths, room acoustic conditions, and hearing losses. To begin with, the idea by vom Hövel (1984) was maintained as far as possible, i.e., the combination of EC and SII and especially the original EC parameters. Only the SII-to-intelligibility mapping function was adjusted to measurement data from normal-hearing subjects without binaural and room acoustic cues, all other parameters were taken from literature and not fitted to speech intelligibility measurement data. Particular attention was paid to which of the listeners' individual characteristics (such as the pure tone audiogram) were necessary as parameters to produce accurate predictions. As a compromise between realistic situations and easy handling, measured manikin head related transfer functions including room impulse responses have been used.

II. METHODS

A. Model of binaural hearing

The model used in this study applies the EC principle (similar to the one proposed by Durlach, 1963), combined with the SII (ANSI, 1997) in order to predict binaural SRTs in noise. Additional masking noises were used to simulate the effects of hearing impairment. The binaural part is shown schematically in Fig. 1.

In the following, the inputs from the left and right ears will be referred to as "left ear channel" and "right ear channel," respectively. Each ear channel includes both speech and noise. Different parts of the interfering noise signal (cf. Sec. II B 2) from the Oldenburg Sentence Test (Wagener *et al.*, 1999a, b, c) filtered with the respective HRTFs were used as speech input signals and as noise input signals. These signals had the same long-term spectrum as the speech material used in the speech intelligibility measurements (important for the SII), speech and noise were uncorrelated (important for the EC model), and the variations of the actual sentences in level and spectrum were avoided. The speech and noise signals were supplied separately to the model to allow for exact signal-to-noise ratio (SNR) calculation. There was no difference between processing the sum of speech and noise or both signals separately and summing afterwards, since all processing steps were linear. The entire model was implemented using MATLAB (The MathWorks, 2002). The SII part was based on a MATLAB implementation of the one-third octave band SII procedure by Müsch (2003).

1. Gammatone filter bank analysis

The input signals were split into 30 frequency bands. Each band was 1 ERB (Glasberg and Moore, 1990) wide with center frequencies from 146 to 8346 Hz using a gammatone filter bank (Hohmann, 2002). Frequency components beyond this range were considered irrelevant for speech intelligibility. The gammatone filter transfer functions are based on the shape and bandwidth of the auditory filters of the basilar membrane (Patterson, 1976). The gammatone filter bank provides complex analytical output signals, which can be resynthesized after the binaural model processing with negligible artifacts.

2. Internal masking noise

Individual hearing thresholds were modeled by adding uncorrelated (between the left and right ear channel) Gaussian noise signals as internal masking noise to the external masking signals. The spectral shape of the internal masking noise was determined by the individual pure tone audiogram for the left and right ear, respectively. In each frequency band of the gammatone filter bank, the total noise energy equaled the energy of a sine tone 4 dB above the individual hearing threshold level at the corresponding band center frequency (Breebaart *et al.*, 2001a; Zwicker and Fastl, 1999).

3. EC stage

The equalization-cancellation processing takes advantage of the fact that signals from different directions result in different interaural time and level differences. It aims at maximizing the SNR in each frequency band. A simple way to maximize the SNR is to choose the ear channel with the largest SNR, but in many cases it is possible to utilize the time and level differences to exceed the SNR obtainable with a monaural signal.

The binaural processing (shown schematically in Fig. 1) is carried out in the model as follows: In each frequency band, the ear channels are attenuated and delayed[1] with respect to each other (*equalization* step), and then the right channel is subtracted from the left (*cancellation* step). The gain and delay parameters for the equalization step are chosen such that after cancellation step the SNR is maximal.[2] Thus there is no need for explicit decision between the two possible strategies of either minimizing the noise level or maximizing the speech level. The actual amplitude equalization is always realized by means of attenuating the correct ear channel, rather than amplifying the other, because in this way a seamless transition to the monaural case is achieved with increasing attenuation.

The noise level is minimized by subtracting one ear channel from the other, because all correlated noise components which are aligned after the equalization step can be eliminated due to destructive interference. Assuming that only the time and level differences of the noise signals are completely compensated for, but not the differences of the speech signals (when noise and speech come from different directions), more speech than noise remains in the resulting signal, which effectively increases the SNR. If the best possible SNR after binaural processing was still lower than the largest SNR of the monaural signal pairs, the best monaural signal pair was used in the SII calculation.

4. Artificial processing errors

Durlach (1963) proposed an artificial variance of the gain and delay parameters used in the EC process in order to model human inaccuracy. The model presented here used a modified way of calculation according to Vom Hövel (1984). The underlying assumption is that the EC processing in a given channel is carried out simultaneously in a number of parallel, equivalent processing units, which only differ in their (time invariant) processing errors. The final result is averaged over the outputs of all processing units (see the following).

The gain errors (ϵ_L, ϵ_R) and delay errors (δ_L, δ_R) of the left and right ear channel were Gaussian distributed, ϵ_L and ϵ_R on a logarithmic scale (level), δ_L and δ_R on a linear scale (time). Their standard deviations, σ_ϵ and σ_δ, depended on the actual gain (α) and delay (Δ) settings in each frequency band of the EC process defined by

$$\sigma_\epsilon = \sigma_{\epsilon 0}\left[1 + \left(\frac{|\alpha|}{\alpha_0}\right)^p\right], \quad \sigma_\delta = \sigma_{\delta 0}\left(1 + \frac{|\Delta|}{\Delta_0}\right) \tag{1}$$

with $\sigma_{\epsilon 0} = 1.5$ dB, $\alpha_0 = 13$ dB, $p = 1.6$, and $\sigma_{\delta 0} = 65$ μs, $\Delta_0 = 1.6$ ms. vom Hövel (1984) calculated these parameters by fitting BMLD predictions to results from measurements with pure tones in noise using a single frequency band ($f_0 = 500$ Hz) of his model with the above-mentioned processing errors. In this way, vom Hövel (1984) was able to predict BMLD data in S_0N_τ and $S_\pi N_\tau$ situations (Langford and Jeffress, 1964) with less deviation from the data than with the original model of Durlach (1963), which only limited the delay values to $|\Delta| < (2f_0)^{-1}$ in order to introduce artifical inaccuracy. Particularly in the S_0N_τ situation, the original model prediction had discontinuities which did not occur in the data of Langford and Jeffress (1964) and in the predictions of vom Hövel (1984). For the gain errors, BMLD data in S_mN_a situations (Blodgett *et al.*, 1962; Egan, 1965) were used, with monaural presentation (m) of the signal and various noise ILDs (a). These, too, could be predicted with the model of vom Hövel (1984) with deviations in the range of about 1 dB, while the original model of Durlach (1963) predicted BMLD values which were way to small and did not even fit qualitatively to the measured data.

In this study, the artificial errors were taken into account using a Monte Carlo method by generating 25 sets of Gaussian distributed random numbers for each of the 30 frequency bands with standard deviations according to Eq. (1) and adding them to the previously found optimal gain and delay values. All subsequent processing steps were carried out repeatedly for each of the 25 sets of errors resulting in a set of SRTs from which a mean SRT was calculated. Each SRT prediction is derived from 750 random values (i.e., 30 frequency channels times 25 Monte Carlo drawings), which supplies a sufficient statistical basis.

TABLE I. Hearing threshold at 500 Hz, pure tone average (mean of the hearing threshold in dB HL over 1, 2, and 4 kHz), hearing loss type and noise level in dB SPL used for the sentence tests of all hearing-impaired subjects participating in this study.

Subject number	Left ear			Right ear			Noise level
	500 Hz	PTA	Type	500 Hz	PTA	Type	
1	10.0	15.0	High freq	50.0	31.7	Flat	65
2	5.0	33.3	Steep	5.0	26.7	Steep	50
3	35.0	40.0	Flat	35.0	35.0	Flat	60
4	45.0	58.3	Flat	5.0	18.3	High freq	65
5	15.0	41.7	High freq	20.0	43.3	High freq	60
6	35.0	50.0	Sloping	25.0	41.7	Sloping	60
7	15.0	46.7	Sloping	50.0	58.3	Sloping	65
8	15.0	43.3	High freq	50.0	63.3	Flat	65
9	30.0	63.3	Sloping	30.0	55.0	Sloping	70
10	45.0	56.7	Sloping	45.0	65.0	Sloping	75
11	25.0	31.7	Flat	55.0	91.7	Steep	65
12	35.0	58.3	Steep	60.0	68.3	Flat	65
13	60.0	68.3	Flat	55.0	66.7	Flat	75
14	30.0	48.3	High freq	75.0	88.3	Flat	70
15	55.0	76.7	Sloping	55.0	60.0	Flat	65

5. Gammatone filter bank synthesis

The resulting speech and noise signals from each frequency band were resynthesized as described in Hohmann (2002) into a broadband speech and noise signal after the EC stage. The resynthesis step consisted of a phase and group delay adjustment in order to equalize the analysis filters according to Hohmann (2002), followed by a simple addition of the frequency bands. The broadband monaural signals were then used in the calculation of the speech intelligibility index. The signals could also be listened to or could be used to examine the benefit of the model binaural processing for human speech intelligibility using SRT measurements.

6. Speech intelligibility index

The SII was calculated from the resulting speech and noise spectra according to ANSI S3.5-1997 using the one-third octave band computational procedure (ANSI S3.5-1997, Table 3) with the band importance function "SPIN" (ANSI S3.5-1997, Table B.2). The hearing threshold level was set to −100 dB HL in the SII procedure, because the effect of hearing threshold was already taken into account by the internal masking noise (cf. Sec. II A 2).

Intelligibility scores for a number of overall speech levels (at constant noise level) were calculated from the corresponding SII values using a mapping function derived from the mapping function for "sentence intelligibility (I)" from Fletcher and Galt (1950, Table III, p. 96, and Fig. 7, p. 99). An adjustment of the SII-to-intelligibility mapping function is necessary to account for differences between the articulation of different speech materials. In this study, the adjustment was based on the anechoic S_0N_0 situation (cf. Sec. II B 3), since in this situation no binaural (same HRTF for speech and noise) or room acoustical effects are involved. First, a suitable analytical function [P(SII), the intelligibility score in percent as a function of the SII, Eq. (2)] was chosen, which described the original mapping function as close as possible,

$$P(\mathrm{SII}) = \frac{m}{a + e^{-b\cdot \mathrm{SII}}} + c, \quad P(0) = 0, \quad P(1) = 1. \tag{2}$$

For the SRT calculation, only the SII at 50% intelligibility is important, therefore only the parameter $a=0.019\,96$ was fitted to the anechoic S_0N_0 measurement data of the normal-hearing subjects, b was set to 20, which yields a slope (at the SRT) of the resulting psychometric function (intelligibility against SNR) close to the one measured by Wagener *et al.* (1999c) for the Oldenburg Sentence Test in noise (17.1%/dB). $m=0.8904$ and $c=-0.019\,96$ are defined by the boundary conditions. The parameters for the original mapping function from Fletcher and Galt (1950) were $a=0.1996$, $b=15.59$, $m=0.2394$, and $c=-0.1996$. The SRT was obtained by a simple search algorithm, which iteratively calculated an estimate of the psychometric function from the previously determined intelligibility scores and stopped, if the difference between the actual intelligibility at the estimated SRT and 50% was below a certain threshold (0.1%).

B. Measurements

1. Subjects

A total number of 10 normal-hearing and 15 hearing-impaired subjects participated in the measurements. Their ages ranged from 21 to 43 years (normal-hearing) and from 55 to 78 years (hearing-impaired).

The hearing levels of the normal-hearing subjects exceeded 5 dB HL at 4 or less out of 11 audiometric frequencies and 10 dB HL at only one frequency. None of the thresholds exceeded 20 dB HL.

The hearing-impaired subjects had various forms of hearing loss, including symmetric and asymmetric, flat, sloping, and steep high frequency losses. They are listed in Table I. Their (monaural) pure tone average (at 1, 2, and 4 kHz) ranged from 15 to 92 dB HL. Twelve hearing losses

TABLE II. Azimuth angles used for the presentation of noise signal. Negative values: left side, positive values: right side, from the subject's viewpoint.

Location	Angles							
Anechoic room & office room	−140°	−100°	−45°	0°	45°	80°	125°	180°
Empty cafeteria	−135°	−90°	−45°	0°	45°	90°	135°	180°

were only sensorineural, three had an additional conductive component. The subjects were paid for their participation.

2. Sentence test procedure

Speech intelligibility measurements were carried out using the HörTech Oldenburg Measurement Applications (OMA), version 1.2. As speech material, the Oldenburg Sentence Test in noise (Wagener *et al.*, 1999a, b, c) was used. Except for the convolution with binaural room impulse responses, the signals complied with the commercial version. A test list of 20 sentences was selected randomly from 45 such lists to obtain each observed SRT value. Each sentence consisted of five words with the syntactic structure *name verb numeral adjective object*. The subjects' task was to repeat every word they recognized after each sentence as closely as possible. The subjects responses were analyzed using word scoring. An instructor marked the correctly repeated words on a touch screen display connected to a computer, which adaptively adjusted the speech level after each sentence to measure the SRT level of 50% intelligibility. The step size of each level change depended on the number of correctly repeated words of the previous sentence and on a "convergence factor" that decreased exponentially after each reversal of presentation level. The intelligibility function was represented by the logistic function, which was fitted to the data using a maximum-likelihood method. The whole procedure has been published by Brand and Kollmeier (2002, A1 procedure). At least two sentence lists with 20 sentences each were presented to the subjects prior to each measurement session for training purposes.

The noise used in the speech tests was generated by randomly superimposing the speech material of the Oldenburg Sentence Test. Therefore, the long-term spectrum of this noise is similar to the mean long-term spectrum of the speech material. The noise was presented simultaneously with the sentences. It started 500 ms before and stopped 500 ms after each sentence. The noise level was kept fixed at 65 dB SPL (for the normal-hearing subjects). The noise levels for the hearing-impaired subjects were adjusted to their individual most comfortable level. They are listed in Table I. All measurements were performed in random order. The measurements with the hearing-impaired listeners were performed in the laboratory of Jürgen Kießling at the University of Gießen, Germany.

3. Acoustical conditions and calibration

Speech and noise signals were presented via headphones (Sennheiser HDA200) using HRTFs (head related transfer functions) in order to simulate different spatial conditions. The speech signals were always presented from the front (0°). The noise signals were presented from the directions shown in Table II. The terminology used here is S_0N_x for a situation where the speech signal was presented from front (0°) and the noise signal from an azimuth angle of x degrees. For example S_0N_{-45} is: speech from front (0°), noise from 45° to the left.

The speech and noise signals had been filtered with a set of HRTFs to reproduce both direction and room acoustics. Three different acoustical environments were used in the measurements: an anechoic room, an office room (reverberation time 0.6 s), and an empty cafeteria (reverberation time 1.3 s).

The headphones were free-field equalized according to international standard (ISO/DIS 389-8), using a FIR filter with 801 coefficients. The measurement setup was calibrated to dB SPL using a Brüel & Kjaer (B&K) 4153 artificial ear, a B&K 4134 $\frac{1}{2}$ in. microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier.

The anechoic HRTFs were taken from a publicly available database (Algazi *et al.*, 2001) and had been recorded with a KEMAR manikin. The office room and cafeteria HRTFs were own recordings with a B&K manikin using maximum length sequences. The sequences were played back by Tannoy System 800a loudspeakers and recorded with a B&K 4128C manikin and a B&K 2669 preamplifier. HRTF calculations were done using MATLAB on a standard PC equipped with an RME ADI-8 PRO analog/digital converter.

In the office room, the loudspeakers were placed in a circle with a radius of 1.45 m around the head center of the manikin which was seated in the middle of the room. The centers of the concentric loudspeaker diaphragms were adjusted to a height of 1.20 m, the height of a sitting, medium-height person's ears. In the cafeteria, a single loudspeaker was placed at different locations around the manikin seated in front of a table. A large window front, tilted from floor to ceiling, was situated at about 3 m distance from the manikin's head, making this situation rather asymmetric.

III. RESULTS AND DISCUSSION

A. Normal-hearing subjects

1. "Anechoic room" condition

Figure 2, left panel, shows predicted SRTs (open circles and crosses) and observed SRT data (closed circles) from eight normal hearing subjects (means and interindividual standard deviations) in anechoic conditions. The observed SRT for 0° noise azimuth (−8.0 dB) differed slightly from the reference value for monaural speech and noise presentation (−7.1 dB, Wagener *et al.*, 1999c). The SRT was approximately 1 dB lower than for noise from the front, if the noise was presented from 180° (from behind), but the difference

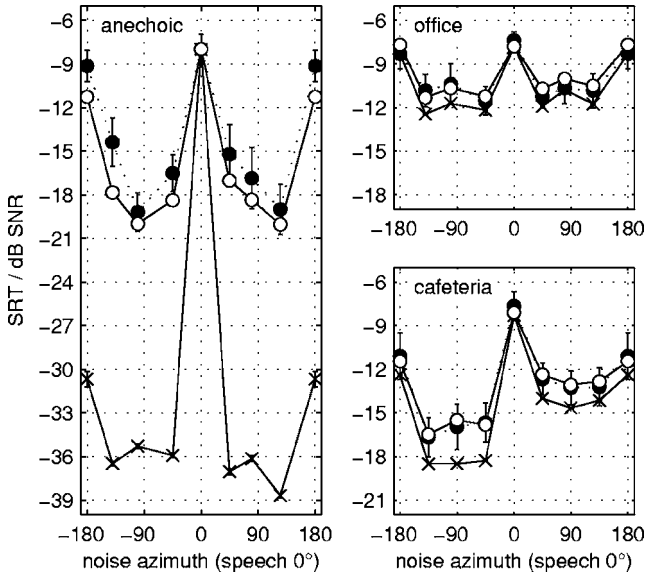


FIG. 2. SRTs for the Oldenburg sentence test with noise from different directions and speech from front (0°) in three room acoustic conditions. Data from eight normal hearing subjects. Closed circles: measurement data, mean, and interindividual standard deviation. Open circles: prediction with internal processing errors. Crosses: prediction without internal processing errors. The SRTs for 180° have been copied to −180° in order to point out the graph's symmetry. Left panel: anechoic room, upper right panel: office room, lower right panel: cafeteria.

was not significant. Lateral noise azimuths led to substantially lower SRTs. Maximum release from masking (difference to reference situation S_0N_0) was reached at a noise azimuth of −100° and could be as large as 12 dB.

The predicted SRT *including* internal processing errors (open circles) are lower than the observed values for all noise azimuths except 0°, which was the reference value for the adjustment of the SII-to-intelligibility mapping function. The prediction error (i.e., the absolute difference between predicted SRT and the corresponding observed SRT) has a mean of 1.9 dB for the individual data and 1.6 dB if both predictions and observed data are averaged across subjects. Although there are differences (≤20 dB) between the normal-hearing subjects in the individual audiograms (which have been taken into account by the model), these are not reflected in the predictions.

The model predictions *without* internal processing errors σ_ϵ and σ_δ [see Eq. (1)] of the EC model (crosses) resulted in SRTs that were much too low.

2. "Office room" conditions

Figure 2, upper right panel, shows predicted SRTs (open circles and crosses) and observed SRT data (closed circles) from eight normal hearing subjects in office conditions. The observed SRTs for noise from front (0°) as well as from behind (180°) did not significantly differ from the corresponding values in anechoic conditions (Fig. 2, left panel), but the release from masking in this situation was reduced to about 3 dB for all other noise azimuths (lateral angles).

The difference between model predictions with (open circles) and without internal processing errors (crosses) decreased compared to anechoic conditions to about 1 dB and less. In this room condition the prediction errors have a mean of 0.9 dB (individual data) and 0.5 dB (data averaged across subjects).

3. "Cafeteria" conditions

Figure 2, lower right panel, shows the predicted (open circles) and observed SRTs (closed circles) in reverberant empty cafeteria conditions. The difference of the observed SRT data compared to the office room and anechoic conditions at 0° noise azimuth was not significant. But there was a clear difference between this room and the others at 180° noise azimuth. The graph also shows a remarkable asymmetry between negative (left) and positive (right from the subject's viewpoint) azimuths. The release from masking at negative azimuths reached about 9 dB, but for positive azimuths the maximal release from masking was only 6 dB. The SRTs for the left side even fall into the range of the corresponding values for anechoic conditions. This asymmetry is probably caused by the asymmetric listening situation with the window front on the left side and the open cafeteria on the other side and will be discussed later.

Like in the office conditions, the difference between model predictions without internal processing errors (crosses) and predictions with internal processing errors (open circles) is much smaller for the cafeteria conditions than for anechoic conditions. The mean prediction error in the cafeteria is 1.1 dB (individual data) and 0.3 dB (data averaged across subjects).

4. Statistical analysis

An analysis of variance (ANOVA) of the observed data for the normal-hearing subjects showed a significant effect (at the 1% level) of both parameters (noise azimuth, room condition) and for interactions of noise azimuth and room condition. In the predicted data for normal-hearing subjects, significant effects (at the 1% level) were found for noise azimuth, room condition and their interaction.

B. Hearing-impaired subjects

In Fig. 3, three examples of individual predictions for hearing-impaired subjects are shown. All examples show a difference between observed (closed circles) and predicted (open circles) SRTs. Possible reasons for this difference will be discussed later. Subjects 7 and 4 have asymmetric hearing losses, with the better ear on the left side for subject 7 and on the right side for subject 4. The influence of these asymmetries can be seen, for instance in the anechoic condition. It leads to a substantial binaural benefit, if the noise source is close to the worse ear, because then the external SNR is larger at the better ear due to the head shadow. Therefore, subject 7 shows a large binaural benefit for noise at the right side and subject 4 for noise at the left side, which can be predicted very well by the model. Due to the large difference of hearing loss between the left and right ear of subject 4, the external SNR at the right, better ear determines most of the speech intelligibility. This is a simple task for the model, which only had to choose the ear with the better internal SNR (in each frequency band), which occurs at the right ear

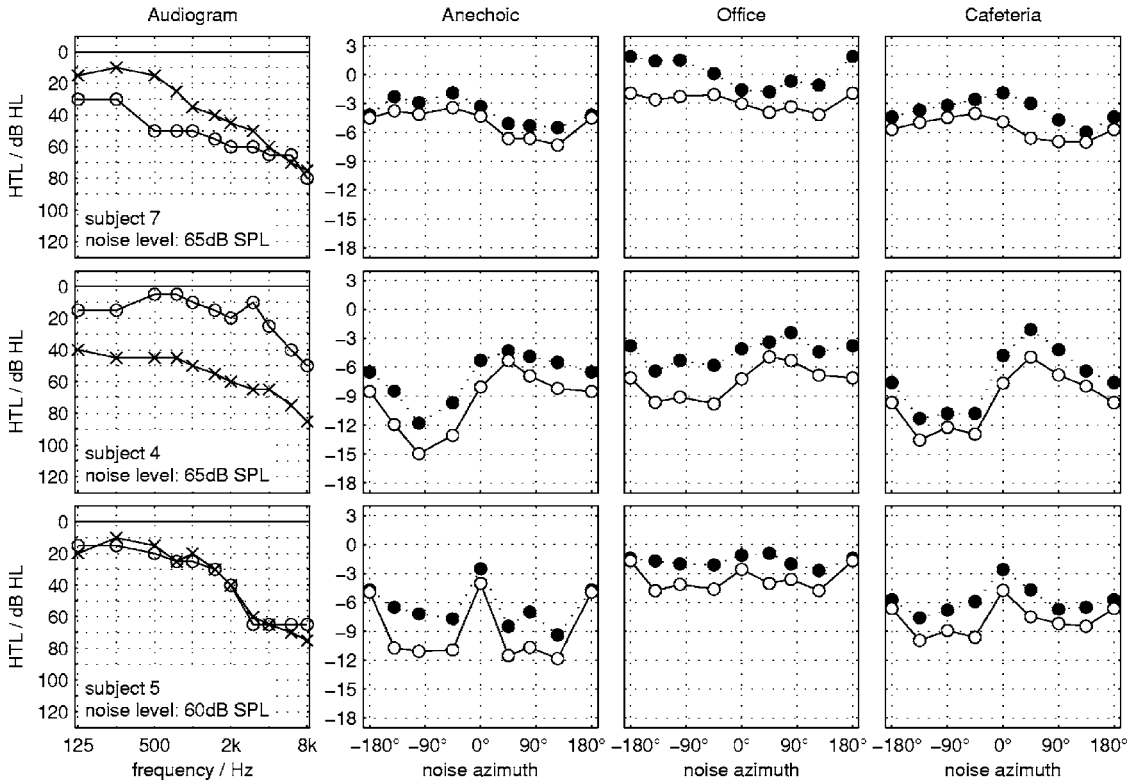


FIG. 3. Three examples of individual predictions of hearing-impaired subject data. Each row contains the results of one subject. The leftmost column shows the individual hearing loss of three listeners and the reference noise level used (crosses: left ear, circles: right ear). The other columns show individual observed SRTs (closed circles) and model predictions (open circles) for each of the three rooms (indicated by the titles). The speech signal was always at 0°. The SRTs for 180° have been copied to −180° in order to point out the graph's symmetry.

in most situations. The predictions for the symmetric hearing loss of subject 5 overestimate the binaural benefit in anechoic conditions. In the office situation, the binaural benefit is very small. For subject 7, the binaural benefit can even be negative at negative azimuths in anechoic and office conditions, which is also found qualitatively in the model predictions, although the prediction error is quite large for some angles. A stronger binaural effect than in the office condition could be found in the cafeteria condition, which is consistent with the results of the normal-hearing listeners.

Figure 4 shows predicted and observed SRTs for all hearing-impaired subjects plotted against each other, with each condition on a separate panel. There are three blocks of panels, each for one of the room acoustic conditions. In each panel, the observed SRTs of all subjects for one of the noise azimuths (indicated in the lower right corner) are plotted against the respective predicted SRTs. The dotted line in each panel represents identity.

The individual observed SRTs in each panel vary due to the different hearing losses and extend from values close to the ones measured in normal-hearing subjects in the corresponding situation to thresholds of almost +6 dB SNR, even in situations where a binaural release of masking should be expected. The maximal increase of SRT due to hearing loss (related to the corresponding mean SRT of all normal-hearing subjects) was 22 dB.

Clear correlations (coefficients greater than 0.9 except for Office/S_0N_{180} and Cafeteria/S_0N_0, >0.8) between predicted and observed SRTs were found. The lower correlations are mainly due to the small variance of observed and predicted data. In anechoic conditions and situations with noise from lateral positions, the binaural benefit was often overestimated by the model, indicated by the wider spread of dots toward lower predicted SRTs at low observed SRTs in the two leftmost columns of Fig. 4. This could not be related to hearing loss and/or noise level. The mean prediction errors for the rooms are 1.7, 1.9, and 1.9 dB (individual data, anechoic, office and cafeteria, respectively).

An ANOVA for the hearing-impaired subjects showed significant main effects (at the 1% level) for all parameters (noise azimuth, room condition, subject) as well as for all interactions of two parameters in both observed and predicted data.

C. Correlations

The overall correlation coefficient between all predicted and observed data shown in this study is 0.95. Regarding individual subjects, the correlation coefficients range from 0.69 to 0.99 with a median of 0.91. There is one subject with nonsignificant correlation (at the 5% level). This is due to the negligible release from masking ($\leq$2 dB) caused by the subject's large hearing losses at both ears (subject 15 in Table I) in combination with a noise level close to the subject's threshold rather than to an insufficient prediction.

The correlation coefficients for the data pooled across room conditions are 0.97, 0.94, and 0.94 (anechoic, office,

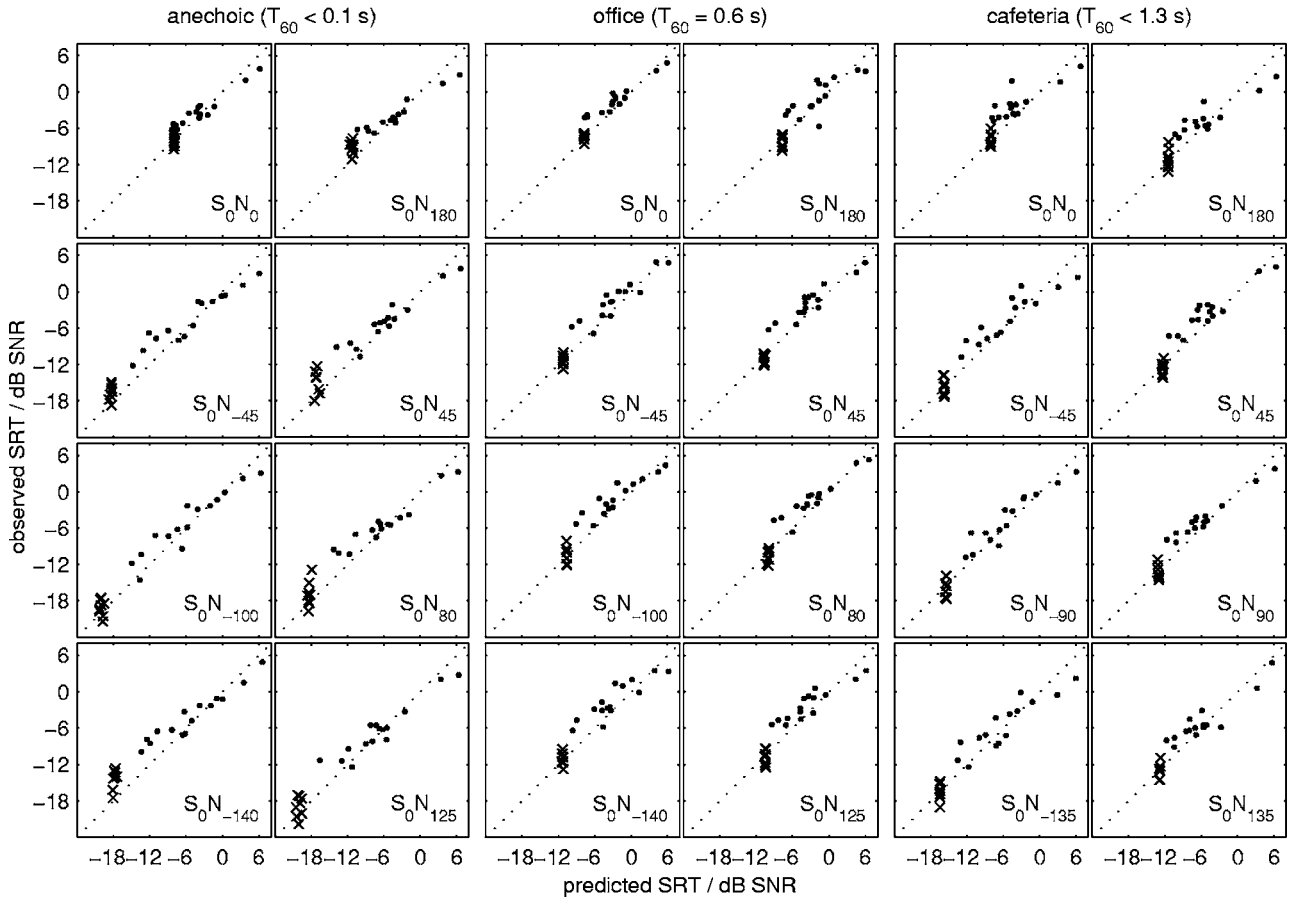


FIG. 4. Predicted and observed SRTs for all hearing-impaired subjects (dots) in this study. The observed SRTs are plotted against the predicted SRT values. Each panel contains the SRTs of 15 hearing-impaired subjects measured at one of the noise source azimuths which are indicated in the lower right corner. There are two columns of panels for each room condition, marked by the respective room names. The SRTs of the normal-hearing subjects (crosses) have been added for comparison.

cafeteria). If the average individual prediction error is subtracted from the predicted SRTs, all correlations increase to 0.98.

Pooled across noise azimuth, the correlation coefficients range from 0.90 to 0.97 with a median of 0.95. With the average individual prediction error subtracted, the median increases to 0.98 (0.94–0.99).

IV. GENERAL DISCUSSION

Although the correlations between model predictions and observed data are high, there are discrepancies between predicted and observed SRTs. A number of reasons for these discrepancies have to be considered and lead to several possibilities to improve the model predictions. Because the goal was to base the whole model on literature data, namely BMLD data of sinusoidals in noise and the standardized SII (ANSI, 1997), only the SII-to-intelligibility mapping function has been adjusted and all other discrepancies have not yet been corrected for in this study. Further work on the model has to include adjustment of internal parameters and possibly the use of further individual external parameters.

The predictions of data in the present study showed an individual average prediction error of −4.1 to +2.5 dB. Although the difference between the mean prediction errors of normal-hearing and hearing-impaired subjects is small (0.5 dB), it is significant (at the 1% level) and the predicted SRTs for hearing-impaired subjects are too low in most cases. It is known from literature (Pavlovic, 1984; Plomp, 1978), that not all of the decrease of monaural speech intelligibility due to hearing loss can be explained only by the individual hearing threshold. The question is whether the binaural part of the model needs to be fed with additional individual data or only the monaural back-end. The latter would mean that binaural processing itself is not affected by the hearing loss, but simply has to deal with the incomplete information coming from the impaired ear. It is still surprising how much of the binaural speech intelligibility measured in this study seems to be determined by audibility. This may be due to the fact that the noise level was adjusted to the individual most comfortable level and was clearly audible, but often close to the hearing threshold, which emphasizes the influence of the threshold.

The predictions for all S_0N_0 situations with and without processing errors are almost equal, which means that an adjustment of the processing error parameters would not change the prediction at S_0N_0. In anechoic conditions, the prediction error for S_0N_0 is smaller than at other noise azimuths, above all S_0N_{180}. 180° and 0° azimuth both result in ITDs and ILDs around zero, and the differences between the HRTFs at 0° and 180° may have been small, but still of use

for the binaural model. Since the HRTFs used for speech and noise in the S_0N_0 situation were exactly the same, not much of an effect of binaural processing could be expected.

The artificial processing errors assumed by the model turn out to be crucial for correct predictions. In reverberant situations there is only a small difference between predictions with and without processing errors. In the anechoic situation, however, the processing errors have a large influence. The differences between the mean prediction errors of the different room conditions (anechoic: 2 dB, office/cafeteria: about 1 dB) for normal hearing subjects appear to be related to the different influence of the processing errors. Moreover, the predictions overestimate the binaural benefit for all subject groups particularly in situations with a strong effect of binaural processing, i.e., when large binaural benefit occurs and for hearing-impaired subjects with symmetric hearing loss, where the better SNR is not necessarily determined by the better ear. Changing the processing error parameters should change the prediction error mainly in the above-mentioned situations where the prediction error is large and thus may improve predictions of absolute SRTs as well as equalize the difference between room conditions. A preliminary study has shown that variation of $\sigma_{\epsilon 0}$ and $\sigma_{\delta 0}$ by a common factor between 0.5 and 2 leads to continuous changes in the predictions of situations with a large influence of the processing error. Nevertheless, there is no quick solution, all error parameters have to be considered.

For normal hearing subjects, no strong dependence of the SRTs on the hearing threshold in both prediction and measurement data would be expected. Although there is only a small difference between individual predicted SRTs, the observed SRTs vary across subjects. The typical standard deviation of the Oldenburg sentence test of about 1 dB (Wagener *et al.*, 1999a, b, c) cannot explain all of this variance. Other factors which cannot be modeled and which are difficult to control experimentally, such as individual attention and motivation, are probably responsible. In this light it is remarkable that the prediction error standard deviations in the different rooms are almost the same for normal-hearing and hearing-impaired subjects.

It is surprising that in the room with the largest reverberation time (cafeteria hall, T_{60}=1.3 s) the release from masking is larger than in the office room, which has only half the reverberation time (0.6 s). Using another room acoustical measure related to the energy in the early parts of the room impulse response, definition or D_{50}, gives a hint why the SRTs are generally lower in the cafeteria than in the office room. The D_{50} is calculated in octave bands and is the ratio between the energy arriving in the first 50 ms and the energy of the whole impulse response. The D_{50} is a common measure used for characterizing rooms in terms of speech perception (ISO 3382; CEN, 2000). Bradley and Bistafa (2002) have shown, that early/late ratios can be a quite good predictor of speech intelligibility in rooms. The D_{50} values averaged over all eight azimuths do not differ significantly between office room and cafeteria at 1–8 kHz (all $>$0.9), but they are generally higher for the cafeteria in the low frequency bands (office/cafeteria 125 Hz: 0.70/0.76, 250 Hz: 0.75/0.89, 500 Hz: 0.84/0.88), which would correctly predict better intelligibility in the cafeteria. The reduced release from masking at positive noise azimuths (to the right of the listener) in relation to negative noise azimuths can be attributed to the reflection of a large window front to the left of the listener. It creates a second, virtual noise source, if the actual noise source is located on the opposite side, which hampers the binaural processing. As can be seen from the predictions, the model is capable of taking these effects into account.

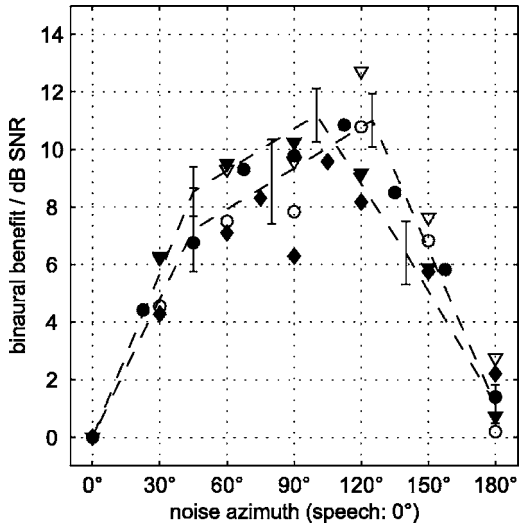


FIG. 5. Release from masking for various noise azimuths with a single noise source and speech presented from the front (0°) relative to the SRT in the S_0N_0 situation. Observed release from masking for eight normal-hearing listeners measured in this study shown with dashed lines (left and right side of the listener) and interindividual standard deviation. The other data points are taken from Platte and vom Hövel (1980, open circles and triangles), Plomp and Mimpen (1981, closed circles), Bronkhorst and Plomp (1988, closed triangles), Peissig and Kollmeier (1997, diamonds) according to Bronkhorst (2000).

A. Comparison with literature data

In Fig. 5, the observed SRT difference compared to the S_0N_0 situation for various noise azimuths and normal-hearing subjects that were obtained in this study are compared to data from a number of similar experiments in literature (Bronkhorst, 2000; Bronkhorst and Plomp, 1988; Peissig and Kollmeier, 1997; Platte and vom Hövel, 1980; Plomp and Mimpen, 1981). All studies used a single, speech-shaped noise source as an interferer. Regardless of the differences in measurement procedures (speech material, noise level, realization of the binaural configuration), the data from literature show a clear trend of release from masking being dependent on noise azimuth. The maximum benefit is found at azimuths of about 105°–120° rather than at 90° where it might be expected. The data from Peissig and Kollmeier (1997) even shows a dip at 90°, due to interference effects. The data from this study fit very well into the range of values found in the literature.

B. Comparison to other models

The model presented here extends the model proposed by vom Hövel (1984). The basic principle, multifrequency band equalization, and cancellation, followed by a monaural

speech intelligibility model, is the same. Extending the model in order to predict data of hearing-impaired subjects was possible by adding a masking noise. It yielded encouraging results without changes in the basic principle, but still needs improvement. The handling of early reflections was left to the EC process instead of explicit division of the room impulse response into useful and detrimental sections like in the model by vom Hövel (1984). Although the effect of room acoustics on the noise signal seems to dominate the binaural perception in the approach of the current study with a rather close speech source and a limited amount of reverberation, care must be taken if the disturbance of the speech signal itself due to reverberation becomes as strong as the effect of the external noise. Solutions to this shortcoming are discussed in the following. The present model's advantage over models like the ones according to vom Hövel (1984) or Zurek (1990) is that it is, in principle, not limited to known HRTFs or spatial configurations, but is still relatively simple.

In the binaural part, the present model is very similar to psychoacoustic models like the ones from Zerbs (2000) or Breebaart *et al.* (2001a), because they are all based on the EC principle. This similarity, and the independence of front-end (EC process) and back-end (SII) in the current model, facilitates the transfer of developments and knowledge between psychoacoustical models and the speech model presented in this study. For example, the present model does not incorporate any peripheral preprocessing like a hair cell model or compression. These could replace the somewhat arbitrary binaural processing errors, because half-wave rectification and low-pass filtering smear the high frequency signal components in a manner similar to the delay processing errors in high-frequency bands. Compression also introduces decorrelation between the ears especially if large ILDs are involved (Breebaart *et al.*, 2001b) and thus acts in a similar manner to the amplitude processing errors.

The present model goes beyond the model by Culling and Summerfield (1995) by actually using the output from the binaural processing to predict speech intelligibility quantitatively. Culling and Summerfield (1995) were able to decide from their recovered spectra (activity in each frequency band after applying the best delay for each band independently), if certain vowel features were present or not. These recovered spectra were an expression of the effects of binaural hearing, but to predict actual speech intelligibility, the frequency dependent weighting of the SII (or similar models) is necessary. For the predictions in the present study, other parameters of binaural coincidence detectors like the shape of temporal integration windows, as Culling and Colburn (2000) mentioned, were obviously not crucial or implicitly included in the internal noise parameters by vom Hövel (1984).

In the same way as Culling and Summerfield (1995), the EC processing in the present model implies little or no interaction between the frequency bands. This is in accordance with the findings by Akeroyd (2004), who has found that binaural detection experiments with complex tones in noise in different binaural configurations yield thresholds which are more consistent with free ITD equalization across different frequency bands than with ITD equalization using the same delay for all frequency bands. Edmonds and Culling (2005) also found that speech intelligibility measurements with opposed ITD of speech and noise ($\pm 500\ \mu s$) yield the same thresholds when the ITDs are fixed over the whole frequency range and when the ITDs of speech and noise are swapped at frequencies exceeding a certain splitting frequency between 750 and 3000 Hz. While this study focused on the binaural processing of different simultaneous spatial cues, another matter is the time needed to switch between binaural processing strategies or to select one of several possibilities (cf. Kohlrausch, 1990). However, it should still be investigated whether the EC parameters are completely independent across frequency bands or if there is a remaining interaction, even when it is weak. Measurements with artificial interfering noise that require gain and/or delay parameters in the EC processing which differ widely between neighboring frequency bands would help in determining the importance of band interaction at the EC stage of the model.

C. Possible extensions

Overall, the results show that the model is capable of predicting the influence of room acoustics on speech intelligibility. Strictly speaking, this only holds for the influence of room acoustics on the noise (for instance, the emergence of additional "mirror" noise sources caused by early reflections). Since the model assumes the whole speech energy as being useful, it only holds for near field speech, because the disturbance of the speech itself caused by reverberation is not taken into account. It might be possible to solve this shortcoming using the speech transmission index (STI, IEC, 1998), which could be used either instead of the SII or as a kind of correction factor. Since the STI considers the modulation transfer function, it is very successful in predicting the influence of room acoustics on speech intelligibility.

In the light of a possible application of the model as a signal processing device, it would be desirable to remove the constraint of separated speech and noise signals. The need for separate speech and noise signals originates only from the way the SNR is calculated in the EC step. Any other way of calculating a sufficiently accurate SNR from the combined speech and noise signals can be principally incorporated into the model and would remove the constraint.

A further step toward a more comprehensive model that takes attention mediated processes into account is probably much more difficult. The fact that the model needs speech and noise in separate recordings implies that the listener is able to distinguish perfectly between speech and noise. Therefore, the experimental setup of this study, using non-modulated speech-shaped noise, certainly supported the accordance between predictions and observations. Maskers that involve informational masking, like competing voices, are clearly much more challenging for models of speech intelligibility.

Nevertheless, even in its present form, the model shows a strong relationship between tone audiogram and binaural speech intelligibility, which might help audiologists to classify clinical results. A recent study (Brand and Beutelmann, 2005) applied the model to a clinical database of 238

hearing-impaired subjects. This large number of different hearing impairments will certainly help in the further development of the model,

V. CONCLUSIONS

(1) A relatively straightforward functional model of binaural speech intelligibility consisting of a gammatone filter bank (Hohmann, 2002), an independent equalization-cancellation process (Durlach, 1963) in each frequency band, a gammtone resynthesis and the speech intelligibility index (SII, ANSI S3.5-1997) yielded high correlations between predictions and measurements of binaural SRT data for spatial arrangement of noise and speech sources (within the horizontal plane) in anechoic as well as reverberant room environments. In order to simulate the limited human accuracy, pure tone in noise BMLD data has been used to determine the maximum precision of the EC-process (vom Hövel, 1984). Only the SII-to-intelligibility mapping function has been adjusted and no other parameters have been fitted to speech intelligibility data, but because it was not possible to predict all absolute SRTs accurately, an adjustment of model parameters to match predictions and measurement data should be considered.

(2) Without changes, the model yields similar correlations between predicted and observed SRTs for both normal-hearing and hearing-impaired subjects and the same order of magnitude in prediction accuracy of relative binaural effects. Regarding absolute SRTs, there is a difference between normal-hearing and hearing-impaired subjects, which probably originates from suprathreshold effects of the hearing impairment, which are not treated by the model.

(3) Early reflections that lead to "mirror" noise sources disrupt binaural unmasking more strongly than long reverberation tails of the room impulse response. This was consistent with the model predictions.

(4) The human processing errors assumed in the EC stage were highly relevant in the anechoic condition. In the conditions with reverberation the predictions were hardly influenced by the processing errors.

ACKNOWLEDGMENTS

We are very grateful to Birger Kollmeier for his substantial support and contribution to this work. We would like to thank Birgitta Gabriel, Daniel Berg, Jürgen Kießling, and Matthias Latzel for organizing and performing the measurements. This work was motivated by helpful discussions with many colleagues, including Kirsten Wagener, Volker Hohmann, and Jesko Verhey. We would also like to thank the editor, Armin Kohlrausch, and two anonymous reviewers for their thorough and helpful reviews. This work was supported by BMBF (Kompetenzzentrum HörTech) and the European 6th Framework Programme "HEARCOM."

[1]The time delay of one channel relative to the other one was realized by means of fast Fourier transformation and multiplication with a phase factor in the frequency domain. This allowed delay times smaller than the sample period. The signals were padded with sufficient zero samples (about 3.5 ms) at both ends to avoid circular aliasing.

[2]A numerical optimization procedure (simplex-based MATLAB function `fminsearch`) was used to find the optimum gain and delay values, which yielded maximum SNR. The SNR was calculated via the rms difference between the resulting speech and noise signal after subtraction of the amplified and delayed left ear channel from the right one. Suitable initial gain and delay values for the optimization procedure were estimated by evaluating a short section of the noise signal: the rms difference between the ear channels was used as gain initial value, the delay was initialized with the lag of the cross-correlation maximum. The SNR as a function of gain and delay exhibits local maxima due to the periodic structure of the bandpass filtered signals. To find the global maximum (assumed that a first search may have only found a local maximum) the optimization procedure was started again with initial parameters close to neighboring local maxima. These could be found at delay intervals calculated from the center frequency of the current bandpass $(1/f_c)$.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**(2), 1135–1148.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001**). "The CIPIC HRTF Database," in Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics.

ANSI (**1969**). "Methods for the Calculation of the Articulation Index," American National Standard S3.5–1969, Standards Secretariat, Acoustical Society of America.

ANSI (**1997**). "Methods for the Calculation of the Speech Intelligibility Index," American National Standard S3.5–1997, Standards Secretariat, Acoustical Society of America.

Blodgett, H. C., Jeffress, L. A., and Whitworth, R. H. (**1962**). "Effect of noise at one ear on the masked threshold for tone at the other," J. Acoust. Soc. Am. **34**(7), 979–981.

Bradley, J. S., and Bistafa, S. R. (**2002**). "Relating speech intelligibility to useful-to-detrimental sound ratios," J. Acoust. Soc. Am. **112**(1), 27–29.

Brand, T., and Beutelmann, R. (**2005**). "Examination of an EC/SII based model predicting speech reception thresholds of hearing-impaired listeners in spatial noise situations," in Proceedings of the 21st Danavox Symposium "Hearing Aid Fitting", edited by A. N. Rasmussen, T. Poulsen, T. Andersen, and C. B. Larsen, ISBN 87–982422–0–2, Center Tryk, Denmark, **2006**, pp. 139–151.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**(6), 2801–2810.

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**(2), 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**(2), 1089–1104.

Bronkhorst, A. W. (**2000**). "The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple Talker Conditions," Acust. Acta Acust. 86, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**(4), 1508–1516.

Bronkhorst, A. W., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**(4), 1374–1383.

CEN (**2000**). "Messung der Nachhallzeit von Räumen mit Hinweis auf andere akustische Parameter," European Standard EN ISO 3382, Europäisches Komitee für Normung.

Colburn, H. S. (**1977a**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise.," J. Acoust. Soc. Am. **61**(2), 525–533.

Colburn, H. S. (**1977b**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Supplementary material," AIP document no. PAPS JASMA-91-525-98.

Culling, J. F., and Colburn, H. S. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise.," J. Acoust. Soc. Am. **107**(1), 517–527.

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common

interaural delays," J. Acoust. Soc. Am. **98**(2), 785–797.
Duquesnoy, A. J., and Plomp, R. (**1983**). "Effect of a single interfering noise or speech source upon the binaural intelligibility of aged persons," J. Acoust. Soc. Am. **74**(3), 739–743.
Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**(8), 1206–1218.
Durlach, N. I. (**1972**). *Binaural Signal Detection: Equalization and Cancellation Theory* (Academic, New York), Vol. **II**, Chap. 10, pp. 371–462.
Edmonds, B. A., and Culling, J. F. (**2005**). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**(5), 3069–3078.
Egan, J. P. (**1965**). "Masking level differences as a function of interaural disparities in intensity of signal and of noise," J. Acoust. Soc. Am. **38**(6), 1043–1049.
Festen, J. M., and Plomp, R. (**1986**). "Speech-reception threshold in noise with one and two hearing aids," J. Acoust. Soc. Am. **79**(2), 465–471.
Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**(2), 89–151.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103–138.
Haas, H. (**1972**). "The influence of a single echo on the audibility of speech," J. Audio Eng. Soc. **20**(2), 146–159.
Hohmann, V. (**2002**). "Frequency analysis and synthesis using a Gammatone filterbank," Acust. Acta Acust. **88**(3), 433–442.
IEC (**1998**). "Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index," International Standard IEC 60268-16.
Irwin, R. J., and McAuley, S. F. (**1987**). "Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation," J. Acoust. Soc. Am. **81**(5), 1557–1565.
Jefress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Kohlrausch, A. (**1990**). "Binaural masking experiments using noise maskers with frequency-dependent interaural phase differences. II. Influence of frequency and interaural-phase uncertainty," J. Acoust. Soc. Am. **88**(4), 1749–1756.
Langford, T. L., and Jeffress, L. A. (**1964**). "Effect of noise crosscorrelation on binaural signal detection," J. Acoust. Soc. Am. **36**(8), 1455–1458.
Levitt, H., and Rabiner, L. R. (**1967**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**(4), 820–828.
Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contrallateral inhibition. I. Simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**(6), 1608–1622.
Mesgarani, N., Grant, K. W., Shamma, S., and Duraiswami, R. (**2003**). "Augmented intelligibility in simultaneous multi-talker environments," in Proceedings of the 2003 International Conference on Auditory Display, International Community for Auditory Display, Boston, MA, pp. 71–74.
Moncur, J. P., and Dirks, D. (**1967**). "Binaural and monural speech intelligibility in reverberation," J. Speech Hear. Res. **10**, 186–195.
Müsch, H. (**2003**). "MATLAB implementation of core aspects of the American National Standard 'Methods for calculation of the Speech Intelligibility Index' ANSI S3.5-1997," Downloadable MATLAB Script: http://www.sii.to.
Müsch, H., and Buus, S. (**2001a**). "Using statistical decision theory to predict speech intelligibility. I. Model structure," J. Acoust. Soc. Am. **109**(6), 2896–2909.
Müsch, H., and Buus, S. (**2001b**). "Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance," J. Acoust. Soc. Am. **109**(6), 2910–2920.
Müsch, H., and Buus, S. (**2004**). "Using statistical decision theory to predict speech intelligibility. III. Effect of audibility on speech recognition sensitivity," J. Acoust. Soc. Am. **116**(4), 2223–2233.
Nábělek, A. K., and Pickett, J. M. (**1974**). "Reception of consonants in a classroom as affected by monaural and binaural listening, noise, reverberation, and hearing aids," J. Acoust. Soc. Am. **56**(2), 628–638.
Osman, E. (**1971**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **6**(2), 1494–1511.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**(3), 640–654.
Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**(4), 1253–1258.
Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**(3), 1660–1670.
Platte, H.-J., and vom Hövel, H. (**1980**). "Zur Deutung der Ergebnisse von Sprachverständlichkeitsmessungen mit Störschall im Freifeld," Acustica **45**(3), 139–150.
Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**(2), 533–549.
Plomp, R., and Mimpen, A. M. (**1981**). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences," Acustica **48**, 325–328.
The MathWorks (**2002**). "MATLAB®6.5."
vom Hövel, H. (**1984**). "Zur bedeutung der übertragungseigenschaften des außenohrs sowie des binauralen hörsystems bei gestörter sprach bertragng," Dissertation, Fakultät für Elektrotechnik, RTWH Aachen.
Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999a**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache. I. Design des Oldenburger Satztests," Zeitschrift für Audiologie/Audiological Acoustics **38**(1), 4–14.
Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999b**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache. II. Optimierung des Oldenburger Satztests," Zeitschrift für Audiologie/Audiological Acoustics **38**(2), 44–56.
Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999c**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache. III. Evaluation des Oldenburger Satztests," Zeitschrift für Audiologie/Audiological Acoustics **38**(3), 86–95.
Zerbs, C. (**2000**). "Modelling the effective binaural signal processing in the auditory system," dissertation, Carl-von-Ossietzky-Universität Oldenburg.
Zurek, P. M. (**1990**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, London), Chap. 15, pp. 255–276.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics - Facts and Models*, 2nd ed. (Springer, Berlin).

Perceptual recalibration in human sound localization: Learning to remediate front-back reversals

Pavel Zahorik[a)]
Department of Psychological and Brain Sciences, University of Louisville, Louisville, Kentucky 40292

Philbert Bangayan, V. Sundareswaran, Kenneth Wang, and Clement Tam
Rockwell Scientific, 1049 Camino Dos Rios, Thousand Oaks, California 91360

(Received 28 March 2005; revised 4 May 2006; accepted 5 May 2006)

The efficacy of a sound localization training procedure that provided listeners with auditory, visual, and proprioceptive/vestibular feedback as to the correct sound-source position was evaluated using a virtual auditory display that used nonindividualized head-related transfer functions (HRTFs). Under these degraded stimulus conditions, in which the monaural spectral cues to sound-source direction were inappropriate, localization accuracy was initially poor with frequent front-back reversals (source localized to the incorrect front-back hemifield) for five of six listeners. Short periods of training (two 30-min sessions) were found to significantly reduce the rate of front-back reversal responses for four of five listeners that showed high initial reversal rates. Reversal rates remained unchanged for all listeners in a control group that did not participate in the training procedure. Because analyses of the HRTFs used in the display demonstrated a simple and robust front-back cue related to energy in the 3–7-kHz bandwidth, it is suggested that the reductions observed in reversal rates following the training procedure resulted from improved processing of this front-back cue, which is perhaps a form of rapid perceptual recalibration. Reversal rate reductions were found to generalize to untrained source locations, and persisted at least 4 months following the training procedure. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2208429]

PACS number(s): 43.66.Qp, 43.66.Pn, 43.66.Lj [AK] Pages: 343–359

I. INTRODUCTION

Unlike vision and touch, where the perceptual representations of space result from topographically oriented sensory receptors, auditory space must be computed from the one-dimensional acoustic waveforms reaching a listener's ears. These waveforms contain various sources of acoustic information, or cues, which are processed by auditory neural pathways in order to produce a perceptual representation of the sound source's spatial location. Because these cues can be affected by a variety of factors, ranging from acoustical aspects of the listening environment to developmental changes in head size and ear morphology, the perceptual processing subserving sound localization must adapt to these cue changes in order to maintain accuracy. Although it seems likely that other sensory systems play a significant role in this process of adaptation and calibration for directional sound localization, questions regarding the precise nature, extent, and time course of the adaptation remain.

Various examples of spatial adaptation to changes or degradations to one or more of the acoustic cues to direction have been reported in the literature. Unilaterally deafened humans and animals are often able to retain or regain relatively accurate directional sound localization abilities without the benefit of binaural directional cues (Florentine, 1976; Hausler *et al.*, 1983; King, 1999; Knudsen, 1999; Slattery and Middlebrooks, 1994; Van Wanrooij and Van Opstal, 2004). Informal observation suggests that humans do not suffer any degradation in directional localization abilities throughout the course of normal development, even though the interaural time difference (ITD) cue is known to change dramatically with the developmental changes in head size (Clifton *et al.*, 1988). Adaptation to the acoustical properties of the listening environment also appears to take place. Repeated exposure to a source with a single echo results in suppression of the spatial attributes of the echo, and releases from this suppression occur when the acoustic environment is changed (Clifton, 1987; Clifton *et al.*, 2002). Listeners can also adapt to other alterations to the binaural directional cues resulting from experimental manipulations including lateral position displacement (Held, 1955; Young, 1928) and lateral position distortion (Shinn-Cunningham *et al.*, 1998a, 1998b).

Relatively little is known about spatial adaptation to altered spectral cues to direction—monaural cues that result from the complex diffraction and reflection patterns of sound about the listener's head and external ear (pinna). Spectral cues are thought to be important for determining the elevation of the sound source, especially on the median plane where the binaural cues provide little directional information (Gardner and Gardner, 1973; Hebrank and Wright, 1974a, 1974b), and for determining whether a source is in front or behind the listener (Kistler and Wightman, 1992; Langendijk and Bronkhorst, 2002; Middlebrooks, 1992; Oldfield and Parker, 1984; Wightman and Kistler, 1997). The effectiveness of spectral cues, however, depends critically on *a priori* information regarding the transmitted signal spectrum.

[a)]Electronic mail: pavel.zahorik@louisville.edu

Important work by Hofman *et al.* (1998) has shown that listeners can effectively adapt or recalibrate to degraded spectral cues to sound-source direction over a period of weeks. In this study, the pinnae of four listeners were fit with thin plastic molds that introduced consistent changes to the spectral cues available to these listeners. Initial localization testing using a matrix of frontal sources located within the ocular-motor range (±30° azimuth and elevation) confirmed that this manipulation profoundly degraded accuracy, although mostly in the vertical dimension where spectral cues are most salient for directional localization. The listeners were then allowed to continue with normal and uncontrolled interaction with the environment over a period of weeks, returning to the laboratory for periodic localization testing to monitor the extent of adaptation. By the end of this period, each listener's localization accuracy had nearly returned to the level observed prior to inserting the mold. This suggests that the spatial maps in the brain that relate spectral cue patterns to physical locations in the environment had been successfully remapped. Surprisingly, no adaptation aftereffects were observed in a final localization test conducted after removing molds. Localization performance returned immediately to baseline levels of accuracy. The authors interpret this interesting finding as evidence of two independent spatial maps: one for normal spectral cues, and one for the modified cues. It is suggested that this "multiple-map" architecture is perhaps similar to that for language processing in multilingual individuals. Corroborating anatomical evidence for this multiple map hypothesis has been observed in the midbrain localization pathway of the barn owl (Knudsen *et al.*, 2000; Linkenhoker *et al.*, 2005).

Although the factors contributing to the spatial remapping in Hofman's study were not explicitly examined, it seems very likely that spatial information from other sensory modalities played an important role. Results from a variety of previous studies support this view. Adaptation to acute perturbations of visual space has been shown to produce compensatory shifts in auditory space in both animals (King *et al.*, 1988; Knudsen and Brainard, 1991) and in humans (Zwiers *et al.*, 2003). The particular importance of visual input for the processing of spectral cues has been demonstrated by reports of severe impairment to localization accuracy in the vertical but not the horizontal dimension for congenitally blind listeners (Zwiers *et al.*, 2001). Proprioceptive inputs have also been shown to affect auditory spatial calibration in blind listeners (Lewald, 2002a). Preliminary results for blindfolded listeners with normal vision also suggest that proprioceptive information can effectively facilitate recalibration to altered spectral cues (Blum *et al.*, 2004). Controlled multimodal feedback may also decrease the time required for recalibration of auditory space. Related studies in which exposure to small amounts of disparity between visual and auditory objects results in a shift of auditory space ("ventriloquism" effects) demonstrate that controlled auditory recalibration can take place within minutes (Lewald, 2002b; Recanzone, 1998).

An additional and important question relates to spectral cue adaptation effects in determining the location of a sound source along a front-back dimension, which was not examined in Hofman's study. Degraded spectral cue information is known to contribute to a particular type of mislocalization in which listeners incorrectly identify whether the source is located in front or behind (Kistler and Winghtman, 1992; Middlebrooks, 1992; Wenzel *et al.*, 1993). Although generally infrequent, these front-back reversals represent dramatic failures of the processes underlying directional sound localization. They are thought to stem from the inherent ambiguity of the binaural cues to sound direction, since certain front and rear hemifield locations produce identical binaural cue values. In these cases, the auditory system must rely on information contained in spectral cues (Burger, 1958), and/or information contained in the changes to binaural cues with listener (Perrett and Noble, 1997a, 1997b; Wallach, 1940) or source movement (Wightman and Kistler, 1999). These latter dynamic cues for resolving front-back ambiguity are only effective for long-duration sounds (greater than approximately 250 ms), however, given the latencies inherent in initiating listener movement (Woodworth and Schlosberg, 1954). For short-duration sounds, spectral cues must be used to resolve front-back ambiguity. These spectral cues are thought to involve relatively simple analyses of the sound level in particular frequency regions, since increased levels in the approximate bandwidth between 3 and 7 kHz have been observed at the ear for front relative to rear hemifield sources (Blauert, 1997; Wightman and Kistler, 1997). In contrast, the spectral cues to source elevation appear to be considerably more complex, likely involving more detailed spectral shape analysis over a broader frequency bandwidth (Bloom, 1977; Macpherson and Middlebrooks, 2003; Zakarauskas and Cynader, 1993). Spectral cues to elevation must also provide continuous information, rather than the more simple binary information regarding front or rear hemifield target location.

The current study seeks to determine if controlled multimodal feedback as to the correct sound-source location can facilitate rapid improvements in sound localization accuracy when spectral cues are inappropriate, presumably through a process of perceptual recalibration. Accuracy improvements in the front-back dimension are of particular interest because the spectral cues to location in this dimension appear to be less acoustically complex than those to elevation, and may require less time for recalibration. We examined the efficacy of a sound localization training procedure that provides listeners with auditory, visual, and proprioceptive/vestibular feedback as to the correct sound-source position. The goal here was simply to provide listeners with multimodal feedback similar to what might be experienced in many real-world situations when orienting to a sound source—not to determine the relative importance of feedback from any particular modality. Spectral cues were modified through the use of a virtual auditory display that used nonindividualized head-related transfer functions (HRTFs), however, instead of the pinna-mold manipulation used by Hofman. Previous results have demonstrated that virtual displays of this type produce distorted spectral cues to sound direction that result in degraded sound localization performance, particularly in the elevation and front-back dimensions (Middlebrooks, 1999b; Wenzel *et al.*, 1993). Sound localization performance using

this virtual display was evaluated before, during, and after a laboratory-controlled active, multimodal feedback training regimen for a wide range of sound sources fully surrounding the listener using a virtual environment system. Results are compared to a control group of listeners that did not receive feedback training. Follow-up testing evaluated the persistence of localization accuracy improvements resulting from the training procedure. Practical implications of the training procedure are also discussed.

II. METHODS

A. Participants

Twelve listeners (all male, age range: 21–41 years) voluntarily participated in the experiment. Three were authors of this article and the remainder were recruited from Rockwell Scientific. All had normal hearing, as verified by a standard audiometric screening procedure, normal or corrected-to-normal vision acuity, and no past histories of any other visual or auditory deficits. All listeners were naive with respect to the sets of spatial positions selected for the experiment.

B. Apparatus

An Intel-based personal computer (PC) 3D sound card (Turtle Beach Montego II A3D, with Aureal Vortex2 chipset) and high-quality stereo headphones (Sennheiser HD265) were used to present the spatialized auditory stimulus to the listener. According to the manufacturer's specifications, this particular sound card used HRTF-based processing to implement 3D sound spatialization in an anechoic environment (room-environment simulation was also available, but was not used in this experiment). This processing used a generalized set of HRTF measurements from a single individual who was not one of the test subjects in this experiment. Although the sound card, which is currently discontinued, was a low-cost ($<$\$100) card that was marketed primarily to computer gaming users, its audio specifications were of high quality. Because this sound card and its associated software did not allow provisions for directly analyzing or modifying the HRTF data used for spatialization, standard system identification techniques were used to estimate the HRTFs implemented on the card for a variety of source locations. These verification techniques and results are described below.

Visual stimuli used in the response and feedback methods of the experiment were presented using a PC-based 3D computer graphics system (SGI 230) coupled to a head-mounted display (Sony Glasstron PLM-A35) with a field-of-view of approximately $30° \times 23°$ (horizontal $\times$ vertical). The orientation of the listener's head was tracked using an ultrasonic 6-degree-of-freedom (DOF) position/orientation sensor (Logitech, Inc.). This device was accurate to 0.1° in orientation angle. The spatial locations of both visual and auditory stimuli were capable of being updated in real time (update rate of at least 30 Hz) as changes in the listener's head orientation were continuously monitored. This allowed for simulation of both auditory and visual sources that appear to remain stationary by compensating for head movements of the listener. Listeners used a standard PC mouse button press to both initiate and terminate the pointing response.

Display integration was handled by a client-server architecture using 2 Intel-based PCs communicating via TCP/IP. The 3D audio and head tracker handling was implemented on the server PC, while a client PC handled both 3D graphics and button press data. This distributed architecture avoids resource competition between 3D graphics rendering processes and those required for 3D sound processing.

C. Stimuli

The auditory stimulus for all portions of the experiment was a sample of Gaussian noise (100 ms, 0.2–20 kHz, 10-ms $\cos^2$ onset/offset ramps, approximately 60 dB SPL), presented from one of a variety of spatial positions using virtual sound-source techniques as implemented on the 3D sound card used in the experiment. The spatial positions were uniformly distributed around an imaginary partial sphere, with origin at the center of the listener's head. This partial sphere included a full 360° of azimuth, and ±40° of elevation relative to ear level and had a radius of 1.5 m.

Three types of visual stimuli were used. All were presented via the head-mounted display (HMD) on a uniform black background. The first stimulus was a head orientation "crosshair" that was presented to the listener at all times. It marked the vector that pointed straight ahead from the center of the listener's head, and as a result was not updated with changes in head orientation. The second visual stimulus was one that, in certain conditions, provided the listener feedback as to the correct sound-source location. This indicator stimulus was a small point of light (subtending approximately 0.5°) with high contrast, also presented via the HMD. When presented, the indicator stimulus was always paired with the auditory stimulus at the same spatial location. During this pairing, the auditory stimulus was repeated at a rate of 1 Hz. The spatial locations of both auditory and visual feedback stimuli were updated in order to compensate for changes in head orientation of the listener. The third visual stimulus indicated a reference location directly in front of the participant (0° azimuth and 0° elevation). This reference stimulus was also a small high-contrast point of light (subtending approximately 0.5°) presented via the HMD and updated to reflect changes in head orientation. Unlike the feedback stimulus, however, the reference point remained illuminated at all times. The feedback and reference stimuli were rendered in different colors (green and blue, respectively). Additional spatial reference information may also have resulted from the compact size of the HMD apparatus that allowed the listener a partial (peripheral) view of the room environment, which remained illuminated at all times.

D. Head-related transfer function measurements

Because the specifics of the HRTF set utilized by the 3D sound card were not provided by the manufacturer, it was important to verify the characteristics of this particular HRTF set. To accomplish this, standard system identification techniques using maximum-length sequence (MLS) signals (Rife and Vanderkooy, 1989) were used to recover the HRTFs

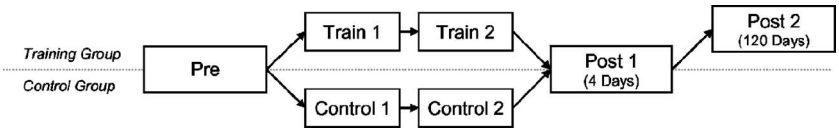


FIG. 1. Phases in the experimental design for both training and control groups of participants.

implemented within the sound card's Aureal Vortex2 chipset for a number of simulated source locations. One set of source locations varied azimuth angle from 0° (directly in front of the listener) to 180° (behind the listener) in 10° steps, all at a fixed 0° elevation angle (ear level). Another set of locations varied elevation from −40° to 40° in 10° steps, with a constant azimuth angle of 0°. Aureal's A3D 2.0 player was used to play an MLS signal ($2^{13}-1$ points) at the specified spatial locations. Left and right channel analog (line-out) responses to this signal were acquired (16 bit, 44.1 kHz) using a second high-quality sound card (Digital Audio Labs CardDeluxe) installed in the same PC (no signal averaging used). Measurement processing to recover the transfer functions from the raw responses was implemented using MATLAB (The Mathworks, Inc.) software. Because the sound card apparatus was operated in a mode designed for headphone presentation of virtual sound sources, we assumed that the card implemented some form of compensation for the acoustical effects of the headphones when coupled to the head of the listener. As a result, all HRTF measurements from the sound card apparatus likely included this location-independent headphone compensation.

In order to more fully evaluate the HRTFs measured from the sound card apparatus, HRTF measurements were also made from a single representative listener (SLO) who participated in the subsequent sound localization experiments. These HRTF measurements, which served as a basis for comparison with the sound card's HRTFs, were conducted in the laboratory of Wightman and Kistler using miniature electret microphones (Sennheiser KE4-211-2) in a blocked-meatus configuration. Details of this measurement procedure have been described elsewhere (Wightman and Kistler, 2005). Source locations were identical to those relating to the sound card HRTF dataset. To further facilitate comparison with the sound card HRTF dataset that likely included compensation for headphone effects, we applied a headphone compensation filter to the SLO HRTF data set that was derived from acoustical measurement of an "open" headphone (Beyerdynamic DT-990 Pro) when coupled to SLO's head using the same microphones and measurement technique as for the HRTF measurements. The details of this headphone measurement procedure have also been previously described (Wightman and Kistler, 2005). Additional compensation was applied to each HRTF measurement in order to remove the response of the loudspeakers (Cambridge SoundWorks, Center/Surround IV) used for presentation of the measurement stimulus.

E. Design and procedure

To assess the effects of the feedback training procedure on sound localization performance, a three-phase experimental design with two groups of listeners was used, as depicted graphically in Fig. 1. Listeners were randomly assigned to either the training group or the control group. After a period of initial familiarization with the experimental response procedure, listeners completed the experimental phases in the order shown in Fig. 1: pretraining/control, training/control, post-training/control. For the pre- and post-training/control phases of the experiment, localization performance was evaluated in the absence of correct location feedback. During the training phase of the experiment, a feedback training procedure was administered to the training group. The control group received no feedback training during the sound localization task in this phase. For some listeners, the pretraining/control and the first portion of the training/control phase ("train 1"or "control 1") were completed on the same day. The second portion of the training/control phase ("train 2" or "control 2") was always completed on a separate day from the first portion. For all listeners, a post-training/control phase was conducted 4 days after completing the training/control phase of the experiment. In order to evaluate any lasting effects of the training procedure, the listeners in the training group also completed a second post-training phase of the experiment approximately 120 days after training that was identical to the first post-training phase. All phases of the experiment were conducted in the same quiet room with the listener seated in a swiveling chair.

A graphical representation of the open-loop localization procedure for all phases of the experiment in which feedback training was not provided is shown in panels 1–3 of Fig. 2. At the start of a given trial (panel 1) the listener oriented to a reference location straight ahead. This reference location was indicated in the HMD by a small point of light. The crosshair was visible to the listener via the HMD and was used to guide the listener to this reference location. Head position in this reference location was verified by the 6-DOF position/orientation sensor. This head position/orientation information was used to update the rendering of the straight-ahead indicator in the HMD, which remained illuminated during the entire procedure. Because the crosshair rendering was not updated with changes in head position/orientation, the crosshair appeared to remain coupled to the head during movement, whereas the straight-ahead indicator appeared to remain stationary during head movement. Once the listener was correctly positioned at this reference orientation, the auditory stimulus was presented at a given spatial location (panel 2). After the presentation of the auditory stimulus, the listener was instructed to point the crosshair in the direction of the perceived sound-source location (panel 3). Depending on the location of the source, the pointing response may have been accomplished via head rotation only, or a combination of head and body rotation. Once the listener placed the crosshair in a position that he/she felt most properly matched the perceived sound-source location, the listener pressed a button to signify the end of the response. During the response phase of the trial, the listener's head orientation was

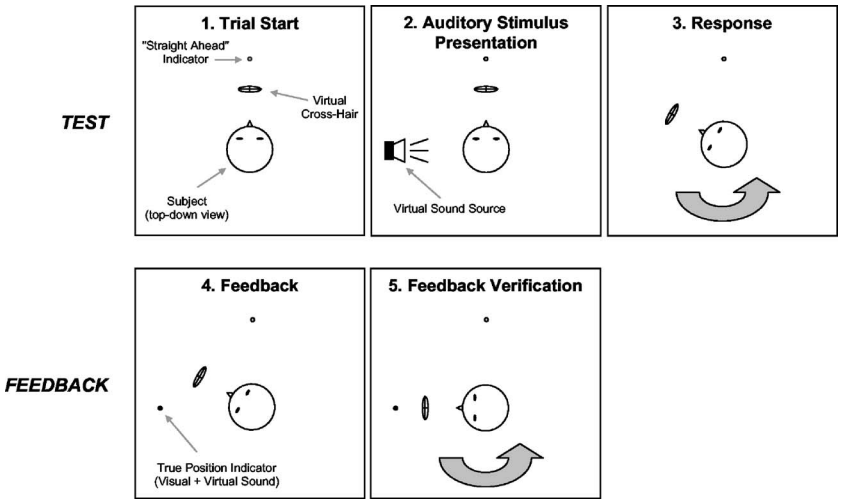


FIG. 2. Experimental procedure for a single trial. During the pre, control 1, control 2, post 1, and post 2 phases of the experiment, only "test" portion of the procedure (steps 1–3) was implemented. During the training phases (train 1 and train 2), all steps of the procedure were implemented: test + feedback.

sampled by the 6-DOF sensor. This allowed for an accurate reading of the final crosshair position (azimuth and elevation angles in a two-pole coordinate system). Listeners were instructed to take as much time as needed to register what they felt to be a good representation of perceived source location. It should be noted that this orienting response procedure is fundamentally similar to other orienting response procedures (e.g., "head pointing") that have frequently been used in the study of human directional sound localization abilities (Bronkhorst, 1995; Makous and Middlebrooks, 1990; Middlebrooks, 1992, 1999b).

The procedure for feedback-training (closed-loop) portions of the experiment was identical to that used in the other phases, except that after each localization response was registered, feedback as to the correct source location was provided to the listeners via a paired auditory/visual stimulus. This procedure is shown graphically in panels 1–5 of Fig. 2, in which the initial portions of a trial were identical to those of the control group localization response (panels 1–3). After the listener had input his/her apparent position response (panel 3), the feedback portion of the trial began. A visual indicator of the correct source position paired with a repeating spatialized auditory stimulus (the same stimulus as in panel 2, only repeating) was then displayed to the subject via the HMD (panel 4). To verify that the listener was able to find this indicator, the listener was asked to aim the virtual crosshair (via head rotation and/or pitch) at the location of the correct position indicator (panel 5). When the listener was confident that he had pointed to the indicator as accurately as possible, a button was pressed, at which point the crosshair position was inferred from the measured head orientation, just as after the source position judgment (panel 3). On average, listeners were able to perform this feedback verification operation to within 0.6° of accuracy. Hence, this feedback procedure forced the listener to actively orient to the correct sound-source position, providing the listener with a combination of proprioceptive and vestibular as well as auditory and visual feedback information.

For the pre- and post-training/control phases of the experiment, 144 spatial positions were tested. The positions were selected such that they were distributed in an approximately uniform fashion throughout the full 360° of azimuth angle surrounding the listener, and over ±40° of elevation angle. Each of the 144 spatial positions was presented once in a randomized order. The post-training/control phase of the experiment was identical to the pretraining/control phase, except that a different sample of 144 spatial positions was used that also satisfied the same spatial distribution criteria. Both pre- and post-training/control phases of the experiment were administered within a single block of trials, and took approximately 45 min to complete. The pre- and post-training/control phases were identical for the two groups of listeners.

For the training phase of the experiment, 48 spatial positions were presented. These positions were also distributed in an approximately uniform fashion throughout the spatial region of interest (360° of azimuth angle and ±40° of elevation angle) and had minimal overlap with the sets of positions used for the pretraining/control and post-training/control phases. Within a block of trials, 24 spatial positions were presented 3 times each, which yielded a total of 72 trials per block, presented in a randomized order. Two blocks of trials were presented on separate days, each with a different sample of 24 spatial positions, yielding a total of 48 spatial positions, and 144 trials in the training/control phase of the experiment. Each block of 72 trials took approximately 30 min to complete. Listeners in the control group received identical stimulus sets, but were not provided with correct location feedback training in this phase.

During a debriefing session following the experiment, none of the listeners in the training group reported being consciously aware of using any type of error compensation strategy during the response when specifically asked, and none reported any negative sound localization effects persisting after experience with the auditory display. All listeners reported that the virtual sound sources appeared to be external to the head.

III. RESULTS

A. Head-related transfer function analyses

Two sets of HRTF measurements were analyzed in terms of various acoustical properties thought to serve as important parameters in human sound localization. One set of HRTFs was measured from the 3D sound card apparatus

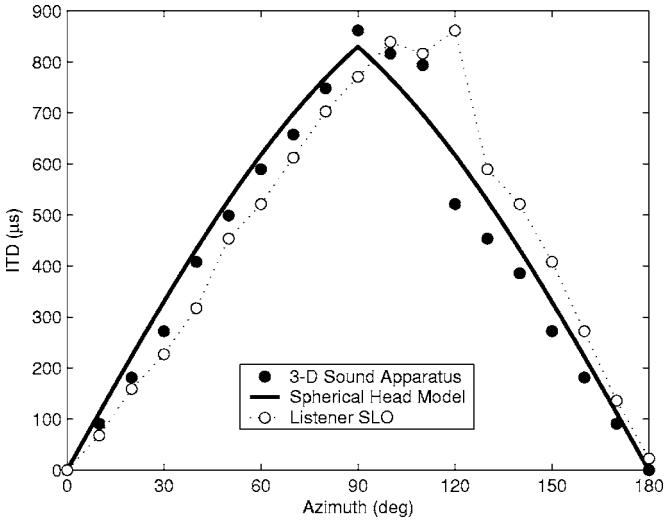


FIG. 3. Interaural time difference (ITD) as a function of source azimuth angle (two-pole coordinates, with source elevation fixed at 0°) based on HRTF verification measurements of the 3D sound card apparatus (solid symbols). These data are well approximated by a model of ITD change based on a spherical head (solid line) with a diameter of 22.7 cm (Woodworth and Schlosberg, 1954). For comparison purposes, ITD data based on HRTF measurements from one of the participants in this study (listener SLO) are also shown (open symbols).

used in all psychophysical sound localization experiments in the current study. A second set of HRTFs was measured, for comparison purposes, from a single representative listener (SLO) who participated in the subsequent sound localization experiments. The acoustical parameters, or cues, that were analyzed in both sets of HRTFs included interaural time differences (ITDs) cues, interaural level differences (ILDs) cues, and monaural spectral cues.

Figure 3 displays ITD as a function of azimuth angle computed via cross-correlation of the left and right head-related impulse responses. The data for the 3D sound card apparatus have been fit (least-squares criterion) with a model that approximates ITD for a spherical head with diameter of 22.7 cm. A relatively good fit may be observed. ITD values from the SLO HRTF data set are also displayed. Although the general pattern of ITD as a function of source azimuth angle is similar in the SLO dataset, differences may be observed at nearly every azimuth angle. The maximum ITD value also appears to be shifted to slightly greater azimuth angles. Although other listeners participating in the localization experiments in this current study also likely experienced related mismatches in the ITD cue displayed by the 3D sound card apparatus and their own natural ITD functions, we expect that this mismatch will in general have less impact on apparent source direction than mismatches in the spectral cues. Results from previous studies support this view (Middlebrooks, 1999b; Wenzel *et al.*, 1993).

ILD and monaural spectral cues may be observed in Figs. 4 and 5, which display contours of constant level in the left and right ear HRTFs as functions of both frequency and source direction for the 3D sound card apparatus (Fig. 4) and SLO (Fig. 5) HRTF datasets, respectively. The top panels of Figs. 4 and 5 display source azimuth angles ranging from 0° to 180° with source elevation fixed at 0° (horizontal plane). The bottom panels of Figs. 4 and 5 display source elevation angles from −40° to 40° with source azimuth fixed at 0° (median plane). Patterns of monaural spectral cues, which are thought to be important for both determining source elevation as well as whether the source is in front or behind the listener, are directly apparent from Figs. 4 and 5. General patterns of interaural level differences (ILDs) may be inferred from Figs. 4 and 5 by comparing the levels between the two ears. The largest ILDs occur at higher frequencies when the source is opposite one ear (i.e., 90° azimuth), and relatively little level difference occurs at low frequencies—both well-known effects (Strutt, 1907). In general, the observed values of these primary sound localization cues in both HRTF datasets are consistent with those reported in a variety of other studies (Mehrgardt and Mellert, 1977; Middlebrooks, 1999a; Shaw, 1974; Wightman and Kistler, 1989).

As is to be expected from HRTF data sets originating from different listeners (Shaw, 1966), substantial differences in the patterns of spectral variation were observed between the two datasets. These differences are readily apparent through visual comparison of the data displayed in Figs. 4 and 5, respectively. The shapes and locations of various prominent spectral features (e.g., spectral peaks and notches) are quite different in the two datasets. This difference is particularly apparent in the elevation dimension, where spectral variation in the SLO HRTF dataset appears to be much more complex (i.e., more prominent peaks and notches) than the spectral variation present in the sound card dataset as a function of elevation. This difference in spectral complexity is also apparent in the azimuth dimension, although perhaps to a somewhat lesser degree. While some of the differences in spectral complexity are to be expected from HRTF datasets derived from measurements of different listeners, it is also possible that the generally more smooth spectral patterns observed in the sound card apparatus HRTFs resulted from signal processing compromises inherent in this low-cost sound card's design.

To more carefully examine the potential front-back information provided by monaural spectral cues, we analyzed the level differences in 1/3-octave bands between front and rear hemifield HRTFs that were symmetrically displaced from 90° azimuth along the horizontal plane. The results of this analysis for the HRTFs measured from the 3D sound card apparatus and from listener SLO are displayed in Figs. 6 and 7, respectively. In these figures, positive level difference values (measured in dB) indicate greater level for the frontal location. Within the 3–7-kHz bandwidth, frontal locations had consistently greater level for all displacements in both datasets. For the sound card HRTF dataset (Fig. 6), level differences as large as +20 dB occurred for certain combinations of displacement and 1/3-octave band frequency within the 3–7-kHz bandwidth. The patterns of level difference were somewhat more complex and of lesser magnitude for the SLO HRTF dataset (Fig. 7). Nevertheless, frontal sources in the SLO HRTF dataset also produced consistently greater level in the 3–7-kHz bandwidth. The consistency of this level difference effect, both across different HRTF datasets and different displacements from 90°, suggests that level in

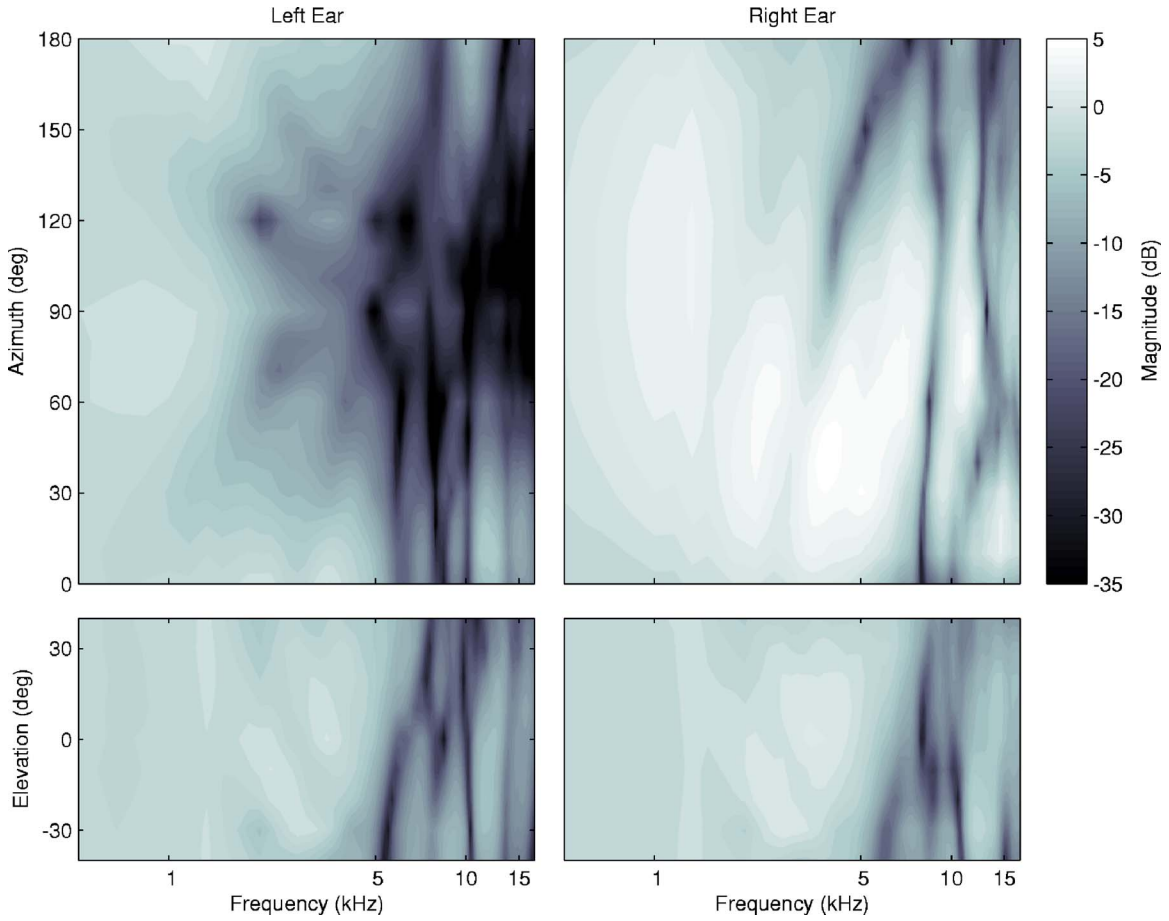


FIG. 4. (Color online) HRTF magnitude spectrum contours for the left and right ear as a function of source azimuth angle (upper panels) and elevation angle (lower panels) based on the HRTF verification measurements of the 3D sound card apparatus. For the azimuth angle analyses, source elevation was fixed at 0°. For elevation angle analyses, source azimuth was fixed at 0°. Note the changes in spectral pattern, especially in the right (ipsilateral) ear, as azimuth angle increases form 0° to 180°. In general, front hemifield sources (0°–90° azimuth) appear to have greater level in the ipsilateral (right) ear in the 3–7-kHz bandwidth than do sources in the rear hemifield (90°–180° azimuth). More complicated changes in spectral patterns are observed for changes in source elevation, where the frequencies of various spectral notches in the 5–10-kHz bandwidth change as a function of source elevation.

the 3–7-kHz bandwidth may serve as a relatively simple and effective cue to the front-back location of sound sources. Similar suggestions of a simple front-back spectral cue have previously appeared in the literature (Blauert, 1997; Wightman and Kistler, 1997).

From these analyses of the HRTF dataset present in the 3D sound card apparatus and the comparison of this dataset to HRTFs measured from a representative listener (SLO), it is clear that substantial differences between the two datasets exist. This implies that if the representative listener were presented with virtual sound sources simulated using the 3D sound card apparatus, the directional cues, including ITD, ILD, and spectral cues, would be inappropriate for this listener. Previous studies have demonstrated that degraded directional localization performance (Bronkhorst, 1995; Middlebrooks, 1999b; Møller *et al.*, 1996; Wenzel *et al.*, 1993) results from inappropriate directional cues caused by nonindividualized HRTFs. Although the HRTFs of the other listeners that participated in the subsequent localization experiments of the current study were not measured, we expect related mismatches between the listener's own HRTFs and the nonindividualized HRTF set present in the display apparatus. The focus of the current study is to determine how listeners may learn to adapt these inappropriate directional localization cues, given feedback on their localization performance.

B. Localization results

The data from all phases of the experiment were transformed from a two-pole coordinate system (azimuth and elevation) to a three-pole coordinate system (Kistler and Wightman, 1992). In this coordinate system, azimuth angle is represented in terms of two angles: a "right-left" angle, which is the angle subtended by the judgment vector and the median plane, and a "front-back" angle, which is the angle subtended by the judgment vector and the coronal plane. Elevation angle is the same in the three-pole transformation as in the two-pole system and is referred to as an "up-down" angle. This transformed coordinate system is preferable in this application to the two-pole system (azimuth and elevation angles) for its ability to obviate certain types of localization errors, such as reversals in the front-back dimension (Kistler and Wightman, 1992).

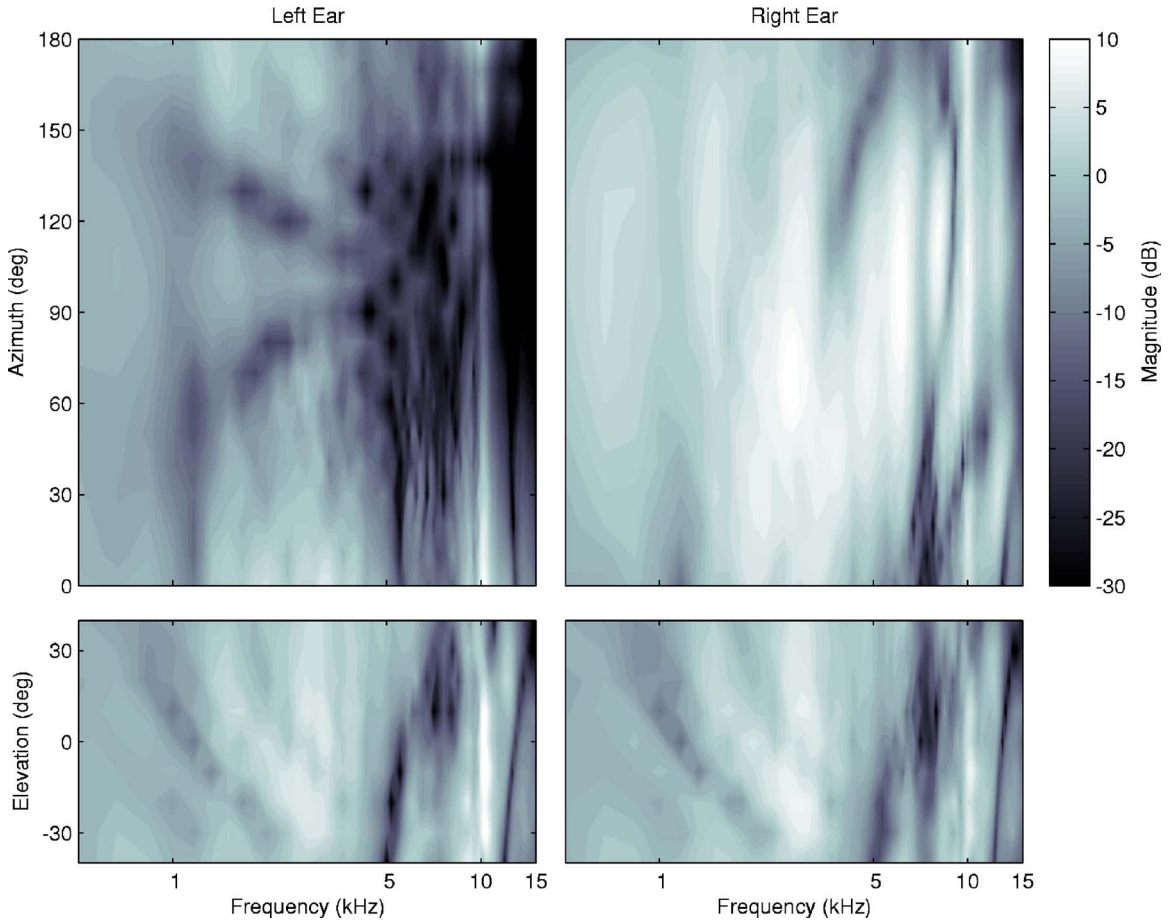


FIG. 5. (Color online) Same as Fig. 4, but for HRTF data from listener SLO. Note generally more complicated patterns of spectral change as a function of both azimuth (front-back location) and elevation relative to the HRTF data from the 3D sound card apparatus (Fig. 4).

1. Apparent sound-source direction and error analysis

Apparent sound-source direction data from all phases of the experiment are shown in Figs. 8–11 for four representative listeners: two from the training group (Figs. 8 and 9) and two from the control group (Figs. 10 and 11). In these figures, apparent source angle is plotted as a function of target angle for each of the dimensions in the three-pole coordinate system. Perfectly accurate responses would lie along the positive diagonal of each dimension. In the right-left and up-down dimensions, localization accuracy was summarized by computing the mean unsigned error, ε, which was defined as the mean unsigned deviation of the response angle from the target angle, in units of degrees. In the front-back dimension, localization accuracy was summarized in two different ways. First, reversals in which the response angle occurred in the incorrect hemifield were identified using the following selection rule:

$$|R - T| > |-R - T| \text{ and } |R - T| > \varepsilon_{\text{est}},$$

where R and T are front-back response and target angles for a given trial, and ε_{est} is an estimate of the mean unsigned error in the front-back dimension that is independent of the number of front-back reversals. We used ε in the right-left dimension for each listener in a given experimental condition for this front-back error estimate, ε_{est}. It is important to note that this reversal metric may be somewhat conservative, because it only labels responses as reversals if they lie outside the estimated region of error (i.e., $\pm\varepsilon_{\text{est}}$) for the front-back dimension. All responses identified as reversals are indicated by open symbols in Figs. 8–11. Front-back localization accuracy was also summarized using a modified ε metric in which all reversed responses were resolved (placed in the correct hemifield) prior to computing ε. As a result, ε for the front-back dimension is a measure of localization accuracy that is independent of front-back reversal rates.

Figure 8 displays data from a listener (SCD) that showed improvement in directional localization accuracy during and following the training procedures, although the improvement was limited entirely to a reduction in the number of front-back reversals. Initially, this listener localized the majority of sound sources to the rear hemifield (front-to-back reversals). Reversal rate decreased substantially during the first and second training phases of the experiment (train 1 and train 2), and transferred to the post-training phases of the experiment. Four days following the training procedure (post 1), reversal rates were greatly reduced relative to baseline performance in the experiment (pre phase). These improvements in general persisted 120 days following the training procedure, although some degradation in accuracy may be observed relative to performance immediately following training (post 2

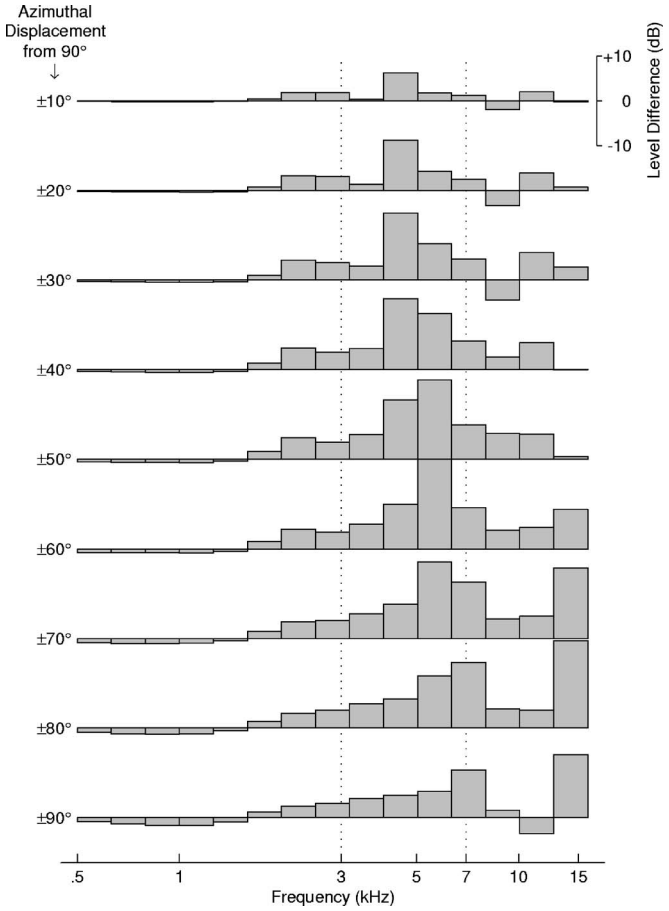


FIG. 6. HRTF level differences (ipsilateral ear only) in 1/3-octave bands between front and rear hemifield locations that are symmetrically displaced from 90° azimuth along the horizontal plane (0° elevation) from the 3D sound card apparatus. Displacements ranging from ±10° to ±90° (directly in front versus directly behind) are shown. Positive level differences indicate greater level for the frontal source. Note for this HRTF set, frontal sources produce consistently greater level in the 3–7-kHz bandwidth for all displacements.

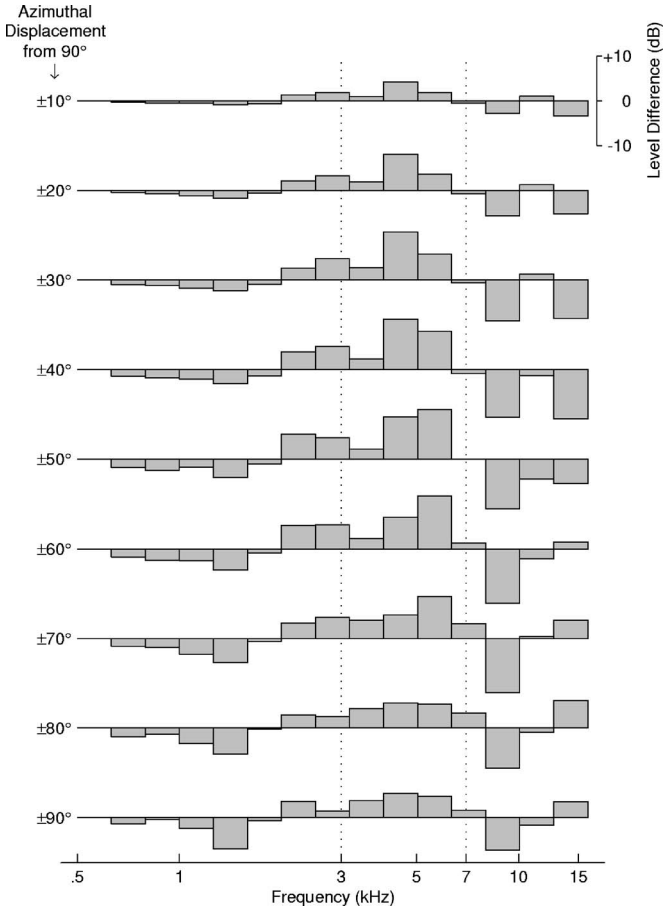


FIG. 7. Same as Fig. 6, but for HRTF data from listener SLO. Note for this HRTF set, frontal sources also produce consistently greater level in the 3–7-kHz bandwidth, although the effect is less pronounced than for the HRTF data from the 3D sound card apparatus (Fig. 6). In other lower and higher frequency regions, rear hemifield locations have greater level.

re: post 1). Very little systematic change in any other patterns of localization responses or error metrics was observed throughout the phases of the experiment for this listener. In the right-left dimension, localization performance was quite accurate in all phases. When reversal responses are resolved in the front-back dimension, performance was relatively accurate in all phases, as indicated by the ε value in each phase. In the up-down dimension, localization performance was consistently quite poor in all phases, considering the more limited range of target angles on which the mean unsigned error metric was based in this dimension. Had a greater range of target elevations been tested, up-down ε would likely have increased for this listener. As discussed in the previous section, the HRTF dataset and/or signal processing specifics related to the 3-D sound card apparatus, which seems to have somewhat limited spectral cue detail in the elevation dimension as compared to the HRTFs from a single representative listener, may have been at least partially responsible for this poor up-down performance.

Data from three of the five remaining listeners in the training group (not shown) were qualitatively similar to the data for listener SCD (Fig. 8), showing consistent reductions in front-back reversal rates throughout the course of the experiment and little to no systematic change in any other aspects of localization accuracy. Description of a more detailed analysis of front-back reversal rates appears in the next section of this report. The remaining two listeners in the training group (SCA and SLO) showed minimal localization accuracy improvements in any localization accuracy metric including rates of front-back reversals. Figure 9 displays data from listener SLO. Because this listener's overall level of accuracy was quite high, ceiling effects may have limited the ability to realize further accuracy increases resulting from the training procedure.

Data from two representative listeners in the control group that did not receive feedback training are shown in Figs. 10 and 11. Little to no systematic change in localization accuracy or the patterns of localization responses may be observed for listener SCI (Fig. 10). Like listener SCD (Fig. 8), listener SCI exhibited large numbers of front-back reversals, as well as substantial error in the up-down dimension. Response variability in the up-down dimension does appear to decrease somewhat over the course of the experiment for this listener, however. Data from most other listeners in the control group (not shown) were qualitatively similar to that of listener SCI: little change in localization accuracy or response patterns throughout the experiment.

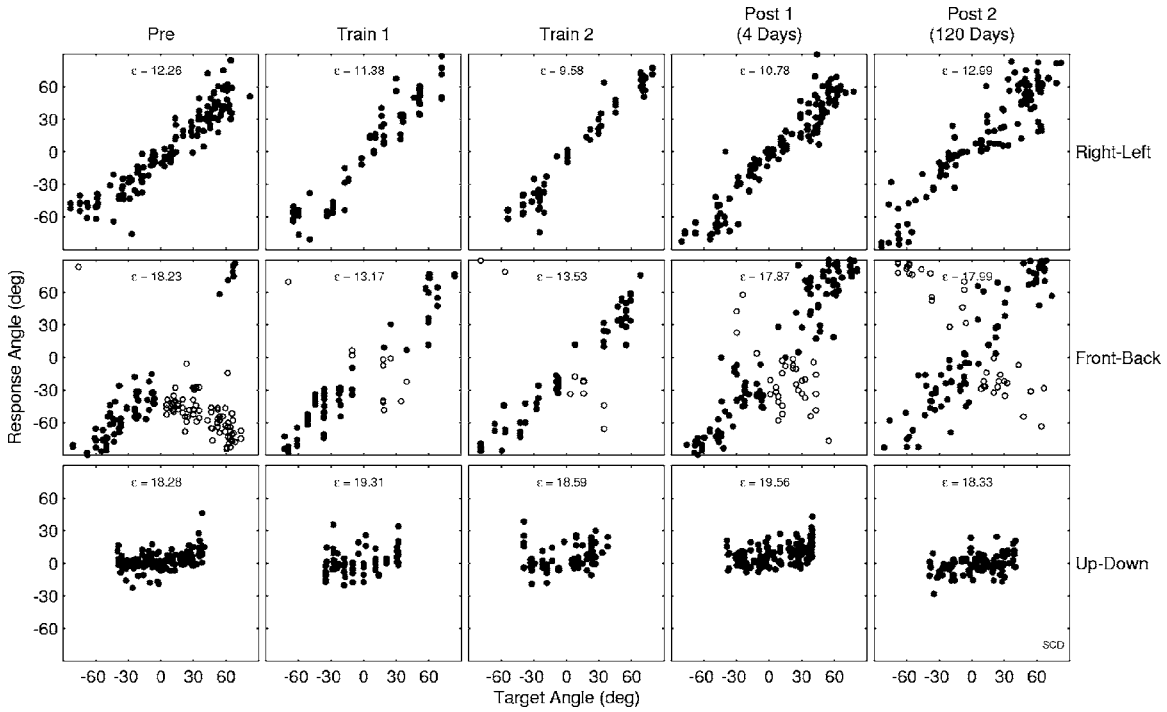


FIG. 8. Scatterplots of target sound source angle versus orienting response angle for a single representative listener in the training group (listener SCD). The data are displayed in the three-pole coordinate system for each phase in the experimental procedure (indicated in column headings). Reversals in the front-back dimension are indicated by open symbols. Mean unsigned error, ε, for the data displayed in each panel (reversals resolved) is also indicated. This listener shows a large reduction in the number of reversals during and following the training procedures.

Some spontaneous changes in the patterns of front-back responses were observed throughout the course of the experiment for listener SCG (Fig. 11). This listener initially (pre phase) made very few reversal responses to the front hemifield (back-to-front reversals), but later (post phase) made more of these types of reversals. This change in the distribution of reversal directions did not affect overall reversal rates, however, which remained nearly constant throughout the experiment. Response variability in the up-down dimension also appears to decrease somewhat over the course of the experiment for this listener, similar to the patterns observed for listener SCI (Fig. 10).

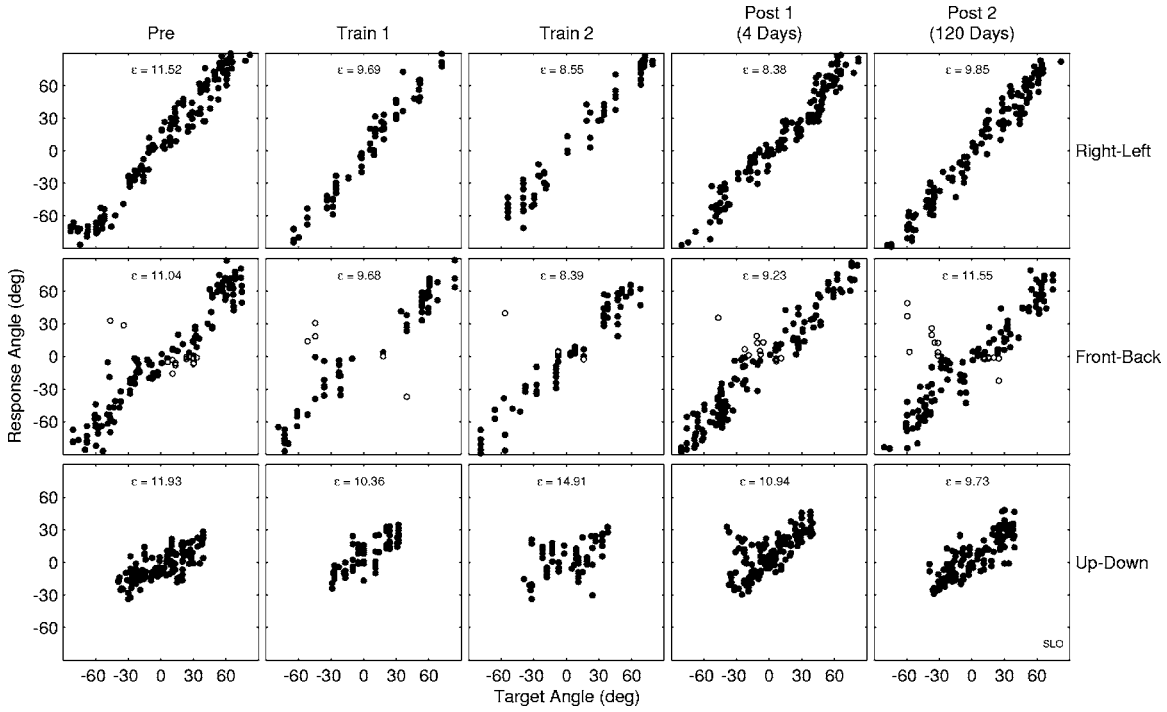


FIG. 9. Same as Fig. 8, but for data from listener SLO in the training group. Note that this listener shows little to no effect of the training procedures, although overall localization accuracy for this listener was quite high initially.

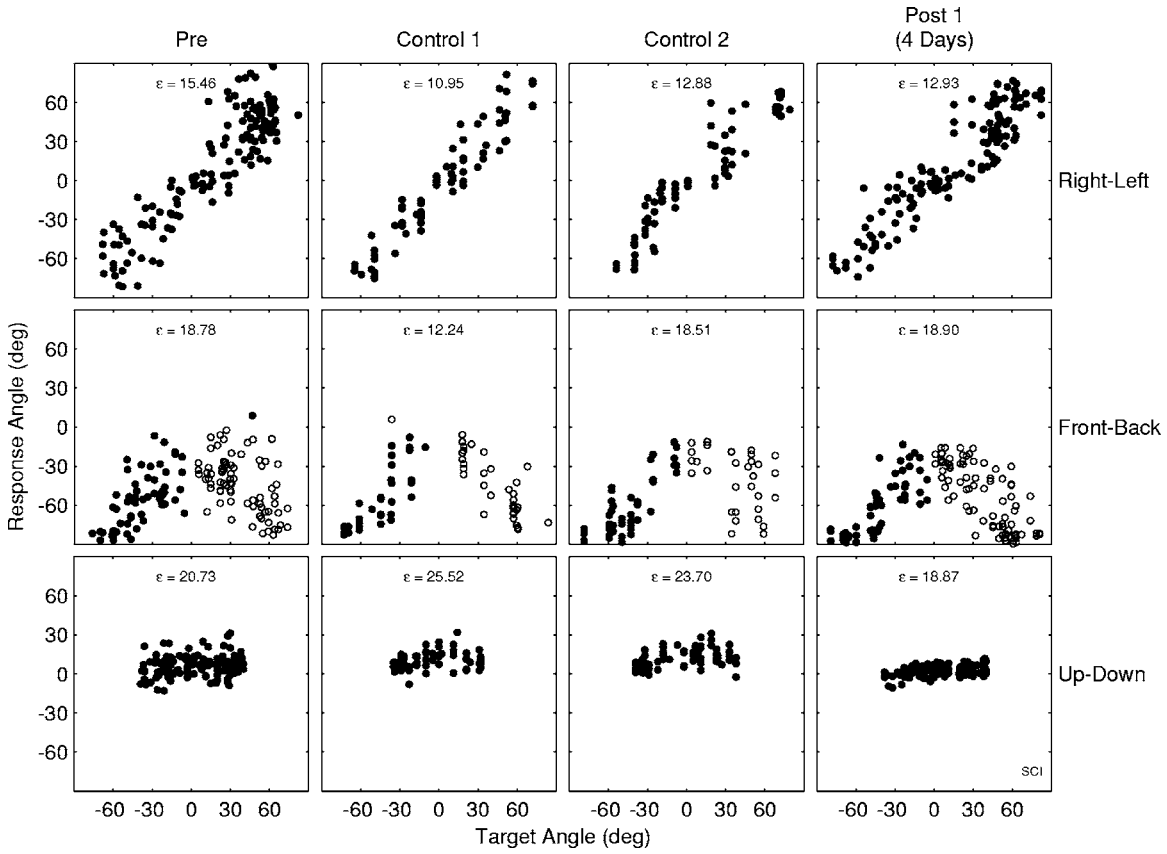


FIG. 10. Same as Fig. 8, but for data from listener SCI in the control group.

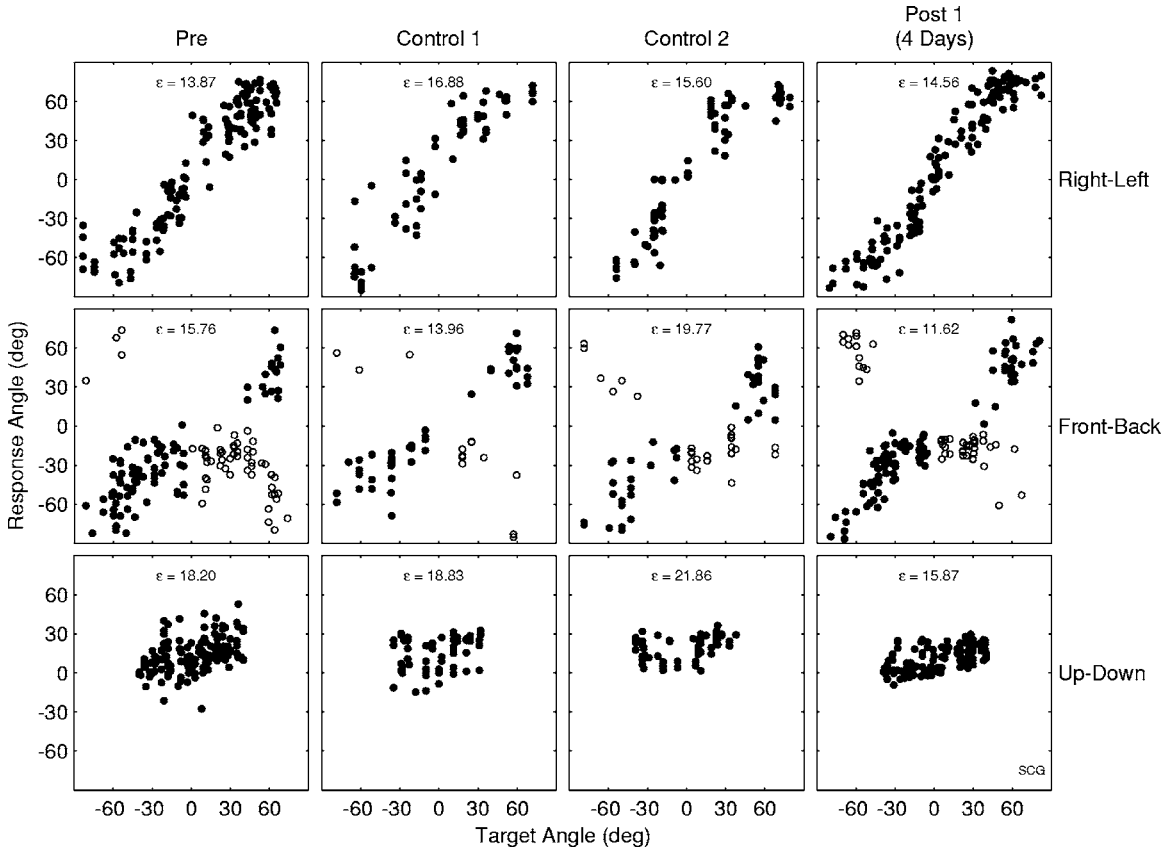


FIG. 11. Same as Fig. 8, but for data from listener SCG in the control group.

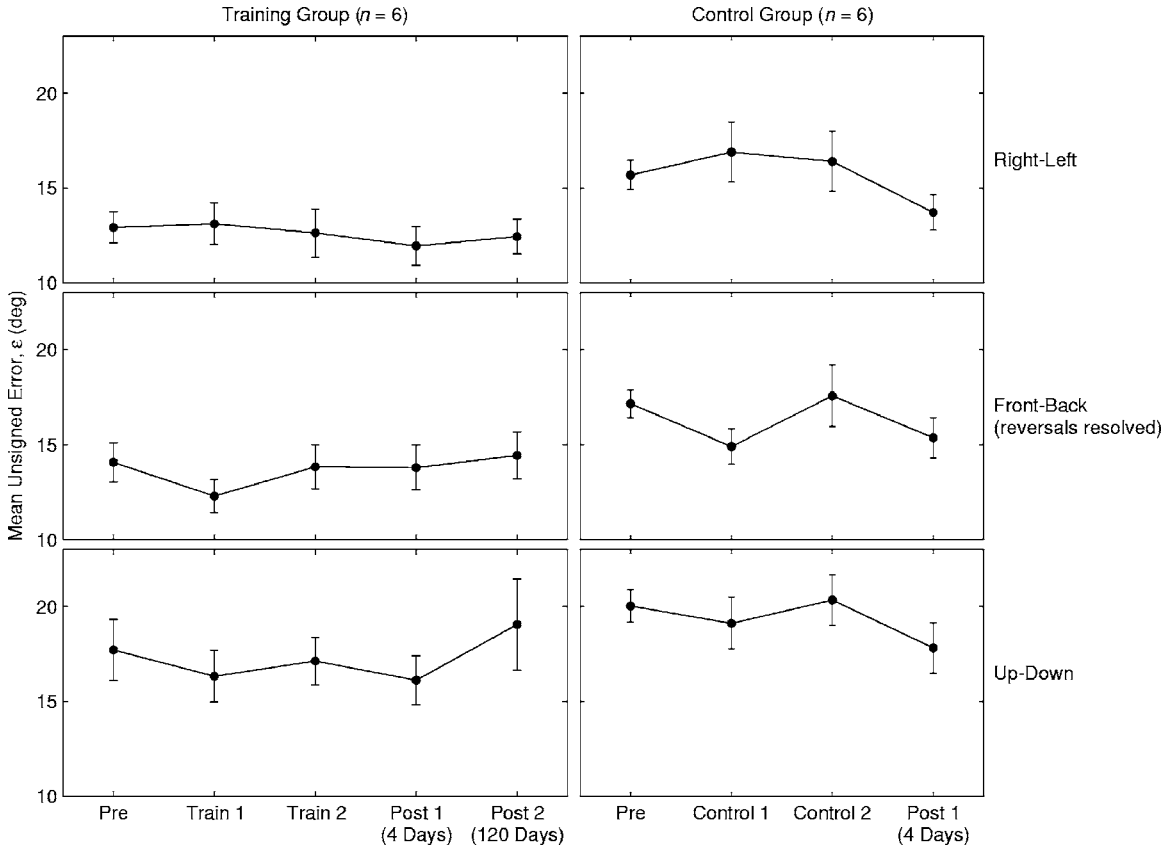


FIG. 12. Summary of mean unsigned error, ε, for each dimension in the three-pole coordinate system. Each data point represents the mean ε across listeners in the training or control groups for a given phase in the experiment. Bars represent one standard error of the mean ε across listeners in each group ($n=6$). All responses identified as front-back reversals were resolved to the correct hemifield prior to ε computation. Within each group, no statistically significant differences are observed in mean ε throughout the phases of the experiment.

The analyses of the individual listener apparent direction data and their associated measures of error all suggest that the localization training procedure selectively impacts the rates of front-back reversals and not other aspects of localization accuracy. Figure 12 displays a summary of localization accuracy independent of front-back reversals as assessed by the ε metric for all phases of the experiment. Error bars indicate one standard error of the mean of the ε scores for individual listeners. Within each group, no statistically significant changes in mean ε were observed throughout the course of the experiment in any of the spatial dimensions. Localization accuracy in the control group was somewhat worse than the training group, however. This is evidenced by the relatively small, but statistically significant, differences in mean ε in the pre phase for both the right-left, $t(10) = -2.45$, $p < 0.05$, and front-back dimensions, $t(10) = -2.41$, $p < 0.05$. Because substantial variability in localization performance between individual listeners is commonly observed, this difference in localization accuracy between the two groups is likely a result of sampling error associated with our relatively small sample sizes. We conducted all subsequent analyses either within groups or within individual listeners to avoid problems associated with this issue. It is also important to note that mean ε across all listeners and experimental phases was generally consistent with previously reported localization results using nonindividualized HRTFs (Bronkhorst, 1995; Middlebrooks, 1999b; Wenzel *et al.*, 1993), which have been shown to be less accurate than corresponding localization performance in the free field (Wenzel *et al.*, 1993), particularly in the up-down and front-back dimensions where spectral cues to source direction are most salient.

2. Front-back reversal analysis

Front-back reversal responses were identified using the selection criterion described in the previous section. Table I displays the proportion of front-back reversal responses for all listeners in both pre and post 1 phases of the experiment. Chi-square (χ^2) statistics (1 degree of freedom) based on the number of reversed and nonreversed responses in each phase are also displayed for all statistically significant differences. Four of six listeners in the training group showed statistically significant decreases in front-back reversal rates between pre and post 1 phases. Of the remaining two listeners in this group that did not show significant differences (SCA and SLO), one (SLO) made very few reversal responses initially in the pre phase. This likely limited the ability to realize any further decrease in reversal rate for this listener. As a result, significant decreases in reversal rates were observed for four of the five listeners that showed high initial reversal rates in the pre phase. No statistically significant differences were observed for any listeners in the control group. This suggests that decreases in reversal rates observed for most listeners in

TABLE I. Proportion of responses identified as front-back reversals before (pre) and 4 days after (post 1) either training or control. Chi-square statistics are indicated for those listeners who showed a significant difference in front-back reversals in the pre versus post 1 phases (* <0.05; ** $p<0.01$; *** $p<0.005$).

Training group				Control group			
Listener	Pre	Post 1	χ^2	Listener	Pre	Post 1	χ^2
SCA	0.38	0.33	NS	SCF	0.41	0.36	NS
SCB	0.48	0.30	9.88 ***	SCG	0.39	0.33	NS
SCC	0.42	0.27	6.72 **	SCH	0.51	0.59	NS
SCD	0.55	0.22	32.38 ***	SCl	0.55	0.60	NS
SCE	0.35	0.18	11.08 ***	SCJ	0.42	0.47	NS
SLO	0.10	0.08	NS	SDV	0.60	0.51	NS
Mean	0.38	0.23	7.92 ***	Mean	0.48	0.48	NS

the training group likely resulted from the training procedure, which decreased reversal rates by nearly 40%, on average (from 0.38 to 0.23).

Reductions in front-back reversal rates persisted 120 days following the training procedure. Table II displays a comparison of front-back reversal proportions in the pre and post 2 (120 days) phases for the training group. Statistically significant decreases in front-back reversal rates were observed for the same four listeners that showed significant decreases between pre and post 1 (see Table I). On average, reversal rates were still approximately 35% less than baseline performance 120 days following the training procedure. This suggests that the processes underlying this reversal rate reduction and resulting localization accuracy improvement do not require recent sensory input with modified spatial cues and do not appear to be affected by the long intervening period of everyday sound localization with normal spatial cues.

Table III displays a breakdown of reversal direction for each listener. Here, reversal responses are characterized as either a "front-to-back" reversal, where the target sound is incorrectly localized to the rear hemifield, or a "back-to-front" reversal, where the target sound is incorrectly localized to the front hemifield. For nearly all listeners in both training and control groups, front-to-back reversals occurred much more frequently than did back-to-front reversals. As a result, the total reversal percentage in most cases reflects a disproportionate amount of front-to-back reversals. Although the cause of this directional bias in reversal responses is unknown, similar results have been reported in other localization studies (Begault and Wenzel, 1993; Oldfield and Parker, 1984; Wightman and Kistler, 1999). Due to the very low occurrence of back-to-front reversals for many listeners, separate chi-square statistics were not computed for the directional breakdown reversal data displayed in Table III.

TABLE II. Proportion of responses identified as front-back reversals before (prc) and 120 days after (post 2) traning. Chi-square statistics are indicated for thsoe listeners who showed a significant difference in front-back reversals in the pre versus post 2 phase (* $p<0.05$; ** $p<0.01$; *** $p<0.005$).

Listener	Pre	Post 2	χ^2
SCA	0.38	0.34	NS
SCB	0.48	0.24	18.51 ***
SCC	0.42	0.27	7.41 **
SCD	0.55	0.32	15.39 ***
SCE	0.35	0.19	9.23 ***
SLO	0.10	0.13	NS
Mean	0.38	0.25	5.80 *

IV. DISCUSSION

This experiment has demonstrated that a relatively brief perceptual training procedure which provides listeners with auditory, visual, and proprioceptive/vestibular feedback as to true target locations can improve localization accuracy for stimulus conditions in which a mismatch in spectral cues to source direction exists. The improvement was realized entirely in the front-back dimension, where the proportions of hemifield reversed responses decreased significantly following the training procedures for four of five listeners who made frequent reversal responses initially. Perhaps most notable was that the training improvements appear to last at least 4 months. These results warrant further discussion of at least three key issues: the issue of what might be taking place during and after the perceptual training procedure, why little change in elevation localization performance was observed, and how these results may be useful for practical applications.

A. Potential processes for front-back reversal remediation

Because relatively simple analysis of sound level in the 3–7-kHz region can provide effective information as to whether the source is located in the front or rear hemifield, we suggest that the improvements observed in front-back reversal rates following the localization training procedure may have resulted from improved processing of this spectral information. Analyses of the HRTFs from the 3D sound card apparatus used in the training experiments confirmed the existence of a sizable and consistent front-back cue in which increased level in the 3–7-kHz region corresponded to frontal source locations (see Fig. 6). Subsequent analyses of the HRTFs from one of the listeners who participated in the experiment revealed a similar front-back level cue in the 3–7-kHz bandwidth, although the magnitude of the cue (level difference) was somewhat different, as was the pattern of 1/3-octave-band level differences both within and outside

TABLE III. Breakdown of front-back reversal direction in initial localization testing (pre phase) for all listeners. Back-to-front reversals represent responses to front hemifield sources that were incorrectly localized to the rear hemifield. Front-to-back reversals represent responses to rear hemifield sources that were incorrectly localized to the front hemifield. The proportions of the total number of responses that fall in each category are displayed.

Traning group			Control group		
Listener	Back-to-front	Front-to-back	Listener	Back-to-front	Front-to-back
SCA	0.12	0.26	SCF	0.04	0.38
SCB	0.03	0.44	SCG	0.03	0.36
SCC	0.25	0.17	SCH	0.00	0.51
SCD	0.01	0.54	SCI	0.00	0.55
SCE	0.03	0.33	SCJ	0.05	0.38
SLO	0.01	0.09	SDV	0.04	0.55
Mean	0.08	0.30	Mean	0.03	0.45

of the 3–7-kHz region (Fig. 7). As with all spectral cues, the effectiveness of this cue depends on the spectrum of the sound source. Although the current experiment only examined a single type of source signal (broadband noise), we would expect that this front-back spectral cue would be effective with a variety of source signals, provided the signal's bandwidth was sufficient to support comparison of level in the 3–7-kHz region to level in some other (likely lower frequency) region of the spectrum. Results from other studies also suggest that level in the approximate region between 3–7-kHz can serve as an effective and robust cue to front-back location (Blauert, 1997; Wightman and Kistler, 1997).

A critical remaining question relates to whether the observed improvements in front-back reversal rate following the training procedure are indicative of underlying changes in the perceived location of the sound sources. Since similar noise-burst signals were used in the training and nontraining phases of the experiment, differences in sound timbre associated with energy in the 3–7-kHz bandwidth could have served as a nonspatial cue to discriminate front and back locations. Of course this question is ultimately intractable, since perception can never be subjected to direct measurement. We do note, however, that three minor lines of evidence are suggestive of a recalibration of perceived space as the causal factor, rather than other nonspatial factors such as timbre. First, because the localization training procedure forced the listener to orient the head in the direction of the sound source, response feedback always ended with the sound source being presented directly in front of the head (0° azimuth, 0° elevation). As a result, it seems unlikely that the feedback procedure would have led listeners to associate particular spectral shapes (i.e., timbre) with correct responses, since the spectrum reaching the ear drum at the ultimate point of correct direction feedback (see Fig. 2) was fixed. Of course, listeners could have associated timbre at the beginning of the stimulus presentation (before feedback) with the final correct response, but this would have required listeners to then ignore the intervening spatial information that resulted from the changes in head orientation, which seems more complicated and less likely. Second, the spatial positions presented during the training phases of the experiment were different than those presented during either the pre or post tests. This suggests that knowledge gained during the training procedures generalized to the different spatial locations and different HRTFs evaluated in the post test. It also suggests that listeners were not simply associating a single fixed-spectrum proximal stimulus (stimulus at the ear drum) with either front or rear hemifield location, since the HRTFs and resulting proximal stimuli in the training and nontraining phases of the experiment were different. This result is therefore more consistent with the idea that the training procedure leads to enhanced processing of spatial information in the front-back dimension, rather than just learning to associate particular spectral shapes (i.e., timbre) with correct responses. Third, participants did not report relying on any response compensation strategies when asked during experimental debriefing. While informal, this is evidence to suggest that participants were responding, as instructed, to the apparent sound source direction and not relying on other nonspatial aspects of the sounds (e.g., timbre) to correct responses to apparent source directions that listeners know to be incorrect.

Although none of these points can be considered proof that spatial recalibration caused the observed changes in localization responses, all appear to be more consistent with this explanation than other potential explanations based on nonspatial cues such as timbre. They are also consistent with the informal observation that changes to the monaural spectral cues to source direction (single source, anechoic space) do not appear to affect the timbre of the sound source. Clearly there is a need for future experiments to examine these causal issues more fully, perhaps through more detailed manipulation of the source spectrum and/or the spatial distributions of source positions.

Finally, it is important to note that the proposed front-back spectral cue may be fundamentally different than other directional localization cues in that it specifies only binary information as to whether the sound source is located in front or behind the listener. In contrast, the other principal directional cues all appear to specify continuous spatial information. This distinction may explain why the presumed recalibration to this front-back cue can occur very rapidly in relationship to the more gradual development of recalibration to altered spectral elevation cues (Hofman *et al.*, 1998). Front-back recalibration may simply involve a switch between two possible percepts, perhaps not unlike switches observed in visual depth percepts related to certain illusory figures, such as the Necker cube (cf. Békésy, 1960).

B. Lack of elevation accuracy improvement

The training procedure described in the current experiments had little effect on elevation localization performance. This result is inconsistent with results reported by Hofman *et al.* (1998), which show clear decreases in elevation localization error during the course of adaptation. We suggest two possible explanations for the difference in results. First, the time course of (presumed) adaptation was very different in the two studies. Whereas participants in the Hofman study required weeks to fully adapt to the altered spectral cues (although considerable adaptation was observed in three of four listeners after 1 week), participants in the current study showed clear performance improvements in the front-back dimension after only two 30-min sessions of training. It is possible that longer periods of time are required to recalibrate to the particular spectral cues important for elevation than for the spectral cues relevant for front-back resolution. The spectral cues that provide discrete front-back information do appear to be relatively simple in relationship to the more complex patterns of spectral peaks and notches that provide continuous elevation information (Bloom, 1977; Macpherson and Middlebrooks, 2003; Zakarauskas and Cynader, 1993). Rates of front-back reversals have also been shown to be independent of elevation errors (Macpherson and Middlebrooks, 2000), which further suggests that the spectral cues underlying front-back and elevation localization are quite different, with perhaps very different adaptation time requirements.

A second possible explanation for the lack of observed elevation accuracy improvement relates to potential limitations in the HRTF dataset present in the sound card apparatus used in the experiment. Comparison of the patterns of spectral change in the elevation dimension for HRTFs from the sound card apparatus (Fig. 5) versus HRTFs from one of the participants in the study (Fig. 6) reveals considerable differences between the two datasets. Not only are the shapes and locations of various prominent spectral features (e.g., spectral peaks and notches) quite different, which is to be expected given known variation of HRTF sets between different listeners: The general pattern of spectral variation in the SLO HRTF dataset also appears to be more complex than the pattern of spectral variation present in the sound card dataset. While some of the differences in spectral complexity are also to be expected from HRTF datasets derived from measurements of different listeners, it is also possible that the generally more smooth spectral patterns observed in the sound card apparatus HRTFs resulted from signal processing compromises inherent in this low-cost sound card's design. As a result, the sound card apparatus simply may not provide sufficient elevation information to support accuracy improvement in this spatial dimension.

C. Practical implications

Virtual auditory display technology holds great promise for many practical applications where relevant spatial information needs to be conveyed or augmented by nonvisual means. Unfortunately, it is often difficult to achieve acceptable spatial localization accuracy with many "off-the-shelf" virtual 3D sound systems. This is because the spatial processing typically employed by these systems, although based on the specification of known acoustical cues to sound-source direction through the use of HRTFs, does not tailor the cues to the individual user. The effect of nonindividualized HRTFs on sound localization accuracy has been well documented (Bronkhorst, 1995; Middlebrooks, 1999b; Wenzel *et al.*, 1993) and also observed in the current study. Because providing for true individualized spatialization is a practical impossibility for commercial 3D sound systems, due to the logistical challenges associated with measuring HRTFs from a large number of source directions for each individual listener, it is of great interest to explore other ways of improving localization accuracy in such systems. This is particularly true for applications that require accurate localization of brief sounds, since the benefits in localization accuracy known to result from listener head movement (Wallach, 1940) cannot be realized given the inherent delays in movement initiation. Although a variety of methods has been proposed for computationally modeling HRTFs with the goal of parametric modification in order to minimize the mismatch with the listener's own HRTFs (Kistler and Wightman, 1992; Middlebrooks, 1999a, 1999b), the results from the current experiment suggest that listeners can rapidly adapt to at least some of the acoustic cue distortions caused by nonindividualized HRTFs, given a short period of multimodal feedback training. This result has great practical significance, since it demonstrates that large improvements in localization accuracy (via front-back reversal reduction) can result without a precise matching of the HRTFs to the display user. Therefore, instead of adapting the display to suit the user, the user can adapt to the display through training. Since the effects of training appear to last a very long time, training sessions need not be frequent. Further, if this training facilitates formation of a secondary spatial map in the brain, as has been suggested in both humans (Hofman *et al.*, 1998) and in animals (Knudsen *et al.*, 2000; Linkenhoker *et al.*, 2005), it seems likely that the training would not interfere or degrade the spatial map used for normal hearing conditions in the real world, which is consistent with the subjective reports of the participants in the current experiment. Of course, for many practical applications, the processes (perceptual or otherwise) underlying the adaptation may be irrelevant, provided increases in localization accuracy are realized. Although the rapid accuracy improvements reported here appear to be limited to reducing front-back reversal responses, these types of reversal errors can be catastrophic in many localization applications, particularly those involving orientation and/or navigation. As a result, methods of improving localization accuracy under nonoptimal stimulus conditions are still of considerable practical significance even when the improvements are limited to correcting front-back reversals.

V. CONCLUSIONS

This experiment has demonstrated that a brief perceptual training procedure (two 30-min sessions) which provides listeners with auditory, visual, and proprioceptive/vestibular

feedback as to the true target locations can improve localization accuracy for stimulus conditions in which a mismatch in spectral cues to source direction exists. The improvement was limited exclusively to the front-back dimension, where the proportions of hemifield reversal responses decreased substantially following the training procedure for four of five listeners that made frequent reversals initially. For these listeners the improvements appear to last at least 4 months after training, and generalize to untrained spatial locations. Because front-back location appears to be coded in the energy contained within the 3–7-kHz bandwidth of the signal at the ear drum, we suggest that the reductions in front-back reversals following training resulted from improved processing of this spectral cue. These results may have important practical implications for accuracy improvements in virtual auditory displays utilizing nonindividualized HRTFs, given that front-back reversals can be particularly problematic in many orientation and navigations tasks.

ACKNOWLEDGMENTS

The authors thank Dr. Fred Wightman and Dr. Doris Kistler for providing HRTF comparison data for listener SLO, and Dr. Armin Kohlrausch, Dr. Nathaniel I. Durlach, and two anonymous reviewers for their comments on earlier versions of this work. Financial support was provided under the Federated Laboratory Program by the U.S. Army Research Laboratory, Cooperative Agreement DAAL01-96-2-0003, and by NIH-NEI (F32EY07010) and NIH-NIDCD (R03DC005709).

Begault, D. R., and Wenzel, E. M. (**1993**). "Headphone localization of speech," Hum. Factors **35**, 361–376.

Békésy, G. v. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).

Blauert, J. (**1997**). *Spatial Hearing* (Revised ed.) (MIT Press, Cambridge, MA).

Bloom, P. J. (**1977**). "Determination of monaural sensitivity changes due to the pinna by use of minimum-audible-field measurements in the lateral vertical plane," J. Acoust. Soc. Am. **61**, 820–828.

Blum, A., Katz, B. F. G., and Warusfel, O. (**2004**). "Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training," in Proceedings of the CFA/DAGA, Strasbourg, France, 1225–1226.

Bronkhorst, A. W. (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **98**, 2542–2553.

Burger, J. F. (**1958**). "Front-back discrimination of the hearing system," Acustica **8**, 301–302.

Clifton, R. K. (**1987**). "Breakdown of echo suppression in the precedence effect," J. Acoust. Soc. Am. **82**, 1834–1835.

Clifton, R. K., Freyman, R. L., and Meo, J. (**2002**). "What the precedence effect tells us about room acoustics," Percept. Psychophys. **64**, 180–188.

Clifton, R. K., Clarkson, M. G., Gwiazda, J., Bauer, J. A., and Held, R. M. (**1988**). "Growth in head size during infancy: Implications for sound localization," Dev. Psychol. **24**, 477–483.

Florentine, M. (**1976**). "Relation between lateralization and loudness in asymmetrical hearing losses," J. Am. Aud Soc. **1**, 243–251.

Gardner, M. B., and Gardner, R. S. (**1973**). "Problem of localization in the median plane: Effect of pinnae cavity occlusion," J. Acoust. Soc. Am. **53**, 400–408.

Hausler, R., Colburn, S., and Marr, E. (**1983**). "Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests," Acta Oto-Laryngol., Suppl. **400**, 1–62.

Hebrank, J., and Wright, D. (**1974a**). "Are two ears necessary for localization of sound sources on the median plane?," J. Acoust. Soc. Am. **56**, 935–938.

Hebrank, J., and Wright, D. (**1974b**). "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834.

Held, R. (**1955**). "Shifts in binaural localization after prolonged exposures to atypical combinations of stimuli," Am. J. Psychol. **68**, 526–548.

Hofman, P. M., Van Riswick, J. G., and Van Opstal, A. J. (**1998**). "Relearning sound localization with new ears," Nat. Neurosci. **1**, 417–421.

King, A. J. (**1999**). "Sensory experience and the formation of a computational map of auditory space in the brain," BioEssays **21**, 900–911.

King, A. J., Hutchings, M. E., Moore, D. R., and Blakemore, C. (**1988**). "Developmental plasticity in the visual and auditory representations in the mammalian superior colliculus," Nature (London) **332**, 73–76.

Kistler, D. J., and Wightman, F. L. (**1992**). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am. **91**, 1637–1647.

Knudsen, E. I. (**1999**). "Mechanisms of experience-dependent plasticity in the auditory localization pathway of the barn owl," J. Comp. Physiol., A **185**, 305–321.

Knudsen, E. I., and Brainard, M. S. (**1991**). "Visual instruction of the neural map of auditory space in the developing optic tectum," Science **253**, 85–87.

Knudsen, E. I., and Zheng, W., and DeBello, W. M. (**2000**). "Traces of learning in the auditory localization pathway," Proc. Natl. Acad. Sci. U.S.A. **97**, 11815–11820.

Langendijk, E. H., and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. **112**, 1583–1596.

Lewald, J. (**2002a**). "Opposing effects of head position on sound localization in blind and sighted human subjects," Eur. J. Neurosci. **15**, 1219–1224.

Lewald, J. (**2002b**). "Rapid adaptation to auditory-visual spatial disparity," Learn. Mem. **9**, 268–278.

Linkenhoker, B. A., von der Ohe, C. G., and Knudsen, E. I. (**2005**). "Anatomical traces of juvenile learning in the auditory system of adult barn owls," Nat. Neurosci. **8**, 93–98.

Macpherson, E. A., and Middlebrooks, J. C. (**2000**). "Localization of brief sounds: effects of level and background noise," J. Acoust. Soc. Am. **108**, 1834–1849.

Macpherson, E. A., and Middlebrooks, J. C. (**2003**). "Vertical-plane sound localization probed with ripple-spectrum noise," J. Acoust. Soc. Am. **114**, 430–445.

Makous, J. C., and Middlebrooks, J. C. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2188–2200.

Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Middlebrooks, J. C. (**1992**). "Narrow-band sound localization related to external ear acoustics," J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C. (**1999a**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.

Middlebrooks, J. C. (**1999b**). "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.

Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?," J. Audio Eng. Soc. **44**, 451–469.

Oldfield, S. R., and Parker, S. P. (**1984**). "Acuity of sound localization: A topography of auditory space. II. Pinna cues absent," Perception **13**, 601–617.

Perrett, S., and Noble, W. (**1997a**). "The contribution of head motion cues to localization of low-pass noise," Percept. Psychophys. **59**, 1018–1026.

Perrett, S., and Noble, W. (**1997b**). "The effect of head rotations on vertical plane sound localization," J. Acoust. Soc. Am. **102**, 2325–2532.

Recanzone, G. H. (**1998**). "Rapidly induced auditory plasticity: The ventriloquism aftereffect," Proc. Natl. Acad. Sci. U.S.A. **95**, 869–875.

Rife, D. D., and Vanderkooy, J. (**1989**). "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–444.

Shaw, E. A. (**1966**). "Ear canal pressure generated by a free sound field," J. Acoust. Soc. Am. **39**, 465–470.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **5**, 1848–1861.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998a**). "Adapting to supernormal auditory localization cues. I. Bias and resolution," J. Acoust. Soc. Am. **103**, 3656–3666.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998b**). "Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response," J. Acoust. Soc. Am. **103**, 3667–3676.

Slattery, W. H., III, and Middlebrooks, J. C. (**1994**). "Monaural sound localization: Acute versus chronic unilateral impairment," Hear. Res. **75**, 38–46.

Strutt, J. W. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Van Wanrooij, M. M., and Van Opstal, A. J. (**2004**). "Contribution of head shadow and pinna cues to chronic monaural sound localization," J. Neurosci. **24**, 4163–4171.

Wallach, H. (**1940**). "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**, 339–368.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening. I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1997**). "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey & T. Anderson (Erlbaum, Mahwah, NJ), pp. 1–24.

Wightman, F. L., and Kistler, D. J. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841–2853.

Wightman, F. L., and Kistler, D. J. (**2005**). "Measurement and validation of human HRTFs for use in hearing research," Acta Acust. (Beijing) **91**, 429–439.

Woodworth, R. S., and Schlosberg, H. (**1954**). *Experimental Psychology* (Holt, Rinehart and Winston, New York).

Young, P. T. (**1928**). "Auditory localization with acoustical transposition of the ears," J. Exp. Psychol. **11**, 399–429.

Zakarauskas, P., and Cynader, M. S. (**1993**). "A computational theory of spectral cue localization," J. Acoust. Soc. Am. **94**, 1323–1331.

Zwiers, M. P., Van Opstal, A. J., and Cruysberg, J. R. (**2001**). "Two-dimensional sound-localization behavior of early-blind humans," Exp. Brain Res. **140**, 206–222.

Zwiers, M. P., Van Opstal, A. J., and Paige, G. D. (**2003**). "Plasticity in human sound localization induced by compressed spatial vision," Nat. Neurosci. **6**, 175–181.

Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients

Robert S. Hong[a] and Christopher W. Turner
Department of Otolaryngology-Head & Neck Surgery, Department of Speech Pathology & Audiology, University of Iowa, 121B WJSHC, Iowa City, Iowa 52242-1012

(Received 20 October 2005; revised 20 April 2006; accepted 21 April 2006)

This study examined the ability of cochlear implant users and normal-hearing subjects to perform auditory stream segregation of pure tones. An adaptive, rhythmic discrimination task was used to assess stream segregation as a function of frequency separation of the tones. The results for normal-hearing subjects were consistent with previously published observations (L.P.A.S van Noorden, Ph.D. dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands 1975), suggesting that auditory stream segregation increases with increasing frequency separation. For cochlear implant users, there appeared to be a range of pure-tone streaming abilities, with some subjects demonstrating streaming comparable to that of normal-hearing individuals, and others possessing much poorer streaming abilities. The variability in pure-tone streaming of cochlear implant users was correlated with speech perception in both steady-state noise and multi-talker babble. Moderate, statistically significant correlations between streaming and both measures of speech perception in noise were observed, with better stream segregation associated with better understanding of speech in noise. These results suggest that auditory stream segregation is a contributing factor in the ability to understand speech in background noise. The inability of some cochlear implant users to perform stream segregation may therefore contribute to their difficulties in noise backgrounds. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2204450]

PACS number(s): 43.66.Ts, 43.66.Mk, 43.66.Lj, 43.71.Ky [JHG] Pages: 360–374

I. INTRODUCTION

Auditory stream segregation is the process used to separate a complex sound into different perceptual streams, often corresponding to the different individual sources from which the sound is derived (Bregman, 1990). For example, it is used when people selectively listen to the melodies (streams) played by different instruments (sources) in the presence of an orchestral accompaniment. It also allows people to listen in on different conversations (streams) with different people (sources), one at a time, at a cocktail party. Compared to normal-hearing individuals, cochlear implant patients have tremendous difficulty accomplishing the tasks described in both of the preceding examples; cochlear implant recipients perform significantly worse than normal-hearing individuals on tests of complex song recognition (Gfeller *et al.*, 2005) and speech recognition in steady and modulated noise (Fu *et al.*, 1998; Nelson *et al.*, 2003; Stickney *et al.*, 2004). Thus, a better understanding of auditory stream segregation in implant users may be relevant for improving their ability to hear in complex acoustic environments.

The study of auditory streaming is commonly performed in the laboratory with sequences composed of two different tones (tones A and tone B) that alternate rapidly in time. When these tones are close together in frequency, they are often heard in a single perceptual stream (fusion) fluctuating between tones A and B. As the frequencies become farther apart, this percept changes to that of two different streams (fission), one composed of repeating tone A's and the other of repeating tone B's. Such a dependence of auditory stream segregation on frequency separation has been observed in both normal-hearing (Miller and Heise, 1950; Van Noorden, 1975) and hearing-impaired individuals (Rose and Moore, 1997; Mackersie *et al.*, 2001).

Van Noorden (1975) varied the frequency separation between tones A and B and found that the boundary between fusion and fission varied depending on the instructions given to the subjects. If the subjects were asked to listen for a single stream, and the frequency difference between the tones was gradually increased until this was no longer possible, one boundary was measured which was termed the temporal coherence boundary. If the subjects instead were asked to listen for two streams, with the frequency difference between tones decreased until this was no longer possible, a different boundary was defined called the fission boundary. In analyzing the data, van Noorden also found a region of frequency separation of the tones between the two boundaries where either fusion or fission could be perceived (depending on the instructions given); this has been referred to as the ambiguity region.

Studies of auditory streaming comparing the ability of normal-hearing and hearing-impaired subjects to stream pure tones have demonstrated that, in general, the hearing-impaired subjects have a reduced ability to perform stream segregation. Rose and Moore (1997) and Mackersie *et al.* (2001) examined the frequency separation for pure tones at the fission boundary and determined that hearing-impaired ears require a greater frequency separation than normal-hearing ears at this boundary, though there were a few exceptions. A number of studies have been performed in an

[a]Electronic mail: robert-hong@uiowa.edu

 0001-4966/2006/120(1)/360/15/$22.50

attempt to explain the reason for this difference in performance between normal-hearing and hearing-impaired listeners. One theory proposed by Beauvois and Meddis (1996) is that auditory streaming can be explained by the degree of overlap of excitation patterns in the cochlea in response to an acoustic stimulus, with less overlap leading to greater stream segregation. This theory implies that the broader auditory filters of hearing-impaired individuals are responsible for their reduced ability to perform stream segregation. However, Rose and Moore (1997) and Mackersie *et al.* (2001) found that frequency selectivity alone is not a good predictor of auditory streaming ability in hearing-impaired subjects. Another theory suggests that it is the clarity of pitch sensations evoked by the pure tones that determines if streaming occurs. It has been suggested that because hearing-impaired individuals have poorer pure-tone frequency discrimination than normal-hearing subjects, they have a weaker pitch sensation associated with pure tones, and thus are less able to stream pure tones (Rose and Moore, 2005). Rose and Moore (2005) found evidence of statistically significant correlations between frequency discrimination and fission boundary in hearing-impaired subjects. However, these correlations were not particularly strong, suggesting that differences in frequency discrimination are also not enough to fully explain much of the variation in streaming ability in the hearing-impaired. The previous studies suggest that frequency selectivity and frequency discrimination ability can contribute to a person's ability to perform pure-tone stream segregation, but are not sufficient for a good prediction of streaming ability. This is not particularly surprising given that central processes, such as attention, are also thought to be important in stream segregation (Carlyon *et al.*, 2001), suggesting that subject-to-subject variation in such central processes is also important to consider in predicting streaming ability.

Cochlear implant recipients have poorer frequency selectivity and frequency discrimination ability via their speech processors than normal-hearing subjects (Dorman *et al.*, 1996; Gfeller *et al.*, 2002). Therefore, we would expect that as a group, their pure-tone stream segregation ability would be "worse" than that of normal-hearing subjects, assuming that the relevant central processing abilities of both groups are similar. (By "worse" streaming ability, we mean that if, for a given frequency difference between two rapidly alternating pure tones, subject 1 hears one perceptual stream but subject 2 hears two streams, then subject 1 has "worse" streaming ability because he cannot utilize the frequency difference cue to segregate the tones into different streams.) We may also see some cochlear implant subjects who have similar (or even better) stream segregation abilities than the worst-streaming normal-hearing subjects. One can imagine that a cochlear implant patient with extremely good central processing abilities could overcome a degradation of frequency cues (if the cues are not too degraded) and outperform a normal-hearing subject with poor central processing but intact frequency cues.

In the present study, we examined the ability of normal-hearing and cochlear implant subjects to perform stream segregation of acoustically presented pure tones as a function of frequency separation. The cochlear implant subjects listened through their cochlear implant speech processor. This allowed a direct comparison with normal-hearing subjects, who also listened in the sound field. Furthermore, the cochlear implant subjects in this study performed all experiments with the MAPs they used on an everyday basis. This permitted us to determine if their performance on the stream segregation task was related to their everyday ability to understand speech in background noise, which was also measured. Additionally, streaming ability was assessed at multiple base frequencies (200, 800, and 2000 Hz) to determine if streaming abilities varied within a subject, as might be expected if there were uneven patterns of nerve survival across the cochlea in cochlear implant users. This also allowed us to determine if the ability to stream in lower frequency regions, such as those corresponding to the fundamental frequency of talkers, was more strongly related to the ability to understand speech in noise than streaming at higher frequencies. This might be expected because fundamental frequency differences between the target and masker speech are thought to be important cues for talker segregation.

II. EXPERIMENT 1. PURE-TONE STREAM SEGREGATION

A. Participants

Seven normal-hearing subjects (ages 21–35) and eight cochlear implant subjects (ages 39–78) participated in this study. All normal-hearing subjects had pure-tone thresholds of less than 20 dB HL across octave audiometric frequencies (0.25–8.0 kHz), with the exception of one subject who had a threshold of 50 dB HL at 8000 Hz. All cochlear implant subjects were tested using their everyday signal processing strategies and had at least 1 year of experience with their device at the time of testing. The age, implant type, signal processing strategies, and stimulation modes of the cochlear implant subjects who participated in experiment 1 are shown in Table I. This study received prior approval from the Institutional Review Board at the University of Iowa.

B. Stimuli and procedures

In this study, we use a rhythmic discrimination task based on the one introduced by Roberts *et al.* (2002) to assess auditory stream segregation. We will refer to this task as the streaming rhythm task. In this method, subjects are presented two sequences of rapidly alternating tones and asked to identify which sequence has an irregular rhythm. The task is based on the premise that the irregular rhythm is most easily identified when tone A and tone B are heard together in the same stream, as opposed to individually in different streams. Thus, this task measures stream segregation at the temporal coherence boundary, which is the boundary assessed when subjects are asked to hear all tones in the same stream.

The sequences used in this experiment are identical in rhythm to those described by Roberts *et al.* (2002). All stimuli were pure tones of 60-ms duration, which included 10-ms linear onset and offset ramps. Subjects were presented with two sequences of tones alternating in an AB fashion, with each sequence containing 12 AB cycles. In one se-

TABLE I. Cochlear implant subject demographics. The age, device, signal processing strategy, and stimulation mode (MP=monopolar; BP=bipolar) of each of the 16 cochlear implant subjects who participated in this study are shown. Also indicated are the specific experiments (Exp. 1, 2, and/or 3) in which each participated.

					Experiment no.		
Subject	Age	Device	Strategy	Mode	1	2	3
CI1	49	Nucleus CI24M	ACE	MP	x		x
CI2	51	Nucleus CI24R	CIS	MP	x	x	x
CI3	78	Nucleus CI24M	ACE	MP	x		x
CI4	39	Nucleus CI24R	ACE	MP	x		x
CI5	75	Nucleus CI24R	ACE	MP	x		x
CI6	76	Nucleus CI24M	ACE	MP	x		x
CI7	65	Clarion CII HF	HiRes	MP	x		x
CI8	48	Clarion CI (spiral)	CIS	MP	x		x
CI9	78	Clarion CII HF	HiRes	MP		x	x
CI10	33	Clarion 90K	HiRes	MP		x	x
CI11	64	Clarion CII HF	HiRes	MP			x
CI12	58	Nucleus CI22	SPEAK	BP[a]			x
CI13	67	Clarion CII HF	HiRes	MP			x
CI14	46	Clarion CII HF	HiRes	MP			x
CI15	54	Clarion CI (spiral)	CIS	MP			x
CI16	44	Clarion CII HF	HiRes	MP			x

[a]Electrode 3 served as the ground for all electrode pairs in bipolar configuration.

quence, tone A and tone B alternated back and forth regularly, separated by a 40-ms silence between each tone. In the other sequence, tone A and tone B alternated such that the first six AB cycles were regularly spaced (identical to the first sequence), the next four AB cycles (transition region) had an increasing delay (with the magnitude of the increase in delay being equal from one cycle to the next) imposed on the onset of tone B, and the final two AB cycles maintained the accumulated delay. The length of each sequence was 2.4 s. In Fig. 1, with "*" = tone A and "#" = tone B, the components of the regular and irregular rhythm sequences are shown.

In the streaming rhythm task, it was to the listener's advantage to hear all the tones in the same perceptual stream, because the delay in the irregular rhythm sequence was more difficult to detect when they were heard in different streams. Thus, as the two tones in the sequence became further apart in frequency, if a listener experienced stronger segregation of the two tones into different streams, then the rhythmic discrimination task became more difficult.

Subjects were presented a two-interval forced choice task (2IFC) task with feedback in which they were asked to identify which of two sequences of alternating tones results in a rhythm that sounds "unsteady and irregular." The irregular sequence was randomly presented in either the first or second interval from trial to trial. Each sequence was numbered according to the order presented ("1" or "2") and displayed as labeled buttons on a touch-screen (MicroTouch). The task measured the smallest delay in tone B that resulted in a detectable irregularity of rhythm (Fig. 1), utilizing a three-down one-up adaptive staircase to converge on the 79.4% correct point on the psychometric function for time delay (Levitt, 1971). The maximum delay that was imposed on tone B to avoid overlap of the tones was 40 ms. The initial size of the accumulated delay was 32 ms. The adaptive step size for the first two reversals was a step size of 8 ms. The last four reversals used a step size of 4 ms. The time delay threshold for each run was taken as the mean of the last four reversals. Thresholds for subjects who made four total incorrect identifications at the maximum time de-

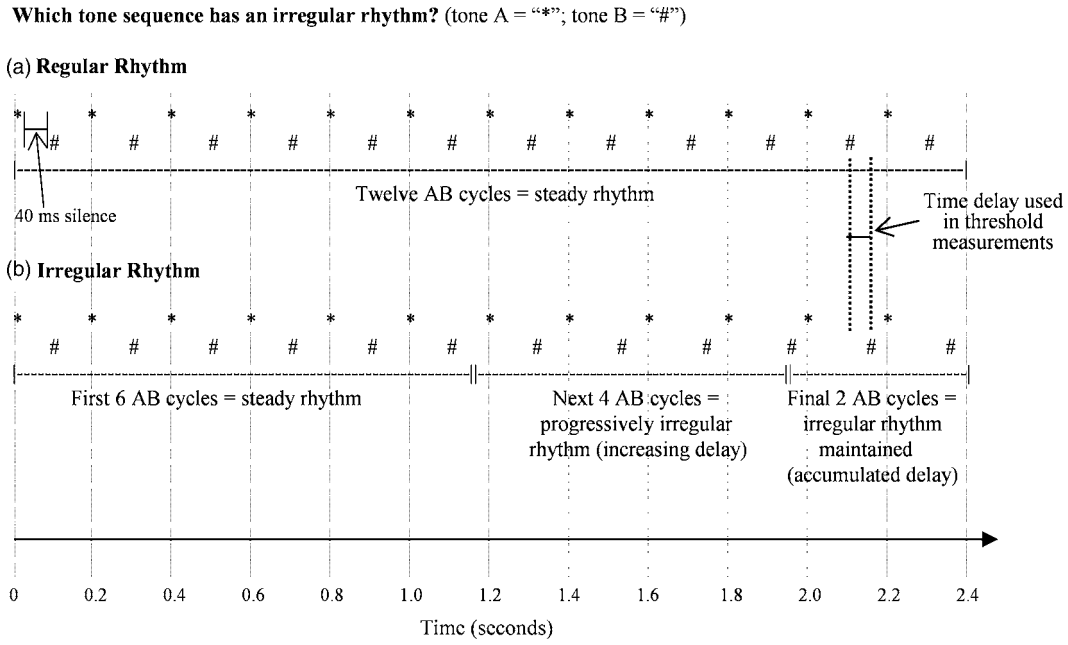


FIG. 1. Test stimuli used in rhythmic discrimination task to assess auditory stream segregation. (a) The regular rhythm is composed of 24 evenly spaced, alternating pure tones in the ABAB format. Tones A and B are always separated by 40 ms of silence. (b) The irregular rhythm is composed of 24 alternating tones that are evenly spaced over the first six cycles, have a progressively increasing delay in tone B over the next four cycles, and maintain the accumulated delay over the final two cycles. The time delay used to calculate threshold measurements for detection of the irregular rhythm is the delay in tone B in the final, accumulated delay section of that sequence.

lay were nominally taken as the 40-ms maximum delay, and the adaptive procedure was terminated in those cases.

The stimuli were grouped into three conditions, based upon the frequency of tone A: (1) low base frequency, with tone A = 200 Hz; (2) medium base frequency, with tone A = 800 Hz; and (3) high base frequency, with tone A = 2000 Hz. For each base frequency (tone A), tone B was chosen to give a Weber fraction [(frequency B–frequency A) / frequency A] of 0, 0.01, 0.1, 0.5, 1.0, and 3.0. The exception was that for the 2000-Hz base frequency, a Weber fraction of 3.0 was approximated with a tone B of 7800 Hz instead of the calculated 8000 Hz, because 8000 Hz is above the frequency range presented by signal processing strategies to cochlear implants. Normal-hearing subjects were tested at all Weber fractions for each base frequency. Cochlear implant subjects were tested at the Weber fractions that appeared to be most relevant for defining the shape of the psychometric function at each base frequency.

Testing was completed for one base frequency condition before proceeding to the next. The order of conditions was randomized across subjects. At each base frequency condition, the first Weber fraction tested was 0 (frequency of tone A = tone B). The order of presentation of all other Weber fractions was randomized within each condition. Prior to the collection of data, each subject was given three practice runs at the first base frequency condition to be tested, with one run corresponding to a Weber fraction of 0, and the other two runs corresponding to a Weber fraction of 0.01. If a subject was not able to obtain an average time delay threshold below 15 ms with the two tones equal in frequency (Weber fraction = 0), the subject was excluded from further testing, since this suggested a lack of ability to perform rhythmic discrimination exclusive of stream segregation. One normal-hearing subject was unable to meet this criterion at any of the base frequencies. Additionally, one cochlear implant subject (subject CI8) was able to meet this criterion for two base frequencies (200 and 2000 Hz) but not the third (800 Hz), and was excluded from testing at 800-Hz base frequency only. At each Weber fraction for each base frequency condition, subjects were tested with three consecutive adaptive runs. A fourth run was added if the standard deviation of the average of the three trials was greater than 6 ms, or if the standard deviation of the reversals of any run was greater than 8 ms. For each subject, the value for time delay threshold at a particular Weber fraction and base frequency was taken as the average of all three (or four) runs performed at that test condition.

All stimuli were generated digitally using MATLAB and stored on a Macintosh G4 computer. Stimuli were output through a 16-bit digital-to-analog converter (Audiomedia III, Digidesign, Inc.) at a sampling rate of 44.1 kHz and smoothed by a 20-kHz antialiasing low-pass filter. The stimuli were presented via a loudspeaker situated directly in front of the listener in a double-walled sound-attenuated booth. For normal-hearing subjects, all stimuli were presented at 80 dB SPL, with their right ear plugged with an ear plug. For cochlear implant subjects, the low base frequency set was presented at 95 dB SPL, and the medium and high base frequency sets were presented at 90 dB SPL. The ear without the cochlear implant was plugged. Cochlear implant subjects were allowed to adjust their microphone settings so that the sound was at a comfortable level.

C. Results and discussion

1. Streaming abilities of normal-hearing and cochlear implant subjects

The normalized results of both normal-hearing and cochlear implant subjects on this task are shown in Fig. 2 for three different base frequencies (tone A = 200, 800, or 2000 Hz). Our approach to eliminating any possible effects of a subject's basic ability to discriminate rhythms upon the question of interest (ability to separate frequencies into different streams) was to normalize the raw scores by dividing each value by the threshold at Weber fraction = 0 (i.e., when all of the tones are identical). This particular method of normalization was selected, because performance on temporal tasks in normal-hearing subjects (such as gap discrimination or detection) has been shown to be relatively linear with respect to frequency separation (of the tones bordering the gap), when both factors are plotted on a logarithmic scale (e.g., Neff *et al.*, 1982; Formby and Forrest, 1991). [We also tested an alternative method of normalization wherein the values of threshold at Weber fraction = 0 were subtracted from the raw scores. This method yielded similar conclusions (data not shown).] Thus, the normalized thresholds may be interpreted as a relative measure of auditory stream segregation, measuring how many times more difficult the streaming rhythm task became compared to baseline (when all the tones were identical) as the frequency difference between alternating tones was increased.

The average normalized results of six normal-hearing (NH) subjects on the task are indicated by the heavy line in Fig. 2. As shown in the figure, the smallest average threshold for normal-hearing subjects was achieved when there was no frequency difference between the tones, which is as expected since all of the tones were identical in frequency and thus heard in the same stream. As the frequency separation increased between tone A and tone B, thresholds increased, suggesting that subjects more strongly heard two streams. A single value for the overall streaming ability of normal-hearing subjects was derived from the slope of a regression line with a y intercept at 1 (since at a frequency difference of 0, normalization by the threshold at Weber fraction = 0 results in a value of 1) that was fit through the data shown in Fig. 2. Data where ceiling effects were reached at larger frequency differences (i.e., when subjects were unable to identify the irregular rhythm at the largest time delay possible) were not included in the regression analysis, since inclusion of such data may result in underestimation of the slope. Higher values for the slope correspond to larger increases in difficulty of the task as the frequency difference between alternating tones is widened; thus, larger slopes correspond to more auditory stream segregation with increasing frequency separation. The value of the slope for the average normal-hearing (NH) subject at each of the three base frequencies and the 95% confidence interval for each slope distribution (±2 standard deviations about the mean) are shown in Fig. 3. As can

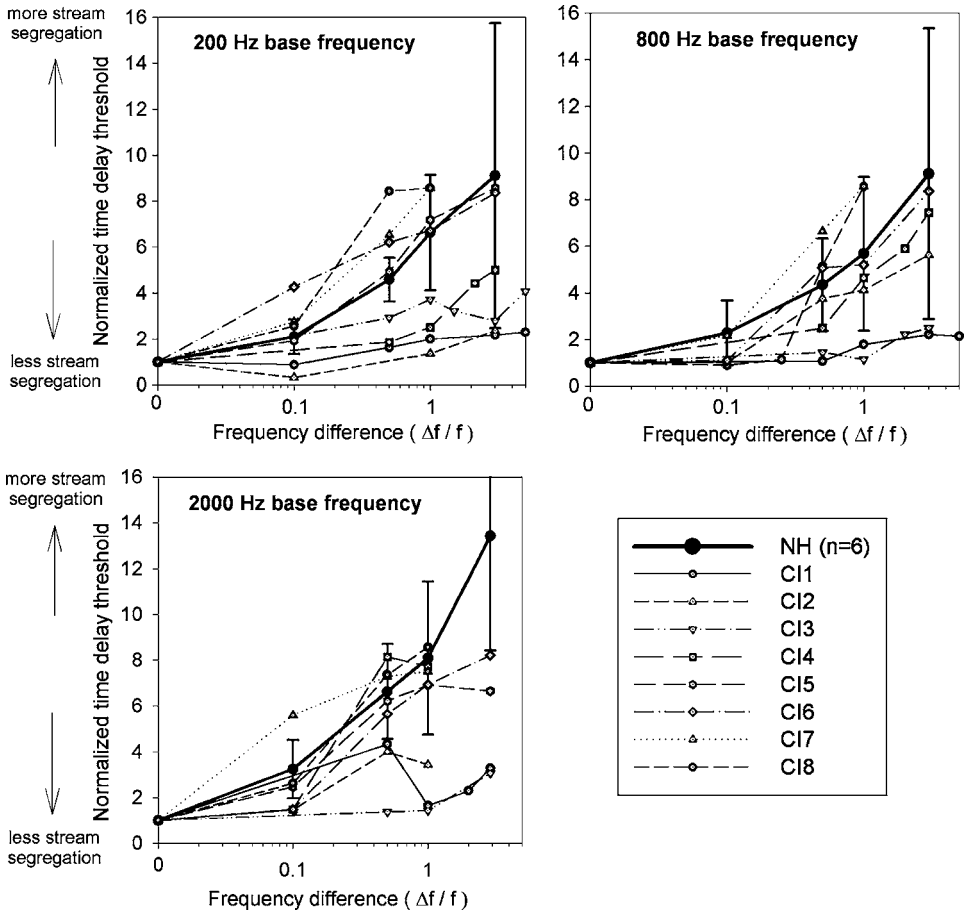


FIG. 2. Normalized performance on the streaming rhythm task for normal-hearing and cochlear implant subjects. The average performance of six normal-hearing subjects (NH) is shown in the heavy line at various frequency differences between tone A and tone B, with the error bars representing one standard deviation above and below the mean. The individual performance of seven to eight cochlear implant (CI) subjects is also shown. Measures are made at each of three base frequencies (tone A=200, 800, or 2000 Hz). The time delay thresholds are normalized by baseline performance, giving a normalized threshold that represents the factor by which the task increases in difficulty at different frequency separations of the alternating tones. A larger value of normalized threshold is consistent with greater auditory stream segregation.

be seen from this figure, there is a relatively large range of pure-tone streaming abilities among normal-hearing subjects as measured by this task.

Figure 2 also shows the individual data from eight cochlear implant (CI) subjects on the streaming rhythm task. With cochlear implant recipients, the baseline performance (Weber fraction=0) was generally comparable to that of normal-hearing subjects on the task, but there was variation in performance among cochlear-implant subjects with increasing frequency separation. Statistical analysis was again performed by fitting a regression line in a similar manner as discussed for normal-hearing subjects through the data for each cochlear implant subject. To determine the performance of cochlear implant relative to normal-hearing subjects, we compared the regression slopes of each cochlear implant subject with that of the average normal-hearing subject at each of the three base frequencies (Fig. 3). Values for individual cochlear implant subjects that fall outside the dotted lines (representing the 95% confidence interval for the normal-hearing slope distribution) are interpreted as significantly different from those of the average normal-hearing subject. At a base frequency of 200 Hz, four cochlear implant subjects have slopes that are significantly lower than the average normal-hearing subject, while at 800 and 2000 Hz, two and three cochlear implant subjects, respectively, have slopes that are significantly lower. These results suggest that a number of cochlear implant subjects experience significantly less auditory stream segregation with increasing frequency separation than normal-hearing listeners. The effect of base frequency on streaming was also analyzed via repeated-measures ANOVA for the regression slopes, with no significant differences observed within cochlear implant subjects between base frequencies [$F_{1.17,6}=4.08$ (Greenhouse-Geisser adjustment for lack of sphericity); $p>0.05$].

2. Influence of electrode separation (place pitch cues) on streaming

The influence of electrode separation on the streaming rhythm task was also examined for the six Nucleus CI24 cochlear implant recipients. The goal of this analysis was to determine if performance on the task could be explained primarily by the distance between the electrodes presenting tones A and B; if this were the case, for example, we would expect two tones presented by adjacent electrodes to result in similar streaming performance whether the base frequency was 200, 800, or 2000 Hz. We focused on the Nucleus subjects because we wanted our measure of electrode separation to reflect the role of place pitch cues for auditory streaming. The low-pass temporal envelope cutoff frequency for the signal processing strategies of the Nucleus subjects was ~125 Hz (personal communication with Bom-Jun Kwon, Cochlear Corporation), which was below the lowest frequency tested (200 Hz), suggesting that temporal pitch cues would not confound our analysis. In contrast, for the Clarion subjects, the low-pass envelope cutoff frequency was high enough that temporal pitch cues may have been available to these subjects.

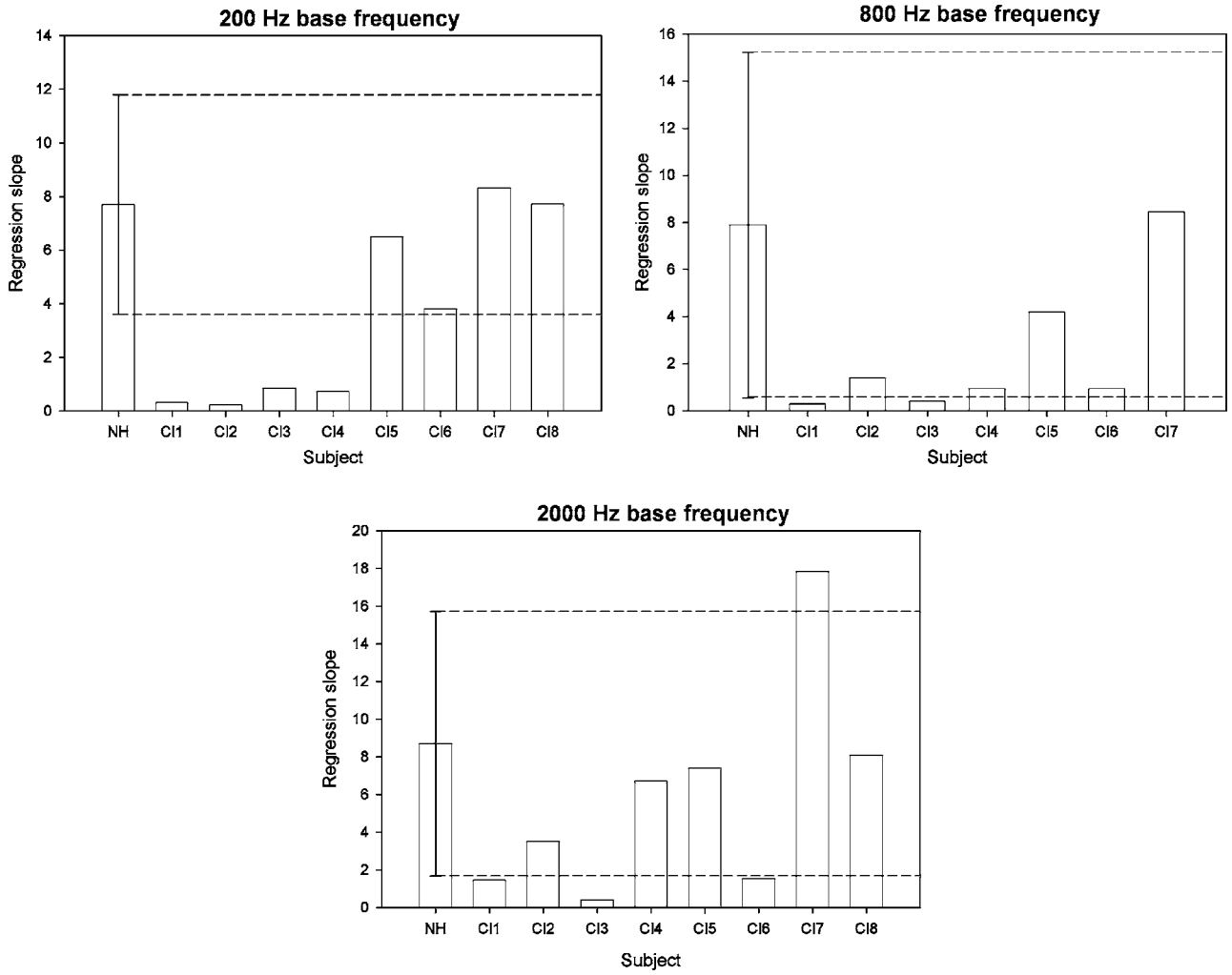


FIG. 3. Overall streaming ability in normal-hearing and cochlear implant subjects. Overall streaming ability is indicated by the slopes of regression lines (with a common y intercept) fit through the data for the streaming rhythm task at each of three base frequencies (tone A=200, 800, or 2000 Hz). NH represents the average slope of six normal-hearing subjects, with the 95% confidence interval for the normal-hearing slope distribution (±2 standard deviations about the mean) denoted by the dashed lines. The slopes of each individual cochlear implant (CI) subject are shown by the other bars.

To analyze the role of electrode separation on streaming, the values for frequency difference shown in Fig. 2 were replotted as electrode separation on separate graphs for each individual subject (with two such examples shown in Fig. 4). To convert frequency difference to electrode separation, each stimulus tone (tone A or tone B) was assigned to the single electrode that was programmed to represent that stimulus frequency in the patient's clinical MAP (according to the cutoff frequencies for each electrode). Then, the number of electrodes separating the stimulating electrodes corresponding to tone A and tone B was determined. This conversion only provided an estimate of the electrode separation for two tones, since the slope of each filter at the cutoff frequency was not infinite. Nevertheless, this was useful in providing a general picture of the influence of electrode separation on streaming. The conversion was performed for each subject at all three base frequencies, with the resulting values plotted versus normalized threshold onto a single graph for each subject. Linear regression was then performed for the data across all three base frequencies for each subject. A statistically significant value for the slope of the regression line was interpreted as evidence that performance on the streaming rhythm task could be explained by electrode separation, regardless of the region of the cochlea that the stimuli were presented.

Four cochlear implant subjects (subjects CI4, CI1, CI5, and CI2) showed evidence of a similar streaming ability based on electrode separation across all frequency ranges. Figure 4(a) shows the data for one of these four subjects, subject CI4. As depicted by the figure, the linear regression line for the normalized time delay thresholds versus electrode separation across the three base frequency conditions was statistically significant ($p<0.001$), with slope=0.278 and $r=0.655$. The linear regressions for the other three subjects (not shown in figure) were also statistically significant—subject CI1 (slope=0.117; $r=0.396$; $p<0.005$), subject CI5 (slope=1.388; $r=0.841$; $p<0.001$), and subject CI2 (slope=0.339; $r=0.792$; $p<0.001$)—suggesting that electrode separation could also explain a significant portion of their performance on the streaming rhythm task in different stimulus frequency ranges.

In contrast, the other two cochlear implant subjects (subjects CI3 and CI6) appeared to have different streaming

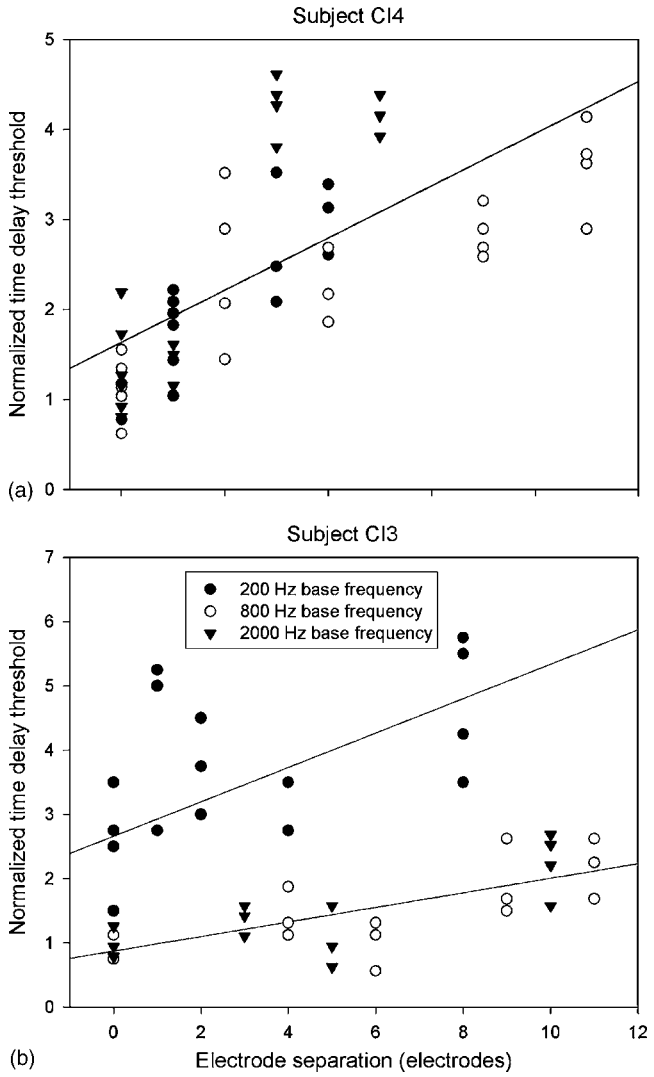


FIG. 4. Performance on streaming rhythm task with respect to electrode separation in cochlear implant subjects. Data from all three base frequencies of the streaming rhythm task are converted into electrode separation and plotted on the same graph for each subject. The resulting graphs of two subjects, which represent the two general types of data patterns observed, are shown. (a) Subject CI4 has a moderate, statistically significant correlation between task performance across all frequency ranges and electrode separation ($r=0.655$; $p<0.001$). (b) Subject CI3 shows no evidence of a correlation between task performance and electrode separation for data combined across all three base frequencies ($r=0.140$; $p>0.1$; regression line not shown). However, correlations were significant when one regression line was drawn through the 200-Hz data ($r=0.570$; $p<0.02$), and a different regression line was drawn through the combined 800- and 2000-Hz data ($r=0.714$; $p<0.001$).

abilities across electrodes in different regions of the cochlea. Figure 4(b) shows an example of one subject (subject CI3) who appeared to have different streaming abilities in different frequency ranges, after accounting for electrode separation, since a single regression line fit the data very poorly ($r=0.140$; $p>0.1$). In fact, if two regression lines are drawn through the data, one through the 200-Hz base frequency data points and the other through the 800- and 2000-Hz base frequency points, the correlations are statistically significant (for 200 Hz line, slope=0.268, $r=0.570$ and $p<0.02$; for the 800–2000-Hz line, slope=0.113, $r=0.714$ and $p<0.001$). The data for subject CI6 also fit a single regression line poorly ($r=0.053$; $p>0.5$). Two regression lines were again required to better describe the data, with one line for the 800- and 2000-Hz base frequencies (slope=0.176; $r=0.723$; $p<0.001$) and another for the 200-Hz base frequency ($r=0.264$; $p>0.1$), though the latter correlation was not statistically significant.

3. General discussion

The results from Fig. 3 suggest that there is a range of streaming abilities for both normal-hearing and cochlear implant users. Nevertheless, despite the variability in performance on the streaming rhythm task within the normal-hearing group, this study demonstrates that some cochlear implant users stream significantly worse than normal-hearing subjects. These differences in pure-tone streaming between cochlear implant subjects likely represent a combination of differences in peripheral (frequency resolution) and central processing. The results from Fig. 4 suggest that in some cases, within a cochlear implant subject, streaming ability based on place pitch cues can vary from one region of the cochlea to another. Within a subject, the contribution of central processing to streaming is presumably similar across all stimulus frequencies, and thus differences in streaming for different electrode locations reflect differences in peripheral processing. One possible explanation for why some subjects may have different streaming abilities for electrode separation in different regions of the cochlea is that there may be different amounts of nerve survival throughout the cochlea, leading to differences in the perceptual distance of pitch attributed to adjacent electrodes and thus differences in pure-tone streaming.

There is a possibility that differences in loudness, as opposed to differences in pitch, could be responsible for some of the auditory stream segregation observed in this experiment. All of the stimuli were presented at equal dB SPL and thus were not loudness balanced. We believe that the effects of loudness differences in streaming in our experiment were minimal for a number of reasons. First, for normal-hearing listeners, there are minimal differences in loudness across much of the intensity and frequency range tested [ISO 226: 2003 (Normal equal-loudness-level contours) of the International Organization for Standardization]. Second, the presence of any loudness cues for all subjects was minimized by the presentation of stimuli in a free field, because the head position of listeners was not fixed with respect to the speaker, which allowed for potential random variations in sound intensity over a few decibels with sporadic changes in head position during the course of testing. Third, none of the subjects reported that loudness differences hindered performance on the task. Fourth, there is no clear evidence that loudness differences can be used to obligatorily segregate sounds in the same powerful way as frequency differences (Bregman, 1990, pp. 126–127). Finally, it is interesting to note that even if loudness cues were present, there were clearly some cochlear implant subjects that had little ability to stream at any frequency difference based on either loudness or frequency cues (e.g., subject CI1 or CI3 in Fig. 3). For the other cochlear implant subjects with better streaming abilities, if they were able to use loudness as a cue

for streaming, this would suggest that our results may overestimate their ability to segregate pure tones into different streams.

There is also the possibility that age differences between the normal-hearing and cochlear implant groups in this experiment may be a confounding factor in the analysis of the results with respect to streaming. The cochlear implant subjects were older than the normal-hearing subjects (ages 39–78 compared to ages 21–35), and there is evidence in the literature that performance on temporal tasks declines with age (for a review, see Pichora-Fuller, 2003). However, there are a number of reasons why we believe that differences in age between the two groups did not affect our overall conclusions. First, as stated earlier, the time delay thresholds from the streaming rhythm task used in our analysis of streaming ability were all normalized with respect to each individual subject's baseline ability to discriminate rhythms (baseline assessed at Weber fraction=0). By doing this, we attempted to control for differences in basic temporal perception ability between individuals, including those due to differences in age. Second, we did not find any statistically significant correlations between age and performance on the streaming rhythm task (for normalized thresholds and regression slopes), both within and across subject groups. Finally, assuming that temporal ability differences due to age were responsible for differences in performance on the task, we would expect that the older cochlear-implant group would have higher thresholds than the younger normal-hearing group. However, our results indicated the opposite: the time delay thresholds of the cochlear-implant group tended to be lower than those of the normal-hearing group. Thus, it does not appear that age differences between the two groups confounded our conclusion that the streaming ability of many cochlear implant subjects was worse than that of normal-hearing individuals.

III. EXPERIMENT 2. AUDITORY STREAM SEGREGATION OR GAP DISCRIMINATION?

A. Rationale

It is possible that the results obtained on the streaming rhythm task may not reflect auditory streaming, but merely gap discrimination. For example, one can imagine a subject who performs the streaming rhythm task by focusing only on the end of the alternating-tone sequence to determine the sequence with the irregular rhythm. Taken to the extreme, it is possible that the subject may ignore the entire sequence except for the final three tones of the sequence, where the largest and smallest gaps are present, turning the task into one that looks very similar to gap discrimination.

It is difficult to determine based on the results of experiment 1 whether gap discrimination or stream segregation is the dominant phenomenon. Studies of gap detection and stream segregation in normal-hearing subjects have demonstrated that both are affected similarly by differences in frequency between tones: it becomes more difficult to perceive a gap and more difficult to hear tones in the same stream as the frequency separation widens between different tones (Neff *et al.*, 1982; Phillips *et al.*, 1997). If we presume that the increasing frequency separation affects performance by increasing the perceptual pitch distance between tones, then a similar relationship might also be seen with gap detection in cochlear implants. It has been found that gap detection in cochlear implants worsens with increasing electrode distance (place pitch effects) and increasing rate differences (temporal pitch effects) between the two tones which border the gap to be detected (Hanekom and Shannon, 1997; Chatterjee *et al.*, 1998; van Wieringen and Wouters, 1999). Furthermore, any correlations found in this study (described in experiment 3) between the streaming rhythm task and performance on a speech perception in noise task also will not lend insight into whether the streaming task measures stream segregation or gap discrimination. While correlations between speech perception in noise have been found with auditory streaming (Mackersie *et al.*, 2001), correlations have also been found between gap detection and speech perception in noise (Tyler *et al.*, 1982; Dreschler and Plomp, 1985), although such correlations with gap detection are not universally present (Strouse *et al.*, 1998; Snell and Frisina, 2000).

Although there are a number of similarities between auditory stream segregation and gap discrimination, there are also a number of ways to distinguish the two. First, auditory stream segregation is affected by the presentation rate of the alternating tones, whereas gap discrimination is not (Neff *et al.*, 1982). If performance on a task designed to assess auditory streaming varies with presentation rate, this lends supportive evidence that the task measures streaming ability. Second, auditory streaming is known to build up over time for alternating tones of moderate frequency separation, such that increasing amounts of segregation are seen over the first 10–30 s of listening to such tones (Anstis and Saida, 1985). This observation can also be used to obtain supportive evidence that differences in performance on the streaming rhythm task employed in this study reflect differences in streaming ability, and this latter approach is the one that we chose to take in experiment 2.

The goal of experiment 2 is to determine if the streaming rhythm task used in experiment 1 measures auditory stream segregation. For this experiment, we use shortened versions of the alternating-tone sequences presented in the streaming rhythm task and again measure the time delay threshold required to hear the irregular rhythm. We refer to this task as the short rhythm task. If there is no difference in relative performance between the short rhythm task and the streaming rhythm task when the frequency separation in tones is increased, this suggests that the streaming rhythm task is merely a measure of gap discrimination. However, if, as we hypothesize, subjects perform relatively worse on the streaming rhythm task, then this suggests that the streaming rhythm task measures streaming ability: the build-up of streaming induced by the longer sequences in the streaming rhythm task makes it more difficult to detect the irregular rhythm.

B. Participants

Three normal-hearing subjects who participated in experiment 1 also participated in this experiment. Additionally,

three cochlear implant (subjects CI2, CI9, and CI10) subjects were randomly selected to participate in this experiment, with their demographics found in Table I.

C. Stimuli and procedures

The task in this experiment is identical in all respects to the streaming rhythm task presented in experiment 1, except for the stimuli. In this experiment, the rhythmic sequences are shortened so that they contain only three pure tones (instead of 24 pure tones). The three pure tones of the regular rhythm sequence in this experiment are derived from the first three tones of the regular rhythm sequence of the streaming rhythm task. The three pure tones of the irregular rhythm sequence in this experiment are identical to the first three tones of the "accumulated delay" region (described in Fig. 1) of the irregular rhythm sequence of the streaming rhythm task. The time delay used for rhythmic discrimination threshold measurements is thus the same for the two experiments. As with the streaming rhythm task, subjects listened to two sequences and are asked to identify the irregular, unsteady rhythm in a 2-IFC adaptive task that converges on the 79.4% correct point for time delay threshold. Subjects were tested at each of three base frequencies (200, 800, and 2000 Hz) at conditions corresponding to a Weber fraction=0 and Weber fraction=0.5. Overall thresholds were determined from the average of the thresholds from three (or four) runs at each test condition. These specific conditions were chosen because they were the ones where correlations were subsequently performed between the streaming rhythm task and speech perception in noise (described in experiment 3), allowing insight into whether any correlations observed at those specific frequency differences reflected stream segregation or gap discrimination.

D. Results and discussion

The baseline ability (Weber fraction=0) of each listener to perform the streaming rhythm task and the short rhythm task at each of three base frequencies (200, 800, and 2000 Hz) is shown in Fig. 5 for both normal-hearing [Fig. 5(a)] and cochlear implant [Fig. 5(b)] subjects. In every case, when all of the tones were identical in frequency, it was easier to detect the irregular rhythm in the streaming rhythm task (24-tone sequences) than in the short rhythm task (three-tone sequences), suggesting that the streaming rhythm task is a fundamentally easier task (irrespective of streaming considerations). There are a number of possible reasons to explain this result. First, the 24-tone sequences contain three repetitions of the irregular rhythm used in the short rhythm task, which may provide more chances for the subject to hear the irregular rhythm in the streaming rhythm task, making the task easier. Second, the 24-tone sequences contain a region of progressively increasing delay in the rhythm (Fig. 1) not present in the three-tone sequences. This may provide an additional cue to subjects taking the streaming rhythm task for detecting the irregular rhythm.

Figure 6 shows the performance of normal-hearing [Fig. 6(a)] and cochlear implant [Fig. 6(b)] subjects on the streaming rhythm task and the short rhythm task at a Weber fraction

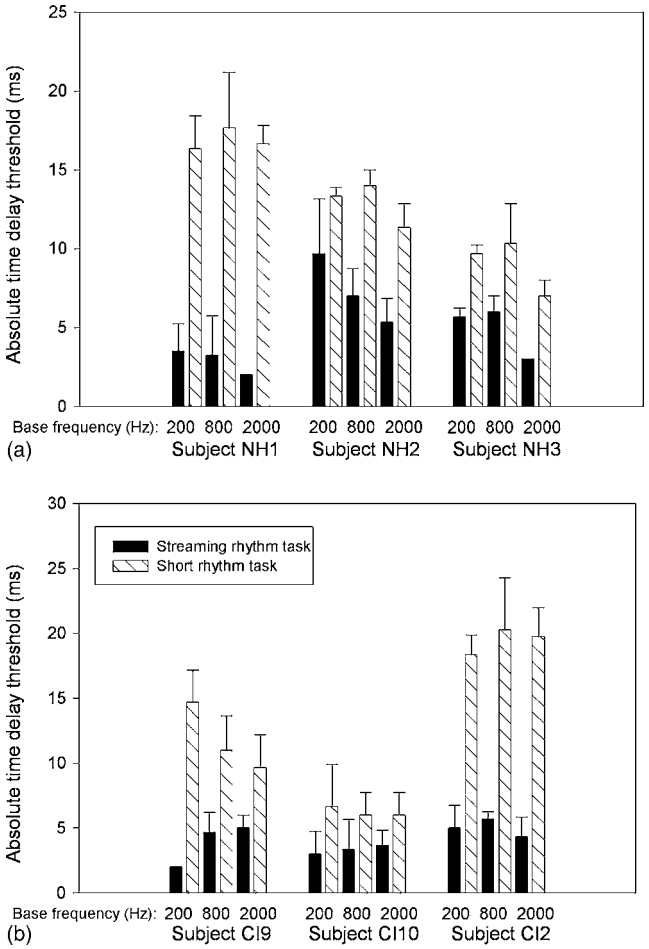


FIG. 5. Baseline performance on streaming rhythm task versus short rhythm task for normal-hearing and cochlear implant subjects. The baseline performance (tone A= tone B) on the streaming rhythm task (filled bars) and short rhythm task (hatched bars) is shown at all three base frequencies (200, 800, and 2000 Hz) for (a) three normal-hearing and (b) three cochlear implant subjects. Error bars represent one standard deviation.

of 0.5. The raw scores for time delay threshold have been divided by baseline performance thresholds (i.e., threshold at Weber fraction=0) for each test to give the resulting normalized thresholds shown on the ordinate in Fig. 6. This normalized threshold represents the performance on each task when the frequency difference of the alternating tones is at a Weber fraction of 0.5, after taking into account the baseline performance on each task shown in Fig. 5. The normalization procedure allows us to directly compare the relative performance on the streaming and short rhythm tasks at a Weber fraction of 0.5 to determine if auditory stream segregation plays a role in performance on the streaming rhythm task. The reason that normalized thresholds as opposed to absolute thresholds are compared between the two tasks is that the normalized threshold is the relevant measure used to assess streaming in the streaming rhythm task; our goal is to determine if a measure derived from the short rhythm task in a similar manner leads to similar results. If normalized performance on the two tasks is the same, this suggests that the difference in sequence length for the two tasks does not affect performance and that the streaming rhythm task does not measure stream segregation. In contrast, if normalized per-

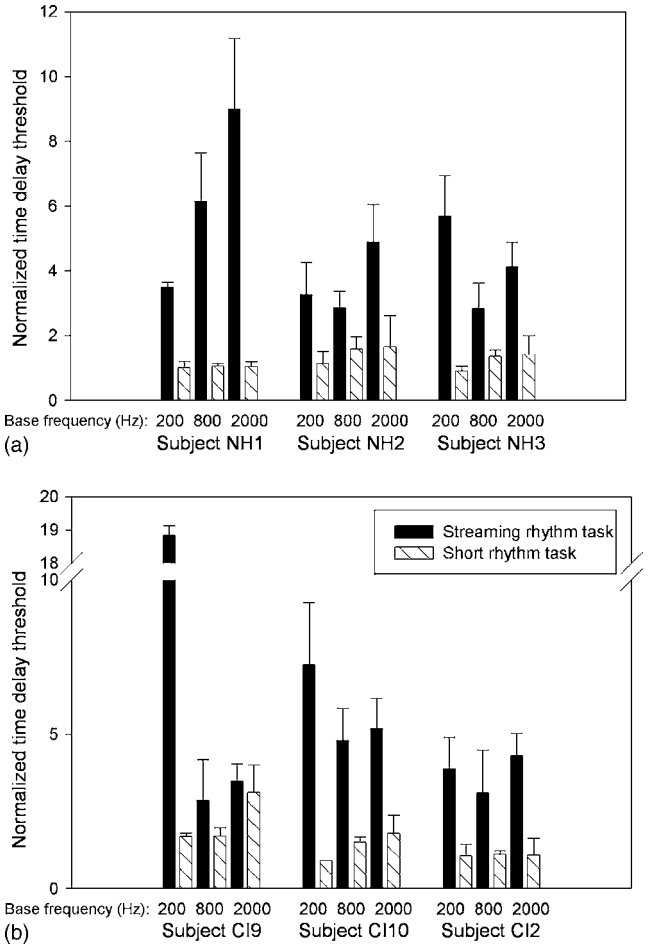


FIG. 6. Performance on streaming rhythm task versus short rhythm task for normal-hearing and cochlear implant subjects. Normalized time delay thresholds for the streaming rhythm (filled bars) and short rhythm (hatched bars) tasks are shown, with the alternating tones separated by a frequency difference corresponding to a Weber fraction of 0.5. Data are shown at all three base frequencies (200, 800, and 2000 Hz) for (a) three normal-hearing and (b) three cochlear implant subjects. Error bars represent one standard deviation.

formance is worse with the streaming rhythm task, this suggests that the longer sequence in the streaming rhythm task has induced a build-up of stream segregation to make it more difficult to detect the irregular rhythm in this task, and that the streaming rhythm task does measure stream segregation.

The values of the normalized threshold for the short rhythm task for both normal-hearing and cochlear implant subjects across all base frequencies generally fell in the range of 1 to 2, suggesting that increasing the frequency separation of the tones from baseline (no difference in frequency) to one corresponding to a Weber fraction of 0.5 resulted in either no change in difficulty of the task to a two times increase in difficulty of the task. For the streaming rhythm task, the normalized threshold values at a Weber fraction of 0.5 ranged from 2.83 to 9 for normal-hearing subjects and from 2.86 to 19 for cochlear implant subjects. This suggests that the streaming rhythm task was at least 2.8 times more difficult for all subjects in this experiment when the frequency separation was increased from baseline to one corresponding to a Weber fraction of 0.5.

To determine the contribution of streaming to this increased difficulty, we compared the normalized thresholds of the streaming rhythm task with those of the short rhythm task. Figure 6 shows that in every case, for both normal-hearing and cochlear implant subjects, the normalized threshold was significantly higher for the streaming rhythm task. For normal-hearing subjects, the mean of the threshold on the streaming rhythm task was 4.69, compared to 1.23 for the short rhythm task ($t_{df=8}=4.85$; $p=0.001$). For cochlear implant subjects, the mean of the threshold on the streaming rhythm task was 5.96, compared to 1.54 for the short rhythm task ($t_{df=8}=2.62$; $p=0.03$). If subjects were performing the streaming rhythm task by ignoring the first part of the 24-tone sequence and only focusing on the end of the sequence, such that streaming was irrelevant to the task, we would expect the normalized thresholds for the two tasks to be similar. This is not the case. Instead, the higher values of normalized threshold for the streaming rhythm task suggest a greater build-up of streaming with the longer sequence, making the streaming rhythm task relatively more difficult than the short rhythm task. Therefore, the results suggest that the normalized threshold in the streaming rhythm task reflects auditory stream segregation and not just gap discrimination abilities.

Our findings are consistent with those in the literature that suggest an increase in separation in pitch between the tones bordering a gap may result in a more difficult gap discrimination task for both normal-hearing (Neff *et al.*, 1982; Phillips *et al.*, 1997) and cochlear implant subjects (Hanekom and Shannon, 1997; Chatterjee *et al.*, 1998; van Wieringen and Wouters, 1999). Furthermore, the differences in performance we observed between the streaming rhythm task and short rhythm task are consistent with the results of Grose and Hall (1996), who showed that the build-up of auditory stream segregation through the use of longer sequences can make a gap discrimination task more difficult; in our experiment, the greater difficulty of the streaming rhythm task compared to the short rhythm task provides evidence that the streaming rhythm task measures auditory stream segregation. Further evidence for the streaming rhythm task as a measure of streaming has been provided by Roberts *et al.* (2002). This study compared the results of the streaming rhythm task with results from a different task that has been used to assess streaming and found similar results for both tasks, suggesting that the present task does measure auditory stream segregation.

In conclusion, experiment 2 provides evidence that the streaming rhythm task is a measure of auditory stream segregation and not merely gap discrimination. Having examined this, we will now move forward to experiment 3 and determine the implications of performance on our measure of stream segregation on the understanding of speech in background noise for cochlear implant users.

IV. EXPERIMENT 3. AUDITORY STREAMING AND SPEECH PERCEPTION IN NOISE

A. Rationale

A better ability to perform auditory stream segregation should be associated with a better ability to understand

speech in various backgrounds; to understand speech in noise, it is helpful to segregate the target speech into one attended perceptual group and the noise into a different, ignored group. Such a relationship between streaming and speech perception in noise has been found in hearing-impaired individuals (Mackersie *et al.*, 2001), but, to our knowledge, has yet to be studied in cochlear implant patients. We are interested in determining if a relationship exists between stream segregation and speech perception in noise in cochlear implant patients, because such a relationship would suggest that measures of stream segregation may be useful for comprehending the problems that implant users have in understanding speech in noise. Our hypothesis is that cochlear implant patients who have difficulty segregating pure tones into separate streams will have more trouble understanding speech in various backgrounds.

Additionally, we have chosen to study two different types of noise—steady-state noise and two-talker babble—to see if streaming ability is more strongly related to one or the other. For example, studies have shown that in normal-hearing subjects, the fundamental frequency of target speech is an important cue for segregating it from background talkers (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990). This leads us to hypothesize that the strongest correlations may occur for cochlear implant subjects between low-frequency pure-tone streaming ability and speech perception in babble, assuming fundamental frequency is also an important cue for talker segregation in cochlear implant users. Furthermore, by studying both types of noise, we can determine if the relative performance on speech perception in steady-state noise versus two-talker babble is related to streaming ability in cochlear implant users. Normal-hearing subjects perform better on tasks of speech perception in a competing talker background compared to steady-state noise (Duqeusnoy, 1983; Festen and Plomp, 1990). In contrast, cochlear implant subjects experience an increase in masking with competing talker(s) compared to steady-state noise (Qin and Oxenham, 2003; Turner *et al.*, 2004), which is the opposite of the trend seen for normal-hearing subjects. One reason that cochlear implant subjects may have more trouble understanding speech in a competing talker background is that they may not be able to segregate the target from background speech in a way that allows them to correctly identify the target speech. Thus, they are not able to take advantage of the temporal and spectral gaps in the competing talker noise to obtain additional glimpses of the target speech and experience the release from masking observed in normal-hearing listeners. This leads us to hypothesize that this inability of cochlear implant subjects to experience a release from masking may also be correlated to pure-tone streaming ability.

B. Participants

All cochlear implant subjects who participated in experiment 1 also participated in this experiment. All subjects were tested on speech perception in steady noise and in multi-talker babble. Eight additional cochlear implant subjects (ages 33–78) were recruited to undertake an abridged version of the pure-tone streaming task described in experiment 1 in order to increase the sample size for the correlations of streaming ability with speech perception in noise. All cochlear implant subjects had at least 1 year of experience with their device at the time of testing. The age, type of implant, signal processing strategy, and stimulation mode of each participating cochlear implant subject are shown in Table I.

C. Stimuli and procedures

1. Pure-tone streaming task (abridged version)

The full version of the streaming rhythm task used in experiment 1 allowed for the measurement of time delay thresholds at many different frequency differences; these data were subsequently fit with a regression line, with the slope interpreted as an overall measure of streaming ability. However, such testing was extremely time consuming, requiring about 6 h per subject. To shorten testing time in experiment 3, we opted to use an abridged version of the streaming rhythm task, where the normalized value of the time delay threshold at a Weber fraction of 0.5 was taken as the streaming metric. This normalized threshold represents how many times more difficult the streaming rhythm task was at a Weber fraction of 0.5 than at baseline (Weber fraction=0). The use of this metric allowed us to more efficiently increase the sample size for the correlation analysis performed in this experiment. The abridged version of the streaming rhythm task is identical to the full version described in experiment 1, except the only conditions tested were those where tones A and B differed by frequency differences corresponding to Weber fractions of 0 and 0.5. The frequency difference corresponding to a Weber fraction =0.5 was chosen to be tested because from experiment 1, it appeared that the results from cochlear implant subjects at this condition demonstrated the largest variation across subjects and exhibited minimal ceiling effects. Large variation across subjects in the rhythmic discrimination threshold was desirable because we wanted to subsequently correlate our measure of streaming ability with speech perception in noise. Furthermore, values for the regression slope streaming metric used in experiment 1 were strongly correlated with normalized thresholds at a Weber fraction of 0.5 in experiment 3 (200 Hz base frequency: $r=0.881$, $p<0.005$; 800 Hz base frequency: $r=0.830$, $p<0.025$; 2000 Hz base frequency: $r=0.906$, $p<0.0025$), suggesting consistency between the two measures of streaming in cochlear implant users.

Testing was once again performed at all three base frequencies (200, 800, and 2000 Hz) in a randomized order. At each base frequency, the first condition tested was that where the Weber fraction was 0, followed by the condition where the Weber fraction was 0.5. The exclusion criterion of an average raw score time delay threshold below 15 ms with the two tones equal in frequency (Weber fraction=0) was maintained for the abridged version of the test. Two subjects did not meet this criterion at one out of three base frequencies: both subjects CI11 and CI15 were unable to meet this criterion at the 2000-Hz base frequency condition.

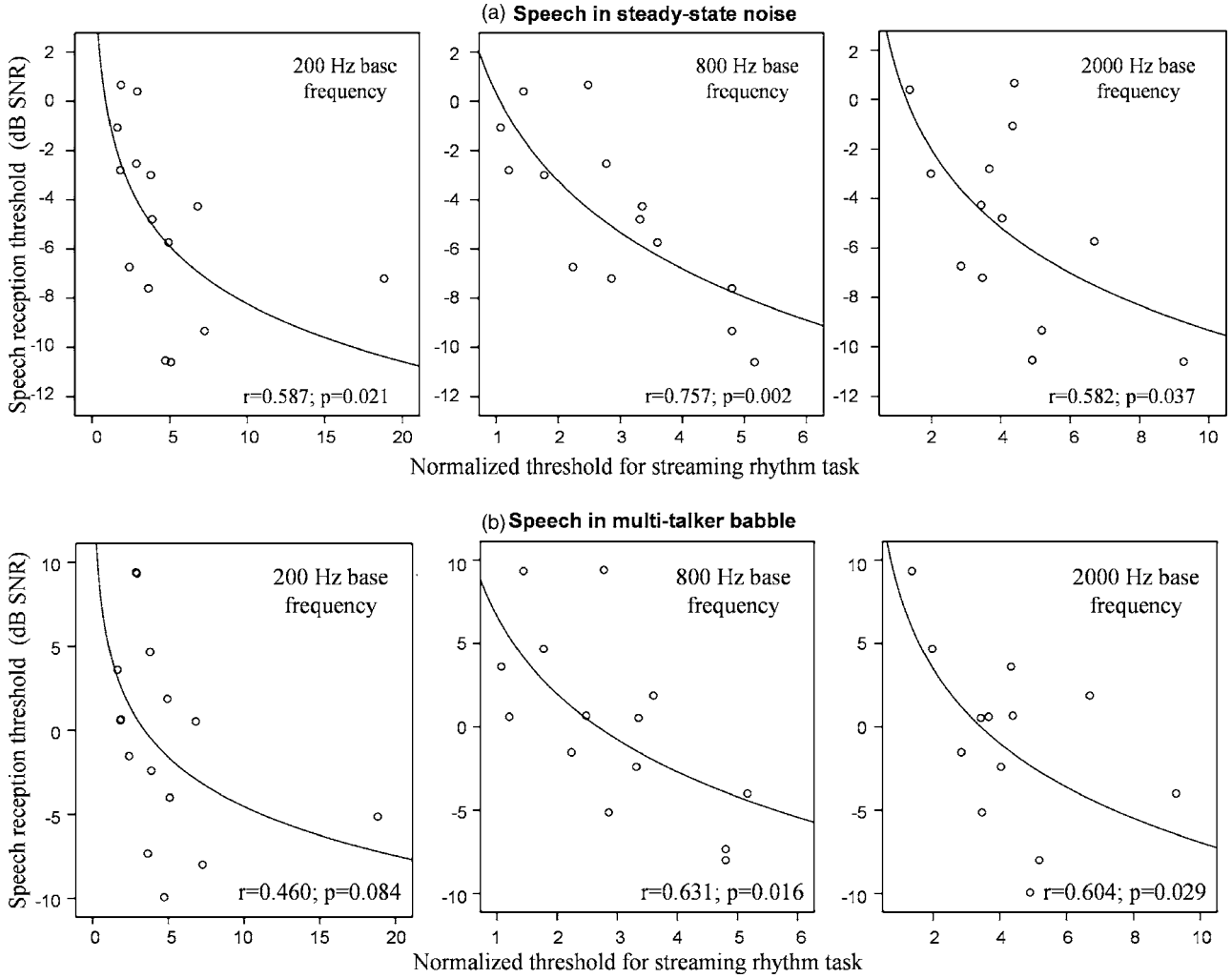


FIG. 7. Correlations between streaming rhythm task and speech in noise for cochlear implant subjects. Normalized time delay thresholds for the streaming rhythm task at a Weber fraction of 0.5 at each of three base frequencies (200, 800, and 2000 Hz) are plotted against performance on (a) speech in steady-state noise and (b) speech in two-talker babble.

2. Speech perception in steady-state noise and multi-talker babble

The tasks of speech perception in steady-state noise and multi-talker babble used in this study are the same as those previously described by Turner *et al.* (2004). The speech in these tasks was presented via a loudspeaker situated directly in front of the listener at 70 dB SPL. The most pertinent details of these tasks will be described in the following paragraph. For a more complete description of the tasks, the reader may refer to Turner *et al.* (2004).

Subjects were asked to identify a randomly chosen, previously recorded spondee from a fixed set of 12 homogeneously difficult spondees. The spondee was spoken by a female talker (fundamental frequency range: 212–250 Hz) in the presence of varying levels of background sound. Two different backgrounds were employed. In the first task, the background was a steady-state white noise, low-pass filtered at −12 dB per octave above 400 Hz to resemble the long-term spectrum of speech. In the second task, the background was two simultaneously presented sentences from the SPIN test (Bilger, 1984), with one sentence spoken by a male talker (fundamental frequency range: 81–106 Hz) and the other sentence spoken by a female talker (fundamental frequency range: 149–277 Hz). Within each task, the same noise sample was used for the background from trial to trial. An adaptive procedure with 2-dB step sizes (for the varying background level) was used to determine the 50%-correct point (in terms of signal-to-noise ratio) for the speech recognition threshold (SRT) for spondees in noise. Each run was composed of 14 reversals, with the mean of the last ten reversals taken as threshold for that run. Each subject completed four or five runs, with the mean of the last three runs taken as SRT for that subject.

D. Results and discussion

This experiment was designed to determine if a correlation exists between pure-tone streaming ability and speech perception in noise. The normalized time delay threshold (calculated in the same manner as for experiment 1; see Fig. 2) for the rhythmic discrimination streaming task was correlated using exponential fits with performance on speech in steady-state noise and speech in multi-talker babble for all cochlear implant subjects. The results are shown in Fig. 7 for

each of the three base frequencies of the rhythm streaming task. Figure 7(a) shows the correlations with speech in steady-state noise, and Fig. 7(b) shows the correlations with speech in multi-talker babble. In each case, there was a negative correlation between performance on the streaming rhythm task and speech in noise, suggesting that cochlear implant subjects who had larger time delay thresholds (and therefore better streaming abilities) performed better on speech recognition in noise. The correlations were of moderate strength, with r values ranging from 0.460 to 0.757, as shown in Fig. 7. All of the correlations were statistically significant ($p<0.05$), except for the correlation between the streaming rhythm task at 200-Hz base frequency and speech in multi-talker babble ($p=0.084$). Exponential fits were chosen to better account for the one data point at the 200-Hz base frequency on the streaming rhythm task that was much higher than all of the others, with a value close to 18 for the normalized threshold. However, all of the data were also fit with linear regression curves (not shown), which yielded similarly moderate correlations and statistically significant results. Additionally, statistical analysis performed to determine if the correlation coefficients were significantly different from each other revealed no significant differences, whether comparing r values within each speech perception in noise task at different base frequencies ($p>0.4$), or comparing r values between the two speech perception tasks at the same base frequency ($p>0.5$).

The cochlear implant subjects in this study perform on average 4.5 dB worse (standard deviation of 3.53 dB) on the speech perception in two-talker babble task than the speech perception in steady-state noise task. For comparison's sake, Turner *et al.* (2004) has reported for the same tasks that normal-hearing subjects perform about 13 dB better on speech perception in babble. Correlations performed between the difference in speech recognition thresholds (i.e., examining the lack of masking release) for the two tasks and the normalized time delay thresholds of the streaming rhythm task at each of three base frequencies revealed no significant correlations ($p>0.35$ for all correlations).

The results of the streaming rhythm task may underestimate the amount of stream segregation when the subject cannot hear the irregular rhythm even at the largest possible time delay. However, this problem occurred only in a few cases. Of the 42 unique values for threshold obtained at a Weber fraction=0.5 (across subjects and base frequencies), only four of them were assigned the nominal (ceiling) value of 40 ms for threshold. Thus, ceiling effects did not appear to be a major problem with our task at the frequency difference where the correlations with speech perception in noise were performed.

The correlation between pure-tone stream segregation and speech perception in noise found in cochlear implant subjects is consistent with findings in the literature that suggest that frequency resolution is an important determinant of the ability to perceive speech in noise (Fu *et al.*, 1998). Diminished frequency resolution likely leads to both poorer pure-tone auditory streaming and poorer speech perception in noise. However, it is likely that other common factors, beyond peripheral processes, are also involved in auditory stream segregation and speech perception in noise. For example, common central processes such as selective attention have been shown to influence auditory stream segregation (Carlyon *et al.*, 2001) and certainly also play a role in the ability to understand speech in noise. We propose that the normalized thresholds from the streaming rhythm task are a measure of the summation of relevant peripheral and central processing abilities within a subject. For example, a cochlear implant subject with poor peripheral frequency resolution and extremely good cognitive processing may perform equally well on pure-tone streaming as a different cochlear implant subject with better frequency resolution but poorer cognitive processing. The shared importance of these peripheral and central processes with speech perception in noise could thus explain why a correlation is observed between streaming and speech perception in noise in cochlear implant subjects.

The observation that the weakest correlation between streaming and speech perception in noise occurred between the 200-Hz base-frequency streaming condition and speech perception in two-talker babble was unexpected. We had hypothesized that we may find the strongest correlation here, because the 200-Hz region is where the fundamental frequencies of the different talkers are found, and the ability to segregate talkers based on fundamental frequencies is thought to be important for understanding speech in noise. However, there may be a number of reasons for the discrepancy between our hypothesis and our results. First, the ability of cochlear implant users to stream pure tones (as measured by the streaming rhythm task) may be different from their ability to stream more complex, speechlike stimuli. This is because cochlear implant users may have differing abilities to extract the fundamental frequencies from more complex tones. Thus, pure-tone streaming at a base frequency of 200 Hz may not be an accurate measure of streaming based on fundamental frequency. Second, a number of recent studies have suggested that current cochlear implant users who rely solely on electric hearing are deficient in or receive little to no benefit from fundamental frequency cues for understanding speech in a competing talker background (Qin and Oxenham, 2005; Kong *et al.*, 2005). This suggests that cues other than fundamental frequency may currently be more important for talker segregation by cochlear implant users. Thus, correlations between speech perception in noise and streaming ability across all frequency regions may be similar, because frequency cues in all regions may be used by cochlear implant users to segregate speech from noise. Finally, the correlations between speech perception and streaming at different base frequencies may be similar because they are correlated with some common underlying factor. For example, central processing abilities may be related to both auditory stream segregation and speech perception in noise.

We also hypothesized that the increase in masking experienced by cochlear implant subjects with speech perception in two-talker babble compared to steady-state noise may be correlated with streaming ability. One possible reason for the lack of such correlations in our results is that the ability to stream speech from different talkers may be different from the ability to correctly identify which speech belongs to the

target and which to the masker(s). In order to take advantage of the temporal and spectral gaps in the competing talker noise to understand speech, one must be able to do both. Thus, variability in the ability to correctly identify the target speech among subjects with similar streaming abilities may account for the lack of correlation between streaming and relative performance on the two speech in noise tasks.

V. CONCLUSIONS

(i) The results of this study suggest that the streaming rhythm task (adapted from Roberts *et al.*, 2002) is a test paradigm that can be used to assess auditory stream segregation in cochlear implant subjects. The task is sensitive to detection of a range of pure-tone streaming abilities that is present within the cochlear implant population.

(ii) There is a range of streaming abilities both within individual cochlear implant subjects (comparing different regions of the cochlea) and between cochlear implant subjects. Some cochlear implant subjects perform comparably to normal-hearing subjects, while others experience much less streaming than normal-hearing subjects as the frequency separation of two alternating tones is increased.

(iii) The variability in pure-tone streaming abilities across a wide range of frequencies among different cochlear implant users correlates moderately well with their ability to perceive speech in both steady-state noise and multi-talker babble.

(iv) For normal-hearing listeners, many cues other than frequency have been shown to be available for stream segregation, including those based on amplitude spectra (Iverson, 1995), temporal envelope (Grimault *et al.*, 2002), and spatial location (Dowling, 1973). It will be important in the future to determine which of these cues available to normal-hearing individuals are also available to cochlear implant users for auditory stream segregation, as well as the utility of these additional streaming cues for the perception of speech in complex acoustical backgrounds.

ACKNOWLEDGMENTS

This research was supported by Grant Nos. R01 DC000377, P50 DC00242, and 5 T32 DC000040-13 from the National Institutes of Health. The authors thank Anna Yao and Sheryl Erenberg for help in collecting the data, and Arik Wald for help with programming the experiments. We also appreciate the thoughtful comments and suggestions on this manuscript from John Grose, Fan-Gang Zeng, and one anonymous reviewer.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Acoust. Soc. Am. **11**, 257–271.

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.

Beauvois, M. W., and Meddis, R. (**1996**). "Computer simulation of auditory stream segregation in alternating-tone sequences," J. Acoust. Soc. Am. **99**, 2270–2280.

Bilger, R. C. (**1984**). "Speech recognition test development," ASHA Reports no. 14, pp. 2–15.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Brokx, J. P. L. and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). "Effects of attention and unilateral neglect on auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **27**, 115–127.

Chatterjee, M., Fu, Q-J., and Shannon, R. V. (**1998**). "Within-channel gap detection using dissimilar markers in cochlear implant listeners," J. Acoust. Soc. Am. **103**, 2515–2519.

Dorman, M. F., Smith, M., Smith, L., and Parkin, J. L. (**1996**). "Frequency discrimination and speech recognition by patients who use the Ineraid and continuous interleaved sampling cochlear-implant signal processors," J. Acoust. Soc. Am. **99**, 1174–1184.

Dowling, W. L. (**1973**). "The perception of interleaved melodies," Cognit Psychol. **5**, 322–337.

Dreschler, W. A., and Plomp, R. (**1985**). "Relations between psychophysical data and speech perception for hearing-impaired subjects. II," J. Acoust. Soc. Am. **78**, 1261–1270.

Duquesnoy, A. J. (**1983**). "Effect of a single interfering noise or speech source on speech intelligibility," J. Acoust. Soc. Am. **74**, 739–743.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Formby, C. and Forrest, T. G. (**1991**). "Detection of silent temporal gaps in sinusoidal markers," J. Acoust. Soc. Am. **89**, 830–837.

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. **88**, 1725–1736.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (**2002**). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," Cochlear Implants Int. **3**, 29–53.

Gfeller, K. E., Olszewski, C., Rychener, M., Sena, K., Knutson, J., Witt, S., and Macpherson, B. (**2005**). "Recognition of 'real-world' musical excerpts by cochlear implant recipients and normal-hearing adults," Ear Hear. **26**, 237–250.

Grimault, N., Bacon, S. P., and Micheyl, C. (**2002**). "Auditory stream segregation on the basis of amplitude-modulation rate," J. Acoust. Soc. Am. **111**, 1340–1348.

Grose, J. H., and Hall, J. W. (**1996**). "Perceptual organization of sequential stimuli in listeners with cochlear hearing loss," J. Speech Hear. Res. **39**, 1149–1158.

Hanekom, J., and Shannon, R. V. (**1997**). "Gap detection as a measure of electrode interaction in cochlear implants," J. Acoust. Soc. Am. **104**, 2372–2384.

Iverson, P. (**1995**). "Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes," J. Exp. Psychol. Hum. Percept. Perform. **21**, 751–763.

Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Mackersie, C. L., Prida, T. L., and Stiles, D. (**2001**). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **44**, 19–28.

Miller, G. A. and Heise, G. A. (**1950**). "The trill threshold," J. Acoust. Soc. Am. **22**, 637–638.

Neff, D. L., Jesteadt, W., and Brown, E. L. (**1982**). "The relation between gap discrimination and auditory stream segregation," Percept. Psychophys. **31**, 493–501.

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.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). "Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," J. Acoust. Soc. Am. **101**, 3694–3705.

Pichora-Fuller, M. K. (**2003**). "Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension," Int. J. Audiol. **42**, 559–567.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech perception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.

Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). "Primitive stream segregation of tone sequences without difference in fundamental frequency or passband," J. Acoust. Soc. Am. **112**, 2074–2085.

Rose, M. M., and Moore, B. C. J. (**1997**). "Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **102**, 1768–1778.

Rose, M. M., and Moore, B. C. J. (**2005**). "The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects." Hear. Res. **204**, 16–28.

Snell, K. B., and Frisina, D. R. (**2000**). "Relationships among age-related differences in gap detection and word recognition," J. Acoust. Soc. Am. **107**, 1615–1626.

Stickney, G. S., Zeng, F-G., Litovsky, R., and Assman, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Wesley, D. G. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.

Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). "Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners." J. Acoust. Soc. Am. **72**, 740–753.

van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Ph.D. dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.

van Wieringen, A., and Wouters, J. (**1999**). "Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees," J. Acoust. Soc. Am. **106**, 1925–1939.

Temporal onset-order discrimination through the tactual sense: Effects of frequency and site of stimulation

Hanfeng Yuan,[a] Charlotte M. Reed, and Nathaniel I. Durlach
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 1 July 2005; revised 21 April 2006; accepted 21 April 2006)

This research extends the study of temporal resolution of the tactual sensory system through measurements of temporal-onset order discrimination for continuous tonal signals addressing (a) the effects of frequency separation of the two stimuli whose onset orders are to be discriminated and (b) the effects of redundant coding of frequency and site of stimulation on performance. Sinusoidal signals were presented either at two separate digits (thumb and index finger of the left hand) or at a single site of stimulation (left index finger) using a multifinger tactual stimulation system. Measurements were obtained using a one-interval two-alternative forced choice procedure in which each interval consisted of the random-order presentation of two different stimuli with roving values of amplitude and duration. Thresholds were estimated from psychometric functions of d' as a function of stimulus-onset asynchrony (SOA). On average, temporal onset-order thresholds were larger for one-finger conditions (mean SOA of 74.8 ms) than for two-finger conditions (mean SOA of 48.5 ms) and decreased as frequency separation increased, particularly for single-site presentation. Redundant coding of frequency and site of stimulation resulted in higher resolution by a factor of 1.5 compared to frequency alone and by a factor of 1.2 compared to site alone.
 [DOI: 10.1121/1.2204452]

PACS number(s): 43.66.Wv [RAL] Pages: 375–385

I. INTRODUCTION

The current research is concerned with the temporal-resolution ability of the human tactual sensory system and extends previous work of Yuan *et al.* (2005a) investigating tactual temporal onset-order resolution. This research is related to the development of optimal encoding schemes for the tactual display of temporal properties of acoustic signals in communication aids for persons with profound hearing impairment (see Yuan, 2003; Yuan *et al.*, 2004, 2005b).

In the tactual modality, much of the previous research has been concerned with the ability to discriminate the *temporal order* of two successive transient mechanical or electrotactile signals (Hirsh and Sherrick, 1961; Sherrick, 1970; Marks *et al.*, 1982; Shore *et al.*, 2002) or brief vibrotactile patterns (Craig and Baihua, 1990; Craig and Busey, 2003). Because of the transient nature of these signals, there is no overlap between signals, and judgments of temporal order can be based on both onset and offset cues. Temporal-order thresholds obtained with such signals are generally on the order of 20–40 ms, with the exception of much larger values on the order of hundreds of ms reported by Marks *et al.* (1982) for electrocutaneous signals. More recently, research has been concerned with *temporal onset-order* measurements through the tactual sense using continuous rather than transient signals (Eberhardt *et al.*, 1994; Yuan *et al.*, 2005a). Measurements with such signals typically involve temporal overlap of the two stimuli whose onset is to be discriminated, and techniques may be introduced to eliminate the use of confounding offset cues (Yuan *et al.*, 2005a).

Eberhardt *et al.* (1994) presented preliminary results of measurements concerned with the ability to determine the temporal-onset order of a "movement" and a "vibratory" signal presented through a haptic display at the index finger where thresholds ranged from roughly 40 to 93 ms across subjects. Yuan *et al.* (2005a) conducted a psychophysical study of tactual temporal-onset order thresholds for the following two signals: a 250-Hz sinewave at the index finger and a 50-Hz sinewave at the thumb. These signals were selected on the basis of a tactual display designed for the presentation of a temporal cue to voicing (Yuan, 2003; Yuan *et al.*, 2004, 2005b). The amplitude and duration of each of these sinewaves were roved independently from trial to trial in a one-interval two-alternative forced choice (1I2AFC) procedure (to approximate the variations in these parameters encountered in real speech signals). Performance, measured as a function of stimulus-onset asynchrony (SOA), averaged roughly 34 ms at threshold (defined as the $|\mathrm{SOA}|$ required for $d' = 1.0$). The current research was undertaken to extend the psychophysical results of Yuan *et al.* (2005a) to include study of the effects of frequency separation and site of stimulation, as well as the effects of redundancy in the coding of these two parameters, on temporal onset-order resolution. Such effects have not been studied systematically for onset-order discrimination of continuous tonal signals and are highly relevant to issues regarding the coding of tactual displays of speech.

Previous psychophysical and neurophysiological research on the tactual sensory system indicates that this system is composed of at least four separate information-processing channels, each with specific receptors and peripheral nerve fibers, associated psychophysical character-

[a]Electronic mail:hfyuan@mit.edu

istics, and distinct perceptual qualities (e.g., see Bolanowski *et al.*, 1988; Gescheider *et al.*, 2002). Three channels contribute primarily to the frequency-response characteristic of the tactual sensory system. The P channel, associated with Pacinian afferents, is the most sensitive one at high frequencies of vibration (60-500 Hz) (Verrillo, 1966a,b; Mountcastle *et al.*, 1972; Verrillo and Gescheider, 1977; Johansson, 1978; Johansson *et al.*, 1982; Bolanowski and Verrillo, 1982; Bolanowski *et al.*, 1988). The perceptual quality associated with these high-frequency low-amplitude (on the order of tens of microns) stimuli is that of a relatively focused vibration. The NPI channel, corresponding to the rapidly adapting (RA) afferent fibers associated with Meissner corpuscles, is primarily responsible for detection of midfrequency vibrations between approximately 10 and 60 Hz (Talbot *et al.*, 1968; Mountcastle *et al.*, 1972; Gescheider *et al.*, 1985; Verrillo and Bolanowski, 1986; and Bolanowski *et al.*, 1988). Stimuli within this region of intermediate frequencies and amplitudes give rise to a qualitative sensation of fluttering. The NPIII channel, associated with the slowly adapting type I (SAI) afferent fibers that innervate Merkel-cell neurite complexes, is the most sensitive at frequencies below roughly 10 Hz; its threshold is relatively independent of frequency (Johansson, 1978; Verrillo, 1979; Bolanowski *et al.*, 1988). The perceptual quality associated with these low-frequency high-amplitude signals (on the order of several mm) is that of slow movements. The NPII channel, associated with the slowly adapting type II (SAII) afferent fibers that innervate Ruffini endings, contributes to the perception of pressure. Its role in vibrotactile processing is less clear than that of the other three channels, but is believed to contribute to the suprathreshold reception of tactual sensations over a wide frequency range (Bolanowski *et al.*, 1993; Gescheider *et al.*, 2002).

For many purposes, the tactual sensory system may be regarded as a continuum from the kinesthetic information provided by high-amplitude low-frequency movements to the cutaneous information provided by low-amplitude high-frequency vibrations. The experiments reported here examined the effects of stimulating frequency within and across the different information-processing channels of the tactual sensory system on the ability to discriminate temporal onset order. The effects of frequency and site of stimulation (and their redundancy) on the ability to make judgments regarding the temporal onset-order of stimulation have not been studied in a systematic manner. These stimulus properties are typically employed in the design of tactual and haptic displays of acoustic or visual information. A more thorough understanding of the effects examined here has important applications to the design of coding systems for the display of information through the sense of touch.

The experiments employed pairs of identical or different frequencies from within each of the three major channels of the tactual sensory system as well as pairs of frequencies from across each of the three channels. The effects of frequency separation were examined at two sites of stimulation (Experiment 1) as well as at one site of stimulation (Experiment 2). The experiments also explored the effects of redundant versus nonredundant coding of site and frequency of stimulation. Redundancy effects were examined in conditions comparing the presentation of the same frequency versus different frequencies in the two-finger condition and conditions comparing the presentation of a given frequency pair at one site versus two sites. In addition, redundancy effects were examined in two conditions in which frequency was independent of site of stimulation but differed as to whether subjects were instructed to respond to temporal-onset order on the basis of frequency (Experiment 3) or site of stimulation (Experiment 4).

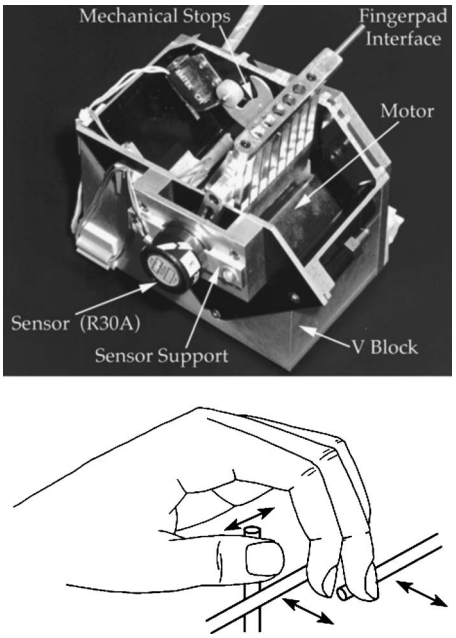


FIG. 1. Illustration of the Tactuator stimulating device. The upper panel is a photograph of one of the three motor assemblies with labeled components. The lower panel is a schematic drawing illustrating finger placement on three vibrating rods of the Tactuator device (right panel). (Taken from Tan, 1996).

II. METHOD

A. Subjects

Four normal-hearing individuals ranging in age from 18 to 40 years (two females) served as subjects in this study. All subjects were screened for tactual threshold detection before participating in the experiments. Audiological testing was conducted to provide a baseline for each subject's hearing level prior to exposure to auditory masking noise and again at the completion of the study. Subjects were paid for their participation in the study. None of the subjects reported any known tactual impairments of the hand. One of the subjects (Subj. 1) participated in a previous study of tactual temporal onset-order (Yuan *et al.*, 2005a) and tactual speech reception (Yuan *et al.*, 2005b). The remaining subjects had no previous experience in experiments with tactual stimulation.

B. Apparatus

The tactual stimulating device used in the experiments (referred to as the Tactuator—see Fig. 1) was initially developed by Tan (1996) for research with multidimensional tactual stimulation. A complete discussion of this system is pro-

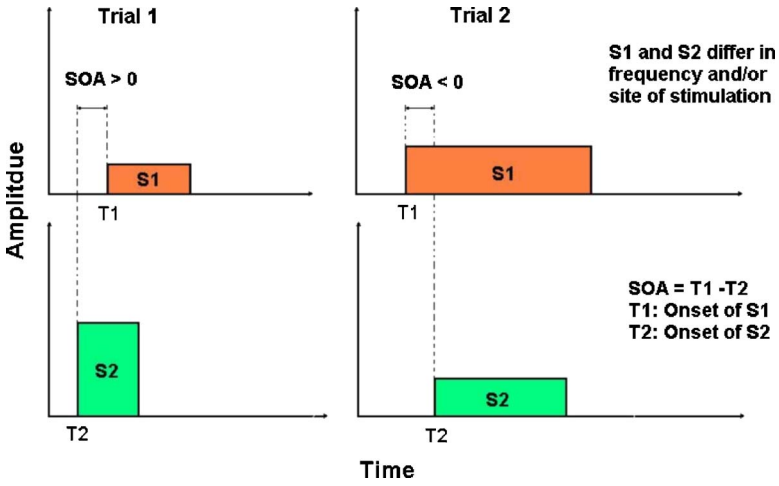


FIG. 2. (Color online) Illustration of the temporal sequence of events for two representative trials in the temporal-onset order discrimination experiment. The upper panel shows S1, and the lower panel shows S2.

vided in Tan and Rabinowitz (1996), which includes a detailed description of the hardware components, controller components, and performance characteristics of the device. For the current research, the original system was upgraded with a new computer, digital signal-processing system, and electronic control system to improve its performance capabilities. A complete description of the characteristics of the modified system is available in Yuan (2003).

The device is a three-finger display capable of presenting a broad range of tactual movement to the human fingers. It consists of three mutually perpendicular rods that interface with the thumb, index finger, and middle finger in a manner that allows for a natural hand configuration (see bottom panel of Fig. 1). A photograph of the motor assembly (with labeled components) associated with one of the rods is provided in the upper panel of Fig. 1. Each rod is driven independently by an actuator that is a head-positioning motor from a hard-disk drive. The position of the rod is controlled by an external voltage source to the actuator and is measured by an angular position sensor that is attached to the moving part of each of the three motor assemblies. The rods are capable of moving the fingers, which rest lightly on the rods, in an outward (extension) and inward (flexion) direction relative to a neutral resting position. The subject is instructed to actively follow the movements of the device by maintaining contact with the rod.

The design of the device was guided largely by the desire to incorporate characteristics of successful "natural" methods of tactual communication used by deaf-blind persons (such as Tadoma) into an artificial display of speech. In Tadoma, the deaf-blind user places his or her thumb on the lips of the talker and fans the other four fingers across the talker's face and neck. By monitoring the movements of the lips and jaw (predominantly kinesthetic cues) and vibrations on the neck (predominantly tactile cues), well-trained deaf-blind individuals can achieve good comprehension of speech presented at slow-to-normal rates (Reed *et al.*, 1985). As Rinker *et al.* (1998) point out (in a study employing a similar type of stimulating device), this type of stimulation is intermediate between (1) the active, unconstrained movements of the fingers involved in natural tactual exploration and (2) the passive delivery of stimuli to a resting body site. Such an intermediate type of design has the advantage of measuring performance in a practical display that permits a natural type of finger movement on the part of the subject, but carries a corresponding limitation in the precision with which the stimulation can be controlled and specified.

The overall performance of the device is nonetheless well suited for psychophysical studies of the tactual sensory system. First, the device is capable of delivering frequencies along a continuum from dc to 300 Hz, allowing for stimulation in the kinesthetic (low-frequency) and cutaneous (high-frequency) regions, as well as in the midfrequency range. Second, the range of motion provided by the display for each digit is roughly 26 mm. This range allows delivery of stimulation at levels from threshold to approximately 50 dB sensation level (SL) throughout the frequency range from dc to 300 Hz. Third, each channel is highly linear, with low harmonic distortion and negligible interchannel crosstalk. Fourth, loading (resulting from resting a finger lightly on a moving bar of the actuator) does not have a significant effect on the magnitude of the stimulation.

C. Experimental Paradigm

Temporal-onset order discrimination thresholds were measured using a 1I2AFC procedure with trial-by-trial correct-answer feedback. The temporal sequence of events for two representative trials in the 1I2AFC procedure is illustrated in Fig. 2. One interval (or trial) of the experiment involved the presentation of two stimuli (S1 and S2) in one of two possible onset orders (with equal *a priori* probabilities): onset of S1 followed by onset of S2 (S1, S2) or onset of S2 followed by onset of S1 (S2, S1). The subject's task was to determine, on each trial, which of the two possible onset orders [(S1, S2) or (S2, S1)] was presented, and performance was examined as a function of stimulus-onset asynchrony SOA. The upper panel of Fig. 2 shows S1, and the lower panel shows S2. Sequential events are shown for two full trials (Trial 1 and Trial 2). The onsets of S1 and S2 of the two trials are marked by T1 and T2, respectively. (The values of the duration and amplitude of each of the two signals are selected randomly and independently from the distributions described below.) Note (as in Trial 2) that it is possible for the signal with an earlier onset to terminate after the second stimulus.

SOA was defined as the time asynchrony between the onset of S1 relative to the onset of S2, i.e., onset time of Signal 1 minus onset time of Signal 2 ($S1_{onset} - S2_{onset}$). The absolute value of SOA was kept constant throughout a given run so that, in effect, the only aspect of the stimulus that varied during the run and needed to be judged by the subject on a trial-to-trial basis was the sign of the SOA. In Fig. 2, Trial 1 represents a positive value of SOA, where the onset time for S2 leads the onset time for S1 ($T1-T2>0$). Trial 2 illustrates a negative value of SOA, where the onset time for S1 leads the onset time for S2 ($T1-T2<0$).

During the end of one trial to the beginning of the next trial, the following series of events occurred: A prompt was presented for the subject to choose one of the two response codes, the response was stored digitally, the text providing correct-answer feedback was displayed on the monitor, the parameters were computed for the next trial, and finally a prompt was presented for a new trial. The duration of this time period varied, depending primarily on the subject's response time.

The experiment was conducted with trial-by-trial independent roving of both the duration and amplitude of each of the two sinewaves. Signals were gated on and off with a rise-fall time of 20 ms shaped with a Hanning window. The duration of the two stimuli in each trial was selected independently from a uniform distribution of the following seven values: 200, 250, 300, 350, 400, 450, 500 ms (leading to $7 \times 7 = 49$ possible duration pairs). Thus, the offset order of the two stimuli on each trial is random, and cannot be used as a cue in the temporal onset-order task. Roving amplitude was also employed to eliminate the potential use of a confounding cue based on possible differences in the perceived magnitude of the two stimuli if they were fixed at a given value of SL. The value of amplitude of the two stimuli in each trial was selected from a uniform distribution of six values of sensation level: 30, 33, 36, 39, 42, and 45 dB SL relative to average thresholds obtained previously with the Tactuator device (Yuan, 2003). These stimulus levels are within the range of comfortable levels of tactual stimulation described by Verrillo *et al.* (1969) and were selected to permit the distinct perception of both stimuli within any given trial.

Threshold measurements were obtained using a two-interval two-alternative forced choice procedure at frequencies of 2, 10, 25, 50, 100, 200, 250, and 300 Hz at the index finger and thumb of four subjects (Yuan, 2003). Thresholds at other frequencies employed in the current study but not in the threshold testing (3, 7, 55, 155, and 315 Hz) were interpolated from curves fit to the data points. Threshold values for the eight frequencies examined in the current study are provided in Table I. The stimulation levels of the signals employed in the experiments are described in terms of dB SL relative to the thresholds reported in Table I.

TABLE I. Mean of the absolute threshold values in dB re 1.0 μm peak at the eight frequencies for the thumb and index finger across four subjects in a previous study (Yuan, 2003), through the Tactuator device.

Frequency (Hz)	3	7	25	50	55	155	250	315
Index	32	25	17	5	3	−15	−22	−19
Thumb	37	30	20	0	2	−20	−25	−25
Mean	34	28	19	3	3	−18	−24	−22

D. Experimental conditions

The experimental conditions employed in the measurement of temporal onset-order discrimination thresholds are summarized in Table II. In Table II, S1 and S2 are described in terms of frequency (Ω) and site of stimulation (F). S1 is notated as (Ω_1, F_1) and S2 as (Ω_2, F_2). Frequency of stimulation Ω could take on one of eight values: two frequencies from the low-frequency region (3 and 7 Hz), three frequencies from the intermediate region (25, 50, and 55 Hz), and three frequencies from the high-frequency region (155, 250, and 315 Hz). The site of stimulation F could assume one of two values: T for thumb and I for index finger.

The conditions shown in Table II are grouped into four different experiments.

In Experiment 1, Ω_1, was always delivered to the thumb, and Ω_2 was always delivered to the index finger. Ten different pairs of Ω_1, Ω_2 were tested: three pairs where $\Omega_1 = \Omega_2$ (SF=same frequency), three pairs where $\Omega_1 \neq \Omega_2$ and both values are from within the same frequency region (WC =within channel), and four pairs where $\Omega_1 \neq \Omega_2$ and both values are taken from different frequency regions (CC=cross channel). Subjects were instructed to respond by typing "T" if the stimulus order was (S1, S2) or "I" if the stimulus order was (S2, S1).

In Experiment 2, only one site of stimulation was employed: the index finger. Three pairs of WC values of Ω_1 and Ω_2 were employed and four pairs of CC values of Ω_1 and Ω_2. Subjects were instructed to respond by typing "L" if the stimulus order was (S1, S2) or "H" if the stimulus order was (S2, S1).

In the four conditions of Experiments 3 and 4, S1 and S2 each could assume one of two different values. In Experiment 3, the subject's task was to determine which of the two *stimulating frequencies* had an earlier onset. The subjects were instructed to press "L" if they perceived the order to be (Ω_1, Ω_2) or "H" if they perceived the order to be (Ω_2, Ω_1). In Experiment 4, the subject's task was to determine which of the two *sites of stimulation* had an earlier onset. The subjects were instructed to press "L" if they perceived the order to be (Ω_1, Ω_2) or "H" if they perceived the order to be (Ω_2, Ω_1). In Experiment 4, the subject's task was to determine which of the two *sites of stimulation* had an earlier onset. The subjects were instructed to press T if they perceived that the stimuli were ordered as (F_1, F_2) or I if they perceived the ordering to be (F_2, F_1).

Experiment 1 was designed to examine the effects of frequency separation on temporal onset-order discrimination for the case where S1 and S2 were presented at different sites of stimulation (i.e., S1 at the thumb and S2 at the index finger). Experiment 2 examined the effects of frequency separation for presentation at a single site of stimulation (index finger). Experiment 3 measured the ability to make temporal onset-order decisions on the basis of frequency in the presence of the irrelevant property of site of stimulation. Ex-

TABLE II. Description of the stimulus pairs (S1 and S2) employed in the temporal onset-order discrimination experiments. The stimuli are defined in terms of frequency (Ω) and finger of stimulation (F). Ω is given in Hz; F can take on one of two values—thumb (T) or index finger (I). S1 is notated as (Ω_1,F_1) and S2 is notated as (Ω_2,F_2). The subject's task was always to determine the order in which the two stimuli were presented, i.e., [S1, S2] or [S2, S1]. The frequency comparison Ω_1 vs Ω_2 for each condition is labeled as one of three types: SF (same frequency), WC (within channel), or CC (cross channel).

Exp.	Ω_1 vs Ω_2	Ω_1[a]	F_1	Ω_2[a]	F_2
1	SF	7	T	7	I
1	SF	55	T	55	I
1	SF	315	T	315	I
1	WC	3	T	7	I
1	WC	25	T	55	I
1	WC	155	T	315	I
1	CC	7	T	55	I
1	CC	55	T	315	I
1	CC	7	T	315	I
1	CC	50	T	250	I
2	WC	3	I	7	I
2	WC	25	I	55	I
2	WC	155	I	315	I
2	CC	7	I	55	I
2	CC	55	I	315	I
2	CC	7	I	315	I
2	CC	50	I	250	I
3[b]	WC	25	I or T	55	T or I
3[b]	CC	7	I or T	55	T or I
3[b]	CC	50	I or T	250	T or I
3[b]	CC	7	I or T	315	T or I
4[c]	WC	25 or 55	T	55 or 25	I
4[c]	CC	7 or 55	T	55 or 7	I
4[c]	CC	50 or 250	T	250 or 50	I
4[c]	CC	7 or 315	T	315 or 7	I

[a]The amplitude and duration of the stimulating frequency Ω_1 and Ω_2 were selected at random (with equal *a priori* probability) on each trial from a distribution of six values of sensation level in the range of 30–45 dB SPL and seven values of duration in the range of 200–500 ms.

[b]In Experiment 3, S1 uses I if and only if S2 uses T; and S1 uses T if and only if S2 uses I. The assignment of I to S1 or S2 is random on each trial with equal *a priori* probabilities.

[c]In Experiment 4, S1 uses the lower frequency value if and only if S2 uses the higher frequency value; and S1 uses the higher frequency value if and only if S2 uses the lower frequency value. The assignment of the low-frequency value to S1 is random on each trial with equal *a priori* probabilities.

periment 4 measured the ability to make such decisions on the basis of site of stimulation in the presence of the irrelevant property of stimulating frequency.

E. Procedure

For each subject, performance was measured at two different values of |SOA| for each of the stimulus pairs (S1, S2) described in the experimental conditions of Table II. As stated previously, for any given value of |SOA|, the subject's task was to determine the sign of the SOA. The particular |SOA| values used for each subject were selected to yield performance in the range of roughly 55–90% correct. Five experimental blocks were run using the two values of |SOA| selected for each subject. Within each block, one 50-trial run was conducted at each of the two values of |SOA| with the order of the two values of |SOA| chosen at random. The subjects were instructed to guess when they were unsure of the order of the stimuli.

A computer program written in C controlled the generation of the stimuli, the presentation of the stimuli to the Tactuator device, and the recording of subject's responses on a DELL computer. During the experiment, the subject sat approximately 0.8 m from the video monitor (DELL Trinitron) and 0.6 m from the Tactuator, and placed the index finger and the thumb of the left hand on the two corresponding rods of the Tactuator. To eliminate any auditory cues from the vibration of the Tactuator, subjects wore foam earplugs that were designed to provide 30-dB attenuation and also wore earphones that delivered pink masking noise at an overall level of roughly 90 dB sound pressure level. The attenuation and the noise were sufficient to mask any air-conducted sounds arising from the device itself. In addition, the masking noise was deemed sufficient for eliminating any bone-conducted sounds that might arise from the highest levels of stimulation.

F. Data analysis

Results for each subject and each experimental run were summarized in terms of a 2×2 stimulus-response confusion matrix. Signal-detection measures of sensitivity (d') and bias (β) (Green and Swets, 1966; Durlach, 1968) were computed for each matrix assuming equal-variance Gaussian distributions. Data were summarized by psychometric plots of d' versus |SOA|. Threshold is estimated as the value of |SOA| corresponding to $d'=1$.

III. RESULTS

A. Experiment 1: Two-finger stimulation, frequency fixed at each finger

Thresholds of the temporal onset-order discrimination task (defined as the value of |SOA| at which $d'=1$) are shown in Table III for each subject and each of the ten frequency pairs. For each subject, thresholds were similar across frequency pairings. Standard deviations across the frequency pairings ranged from 4.7 ms (Subj. 1) to 9.4 ms (Subj. 2). Average threshold values, however, varied substantially across subjects, ranging from 23.1 ms (Subj. 1) to 69.9 ms (Subj. 2).

Results are further grouped into three types of frequency pairs: SF, WC, and CC frequency pairs (see Table II, Experiment 1). Mean and standard deviation of performance on the pairs within each group are shown for each of the four individual subjects in the top panel of Fig. 3. For each subject, thresholds were similar within each of the three groups of frequency pairings. Standard deviations were small within each of the groups, and never exceeded 10 ms for any sub-

TABLE III. Temporal onset-order thresholds (in ms) for each subject and each frequency pair in Experiment 1.

	SF			WC			CC					
	(7,7)	(55,55)	(315,315)	(3,7)	(25,55)	(155,315)	(50,250)	(55,315)	(7,55)	(7,315)	Mean	S.D.
Subj. 1	22.16	26.81	29.98	15.36	18.30	18.04	25.06	25.54	23.24	26.94	23.14	4.66
Subj. 2	84.24	67.12	78.86	66.29	71.04	75.63	65.30	73.60	67.26	49.66	69.90	9.38
Subj. 3	54.66	63.01	58.47	50.60	50.39	57.40	55.40	53.09	42.61	46.37	53.20	5.98
Subj. 4	50.94	53.71	57.00	52.34	45.61	45.23	41.99	38.05	48.48	45.27	47.86	5.75
Mean	53.00	52.66	56.08	46.15	46.34	49.08	46.94	47.57	45.40	42.06	48.53	
S.D.	25.39	18.12	20.06	21.69	21.70	24.17	17.43	20.69	18.13	10.25		18.15

ject or frequency grouping. Thresholds averaged across subjects were 53.9 ms for SF pairs, 47.2 ms for WC pairs, and 47.8 ms for CC pairs. A two-way ANOVA was performed on the threshold values using type of frequency pair and subject as the two main factors. The effects of subject and frequency grouping on performance were both significant with $[F(3,34)=116.88;p<0.001]$ and $[(F(2,34)=8.07;p<0.001]$, respectively. A posthoc Scheffe analysis was also conducted to examine the significance of differences within pairs of means on each of the main factors. For the factor of subject, the performance of Subj. 1 and Subj. 4 was significantly different from each other ($\alpha=0.05$) and from the performance of Subj. 2 and Subj. 3 ($\alpha=0.05$), who did not differ significantly. For the factor of frequency grouping, performance under SF grouping was significantly ($\alpha=0.05$) different from the performance under the other two frequency groupings (WC and CC groupings).

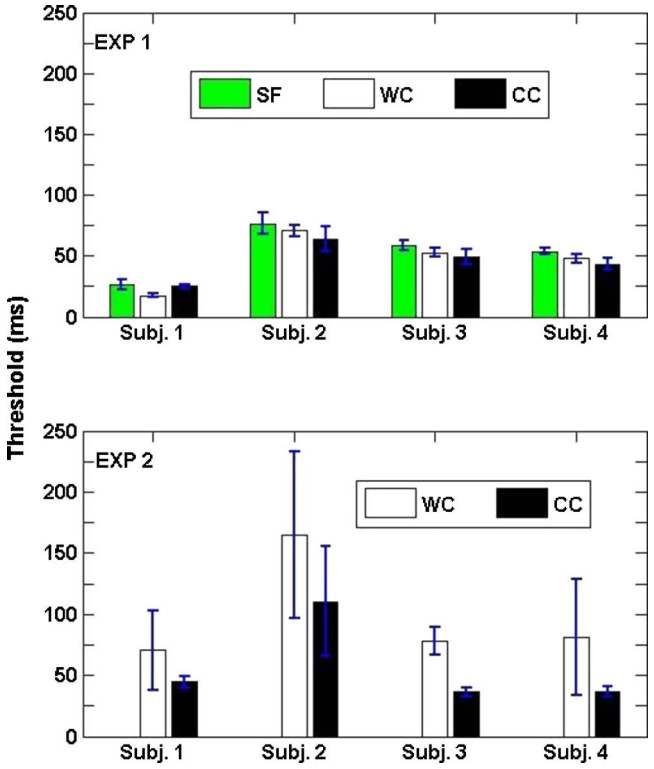


FIG. 3. (Color online) Summary of results for Experiments 1 and 2. In the top panel, results of Experiment 1 are grouped into three types of frequency pairs: SF pairs, WC frequency pairs, and CC frequency pairs. Mean and ±1 S.D. of performance within each group are shown for each of the four individual subjects. In the bottom panel, results of Experiment 2 are grouped into two types of frequency pairs: WC and CC frequency pairs. Mean and ±1 S.D. of performance within each group are shown for each of the four individual subjects.

B. Experiment 2: One-finger stimulation

Thresholds for each subject and each frequency pair are shown in Table IV. Mean threshold values across the frequency pairings were similar for Subj. 1, Subj. 3, and Subj. 4 (54.2 to 55.6 ms) compared to the much larger mean threshold of 133.6 ms for Subj. 2. Standard deviations were larger across the frequency pairings than in Experiment 1 and ranged from 23.4 ms (Subj. 3) to 58.3 ms (Subj. 2).

Results are further grouped into two types of frequency pairs: WC and CC frequency pairs. Mean and standard deviation of performance within each group are shown for each of the four individual subjects in the bottom panel of Fig. 3. For each subject, effects of frequency separation are obvious. Thresholds for the three WC pairs (averaging 98.5 ms across subjects) are larger than for the four CC frequency pairs (averaging 56.9 ms). The ratio of WC/CC threshold values is in the range of 1.5 to 2.2 across the four subjects. A two-way ANOVA was performed on the threshold values using type of frequency pair and subject as the two main factors. The results of the ANOVA indicate that both factors are significant $[F(1,23)=11.83;p<0.002;F(3,23)=10.75;p<0.0001]$. A posthoc Scheffe analysis on the Subject effect indicated that the performance of Subj. 2 was significantly different ($\alpha=0.05$) from that of the other subjects.

C. Experiment 3: Two-finger stimulation, frequency not tied to finger, discriminate on basis of frequency

Thresholds for each subject and each of the four frequency pairs are provided in Table V. Again, mean threshold values were similar for Subj. 1, Subj. 3, and Subj. 4 (ranging from 41.1 to 49.3 ms) and substantially lower than the 146.4 ms threshold obtained by Subj. 2. Standard deviations across the four frequency pairings ranged from 1.7 ms (Subj. 4) to 81.0 ms (Subj. 2). Results are further grouped into two types of frequency pairs: one WC pair (25,55) and three CC frequency pairs [(50,250), (7,55), and (7,315)]. Performance within each group is shown for the four individual subjects in the upper panel of Fig. 4. For each subject (except Subj.

TABLE IV. Temporal onset-order thresholds (in ms) for each subject and each frequency pair in Experiment 2.

	WC			CC				
	(3,7)	(25,55)	(155,315)	(50,250)	(55,315)	(7,55)	(7,315)	Mean	S.D.
Subj. 1	106.98	59.69	44.47	39.07	49.28	47.89	41.90	55.61	23.60
Subj. 2	243.32	125.40	125.58	119.95	167.96	90.26	63.02	133.64	58.31
Subj. 3	73.17	90.83	70.37	40.98	34.17	32.76	36.96	54.18	23.44
Subj. 4	135.57	49.24	57.87	37.11	35.36	42.21	32.06	55.63	36.35
Mean	139.76	81.29	74.57	59.28	71.69	53.28	43.49	74.77	
S.D.	73.60	34.30	35.61	40.48	64.54	25.43	13.63		49.93

4), threshold for the WC pair (25,55) is larger than thresholds for the three CC pairs. Mean threshold values across the four subjects averaged 111.1 ms for the WC pair and 57.4 ms for the CC pairs. A two-way ANOVA was performed on the threshold values using type of frequency pair and subject as the two factors. The results of the ANOVA indicate that both factors are significant $[F(1,11)=7.46, p<0.02; F(3,11)=8.82, p<0.002]$. A posthoc Scheffe analysis of the Subject effect indicated that the performance of Subj. 2 was significantly different ($\alpha=0.05$) from that of the other three subjects.

D. Experiment 4: Two-finger stimulation, frequency not tied to finger, discriminate on basis of finger

Thresholds for each subject and each of the four frequency pairs are shown in Table VI. Threshold values were similar for Subj. 1, Subj. 3, and Subj. 4 (ranging from 46.8 to 56.2 ms) and somewhat higher for Subj. 2 (92.0 ms). Standard deviations across the four frequency pairings ranged from 5.5 ms (Subj. 3) to 36.7 ms (Subj. 2). Results are further grouped into two types of frequency pairs (WC and CC). Performance within each group is shown for the four individual subjects in the bottom panel of Fig. 4. For each subject, no significant effects of frequency separation were observed. A two-way ANOVA was performed on the threshold values using type of frequency pair and subject as the two main factors. The results of the ANOVA indicate significance only for the factor of subject $[F(3,11)=3.96, p<0.04]$.

TABLE V. Temporal onset-order thresholds (in ms) for each subject and each frequency pair in Experiment 3.

	WC	CC				
	(25, 55)	(50, 250)	(7, 55)	(7, 315)	Mean	S.D.
Subj. 1	76.69	38.34	32.59	49.56	49.30	19.58
Subj. 2	258.66	133.63	128.11	65.35	146.44	80.97
Subj. 3	66.38	37.26	41.56	40.48	46.42	13.43
Subj. 4	42.47	39.35	42.65	39.99	41.12	1.69
Mean	111.05	62.15	61.23	48.85	70.82	
S.D.	99.45	47.66	44.82	11.85		60.15

IV. DISCUSSION

A. Effects of frequency separation

The effects of frequency separation on temporal onset-order discrimination were much larger for single-site stimulation than for two-finger stimulation. For stimuli delivered solely to the left index finger, thresholds for within-channel

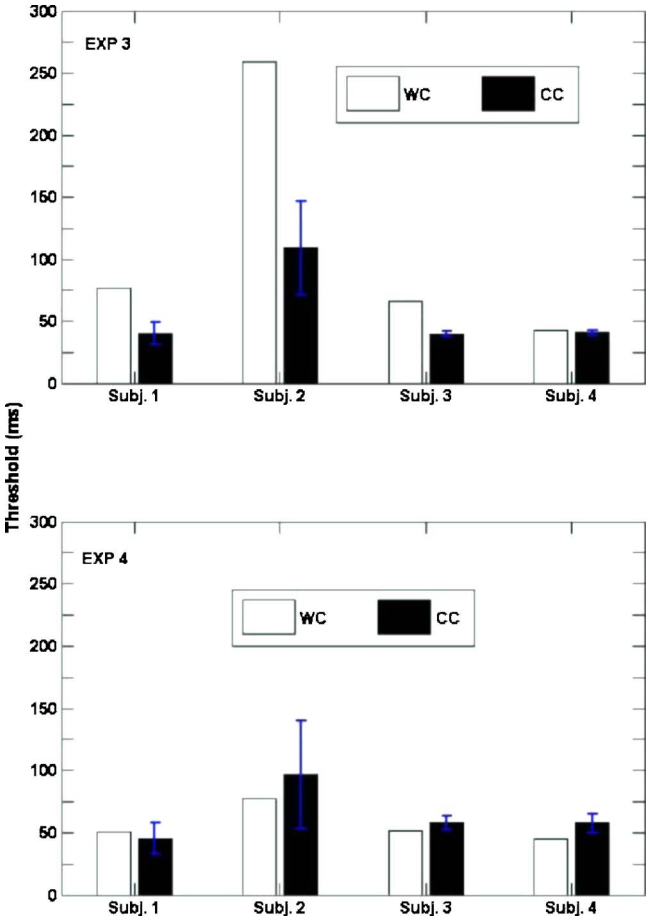


FIG. 4. (Color online) Summary of results for Experiments 3 and 4. In the top panel, results of Experiment 3 (Two-finger stimulation: onset order discrimination based on frequency) are grouped into two types of frequency pairs: WC and CC frequency pairs. Mean and ±1 S.D. of performance of each group are shown for the four individual subjects. In the bottom panel, results of Experiment 4 (Two-finger stimulation: onset-order discrimination based on finger) are grouped into two types of frequency pairs: WC and CC frequency pairs. Mean and ±1 S.D. of performance of each group are shown for four individual subjects.

TABLE VI. Temporal onset-order thresholds (in ms) for each subject and each frequency pair in Experiment 4.

	WC	CC				
	(25,55)	(50,250)	(7,55)	(7,315)	Mean	S.D.
Subj. 1	50.94	57.25	32.68	46.21	46.77	10.43
Subj. 2	77.72	146.59	76.83	66.82	91.99	36.73
Subj. 3	51.61	59.11	51.55	62.56	56.21	5.53
Subj. 4	44.78	61.95	48.73	62.60	54.52	9.11
Mean	56.26	81.23	52.45	59.55	62.37	
S.D.	14.63	43.62	18.26	9.11		25.29

frequency pairs (averaging 98.5 ms across subjects) were roughly 1.7 times greater than those observed for cross-channel pairs (56.9 ms). Among the within-channel pairs for one-finger stimulation, thresholds were substantially larger for the low-frequency pair (139.8 ms) than for the intermediate-frequency (81.29 ms) or high-frequency (74.6 ms) pair. The length of the periods associated with the low-frequency stimuli (i.e., periods of 333.33 ms and 142.9 ms for 3 Hz and 7 Hz, respectively) may require a longer time in which to identify and discriminate low-frequency compared to higher-frequency signals. For temporal gap detection of vibrotactile sinusoidal signals, Gescheider *et al.* (2003) did not observe any difference in resolution for stimuli selected to activate the Pacinian (250 Hz, large contactor), NPII (250 Hz, small contactor), and NPI (62 Hz, small contactor) channels. The current data also show a trend for similar temporal resolution for intermediate- and high-frequency signals. Because Gescheider *et al.* (2003) did not include motional stimulation in their study, it is not known whether the increased temporal-onset order thresholds observed here would be reflected by increased gap-detection thresholds.

Although the thresholds for WC pairs are clearly higher than those at for CC pairs in the single-finger condition, it seems that there is no strong evidence that the separation [in terms of the number of just-noticeable-difference (JND) steps] between the two frequencies in a pair plays a significant role in determining the size of the temporal onset-order threshold. The numbers of JND steps was calculated for each frequency pair using previous estimates of tactile frequency-discrimination data available in the literature (see Fig. 1.2 of Verrillo and Gescheider, 1992). The calculations assumed that the Weber fraction ($\Delta F/F$) increases with frequency and is 0.1 in the low-frequency range, 0.3 for intermediate frequencies, and 0.5 in the high-frequency range. Estimates of the number of JND steps between nonidentical pairs of frequencies were calculated assuming a logarithmic relationship to describe the growth of the JND in Hz within each region. For WC pairs, our estimates of the number of JND steps decrease from roughly 9 for the low-frequency pair, to 3.0 for the midfrequency pair, to 1.7 for the high-frequency pair. Thus, if thresholds are dependent on JND steps, we would expect to observe a decrease in performance with an increase in frequency. For Experiment 1, the mean temporal onset-order thresholds across subjects were quite similar for the three WC pairs, as well as for the four CC pairs (where the number of JND steps are estimated to be as high as 20 for the low-high pair), strongly suggesting that performance is not closely related to the number of JND steps. In Experiment 2, the mean performance on WC pairs, in fact, improved with increasing frequency [with the largest threshold for the pair (3,7)]. Thus, temporal onset-order discrimination thresholds appear not to be determined by the number of JND steps, per se, but rather to be more closely linked to whether or not frequencies are from within the same or across different channels of the tactual system.

The increase in thresholds for WC pairs compared to CC pairs in the single-finger condition may be explained by an increase in signal interactions for stimulus frequencies within the same channel (compared to across channels). This observation is consistent with that of previous psychophysical studies of the interactions of signals from within and across the various channels of the tactual system (e.g., see review by Gescheider *et al.*, 2004). For example, psychophysical measurements of tuning curves for the high frequency Pacinian channel, obtained using same-site forward-masking or adaptation techniques, demonstrate threshold elevations of probe signals only in the presence of masking and adaptation signals located within the same channel as the probe. Similarly, results from same-site loudness-matching studies indicate that sensation magnitudes differ for combinations of two signals within a given channel compared to signals across channels. The results of these studies have been interpreted as suggesting that perceptual interactions between WC signals occur at a more peripheral level than those of CC signals.

For the two-finger conditions employed in Experiment 1, the use of the same stimulating frequency at both sites led to somewhat higher thresholds (averaging 53.9 ms) than the use of different frequencies (either within or across channels, where thresholds averaged roughly 47.5 ms) at the two sites. Generally, the site of stimulation appears to provide a strong cue for onset-order independent of frequency, although the use of different frequencies at the two sites does lead to a slight reduction in threshold. When subjects were asked to determine temporal-onset order on the basis of frequency alone (ignoring place of stimulation in a two-finger condition, as in Experiment 3), performance on the within-channel pair of frequencies (111.0 ms) was inferior to that obtained on the cross-channel frequencies (64.4 ms), suggesting that a greater frequency separation facilitates onset-order discrimination. On the other hand, frequency separation had a much smaller effect on thresholds when the task was performed on the basis of site alone (as in Experiment 4).

B. Effects of redundancy

Conditions employing redundant coding of frequency and site of stimulation led to lower thresholds than those obtained in conditions where performance was based solely on frequency or site. Under the redundant coding conditions of Experiment 1 (two-finger stimulation with WC and CC pairs), thresholds averaged roughly 47.5 ms over subjects and pairs. Site-alone performance was measured in the two-

finger conditions of Experiment 1 employing SF pairs and in Experiment 4 (where the subject was instructed to determine temporal-onset order on the basis of site alone and to ignore the irrelevant dimension of frequency). Thresholds for these two cases described above averaged 53.9 ms and 62.4 ms, and were greater than the thresholds for redundant coding by factors of 1.1 and 1.3, respectively. Frequency-alone performance was measured through the presentation of WC and CC frequency pairs to a single site (Experiment 2) and in Experiment 3 (where the subject's task was to determine temporal-onset order on the basis of frequency and to ignore the irrelevant dimension of site). Thresholds on these two tasks averaged 74.8 ms and 70.8 ms, respectively, and were greater than the redundant threshold value by factors of 1.6 and 1.5, respectively. Thus, the current results indicate that the discrimination of temporal-onset order is more difficult for frequency-based cues than site-based cues.

Taylor (1978) investigated redundancy effects for a vibrotactile temporal-order identification task which required subjects to judge the order of three sequential stimuli presented with a duration of 120 ms and SOA of 120 ms (a more complex and difficult task than the discrimination procedure employed in the current study). The task was performed on the basis of frequency alone (using three values of frequency balanced for loudness at the index finger), site of stimulation alone (using 450-Hz stimulation balanced for loudness at the index, middle, and ring fingers), and redundant frequency-site pairings. Performance averaged 33% correct for the frequency-alone task, 42% for the site-alone task, and 52% for the redundant-coding condition. As in the current study, site of stimulation appears to provide a stronger cue than frequency for temporal-order judgments. The redundant-coding score was greater than the frequency-alone and the site-alone scores by factors of 1.6 and 1.2, respectively, similar to the threshold ratios observed in the current data. Taylor (1978) also examined the effect of roving an irrelevant stimulus dimension on performance (e.g., judgments were based on site, but the frequency delivered to each finger was selected at random on each trial or judgments were based on frequency with random selection of site of stimulation) and found that performance was equivalent to that obtained with a fixed value of the irrelevant dimension. The conditions employed in Experiments 3 and 4 also examined performance along one dimension (frequency or site) with a roving value of the other dimension and resulted in thresholds that were similar to those obtained with a fixed value of the irrelevant dimension.

Craig and Busey (2003) also observed positive effects of redundancy on the ability to discriminate temporal order in measurements employing brief vibrotactile patterns presented at two sites (the index and middle fingers of one hand or of two separate hands). Performance was measured at five values of SOA in the range of 13 to 200 ms for a fixed 65-ms stimulus duration. Stimulus overlap occurred at the three shortest values of SOA; however, the second stimulus always ended later than the first stimulus, providing an offset cue that was confounded with the onset cue. Temporal order thresholds were roughly 45 ms for static presentation of the 65-ms patterns. When the temporal patterns were scanned across the fingertip, the ability to judge the onset order of the two sites of stimulation was found to be dependent on the direction of movement of the scanned patterns. When the temporal order of stimulation at the two different sites was consistent with the direction of movement in which the patterns were scanned across the fingertip, thresholds were substantially lower than those observed when movement was inconsistent with onset-order presentation (roughly 13 ms versus 187 ms). Similar trends were observed with and without correct-answer feedback as well as for conditions involving the presentation of one static and one scanned stimulus and for bilateral presentation, suggesting that the effect most likely does not arise from a bias derived from the natural process of haptic exploration, but rather from higher-level mechanisms of attention.

C. Overlapping versus nonoverlapping trials

The employment of roving in duration leads to two types of experimental trials: (1) overlap trials where the onset of the second stimulus occurs earlier than the offset of the first stimulus; and (2) nonoverlap trials where the offset of the first stimulus occurs earlier than the onset of the second stimulus. Overlap trials occur when the duration of the first stimulus is larger than the SOA for a given run of trials; nonoverlap trials occur when the duration of the first stimulus is shorter than the SOA. Due to the distribution of durations and the values of SOA employed in the measurements for each subject, overlap trials accounted for roughly 70% of all the experimental trials. In addition to the onset-order cue present in both types of trials, an offset-order cue is also introduced in the nonoverlap trials (that is, the second stimulus necessarily ends later than the first stimulus).

To determine whether performance differed for the two types of trials, data for each frequency pair under each condition were broken down into each of these two categories and percent-correct scores were calculated for each of the two types. This analysis was applied primarily to data obtained in Experiment 2 (and primarily for Subj. 2), where the SOA values required to measure threshold resulted in both types of trials. A comparison of mean percent-correct scores for the two types of trials in Experiment 2 (80% for nonoverlap trials versus 82% for overlap trials) indicates that there is no strong evidence for a difference in performance between the two types of trials. Thus, subjects seem not to have benefited from the additional offset-order cue available on the nonoverlap trials, perhaps due to the fact that they were instructed to attend only to the onset of the stimuli and may have ignored any information contained in the offset order.

D. Intersubject differences

The most notable difference in performance across the four subjects is observed in the data of Subj. 2. The mean thresholds of Subj. 2 exceed those of the other three subjects in each of the four experiments. These differences are most pronounced in Experiments 2 and 3, where temporal onset-order decisions are made on the basis of frequency. The reduced resolution of Subj. 2 may be related to his age of 40 years compared to an age range of 18 to 26 years across the

other three subjects. Effects of aging have been observed previously in a variety of psychophysical tasks employing tactual stimulation, including elevated thresholds for the detection of vibrotactile signals at higher frequencies (Gescheider *et al.*, 1994) greater susceptibility to masking (Gescheider *et al.*, 1992), and a decrease in subjective magnitude of vibration (Verrillo *et al.*, 2002). In the temporal domain, Van Doren *et al.* (1990) studied temporal gap detection for vibrotactile stimuli as a function of aging. The signals consisted of two equal-amplitude, 350-ms sinusoidal (250 Hz) or bandpass noise (250–500 Hz) bursts separated by a temporal gap which took on values in the range of 8 to 256 ms. Threshold was defined as the signal level in dB SL at which a given gap duration was detected with 75% accuracy. Effects of aging were observed with bandpass noise (but not with sinusoids) where the thresholds increased with age primarily at the short (i.e., less than 32 ms) values of gap duration; additionally, the function relating threshold to gap duration was different for tonal and noise signals in older subjects, leading the authors to conclude that different processes may be responsible for gap detection in the two types of signals.

In the current study, it appears that the data of the older subject (Subj. 2) diverge more from those of the younger subjects for onset-order discriminations based on frequency as opposed to site of stimulation. In Experiment 2 (one-finger stimulation) and Experiment 3 (which required judgments based on frequency independent of site of stimulation), the mean thresholds of Subj. 2 were nearly 2.5 to 3.0 times as large as that of the other three subjects. Peripheral factors related to a decrease in the structure and number of sensory cells with age (P cells, in particular) may play a role in the reduced performance of Subj. 2. For example, it is possible that his tactual detection thresholds are higher than the average data (from a separate group of subjects see Table I) used to establish the sensation levels employed in the current study. The poorer performance of Subj. 2 may also be related in part to a decline in central-processing ability.

An additional instance of intersubject variability is observed in Experiment 1, where the performance of Subj. 1 is significantly better than that of the other subjects. This subject's superior performance on the two-finger task may be related to her previous experience as a subject in the study of Yuan *et al.* (2005a) which involved temporal onset-order discrimination in a redundant-coding condition employing a 250-Hz signal presented at the index finger and a 50-Hz signal at the thumb. This condition was also tested in the current study (see Table IV), where Subj. 1 achieved a threshold value of 25 ms compared to her performance of 43 ms in the previous study (Yuan *et al.*, 2005a, where she is identified as S2). Her previous experience on the two-finger task appears to have resulted in a transfer of training to all of the fixed-frequency two-finger conditions of Experiment 1, but not to the single-finger (Experiment 2) or roving-parameter two-finger conditions (Experiments 3 and 4), where her performance was quite similar to that of the other two young-adult subjects.

E. Relation to design of tactual displays of speech

The results of the current experiments can be applied to the design of tactual displays of temporal onset-order cues in speech. The data reported here suggest that decisions based solely on site of stimulation result in lower threshold values than those based solely on frequency of stimulation. An advantage for redundant coding is seen over coding through each cue alone; however, this advantage is smaller for site than for frequency. The results obtained with a site cue in the presence of an irrelevant rove of stimulating frequency are similar to those obtained with a site-alone cue in the presence of a fixed value of frequency. This finding suggests that site may be used to provide a temporal-onset order cue for one speech cue (such as voicing) while frequency is used to encode an additional parameter of speech (e.g., manner of production).

V. CONCLUSIONS

Tactual temporal onset-order thresholds ranged from roughly 20 to 200 ms across subjects and conditions. Effects of frequency separation were greater for stimuli presented to one finger compared to two fingers. For presentation at one finger, thresholds were substantially higher for WC frequency pairs compared to CC pairs. Performance for two-finger experiments employing redundancy of frequency and finger was superior to that of one-finger experiments for WC pairs. For two-finger experiments, performance with redundant coding was generally superior or similar to the nonredundant coding schemes. Onset-order cues can be usefully encoded through site of stimulation or through the use of frequencies from different regions of the tactual sensory system at the same site. Age also seems to play a role in onset-order discrimination performance with higher thresholds for older compared to younger subjects. These results provide guidelines to the design of tactual displays of temporal properties that are important to the perception of speech, music, and environmental sounds.

ACKNOWLEDGMENTS

This research was supported by Research Grant No. R01–DC00126 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We wish to thank Bob Lutfi and two anonymous reviewers for their helpful advice in revising the manuscript.

Bolanowski, S. J., and Verrillo, R. T. (**1982**). "Temperature and criterion effects in a somatosensory subsystem: A neurophysiological and psychophysical study," J. Neurophysiol. **48**, 836–855.

Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (**1988**). "Four channels mediate the mechanical aspects of touch," J. Acoust. Soc. Am. **84**, 1680–1694.

Bolanowski, S. J., Checkosky, C. M., and Wengenack, T. M. (**1993**). "And now, for our two senses," *in Sensory Research: Multimodal Perspectives*, edited by R. T. Verrillo and J. J. Zwislocki (L. Erlbaum Assoc., Hillsdale, NJ), pp. 119–139.

Craig, J. C., and Baihua, X. (**1990**). "Temporal order and tactile patterns," Percept. Psychophys. **47**, 22–34.

Craig, J. C., and Busey, T. A. (**2003**). "The effect of motion on tactile and visual temporal order judgments," Percept. Psychophys. **65**, 81–94.

Durlach, N. I. (**1968**). "A decision model for psychophysics," Communication Biophysics Group, Research Laboratory of Electronics, MIT, MA.

Eberhardt, S. P., Bernstein, L. E., Barac-Cikoja, D., Coulter, D. C., and Jordan, J. (**1994**). "Inducing dynamic haptic perception by the hand: System description and some results," Proceedings of the ASME Dynamic Systems and Control Division, New York, Vol. **1**, pp. 345–351.

Gescheider, G. A., Sklar, B. F., Van Doren, C. L., and Verrillo, R. T. (**1985**). "Vibrotactile forward masking: psychophysical evidence for a triplex theory of cutaneous mechanoreception," J. Acoust. Soc. Am. **78**, 534–543.

Gescheider, G. A., Valetutti, A. A., Padula, M. C., and Verrillo, R. T. (**1992**). "Vibrotactile forward masking as a function of age," J. Acoust. Soc. Am. **91**, 1690–1696.

Gescheider, G. A., Bolanowski, S. J., Hall, K. L., Hoffmann, K. E., and Verrillo, R. T. (**1994**). "The effects of aging on information-processing channels in the sense of touch. 1. Absolute sensitivity," Somatosens. Mot. Res. **11**, 345–357.

Gescheider, G. A., Bolanowski, S. J., Pope, J. V., and Verrillo, R. T. (**2002**). "A four-channel analysis of the tactile sensitivity of the fingertip: frequency selectivity, spatial summation, and temporal summation," Somatosens. Mot. Res. **19**, 114–124.

Gescheider, G. A., Bolanowski, S. J., and Chatterton, S. K. (**2003**). "Temporal gap detection in tactile channels," Somatosens. Mot. Res. **20**, 239–247.

Gescheider, G. A., Bolanowski, S. J., and Verrillo, R. T. (**2004**). "Some characteristics of tactile channels," Behav. Brain Res. **148**, 35–40.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hirsh, I. J., and Sherrick, C. E. (**1961**). "Perceived order in different sensory modalities," J. Exp. Psychol. **62**, 423–432.

Johansson, R. S. (**1978**). "Tactile sensibility in the human hand: receptive field characteristics of mechanoreceptive units in the glabrous skin area," J. Physiol. (London) **281**, 101–123.

Johansson, R. S., Landstrom, U., and Lundstrom, R. (**1982**). "Responses of mechanoreceptive afferent units in the glabrous skin of the human hand to sinusoidal skin displacements," Brain Res. **244**, 17–25.

Marks, L. E., Girvin, J. P., O'Keefe, M. D., Ning, P., Quest, D. O., Antunes, J. L., and Dobelle, W. H. (**1982**). "Electrocutaneous stimulation 3: The perception of temporal-order," Percept. Psychophys. **32**, 537–541.

Mountcastle, V. B., LaMotte, R. H., and Carli, G. (**1972**). "Detection thresholds for stimuli in humans and monkeys: comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand," J. Neurophysiol. **32**, 453–484.

Reed, C. M., Rabinowitz, W. M., Durlach, N. I., Braida, L. D., Conway-Fithian, S., and Schultz, M. C. (**1985**). "Research on the Tadoma method of speech communication," J. Acoust. Soc. Am. **77**, 247–257.

Rinker, M. A., Craig, J. C., and Bernstein, L. E. (**1998**). "Amplitude and period discrimination of haptic stimuli," J. Acoust. Soc. Am. **104**, 453–463.

Sherrick, C. E. (**1970**). "Temporal ordering of events in haptic space," IEEE Trans. Man-Machine Systems **11**, 25–28.

Shore, D. I., Spry, E., and Spence, C. (**2002**). "Confusing the mind by crossing the hands," Cognit. Brain Res. **14**, 153–163.

Talbot, W. H., Darian-Smith, I., Kornhuber, H. H., and Mountcastle, V. B. (**1968**). "The sense of flutter-vibration: Comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand," J. Neurophysiol. **31**, 301–324.

Tan, H. Z. (**1996**). "Information transmission with a multi-finger tactual display," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Tan, H. Z., and Rabinowitz, W. M. (**1996**). "A new multi-finger tactual display," in *Proceedings of the Dynamic Systems and Control Division*, DSC-Vol. **58**, pp. 515–522.

Taylor, B. (**1978**). "Dimensional redundancy in the processing of vibrotactile temporal order," Ph.D. dissertation, Princeton University, Princeton, NJ.

Van Doren, C. L., Gescheider, G. A., and Verrillo, R. T. (**1990**). "Vibrotactile temporal gap detection as a function of age," J. Acoust. Soc. Am. **87**, 2201–2206.

Verrillo, R. T. (**1966a**). "Specificity of a cutaneous receptor," Percept. Psychophys. **1**, 149–153.

Verrillo, R. T. (**1966b**). "Vibrotactile sensitivity and frequency response of the Pacinian corpuscle," Psychonomic Sci. **4**, 135–136.

Verrillo, R. T. (**1979**). "The effect of surface gradients on vibrotactile thresholds," Sens. Processes **3**, 27–36.

Verrillo, R. T., Fraioli, A. J., and Smith, R. L. (**1969**). "Sensation magnitude of vibrotactile stimuli," Percept. Psychophys. **6**, 366–372.

Verrillo, R. T., and Gescheider, G. A., (**1977**). "Effect of prior stimulation on vibrotactile thresholds," Sens. Processes **1**, 292–300.

Verrillo, R. T., and Bolanowski, S. J. (**1986**). "The effects of skin temperature on the psychophysical responses to vibration on glabrous and hairy skin," J. Acoust. Soc. Am. **80**, 528–532.

Verrillo, R. T., and Gescheider, G. A., (**1992**). "Perception via the sense of touch," in *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr Publishers, London), pp. 1–36.

Verrillo, R. T., Bolanowski, S. J., and Gescheider, G. A. (**2002**). "Effect of aging on the subjective magnitude of vibration," Somatosens. Mot. Res. **19**, 238–244.

Yuan, H. F. (2003). "Tactual display of consonant voicing to supplement lipreading," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (**2004**). "Envelope-onset asynchrony as a cue to voicing in initial English consonants," J. Acoust. Soc. Am. **116**, 3156–3167.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (**2005a**). "Temporal onset-order discrimination through the tactual sense," J. Acoust. Soc. Am. **117**, 3139–3148.

Yuan, H. F., Reed, C. M., and Durlach, N. I. (**2005b**). "Tactual display of consonant voicing as a supplement to lipreading," J. Acoust. Soc. Am. **118**, 1003–1015.

Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration

Soren Y. Lowell[a] and Brad H. Story
Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona, 85721-210071

(Received 18 January 2005; revised 3 April 2006; accepted 20 April 2006)

Adjustments to cricothyroid and thyroarytenoid muscle activation are critical to the control of fundamental frequency and aerodynamic aspects of vocal fold vibration in humans. The aerodynamic and physical effects of these muscles are not well understood and are difficult to study *in vivo*. Knowledge of the contributions of these two muscles is essential to understanding both normal and disordered voice physiology. In this study, a three-mass model for voice simulation in adult males was used to produce systematic changes to cricothyroid and thyroarytenoid muscle activation levels. Predicted effects on fundamental frequency, aerodynamic quantities, and physical quantities of vocal fold vibration were assessed. Certain combinations of these muscle activations resulted in aerodynamic and physical characteristics of vibration that might increase the mechanical stress placed on the vocal fold tissue.
[DOI: 10.1121/1.2204442]

PACS number(s): 43.70.Bk, 43.70.Aj, 43.70.Gr [AL] Pages: 386–397

I. INTRODUCTION

Intelligible and efficient production of speech and song relies, in part, on the ability to precisely control the fundamental frequency (F_0) of the voice. Average F_0 and changes in F_0 during conversational speech signal critical linguistic information such as questions versus statements (Thorsen, 1980) and syllable and word emphasis, as well as personal information such as personality traits of the speaker (Brown *et al.*, 1974), sex (Bachorowski and Owren, 1999; Gelfer and Schofield, 2000), and age (Jacques and Rastatter, 1990). The primary control of F_0 is thought to be based on the neuromuscular activation of two sets of intrinsic laryngeal muscles: the cricothyroid (CT) and the thyroarytenoid (TA). According to the cover-body theory of vocal fold vibration (Hirano, 1974), physical quantities of the vocal folds such as length, mass, and stiffness are adjusted by the complex activations of these two muscles. Adjustment of these muscles can therefore have a substantial effect on the aerodynamics and mechanics of vocal fold vibration. The relationships between several intrinsic laryngeal muscles and acoustic characteristics of voice (Gay *et al.*, 1972; Kempster *et al.*, 1988; Tanaka and Tanabe, 1986) as well as vocal fold movement (Nasri *et al.*, 1994; Poletto *et al.*, 2004) have been studied in humans and in canine models. Less is known about the effects of intrinsic laryngeal muscle activation on aerodynamic and mechanical properties of vocal fold vibration. This knowledge is critical for understanding normal vocal fold physiology and pathophysiology in voice disorders. Vocal fold vibration during speech production often involves co-contraction of the CT and TA muscles (Poletto *et al.*, 2004; Shipp and McGlone, 1971; Titze *et al.*, 1989). Thus, determining the aerodynamic and mechanical effects of a range of co-activation combinations for CT and TA is important to study and has not been addressed in research to date. This study investigated the simulated effects of increased cricothyroid and thyroarytenoid muscle activation on acoustic, aerodynamic, and mechanical characteristics of adult-male vocal fold vibration.

Through their ability to lengthen or shorten the vocal folds, the CT and TA muscles provide the greatest mechanical advantage in altering vocal fold stiffness and thereby achieving changes in F_0. Whereas the apparent importance of other intrinsic laryngeal muscles such as the lateral cricoarytenoid (LCA) and posterior cricoarytenoid (PCA) in F_0 control has varied between studies (Atkinson, 1978; Gay *et al.*, 1972), the CT and TA muscles consistently emerge as the primary contributors to F_0 change (Gay *et al.*, 1972; Hirano *et al.*, 1970; Shipp and McGlone, 1971). Gay *et al.* (1972) used intramuscular electromyography (EMG) to measure laryngeal muscle activity relative to F_0 change and found that increases in F_0 were correlated with predominant, progressive increases in CT and TA activity. Other investigators systematically varied CT and TA activity to observe the resulting effect on F_0. Van den Berg and Tan (1959) conducted studies on excised human larynges, in which they artificially induced tension to mimic the action of the CT, TA, and LCA, and found that F_0 increased as a result of these manipulations. Direct electrical stimulation of the CT and TA muscles has been shown to increase F_0 (Kempster *et al.*, 1988), thus demonstrating the causal contribution of CT and TA activity to control of F_0.

Simultaneous contraction of the CT and TA muscles during human phonation offers a great deal of flexibility in producing changes in fundamental frequency. Male subjects show both intra- and intersubject variability in the levels of CT and TA activation used to produce a given F_0 (Titze *et al.*, 1989). Differing levels of TA and CT activation will

[a]Current affiliation: The National Institutes of Health, National Institute of Neurological Disorders and Stroke, Laryngeal and Speech Section, Bethesda, MD 20892.

influence aerodynamic and physical characteristics of the vocal folds during vibration. Based on Hirano's (1974) cover-body theory, Titze *et al.* (1989) proposed that TA activation can increase the stiffness of the vocal fold body while simultaneously decreasing the stiffness of the cover. Depending on the degree to which the body participates in vocal fold vibration, TA activation could raise or lower F_0. Increased F_0 would be expected when the body participates significantly in vibration, such as in the low-frequency range or for loud phonation. At the mid-frequency range, TA activation could result in either increased or decreased F_0. At high fundamental frequencies such as in falsetto, and during soft phonation when the cover contributes more to vibration than the body, TA activation would be expected to lower F_0 because it creates a slack cover (Titze, 2000). Changes to the stiffness of the body and cover not only affect F_0, but can also affect dimensions such as amplitude of vibration, degree of excursion, and vertical phase difference between the body and cover during vocal fold vibration, as well as aerodynamic characteristics.

Support for the complex effects of TA activation on F_0 is based on investigations of muscle activation levels of the CT and TA during F_0 change. Hirano and colleagues have demonstrated frequency-dependent, differential activation levels of the CT and TA with F_0 changes (Hirano, 1988; Hirano *et al.*, 1969, 1970). Titze *et al.* (1989) used electromyography (EMG) to study CT and TA muscle activations associated with volitional F_0 change. Results showed that speakers generally raised F_0 by increasing both CT and TA activation levels. However, for a given speaker, several combinations of CT and TA activation could yield the same F_0. In the second portion of the study, Titze *et al.* (1989) used the electrodes for direct stimulation of the TA muscle. Electrical stimulation in one subject resulted in consistent increases in F_0 for low to medium F_0 ranges. In the medium F_0 range, TA stimulation resulted in both increased and decreased F_0. Decreasing F_0 changes were produced with TA stimulation in the falsetto range and for soft phonation. EMG stimulation in the other three subjects, however, yielded less clear results that may have been affected by electrode placement difficulty or measurement error (Titze *et al.*, 1989). Overall, these findings highlight the complexity of the TA muscular role in F_0 control and support the theoretical notion that TA activation can result in both increased and decreased F_0.

Beyond the control of F_0, less is known about the mechanical and aerodynamic consequences of intrinsic laryngeal muscle activation. Physical movement of the vocal folds has been correlated to the activity of several intrinsic laryngeal muscles. Poletto *et al.* (2004) found posterior cricoarytenoid EMG activity to correlate consistently with vocal fold opening, whereas CT and TA activity were correlated with both opening and closing of the vocal folds during speech production. An *in vivo* canine model (Nasri *et al.*, 1994) supported the adductory roles of the TA, lateral cricoarytenoid (LCA), and interarytenoid (IA) muscles.

Few studies have quantified aerodynamic changes associated with intrinsic laryngeal muscle activation changes. Investigators who have included estimations or measures of tracheal pressure or airflow and intrinsic laryngeal muscle activity (Atkinson, 1978; Baer, 1979; Baker *et al.*, 2001; Shipp and McGlone, 1971) have generally focused on the relationship between these aerodynamic variables and the acoustic variables, versus laryngeal muscle activation. In the few studies that have correlated intrinsic laryngeal muscle activity to aerodynamic parameters, conflicting results regarding the correlation of TA activity and laryngeal resistance have been reported (Finnegan *et al.*, 2000; Tanaka and Tanabe, 1986). CT and TA activity were positively correlated to subglottal pressure (Baer, 1979; Shipp and McGlone, 1971), and CT activity was not significantly correlated to airflow (Faaborg-Andersen *et al.*, 1967). In all of these studies, subjects were asked to produce speech utterances at specific frequencies or intensities, while EMG activity and aerodynamic parameters were measured. However, data are lacking on the aerodynamic and mechanical effects that result from changes to intrinsic laryngeal muscle activation levels.

Measuring the isolated effects of intrinsic laryngeal muscle activation is difficult *in vivo*. Although speakers can reliably produce speech at a specified target frequency or intensity, especially if given a means of external monitoring and correction, they typically cannot volitionally control the level of TA or CT activation as targets. Changes to multiple extralaryngeal variables often occur simultaneously with intricate, synergistic activation of intrinsic laryngeal muscles in the control of F_0. These co-occurring changes complicate the interpretation of individual, intrinsic laryngeal muscle effects. One way in which the isolated effects of these muscle activations have been studied is by electrically stimulating the muscle and then observing the response to the stimulation. As summarized above, electrical stimulation of the TA muscle and resulting effects on F_0 were assessed by Titze *et al.* (1989) in four males. Stimulation either increased or decreased F_0, depending on the F_0 at which the subject was vocalizing. Kempster *et al.* (1988) found that electrical stimulation of the TA and CT increased F_0 in the four human subjects that they studied. Intrinsic laryngeal muscle activation was simulated in anesthetized canines using mechanical retraction of cartilages and by electrically stimulating the thyroarytenoid muscle (Tanaka and Tanabe, 1986). The effects of intrinsic laryngeal muscles on subglottal pressure, flow, and voice intensity were then observed. TA muscle contraction resulted in increased subglottal pressure and decreased airflow, with no substantial change to voice intensity. Thus, preliminary work in humans and canines has suggested some causal effects of intrinsic laryngeal muscle activation on acoustic and aerodynamic parameters. However, investigation in humans is hampered by the difficult and invasive techniques of intramuscular stimulation, and the need to minimize the number of stimulations given to any one subject.

The effects of intrinsic laryngeal muscle activity on F_0 can also be studied with computational models that are based on approximations of the physical properties of the vocal folds (Alipour-Haghighi and Titze, 1983, 1991; Ishizaka and Flanagan, 1972; Story and Titze, 1995). Such models cannot necessarily explain how any particular speaker controls their voice, but rather allow for predictive simulations of possible

voice productions. Titze *et al.* (1989) and Titze (1991) developed a model of male F_0 control based on a physiologically motivated and empirically determined relation between vocal fold strain (i.e., length change) and normalized activation levels of the CT and TA muscles. This relation resulted from experiments with excised larynges, *in vivo* animal preparations, and EMG recordings of human speakers, and was subsequently used in the model to specify the passive stress developed in the various tissue layers within the vocal folds. These stresses, along with the active stress contributed by the TA activity, were then combined to predict the vibrational frequency of the vocal folds. This particular model did not, however, include an actual simulation of the self-sustained oscillation of the vocal folds.

More recently, Titze and Story (2002) have incorporated many aspects of the strain-based F_0 control model into a system in which the activations of the CT and TA muscles, specified as input parameters, are transformed into the mechanical parameters (i.e., stiffness, mass, damping) of a low-dimensional self-oscillating vocal fold model. This provides a means by which the vibration of the vocal folds and resulting output quantities such as pressure and airflow can be simulated relative to intrinsic laryngeal muscle activation. Simulations showed that a continuum of muscle activation levels for TA and CT could theoretically produce a constant F_0, with isofrequency contour lines generated to depict these muscle activation combinations (Story and Titze, 1995; Titze, 2000; Titze *et al.*, 1989; Titze and Story, 2002). These simulations suggested that, at certain ranges of CT and TA activation levels, a male speaker would have various options for increasing or decreasing F_0. Thus, computational modeling provides a means to explore the predicted, isolated contributions of intrinsic laryngeal muscles to vocal biomechanics. Follow-up studies in humans are then needed to validate the simulated effects of these muscle activations in humans.

Determining how simulated changes to CT and TA activation may alter parameters such as vocal fold configuration, airflow, and F_0 during phonation is a critical step toward understanding normal and disordered voice physiology. Changes in certain aerodynamic characteristics have been associated with voice disorders. Maximum flow declination rate (MFDR) is indicative of the velocity of vocal fold closure, and an increase in MFDR may result in a greater degree of vocal fold collision forces (Hillman *et al.*, 1989). Sound pressure level (SPL), tracheal pressure, laryngeal resistance, and MFDR are positively correlated (Holmberg *et al.*, 1988, 1989, 1994). Increased MFDR values have been reported in subjects with vocal nodules and vocal polyps (Hillman *et al.*, 1989, 1990)—vocal pathology that is thought to be related to vocal hyperfunction and increased vocal fold impact stress (Boone and McFarlane, 2000; Case, 2002; Gray and Titze, 1988). Higher levels of impact stress occur at the midmembranous portion of the vocal folds during vocal fold vibration, which corresponds with the location at which vocal nodules often develop in humans (Jiang and Titze, 1994).

In addition to collision forces and impact stress, shearing forces are considered potentially harmful to vocal fold tissue. Shearing forces occur during vibration of the vocal folds, and prolonged or excessive phonation may result in excessive shearing that can cause damage to the vocal fold tissue (Courey *et al.*, 1996; Gray and Titze, 1988). Changes in muscle activation levels of the CT and TA could differentially affect mass and tension of the vocal fold layers, and could therefore result in changes to tissue displacement that might be associated with increased or decreased shearing forces.

In this study, the effects of controlled change of CT and TA muscle activity were investigated with the low-dimensional vocal fold model reported by Titze and Story (2002) and Story and Titze (1995) and applicable to the adult-male voice. The first purpose of this study was to determine the effects of CT and TA muscle activation on F_0 when the level of each muscle was independently manipulated. The second purpose was to determine whether increased CT or TA activation produced aerodynamic and mechanical changes to vocal fold vibration that might be harmful to vocal fold tissue. The aerodynamic quantities of maximum intraglottal pressure, maximum glottal flow, and MFDR, and the physical quantities of amplitude ratio (lower to upper cover mass) as well as vertical phase difference, were chosen due to their influence on mechanical stress in phonation and for their associated potential for increasing the risk of vocal fold damage in humans.

II. METHOD

Vocal fold vibration was simulated with a model designed to approximate the body-cover structure of the vocal folds, where upper and lower masses represent the cover, and a third, laterally positioned mass represents the body (Story and Titze, 1995). A schematic diagram of the model is shown in Fig. 1(a); the stiffness and damping elements have been combined to simplify the picture. The upper and lower masses are coupled to each other and to the body with spring and damping elements. The springs account for shearing forces and stiffness of the tissue, whereas the damping elements account for the energy losses that occur in the system. The two-mass representation of the cover allows the vertical phase difference of the mucosal wave to be represented. In addition, the separation of cover and body tissue in the model allows for individual specification of the mechanical properties of each tissue layer. For all of the simulations in this study, bilateral symmetry was assumed such that identical vibrations occur within the right and left vocal folds.

The vocal fold model was coupled to the pressures in the trachea and the vocal tract [see Fig. 1(b)] according to the aerodynamic and acoustic considerations specified by Titze (2002), thus allowing for self-sustained oscillation. Acoustic wave propagation in both the trachea and vocal tract was simulated in time-synchrony with the vocal fold model. This was carried out with a wave-reflection approach (digital waveguide) (e.g., Liljencrants, 1985) that included energy losses due to yielding walls, viscosity, and radiation at the lips (Story, 1995). The shape of the trachea and the vocal tract shown in Fig. 1(b) were maintained for all simulation cases in this study. The cross-sectional area of the epilaryngeal portion was set to be 0.5 cm^2, whereas the uniform tube representation of the pharynx and oral cavity was set at

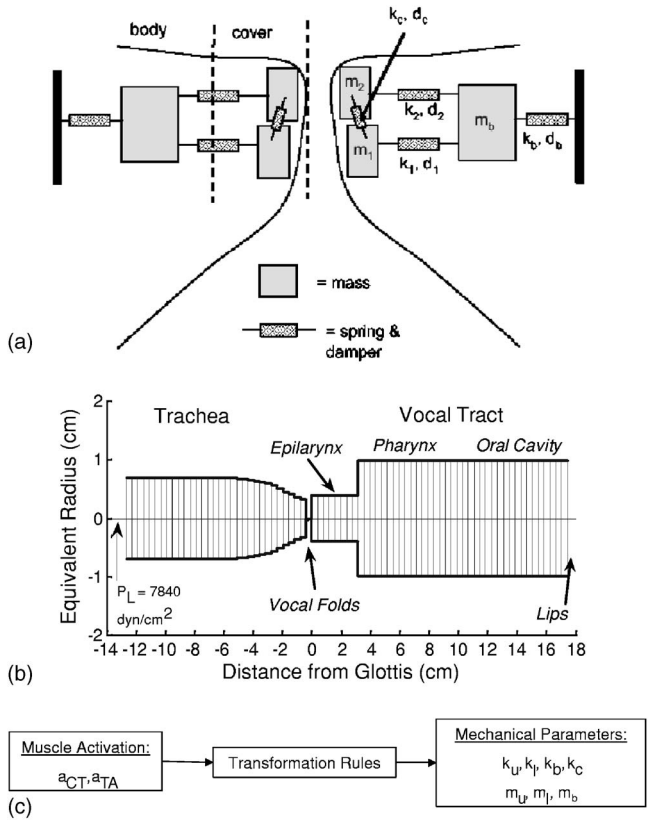


FIG. 1. Schematic representation of the simulation model: (a) three-mass model of the vocal folds, (b) tubular configuration of the trachea and vocal tract, and (c) transformation of CT and TA muscle activation levels into the mechanical parameters of the three-mass model.

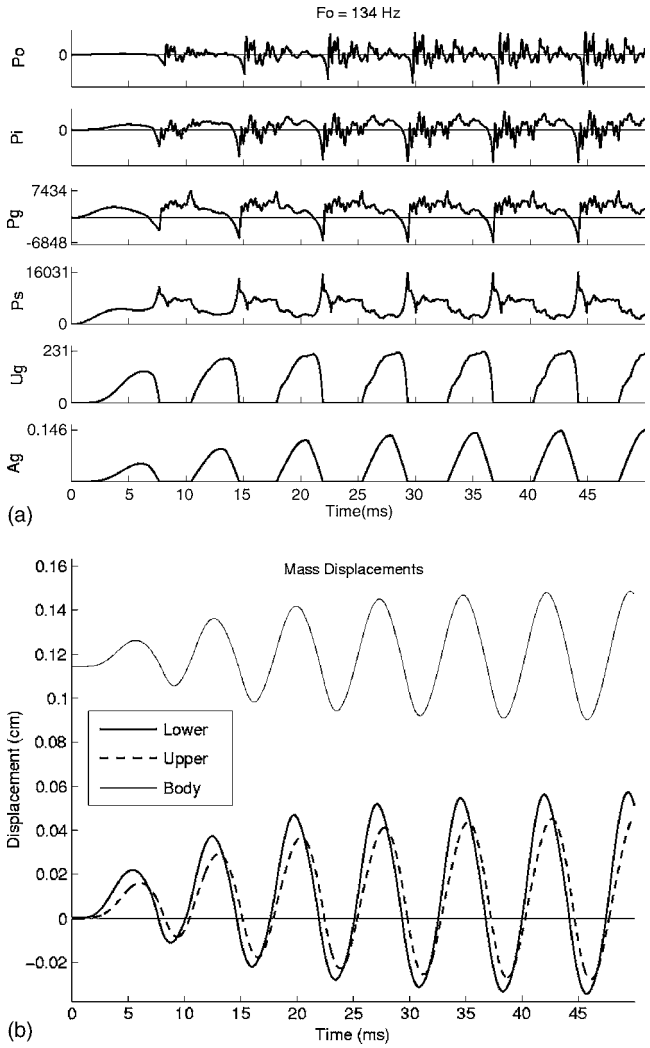


FIG. 2. Waveforms for the condition of a_{CT}=0.24, a_{TA}=0.24, depicting time-varying output pressure (Po), input pressure (Pi), intraglottal pressure (Pg), subglottal pressure (Ps), maximum glottal flow (Ug), and glottal area (Ag) as a function of time (a), and depicting changes in displacement of the lower and upper cover masses (bottom traces) and body mass (upper trace) as a function of time (b).

3 cm^2. A standard input pressure (lung pressure) of 7840 dyn/cm^2 was used for all simulations. A very slight prephonatory adduction angle was used (0.0001 cm). Lateral cricoarytenoid activation was set at a normalized level of 0.5 for all simulations.

The muscle activation levels of the CT and TA were specified in a normalized range from 0.0 to 1.0 and were intended to represent minimum to supramaximal activity within each muscle. As shown schematically in Fig. 1(c), these muscle activation levels were transformed, according to the "rules" developed by Titze and Story (2002), into the mechanical parameters of mass and stiffness. The rest dimensions of each of the vocal fold layers (length, thickness, depth) were also set to be the same as specified by Titze and Story (2002). Based on a comparison of EMG recordings and measurements of voice fundamental frequency reported by Titze *et al.* (1989) to the range of fundamental frequencies produced by the simulation model in Titze and Story (2002), it can be concluded that the normalized muscle activities produce physiologically realistic results.

The output waveforms produced by this model consisted of output pressure (radiated pressure at the lips), input pressure, glottal pressure, subglottal pressure, glottal flow, glottal area, and displacements of the upper and lower masses (Story and Titze, 1995; Titze, 2000; Titze *et al.*, 1989; Titze and Story, 2002). Shown in Figs. 2 and 3 are two different cases of CT muscle activation (a_{CT}) and TA muscle activation (a_{TA}) levels (a_{CT}=0.24, a_{TA}=0.24 for Fig. 2, and a_{CT}=0.24, a_{TA}=0.74 for Fig. 3). Note that the displacements of the upper and lower cover masses are strongly affected by levels of TA muscle activation and result in different waveform shapes for glottal flow (Ug) and glottal area (Ag).

To observe the output quantities of the model over a large range of muscle activation levels, simulations were generated for 2500 settings of a_{CT} and a_{TA}; the ranges of both a_{CT} and a_{TA} were divided into 50 evenly spaced increments from 0 to 1.0. For each simulation, the F_0 was determined with a zero-crossing detector and interpolation applied to the resulting glottal area signal. Additionally, vertical phase difference between the upper and lower masses, amplitude ratio of lower to upper mass displacement, maximum glottal flow, and maximum flow declination rate were computed for each simulation.

A contour plot showing lines of constant F_0 was produced with the "contour" function in MATLAB 7 (Mathworks, 2004), based on the collection of fundamental frequencies from each of the 2500 simulations. The resulting

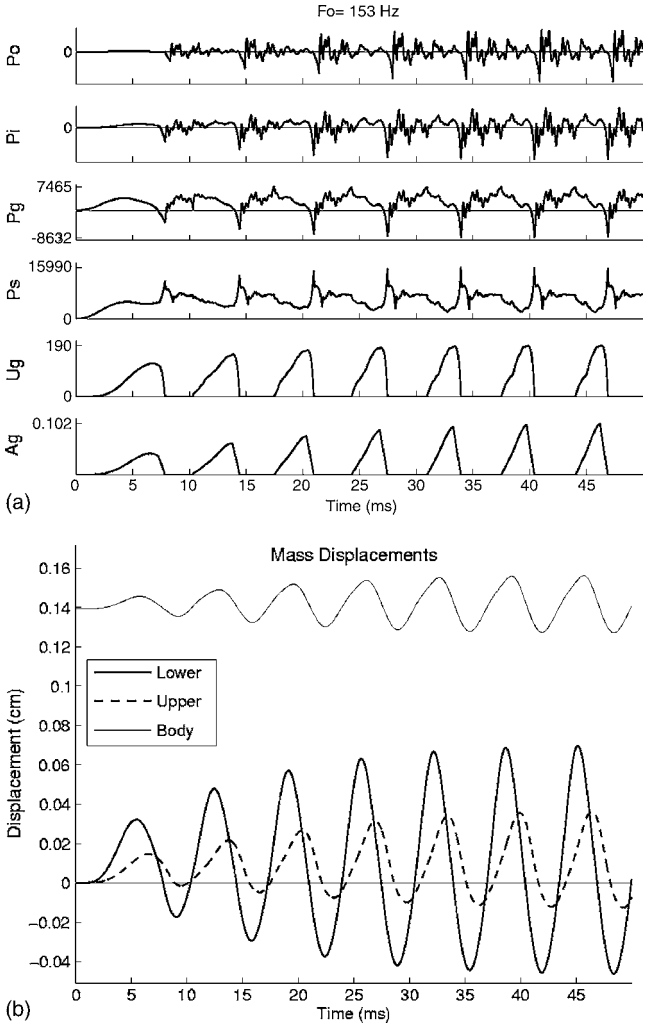


FIG. 3. Waveforms for the condition of a_{CT}=0.24, a_{TA}=0.74, depicting time-varying output pressure (Po), input pressure (Pi), intraglottal pressure (Pg), subglottal pressure (Ps), maximum glottal flow (Ug), and glottal area (Ag) as a function of time (a), and depicting changes in displacement of the lower and upper cover masses (bottom traces) and body mass (upper trace) as a function of time (b).

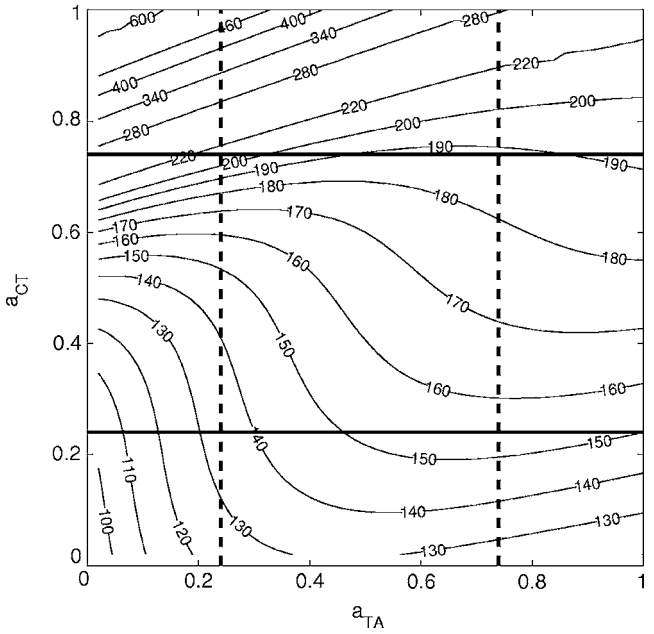


FIG. 4. Muscle activation plot (MAP) depicting isofrequency contour lines for normalized minimum to maximum thyroarytenoid (TA) activation levels and cricothyroid (CT) activation levels.

contour plot is shown in Fig. 4 and indicates how F_0 either changes or remains constant as the CT and TA activation levels change. Hence, it is referred to as a muscle activation plot (MAP).

The lower left portion of Fig. 4 represents CT and TA activation levels that are both low, and fundamental frequencies that are typical of conversational speech (Titze *et al.*, 1989). The upper right portion of Fig. 4 represents high levels of CT and TA activation. A speaker would presumably use muscle activation levels in this region for production of high fundamental frequencies. High CT activation levels and low TA activation levels are represented in the upper left portion, which encompasses the highest F_0 in the MAP. Four lines were drawn on the MAP to represent constant CT activation levels with progressively increasing TA activation (solid lines), and constant TA activation levels with progressively increasing CT activation (dashed lines). These lines were selected to represent low and high levels of constant CT activation (0.24 and 0.74, respectively) with TA activation varying from 0.0 to 1.0, and low and high levels of constant TA activation (0.24 and 0.74, respectively) with CT activation varying from 0.0 to 1.0. Thus, these four lines represent midpoints of the lower and upper portions of the MAP (constant CT) and midpoints of the right and left portions of the MAP (constant TA). Table I shows the values of the mechanical parameters of the three-mass model at the intersection points of the four lines indicated in Fig. 4. These values result from the rule-based transformation implemented by Titze and Story (2002).

By systematically increasing TA or CT activation levels from low to high along the selected lines, the isolated effects of these simulated muscle activations could be assessed. F_0, aerodynamic and physical quantities were assessed for each data line that represented a range from low to high TA activation levels (with constant CT) or low to high CT activation levels (with constant TA). From the output quantities described by the waveforms shown previously in Figs. 2 and 3 (or derivations from those quantities), changes to the following acoustic, aerodynamic, and physical quantities were assessed in response to systematic manipulation of TA and CT

TABLE I. Mechanical parameter values of the three-mass model at four settings of TA and CT activation levels. These result from the rule-based transformation implemented by Titze and Story (2002).

	$[a_{TA}, a_{CT}]$			
Mechanical parameters	[0.24, 0.24]	[0.24, 0.74]	[0.74, 0.24]	[0.74, 0.74]
Lower cover mass: m_1	0.0619	0.0612	0.0867	0.0864
Upper cover mass: m_2	0.0879	0.0869	0.0628	0.0626
Body mass: M	0.0978	0.0968	0.1973	0.1967
Lower cover stiffness: k_1	102 940	487 470	78 484	393 010
Upper cover stiffness: k_2	146 100	691 900	56 833	284 590
Body stiffness: K	188 210	430 190	529 310	481 360
Cover coupling stiffness: k_c	6739	11 857	5383	10 275

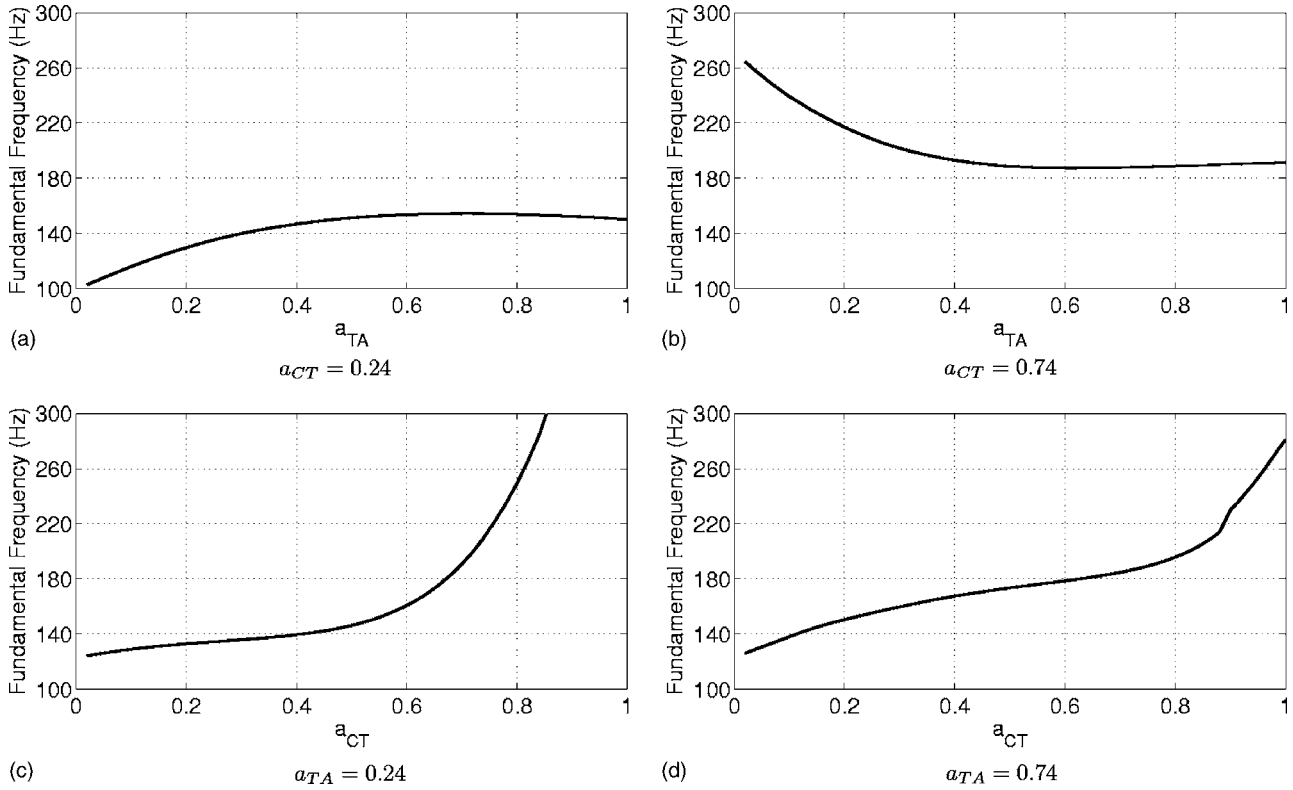


FIG. 5. Changes to fundamental frequency (F_0) as a function of increasing TA activation level when CT was constant at 0.24 (a) and 0.74 (b), and as a function of increasing CT activation when TA was constant at 0.24 (c) and 0.74 (d).

activation levels: fundamental frequency, maximum glottal flow, maximum flow declination rate (MFDR), vertical phase difference (number of degrees within a vibratory cycle that the lower mass leads the upper mass), and amplitude excursion ratios of the lower to upper cover masses. These quantities were chosen due to their potential impact on mechanical stress during vocal fold vibration. MFDR was measured as the most negative portion of the flow derivative within a glottal cycle. Maximum intraglottal pressure was analyzed but was found to be nearly constant at approximately 7300 dyn/cm^2 for all muscular settings, hence it was not included in the Secs. III and IV.

III. PREDICTIONS

To simplify presentation of the simulations, four different muscle activation cases were summarized for each output quantity: TA activation starting at a low normalized level (0.0) and progressively increasing to a high level (1.0) while (1) CT activation was held at a constant, low level of 0.24 (case 1) and (2) a constant, high level of 0.74 (case 2), and CT activation starting at a low normalized level (0.0) and progressively increasing to a high level (1.0) while (3) TA activation was held at a constant, low level of 0.24 (case 3) and (4) a constant, high level of 0.74 (case 4). All plots depict the change that occurred for each quantity as muscle activation was systematically increased.

A. Fundamental frequency

When CT activation level was held at 0.24, F_0 increased with increasing TA activation to an approximate TA activation level of 0.68, as shown in Fig. 5(a). For TA activation levels higher than 0.68, changes to F_0 plateaued. As can be seen by comparing the lower solid line in Fig. 4 and the fundamental frequencies in Fig. 5(a) for this same set of muscle activation combinations, F_0 started low at approximately the 100-Hz isofrequency contour and increased to cross the 150-Hz isofrequency contour as TA activation was maximally increased. At a constant CT activation level of 0.74 (case 2), F_0 decreased from 265 Hz to about 190 Hz as TA activation was increased to about 0.50, and then remained nearly constant, as shown in Fig. 5(b).

When TA activation was held at 0.24 (case 3), F_0 increased from 125 to 580 Hz as CT activation level increased, as shown in Fig. 5(c). The most substantial increase, however, occurred when CT activation exceeded the level of 0.60. To keep all plots in Fig. 5 on the same scale, the increase in F_0 is only shown up to 300 Hz. This increase in F_0 can also be observed in Fig. 4 along the vertical dashed line denoting TA activation level of 0.24. Here the F_0 levels begin below the 130-Hz isofrequency contour, and extend to the 460-Hz isofrequency contour (see Fig. 4). For a constant TA level of 0.74 (case 4), F_0 showed a small, gradual increase as CT activation was increased throughout the range of 0 to 1.0, as shown in Fig. 5(d).

B. Maximum glottal flow (Max Ug)

Max Ug decreased with increased TA activation and CT activation constant at 0.24 (case 1), as demonstrated in Fig. 6(a). As previously shown in Figs. 2(a) and 3(a), changes to the shape of the flow waveform were substantial when TA

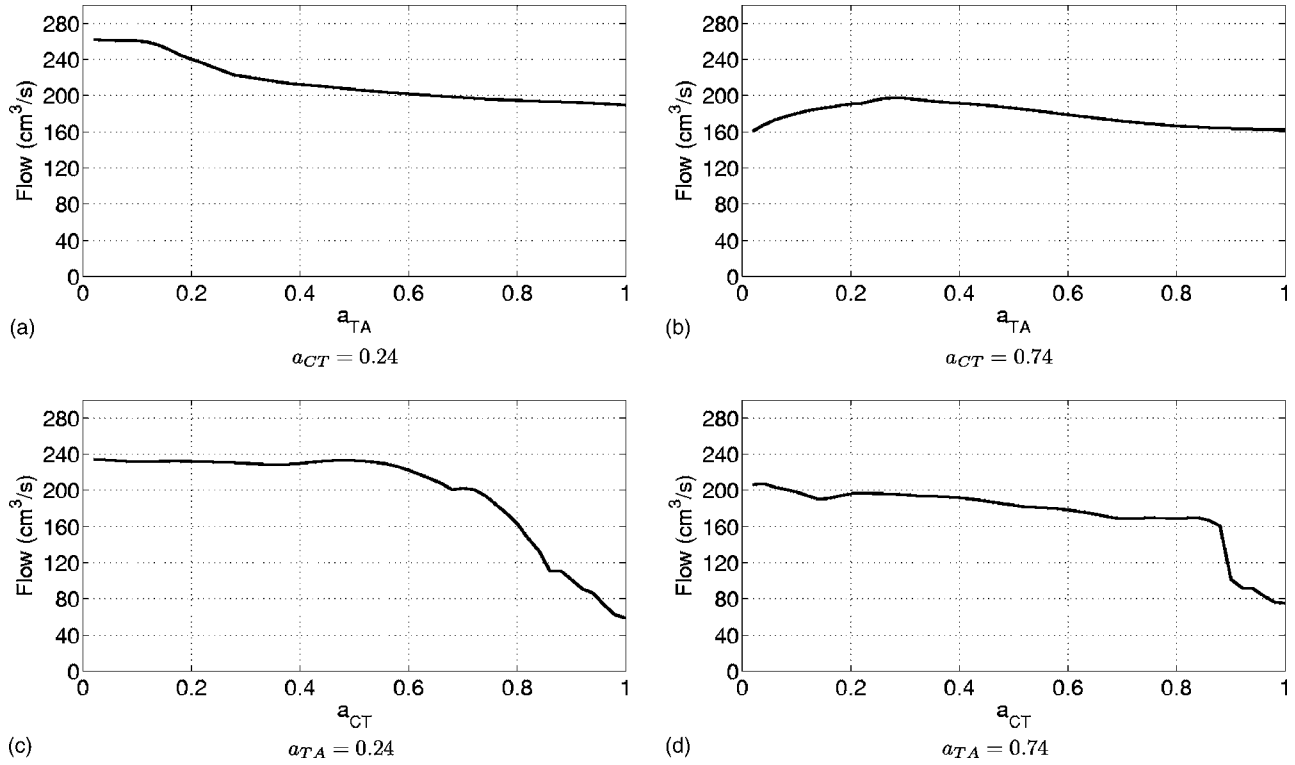


FIG. 6. Changes to maximum glottal flow as a function of increasing TA activation level when CT was constant at 0.24 (a) and 0.74 (b), and as a function of increasing CT activation when TA was constant at 0.24 (c) and 0.74 (d).

was increased and CT was held at a low level. At low TA activation levels, a rounded, more gradual flow waveform was produced that was slightly skewed to the right relative to the glottal area waveform. At high TA levels, the flow waveform had a sharp upper cutoff, with a triangular shape that followed the changes to the glottal area waveform. Changes in Max Ug as a function of increasing TA activation with CT activation constant at 0.74 (case 2) were variable and dependent on TA activation level, as shown in Fig. 6(b). Max Ug generally increased between the TA activation levels of 0.0 to approximately 0.28, and then generally decreased for TA values greater than 0.28.

When TA activation was held at 0.24 (case 3), Max Ug decreased as CT activation was increased beyond 0.60, as demonstrated in Fig. 6(c). Changes in Max Ug as CT activation was increased and TA activation was held at 0.74 (case 4) were small but generally in a decreasing direction, as shown in Fig. 6(d).

C. Maximum flow declination rate (MFDR)

MFDR changes varied when CT activation was held at a constant, low level and TA activation was increased (case 1), as shown in Fig. 7(a). MFDR showed an initial decrease as TA activation was increased from 0 to approximately 0.24. Beyond those low TA values, MFDR showed a gradual, steady increase as TA activation was increased. As demonstrated in Fig. 7(b), MFDR showed a gradual, small decrease as TA activation increased and CT activation was held at a constant, high level (case 2). This decrease leveled off at TA values of approximately 0.70 to 0.80.

As shown in Fig. 7(c), when CT activation was increased and TA activation was held at a constant, low value (case 3), changes to MFDR varied by the level of CT activation. Little change in MFDR was evidenced as CT activation was increased to approximately 0.60. As CT values were increased from 0.60 to 1.0, MFDR showed a fluctuating but substantial decrease. MFDR decreased substantially (from approximately 1 400 000 to 350 000 cm^3/s^2) as CT activation was increased and TA activation was held constant at 0.74 [case 4, Fig. 7(d)].

D. Amplitude ratio (of lower to upper mass)

The amplitude ratios of lower to upper cover masses gradually increased as TA activation increased when CT was held at 0.24 (case 1), as demonstrated in Fig. 8(a). The displacement waveforms shown in Figs. 2(b) and 3(b) exemplify the contrast in amplitude ratios between low and high TA activation conditions. Amplitude ratios minimally changed as TA activation increased and CT activation stayed constant at 0.74 (case 2), as shown in Fig. 8(b).

Amplitude ratios of the lower to upper cover masses showed minimal change as CT activation increased and TA was held at 0.24 [case 3, Fig. 8(c)]. Amplitude ratios decreased as CT activation increased beyond 0.3 and TA was held at 0.74 (case 4), as shown in Fig. 8(d).

E. Vertical phase difference (of the upper and lower cover masses)

With CT activation constant at 0.24 (case 1), vertical phase difference increased substantially (from approximately

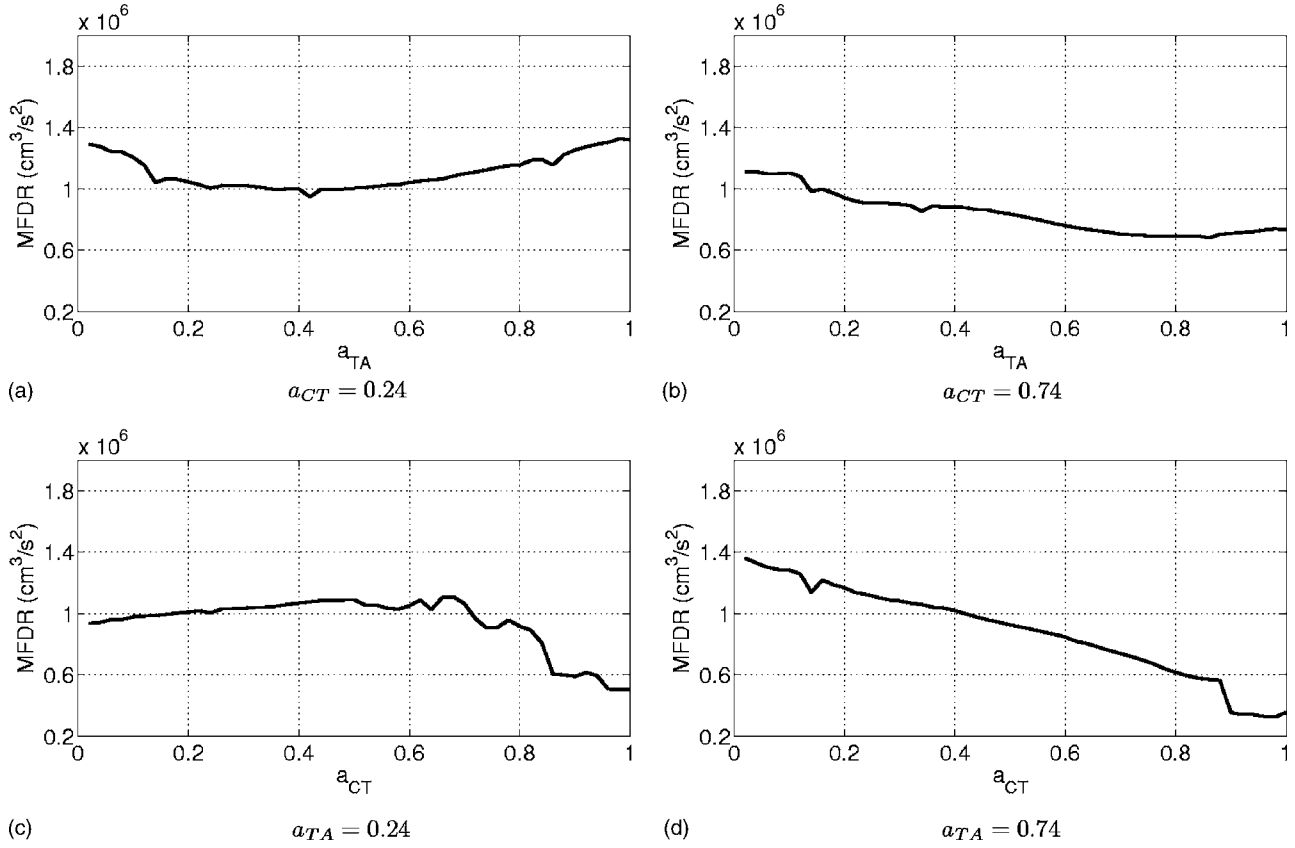


FIG. 7. Changes to maximum flow declination rate as a function of increasing TA activation level when CT was constant at 0.24 (a) and 0.74 (b), and as a function of increasing CT activation when TA was constant at 0.24 (c) and 0.74 (d).

5 to 81 deg) as TA activation increased [Fig. 9(a)]. Figure 9(b) shows that when CT activation was held at 0.74 (case 2), little change in vertical phase difference was demonstrated as TA activation was increased.

At a constant TA activation level of 0.24 (case 3), vertical phase difference decreased (from approximately 58 deg to approximately 17 deg) as CT activation increased to approximately 0.58 [Fig. 9(c)]. Between CT levels of 0.58 and 1.0, vertical phase difference leveled off and then increased slightly. When TA activation was held at 0.74 (case 4), vertical phase difference decreased greatly (approximately 84 deg to as low as 6 deg) as CT activation was increased [Fig. 9(d)].

IV. DISCUSSION

The purpose of this study was to investigate the predicted effects of independent manipulations to CT and TA muscle activation levels on F_0 and vocal fold vibration characteristics of adult males. These intrinsic laryngeal muscles are vital to F_0 control. Understanding the effects of these muscles on aerodynamic and physical quantities of vocal fold vibration, without the complications of other intrinsic and extrinsic laryngeal factors, can provide important insights regarding normal and disordered voice physiology. However, systematic increase of one muscle only, while controlling activation of the antagonist muscle, is difficult for a real speaker to achieve. Use of the three-mass model for vocal fold vibration provided a tool for predicting the independent, simulated effects of manipulating the CT and TA muscles during vocal fold vibration. This study is the first to provide predicted, causal effects of CT and TA activation on vocal fold aerodynamics and biomechanics.

A. Effect on fundamental frequency

Simulated effects of CT activation on F_0 highlighted the contrasting degree to which CT affects F_0, dependent on the level of TA activation present. Whereas increased CT activation consistently resulted in increased F_0, the greatest degree of F_0 change occurred with low TA activation levels when CT activation levels exceeded 60% of the maximum. When TA activation was high, changes in F_0 that resulted from increased CT were much smaller. This can be explained through the cover-body theory of vibration by the notion that at high TA levels, the vocal fold body is already quite stiff. Increases in CT activation would therefore be less effective in increasing overall vocal fold tension, and the increase in F_0 would be less pronounced. The simulated effects of CT activation in this study support the findings from electromyography studies showing the CT muscle to be a primary controller of F_0 (Atkinson, 1978; Faaborg-Andersen *et al.*, 1967).

The simulated effects of TA activation support the notion that TA activation can either raise or lower F_0, as predicted by Hirano (1974) and Titze *et al.* (1989). Electromyography recordings from a small number of subjects (Titze *et al.*, 1989) provided preliminary evidence supporting the bio-

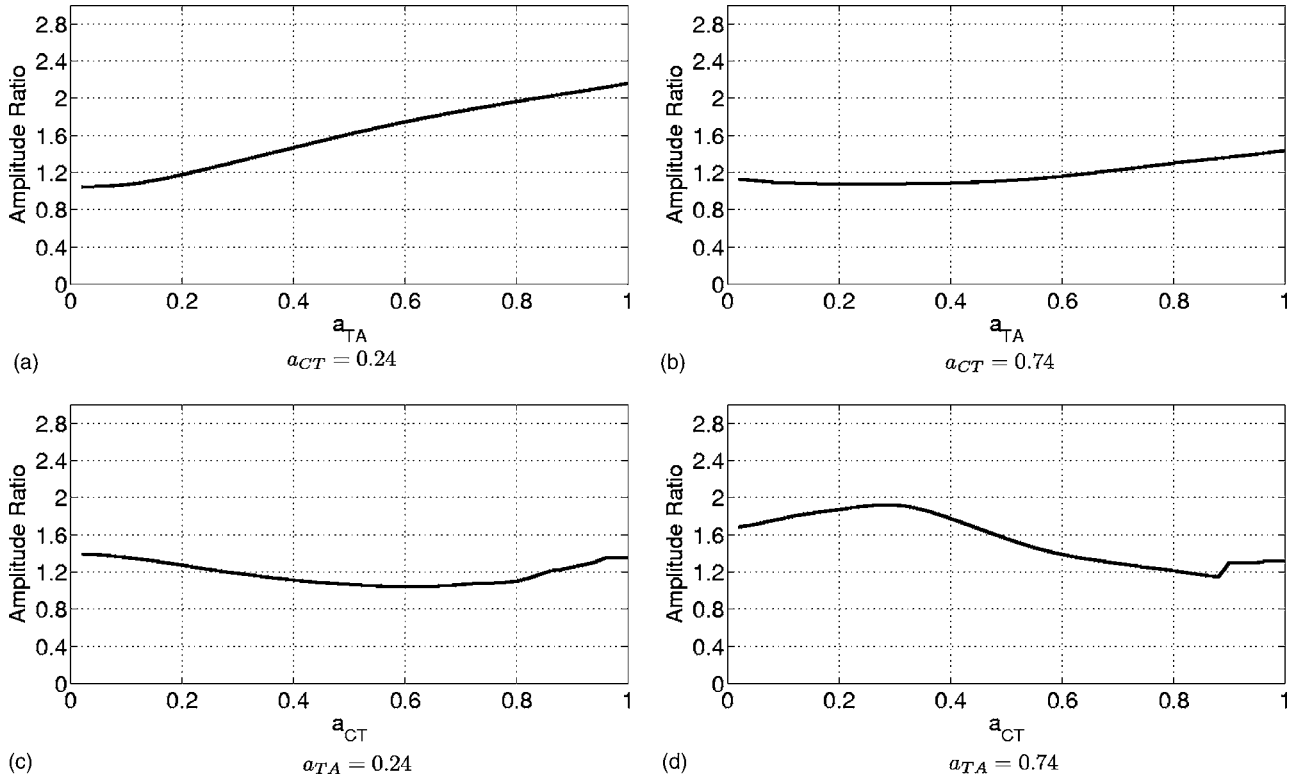


FIG. 8. Changes to amplitude ratio as a function of increasing TA activation level when CT was constant at 0.24 (a) and 0.74 (b), and as a function of increasing CT activation when TA was constant at 0.24 (c) and 0.74 (d).

mechanical theory outlined by Titze *et al.* (1989). However, the findings from this study expanded on that data by running simulations for thousands of muscle activation combinations and determining the precise conditions in which TA activation was predicted to raise or lower F_0, dependent on concurrent CT activation. With low CT activation levels, TA activation increased F_0 up to approximately 60% of its maximum activation, without further changes to F_0 above that activation level. In contrast, when CT activation was high, increasing TA activation resulted in *decreased* F_0 until about mid TA activation levels, after which no further F_0 change was realized. At low CT and TA activation levels, increased TA activation would increase the tension in the relatively slack vocal folds, resulting in an increased F_0 for the fundamental frequencies of approximately 100 to 150 Hz in this adult-male model. In contrast, with high CT activation and low TA activation levels (yielding high fundamental frequencies), the vocal folds would be stiff and tense, and any increased TA activation would shorten the vocal folds and produce sufficient reduction in tension to result in a lowering of F_0.

B. Effect on aerodynamic and physical quantities

Modification of CT and TA muscle activation levels had several pronounced effects on aerodynamic and physical quantities of vocal fold vibration. These quantities influence vocal fold impact stress and shearing stress, parameters that contribute to overall mechanical stress during vocal fold vibration (Titze, 1994). Glottal area and glottal airflow generally decreased as both TA and CT muscle activation levels were increased. Decreased airflow during vibration may be optimal for a speaker who is trying to conserve airflow and driving pressure, and may allow the speaker to limit their frequency of respiratory replenishment or depth of inspiration. However, if this airflow conservation occurs as a result of increased intrinsic laryngeal muscle activation (TA and/or CT), the cost to the speaker relative to muscle expenditure may outweigh the airflow conservation benefits.

Maximum flow declination rate (MFDR) has been used as an indicator of velocity of vocal fold closure (Hillman *et al.*, 1989), and increased MFDR may be associated with increased vocal fold collision forces. MFDR was generally at its highest when the simulated difference in levels of CT and TA activation was greatest. Specifically, when muscle combinations of high TA activation were coupled with low CT activation, MFDR was high. As CT and TA activation levels approached each other, MFDR generally decreased. Therefore, when high TA muscle activation is used with low CT muscle activation, velocity of vocal fold closure may be increased and may result in increased collision forces or impact stress of the vocal folds. When simulated activation for both the CT and TA were high, MFDR values were quite low. Interestingly, speakers have been observed to use approximately equal increases in both CT and TA activation when increasing F_0 (Titze *et al.*, 1989), effectively utilizing a diagonal from the lower left and upper right corners in the muscle activation plot (Fig. 4). This muscle use strategy would apparently result in lower MFDR values. It should be noted that, at times, an increase in MFDR is desirable, such as when increased voice intensity is needed. Two trained male singers showed different muscle use strategies for increasing F_0, primarily relying on increased CT activation

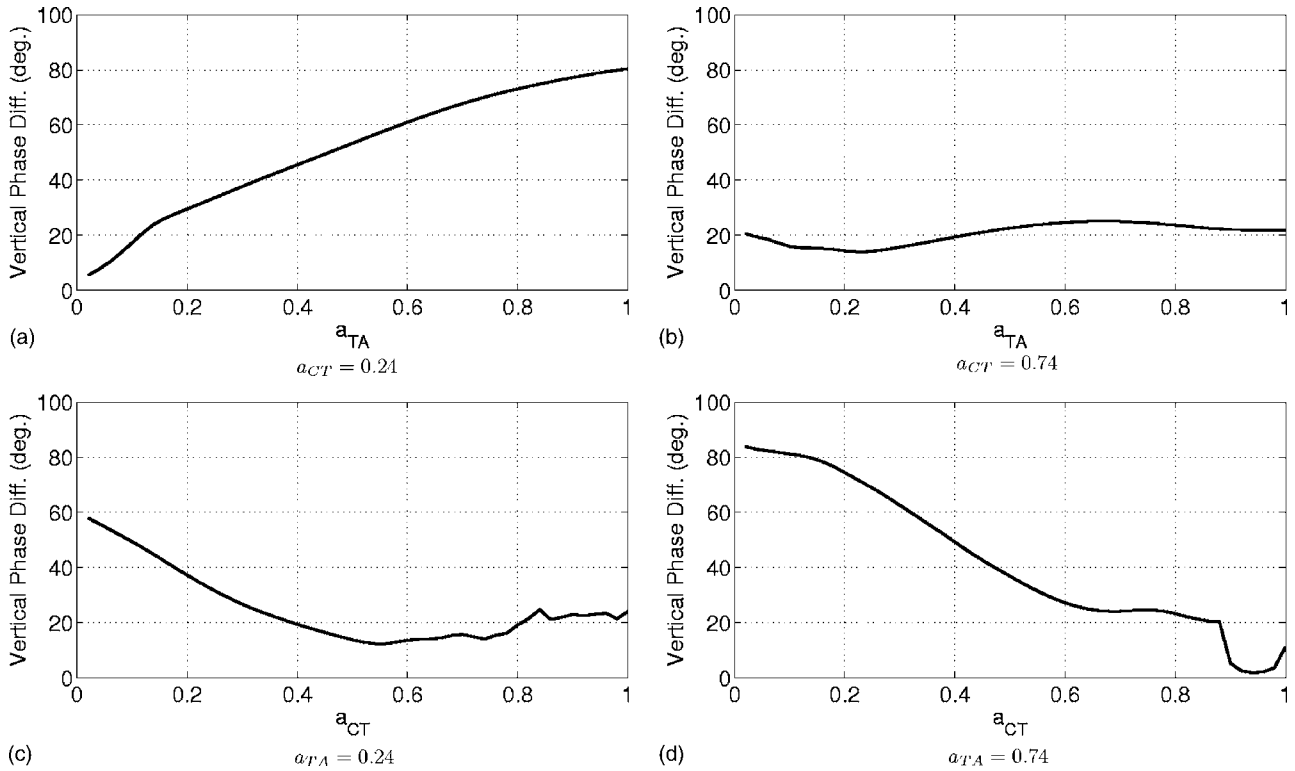


FIG. 9. Changes to vertical phase difference as a function of increasing TA activation level when CT was constant at 0.24 (a) and 0.74 (b), and as a function of increasing CT activation when TA was constant at 0.24 (c) and 0.74 (d).

while maintaining lower TA activation (Titze *et al.*, 1989). This strategy might result in increased MFDR, which would be important for a singer who needs to achieve both increased F_0 and increased intensity.

The amplitude ratio and vertical phase difference of the lower to upper cover masses are important to the self-oscillating nature of vocal fold vibration (Titze, 1988), but excessive vertical phase differences can also contribute to shearing forces (Titze, 1994). No previous studies have documented the effects of varying levels of intrinsic laryngeal muscle activation on these quantities. In this simulation study, substantial changes in the amplitude ratio occurred when TA activation was increased and CT muscle activation was low, creating a large CT to TA activation differential. In this case, the amplitude ratio more than doubled. Thus, at a low F_0 range, when the vocal folds would be relatively slack, increased TA activation resulted in increased excursion of the lower cover mass and decreased excursion of the upper cover mass. These changes were accompanied by a large increase in vertical phase difference between the upper and lower cover masses, due to the upper mass not paralleling the increased excursion of the lower mass, with resulting differences in phase as TA muscle activation was increased. These increases in amplitude ratio and vertical phase difference resulted in substantial changes to the glottal area and glottal flow waveforms. When TA activation was high and CT activation was low, the glottal area waveform became very sharp, with an abrupt cutoff point between the opening and closing phases of vibration.

Abrupt changes to vocal fold dynamics can result in increased mechanical stress during vibration. An increase in vertical phase difference of the upper and lower cover masses during vibration will result in an increase in shearing forces on the vocal folds; the greater phase asymmetry of the upper and lower masses would mean increased shearing of the cover tissue that connects these masses. If shearing forces are harmful to vocal fold tissue, increased levels of TA activation may place an individual at risk for damage or change to the vocal fold tissue due to an increase in shearing forces that occurs as TA activation is increased. The simulated effects of muscle activation support Hillman *et al.* (1989), who suggested that increased levels of muscle activation associated with adducted hyperfunction might result in increased vocal fold stiffness, high velocity of tissue movement, and an increase in vocal fold collision forces, all contributing to an increased risk of vocal fold pathology. These authors theorized that the vocal fold dynamics associated with increased muscle activation would also result in increased amplitude of vocal fold excursion, in turn contributing to increased collision forces. The simulations from the present study indicated that with increased TA activation and relatively low CT activation, the largest increase in amplitude excursion will occur in the lower cover mass, with a greater amplitude differential at these muscle activation levels between the lower and upper cover mass movement. Therefore, in addition to the increased impact forces, increased shearing forces may compound the mechanical stress that is placed on the vocal folds.

There are few studies documenting intrinsic laryngeal muscle activation levels in people with voice disorders. Most of the available literature has been conducted with individuals presenting with spasmodic dysphonia, a voice disorder of

neurologic origin. Several investigators have not found differences in intrinsic laryngeal muscle activation between control subjects and those with spasmodic dysphonia (Van Pelt *et al.*, 1994; Watson *et al.*, 1991). Interestingly, the studies that have found group differences have consistently reported increased activity of the thyroarytenoid muscle in people with spasmodic dysphonia as compared to control subjects (Cyrus *et al.*, 2001; Nash and Ludlow, 1996; Schaefer *et al.*, 1992). Speech samples that elicited these group differences varied from phonation breaks only (Nash and Ludlow, 1996) to speech with and without breaks (Cyrus *et al.*, 2001) to repeated consonant-vowel-consonant tasks (Schaefer *et al.*, 1992). One hypothesis based on the predictions obtained from this study is that increased TA activation may alter displacement characteristics of the vocal fold cover and body, contributing to decreased phonatory stability during vibration. In voice disorders that occur in the absence of neurologic or structural laryngeal pathology (often referred to as functional voice disorders), increased intrinsic and extrinsic laryngeal muscle activation is frequently assumed but has generally not been objectively documented (Eustace *et al.*, 1996; Morrison and Rammage, 1993; Morrison *et al.*, 1983; Stemple *et al.*, 1995). The simulations produced in this study indicate that certain combinations of intrinsic laryngeal muscle activation may result in aerodynamic and physical characteristics of vocal fold vibration that could place an individual at risk for voice problems. However, modeling studies must be followed by *in vivo* speech recordings in people with and without voice disorders to validate these predictions.

There are several limitations to the present study, and critical future directions for research in this area. To simplify the interpretation of results, only a realistic epilaryngeal and tracheal configuration was included, with the remaining vocal tract modeled as an open tube. To more realistically depict the acoustic, aerodynamic, and physical changes associated with variation of muscle activation and epilaryngeal area, modeling of particular vocal tract configurations representing vowels such as /a/ or /i/ would be useful. Furthermore, this study controlled the configuration of the trachea and held input pressure (lung pressure) at a constant value to isolate the effects of CT and TA muscle activation. Manipulation of parameters such as driving pressure would be expected to influence F_0 (Baer, 1979; Hixon *et al.*, 1971). An important step in future studies would be manipulating such parameters in conjunction with CT and TA muscle activations and determining the predicted outcomes. Likewise, the output parameters assessed in this study were limited so that the length and interpretability of the predictions would not be too unwieldy. Due to the finding of increased skewing of the glottal pulse under simulated conditions of high TA activation and low CT activation in this study, future studies might include the speed quotient as a measure of the symmetry of the open phase (Baken and Orlikoff, 2000).

A major limitation to all vocal fold modeling studies is that the effects that are obtained by experimentally manipulating various parameters may not be evidenced in the human with those same manipulations. In this study, it was possible to manipulate intrinsic laryngeal muscles independent of other factors. In humans, these muscle changes would occur with a probable concomitant increase in other laryngeal and pharyngeal muscles, and the effects of those muscle changes would not be reflected in the present modeling study. Thus, modeling studies can provide predictions regarding the effects of controlled manipulation of variables, but must be followed by *in vivo* studies to validate these predictions. In attempting to draw inferences regarding changes to aerodynamic and physical quantities and their implications for risk of vocal fold damage, it is also important to note that the three-mass model of vibration cannot depict tissue damage. Thus, the actual risk for tissue damage associated with these changes to vocal fold dynamics is unknown. Finally, the current model has been developed on male speakers and may have limited applicability to female voice production. Future modifications to this computational model are therefore needed to adequately represent female voice biomechanics, as many voice disorders occur more frequently in women.

C. Conclusions

The number of variables that affect F_0 in speech and the interdependence of these variables make the study of F_0 control difficult *in vivo*. The three-mass model of adult-male vocal fold vibration allows for the isolated simulation of several variables that are critical to F_0 control, such as CT and TA muscle activation. By manipulating activation of one intrinsic laryngeal muscle while holding other variables constant, the simulated effects of that muscle on F_0, as well as on aerodynamic and physical characteristics of vibration, can be studied. The aerodynamic and physical quantities analyzed in this study were chosen due to their contribution to vocal fold dynamics and their influence on various forms of mechanical stress during vibration. F_0 was greatly affected by the simulated manipulation of CT and TA muscle activation, as were the aerodynamic quantities of glottal airflow and MFDR. Physical quantities of amplitude ratio and vertical phase difference were also affected by simulated muscle activation. A simulated increase in TA activation with relatively low CT activation substantially increased both the amplitude ratio and vertical phase difference. These aerodynamic and physical changes would be expected to increase both vocal fold collision forces and shearing forces, which may increase the potential for vocal fold tissue damage.

Alipour-Haghighi, F., and Titze, I. R. (**1983**). "Simulation of particle trajectories of vocal fold tissue during phonation," *in Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, CO), pp. 183–190.

Alipour-Haghighi, F., and Titze, I. R. (**1991**). "Elastic models of vocal fold tissues," J. Acoust. Soc. Am. **90**, 1326–1331.

Atkinson, J. E. (**1978**). "Correlation analysis of the physiological factors controlling fundamental voice frequency," J. Acoust. Soc. Am. **63**, 211–222.

Bachorowski, J.-A., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.

Baer, T. (**1979**). "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes," J. Acoust. Soc. Am. **65**, 1271–1275.

Baken, R. J., and Orlikoff, R. F. (**2000**). *Clinical Measurement of Speech and Voice, 2nd ed.* (Singular, Thomson Learning, San Diego, CA).

Baker, K. K., Ramig, L. A., Sapir, S., Luschei, E. S., and Smith, M. E.

(**2001**). "Control of vocal loudness in young and old adults," J. Speech Lang. Hear. Res. **44**, 297–305.
Boone, D. R., and McFarlane, S. C. (**2000**). *The Voice and Voice Therapy, 6th ed.* (Allyn and Bacon, Needham Heights, MA).
Brown, B. L., Strong, W. J., and Rencher, A. C. (**1974**). "Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech," J. Acoust. Soc. Am. **55**, 313–318.
Case, J. L. (**2002**). *Clinical Management of Voice Disorders, 4th ed.* (Pro-ed, Austin, TX).
Courey, M. S., Shohet, J. A., Scott, M. A., and Ossoff, R. H. (**1996**). "Immunohistochemical characterization of benign laryngeal lesions," Ann. Otol. Rhinol. Laryngol. **105**, 525–531.
Cyrus, C. B., Bielamowicz, S., Evans, F. J., and Ludlow, C. L. (**2001**). "Adductor muscle activity abnormalities in abductor spasmodic dysphonia," Otolaryngol.-Head Neck Surg. **124**(1), 23–30.
Eustace, C. S., Stemple, J. C., and Lee, L. (**1996**). "Objective measures of voice production in patients complaining of laryngeal fatigue," J. Voice **10**(2), 146–154.
Faaborg-Andersen, K., Yanagihara, N., and von Leden, H. (**1967**). "Vocal pitch and intensity regulation: A comparative study of electrical activity in the cricothyroid muscle and the airflow rate," Arch. Otolaryngol. **85**, 122–128.
Finnegan, E. M., Luschei, E. S., and Hoffman, H. T. (**2000**). "Modulations in respiratory and laryngeal activity associated with changes in vocal intensity," J. Speech Lang. Hear. Res. **43**, 934–950.
Gay, T., Hirose, H., Strome, M., and Sawashima, M. (**1972**). "Electromyography of the intrinsic laryngeal muscles during phonation," Ann. Otolaryngol. **81**, 401–409.
Gelfer, M. P., and Schofield, K. J. (**2000**). "Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male," J. Voice **14**, 22–33.
Gray, S. D., and Titze, I. R. (**1988**). "Histologic investigation of hyperphonated canine vocal cords," Ann. Otol. Rhinol. Laryngol. **97**, 381–388.
Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., and Vaughan, C. (**1989**). "Objective assessment of vocal hyperfunction: An experimental framework and initial results," J. Speech Hear. Res. **32**, 373–392.
Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., and Vaughan, C. (**1990**). "Phonatory function associated with hyperfunctionally related vocal fold lesions," J. Voice **4**(1), 52–63.
Hirano, M. (**1974**). "Morphological structure of the vocal cord and its variations," Folia Phoniatr. **26**, 89–94.
Hirano, M. (**1988**). "Behavior of laryngeal muscles of the late William Vennard," J. Voice **2**(4), 291–300.
Hirano, M., Ohala, J., and Vennard, W. (**1969**). "The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation," J. Speech Hear. Res. **12**, 616–628.
Hirano, M., Vennard, W., and Ohala, J. (**1970**). "Regulation of register, pitch and intensity of voice: An electromyographic investigation of intrinsic laryngeal muscles," Folia Phoniatr. **22**, 1–20.
Hixon, T. J., Klatt, D., and Mead, J. (**1971**). "Influence of forced transglottal pressure changes on vocal fundamental frequency," J. Acoust. Soc. Am. **49**, 105(A).
Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84**, 511–529.
Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1989**). "Glottal airflow and transglottal air pressure measurements for male and female speakers at low, normal, and high pitch," J. Voice **3**(4), 294–305.
Holmberg, E. B., Hillman, R. E., Perkell, J. S., and Gress, C. (**1994**). "Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings," J. Speech Hear. Res. **37**, 484–495.
Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.
Jacques, R. D., and Rastatter, M. P. (**1990**). "Recognition of speaker age from selected acoustic features as perceived by young and older listeners," Folia Phoniatr. **42**(3), 118–124.
Jiang, J., and Titze, I. R. (**1994**). "Measurement of vocal fold intraglottal pressure and impact stress," J. Voice **8**(2), 132–144.
Kempster, G. B., Larson, C. R., and Kistler, M. K. (**1988**). "Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency," J. Voice **2**(3), 221–229.
Liljencrants, J. (**1985**). "Speech synthesis with a reflection-type analog," DS dissertation, Dept. of Speech Commun. and Music Acoust., Royal Inst. of Tech., Stockholm, Sweden.
Mathworks (**2004**). MATLAB 7 Version 4; (The MathWorks, Inc., Natick, MA).
Morrison, M. D., and Rammage, L. A. (**1993**). "Muscle misuse voice disorders: Description and classification," Acta Oto-Laryngol. **113**, 428–434.
Morrison, M. D., Rammage, L. A., Belisle, G. M., Pullan, C. B., and Nichol, H. (**1983**). "Muscular tension dysphonia," J. Otolaryngol. **12**(5), 302–306.
Nash, E. A., and Ludlow, C. L. (**1996**). "Laryngeal muscle activity during speech breaks in adductor spasmodic dysphonia," Laryngoscope **106**(4), 484–489.
Nasri, S., Sercarz, J. A., Azizzadeh, B., Kreiman, J., and Berke, G. S. (**1994**). "Measurement of adductory force of individual laryngeal muscles in an in vivo canine model," Laryngoscope **104**(10), 1213–1218.
Poletto, C. L., Verdun, L. P., Strominger, R., and Ludlow, C. L. (**2004**). "Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures," J. Appl. Physiol. **97**, 858–866.
Schaefer, S. D., Roark, R. M., Watson, B. C., Kondraske, G. V., Freeman, F. J., Butsch, R. W., *et al.* (**1992**). "Multichannel electromyographic observations in spasmodic dysphonia patients and normal control subjects," Ann. Otol. Rhinol. Laryngol. **101**, 67–75.
Shipp, T., and McGlone, R. E. (**1971**). "Laryngeal dynamics associated with voice frequency change," J. Speech Hear. Res. **14**, 761–768.
Stemple, J. C., Stanley, J., and Lee, L. (**1995**). "Objective measures of voice production in normal subjects following prolonged voice use," J. Voice **9**(2), 127–133.
Story, B. H. (**1995**). "Physiologically-based speech simulation using an enhanced wave-reflection model of the vocal tract," Ph.D. dissertation, University of Iowa.
Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.
Tanaka, S., and Tanabe, M. (**1986**). "Glottal adjustment for regulating vocal intensity," Acta Oto-Laryngol. **102**, 315–324.
Thorsen, N. G. (**1980**). "A study of the perception of sentence intonation-evidence from Danish," J. Acoust. Soc. Am. **67**, 1014–1030.
Titze, I. R. (**1988**). "The physics of small amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1551.
Titze, I. R. (**1991**). "Mechanisms underlying the control of fundamental frequency," *in Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by J. Gauffin and B. Hamarberg (Singular, San Diego, CA), pp. 129–138.
Titze, I. R. (**1994**). "Mechanical stress in phonation," J. Voice **8**(2), 99–105.
Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Iowa City).
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., and Story, B. H. (**2002**). "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am. **112**, 1064–1076.
Titze, I. R., Luschei, E. S., and Hirano, M. (**1989**). "Role of the thyroarytenoid muscle in regulation of fundamental frequency," J. Voice **3**(3), 213–224.
Van den Berg, J., and Tan, T. S. (**1959**). "Results of experiments with human larynxes," Pract. Otorhinolaryngol. (Basel) **21**, 425–450.
Van Pelt, F., Ludlow, C. L., and Smith, P. J. (**1994**). "Comparison of muscle activation patterns in adductor and abductor spasmodic dysphonia," Ann. Otol. Rhinol. Laryngol. **103**, 192–200.
Watson, B. C., Schaefer, S. D., Freeman, F. J., Dembowski, J., Kondraske, G. V., and Roark, R. M. (**1991**). "Laryngeal electromyographic activity in adductor and abductor spasmodic dysphonia," J. Speech Hear. Res. **34**, 473–482.

Application of spectral subtraction method on enhancement of electrolarynx speech

Hanjun Liu, Qin Zhao, Mingxi Wan,[a] and Supin Wang
The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, People's Republic of China

(Received 7 December 2004; revised 11 April 2006; accepted 11 April 2006)

Although electrolarynx (EL) serves as an important method of phonation for the laryngectomees, the resulting speech is of poor intelligibility due to the presence of a steady background noise caused by the instrument, even worse in the case of additive noise. This paper investigates the problem of EL speech enhancement by taking into account the frequency-domain masking properties of the human auditory system. One approach is incorporating an auditory masking threshold (AMT) for parametric adaptation in a subtractive-type enhancement process. The other is the supplementary AMT (SAMT) algorithm, which applies a cross-correlation spectral subtraction (CCSS) approach as a post-processing scheme to enhancing EL speech dealt with the AMT method. The performance of these two algorithms was evaluated as compared to the power spectral subtraction (PSS) algorithm. The best performance of EL speech enhancement was associated with the SAMT algorithm, followed by the AMT algorithm and the PSS algorithm. Acoustic and perceptual analyses indicated that the AMT and SAMT algorithms achieved the better performances of noise reduction and the enhanced EL speech was more pleasant to human listeners as compared to the PSS algorithm.
 [DOI: 10.1121/1.2203592]

PACS number(s): 43.70.Dn [DOS] Pages: 398–406

I. INTRODUCTION

Electrolarynix (EL) is used by people who are unable to use their natural voice production due to the total removal of the larynx for the treatment of laryngeal cancer. The most common EL is the electromechanical vibrator that is typically held against the neck at the level of the former glottis to excite the vocal tract acoustically (Barney *et al.*, 1959). The advantages of EL are that it is easier to use, produces longer sentences without special care, and is more effective for communication in many situations as compared to other methods of voice rehabilitation (Lauder, 1970; Rothman, 1982). However, EL speech has several serious shortcomings including artificial quality, reduced intelligibility, and poor audibility. Some of the sound produced by the vibrating diaphragm is radiated directly from the instrument, its interface with the neck and the surrounding neck tissues. Barney *et al.* (1959) reported that the intensity of the radiated noise was about 20–25 dB (SPL) when the mouth was closed. Weiss *et al.* (1979) found that this value varied over 4–15 dB SPL across the subjects for the same device. Previous researchers suggest that the radiated noise may result in a loss of intelligibility, especially at the low signal-to-noise ratio (SNR) (Knox and Anneberg, 1973). An acoustic cue that distinguishes voiced and voiceless stops is the presence of a periodic low-frequency signal during the closed portions of voiced stops. Due to continuous operating of EL throughout the utterance, however, the closure portion consisting of the periodic radiated noise leads to the confusions between voiced and voiceless stops (Espy-Wilson *et al.*, 1998). Weiss *et al.* (1979) reported that most of the direct-radiated noise energy was concentrated in the frequency region 400–800 Hz, which could lead to vowel identification errors due to a significant auditory masking of the vowel formants.

Although EL has been clinically used for laryngectomees for over 50 years, there have been few scientific efforts to improve EL devices or the resulting speech. Norton and Bernstein (1993) tried to improve EL speech by changing the driving signal of the vibration source. By applying a one-inch-thick foam shield around the EL, they found that listeners judged the modified EL speech as sounding more natural. However, Espy-Wilson *et al.* (1998) noted that the soundproof shield failed to provide effective noise isolation and increased the size of EL. Qi and Weinberg (1991) designed a digital filter to compensate the low-frequency deficit, minimizing differences between spectra of normal and EL vowels below 550 Hz using a least-squares estimation procedure. They found that the low-frequency-enhancing EL speech was judged more intelligible than the original speech.

The limited effectiveness of the above-mentioned efforts led researchers to consider the use of signal enhancement techniques to improve the resulting speech (Espy-Wilson *et al.*, 1996, 1998; Cole *et al.*, 1997; Pandey *et al.*, 2002; Niu *et al.*, 2003; Pratapwar *et al.*, 2003; Liu *et al.*, 2006). One method is adaptive noise canceling (Espy-Wilson *et al.*, 1996, 1998; Niu *et al.*, 2003). Espy-Wilson *et al.* (1996, 1998) used a two-input least mean squares (LMS) algorithm, which removes the noise components of the primary input signal that depend on the reference input signal and are based on second-order statistics. Niu *et al.* (2003) proposed an adaptive noise canceling method-based independent component analysis (ICA) and found a better performance than the

[a]Electronic mail: mxwan@mail.xjtu.edu.cn

 0001-4966/2006/120(1)/398/9/$22.50

LMS algorithm. However, there may be many other noise components existing in the primary input signal that depend on the noise reference signal through higher-order statistics. The subtractive-type algorithm is the other method for EL speech enhancement, including power spectral subtraction (PSS) (Cole *et al.*, 1997), magnitude spectral subtraction (MSS) (Pandey *et al.*, 2002), spectral subtraction with quantile based noise estimation (SS-QBNE) (Pratapwar *et al.*, 2003), or improved spectral subtraction (ISS) (Liu *et al.*, 2006). This method is based on two assumptions: one is that speech and additive noise are uncorrelated; and the other is that the noise is a stationary or a slowly varying process so that the noise spectrum does not change significantly during the update periods. The key idea is to estimate the background noise and then to subtract the estimation value from the noisy speech in the frequency domain (Boll, 1979). Subtractive-type algorithms have been chosen for their simplicity of implementation and relatively inexpensive computational complexity. However, the subtraction parameters of the PSS and the MSS algorithms used for EL speech enhancement are fixed, limiting the enhancement effects of noise reduction. Moreover, spectral subtraction can result in negative estimates of the magnitude or power spectrum that are non-negative variables, in which a musical noise is introduced. Although many solutions have been proposed to reduce the musical noise in the subtractive-type algorithms-(Boll, 1979; Berouti *et al.*, 1979; McAulay and Malpass, 1980; Ephraim and Malah, 1984; Lockwood and Boudy, 1992; Hansen, 1994), results performed with these algorithms show that there is a need for further improvement.

The purpose of this investigation was motivated by the need of improving EL speech in electronically mediated environments, when speakers and listeners cannot talk face to face. For example, during the use of a telephone, when addressing public gatherings, or in any situation in which electronic media could reasonably be employed, EL speech has to be enhanced extensively for correct understanding. Additionally, previous research was focused on the radiated noise reduction in quiet environments. Most speech communication takes place in noisy environments, and the low-energy EL speech is easily masked by the different environment noises. The reduction of speech quality due to environment noise causes listeners fatigue. In the perceptual study of EL speech in noise, it decreased in the intelligibility as SNR decreased (Weiss *et al.*, 1979; Holly *et al.*, 1983). Hence, it is important to investigate the methods to eliminate the additive noise, although few efforts are put into EL enhancement in the case of additive noise (Niu *et al.*, 2003; Liu *et al.*, 2006). It is necessary to find a new way to efficiently eliminate both additive noise and radiated noise, which will be helpful to improve the life quality of the laryngectmees.

In review of normal speech enhancement, efforts have been made to reduce the musical noise by using a human auditory masking model, which is widely used in wideband audio coding (Johnston, 1988; Brandenburg and Stoll, 1994; Painter and Spanias, 2000). This is concerned with the critical band (CB) analysis, which is a central notion because the auditory perception is based on a similar analysis in the inner ear. The auditory masking model is used for speech enhancement by calculating an auditory masking threshold (AMT), below which any noise components will not be detectable by the human listener and so perceptually are not important components to suppress. The goal, then, is to minimize only the audible portion of the noise spectrum. Some methods (Tsoukalas *et al.*, 1993, 1997; Usagawa *et al.*, 1994; Nandkumar and Hansen, 1995; Hansen and Nandkumar, 1995; Virag, 1999) by modeling several aspects of the enhancement mechanism present in the auditory system have been developed.

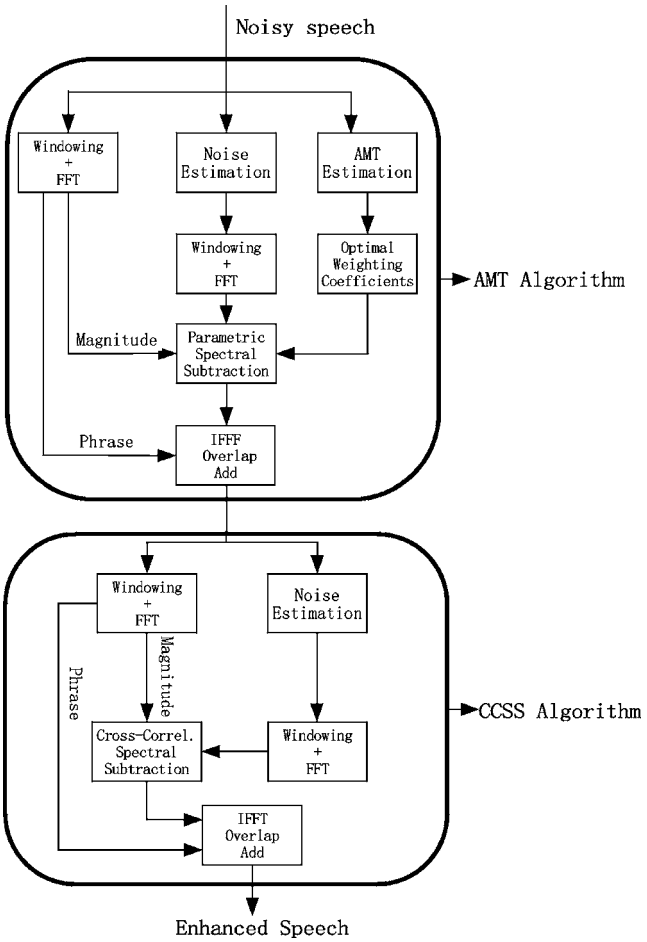


FIG. 1. The proposed speech enhancement scheme.

Therefore, the purpose of the present study was to investigate the enhancement of EL speech based on the spectral subtraction by incorporating the AMT, not only eliminating the radiated noise but also reducing the additive noise. Our goals are described as follows: (1) adopt the AMT algorithm for the enhancement of EL speech; (2) apply a supplementary AMT (SAMT) algorithm, incorporating a cross-correlation spectral subtraction (CCSS) approach as a post-processing scheme, to improving EL speech; and (3) evaluate the quality of EL speech enhanced by these two algorithms as compared to the PSS algorithm.

II. METHOD

The proposed enhancement scheme is presented in Fig. 1. It consists of two parts: the AMT algorithm and the CCSS algorithm. The whole block diagram describes the SAMT

algorithm. The key of the proposed enhancement scheme can be divided into two parts: noise estimation and spectral subtraction.

A. Noise estimation

Since babble noise is highly nonstationary noise, it is imperative to update the estimate of the noise spectrum frequently. We adopted the minimum-statistics method proposed by Cohen and Berdugo (2002) which was found to work well for nonstationary environments. The minimum tracking is based on a recursively smoothed spectrum which is estimated using first-order recursive averaging

$$|\hat{D}_{(k,l)}(\omega)|^2 = \lambda_D |\hat{D}_{(k-1,l)}(\omega)|^2 + (1-\lambda_D)|\hat{Y}_{(k,l)}(\omega)|^2 \qquad (1)$$

$$0 < \lambda_D < 1,$$

where $|\hat{D}_{(k,l)}(\omega)|^2$ and $|\hat{Y}_{(k,l)}(\omega)|^2$ are the kth components of noise spectrum and noisy speech spectrum at the frame l, and λ_D is a smooth parameter. Let $p'(k,l)$ denote the conditional signal presence probability in Cohen and Berdugo (2002), then Eq. (1) implies

$$|\hat{D}_{(k,l)}(\omega)|^2 = \hat{\lambda}_D(k,l)|\hat{D}_{(k-1,l)}(\omega)|^2 + (1-\hat{\lambda}_D(k,l))|\hat{Y}_{(k,l)}(\omega)|^2, \qquad (2)$$

where $\hat{\lambda}_D(k,l) \triangleq \lambda_D + (1-\lambda_D)p'(k,l)$ is a time-varying smoothing parameter. Therefore, the noise spectrum can be estimated by averaging past spectral power values.

B. Spectral subtraction

1. Modified spectral subtraction

Spectral subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noisy signal model in the frequency domain is expressed as follows:

$$Y(\omega) = S(\omega) + L(\omega), \qquad (3)$$

where $Y(\omega)$, $S(\omega)$, and $L(\omega)$ are the fast Fourier transforms (FFT) of the noisy speech, clean speech, and additive stationary background noise.

In order to reduce the musical noise, various existing subtraction rules are derived and most of them have a parametric form allowing for a greater flexibility in the variation of the suppression curves (Lim and Oppenheim, 1979). According to the subtractive-type algorithm proposed by Berouti *et al.* (1979), the generalized spectral subtraction scheme is described as follows:

$$|\hat{S}(\omega)|^\gamma = \begin{cases} |Y(\omega)|^\gamma - \alpha|\hat{L}(\omega)|^\gamma, & \text{if } \dfrac{|\hat{L}(\omega)|^\gamma}{|Y(\omega)|^\gamma} < \dfrac{1}{\alpha+\beta} \\ \beta|\hat{L}(\omega)|^\gamma, & \text{otherwise,} \end{cases} \qquad (4)$$

where $\alpha(\alpha > 1)$ is the subtraction factor, $\beta(0 \le \beta \le 1)$ is the spectral floor, and γ is the exponent determining the transition sharpness. Here we set $\gamma=2$ but α and β are adapted frame by frame.

2. Cross-correlation spectral subtraction

As a post-processing scheme, the input signal of the CCSS algorithm is EL speech enhanced by the AMT algorithm (see Fig. 1). For this noisy signal, the assumption that speech and noise are uncorrelated is not valid. Therefore, we cannot neglect those cross terms between speech and noise. The CCSS algorithm proposed by Hu *et al.* (2001) can be expressed

$$|\hat{S}'(\omega)|^2 = \begin{cases} |Y'(\omega)|^2 - \alpha|\hat{L}'(\omega)|^2 - \delta|Y'(\omega)| \cdot |\hat{L}'(\omega)| & \text{if } |Y'(\omega)|^2 > \alpha|\hat{L}'(\omega)|^2 \\ \beta|\hat{L}'(\omega)|^2 & \text{otherwise,} \end{cases} \qquad (5)$$

where $\gamma'(\omega)$, $S'(\omega)$, and $L'(\omega)$ are the FFT of the input signal of the CCSS algorithm, and δ is the cross-correlation coefficient, which provides an estimate of the correlation between corrupted speech and noise in the current window frame. The value of δ determines the factor of subtraction and is proportional to the degree of correlation between speech and noise.

Considering the fact that the uncorrelated noise ($|\hat{L}'(\omega)|^2$) has been eliminated during the enhancement of the AMT algorithm and it will result in the oversubtraction if it is subtracted in the post-processing scheme, we only subtract the correlated noise ($\delta|Y'(\omega)| \cdot |\hat{L}'(\omega)|$) in the CCSS algorithm

$$|\hat{S}'(\omega)|^2 = |Y'(\omega)|^2 - \delta|Y'(\omega)| \cdot |\hat{L}'(\omega)|. \qquad (6)$$

3. Adaptation of subtraction parameters

The AMT is obtained through modeling the frequency selectivity of the human ear and its masking properties (Johnston, 1988; Schroeder *et al.*, 1979; Arehart *et al.* 2003). Figure 2 shows an example of the AMT from the vowels /a/ of normal speech and EL speech in the quiet environment. It can be found that the AMT values in the low-frequency regions are higher than those in the high-frequency regions. On the other hand, the AMT values of EL speech are lower than those of normal speech.

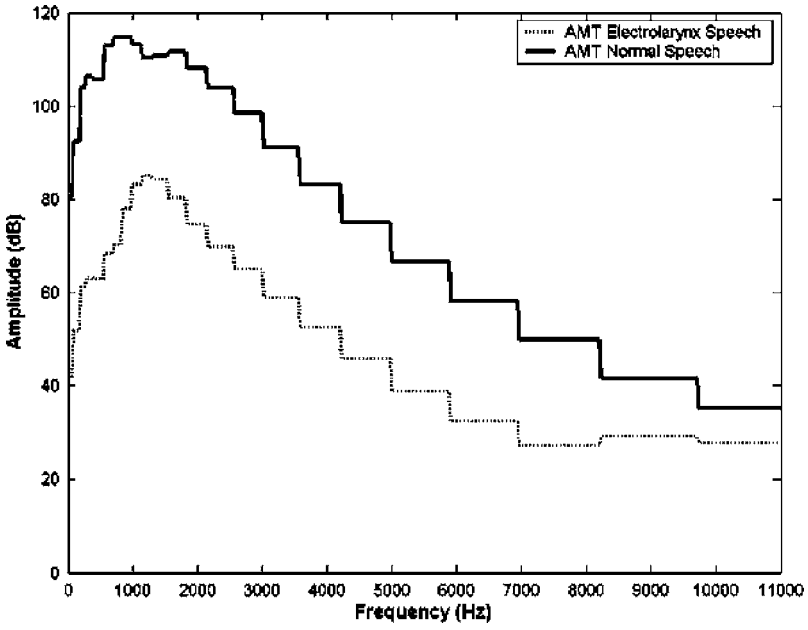


FIG. 2. Examples of the auditory masking threshold of the vowel /a/ from the clean normal speech and electrolarynx speech. The sampling frequency is 20 kHz and the total number of critical bands is 24.

The AMT should be computed from the clean speech signal. However, in the proposed enhancement scheme, it is impossible to obtain clean EL speech. Therefore the AMT is calculated with the original EL speech without additive noise. For the noisy speech with low SNR, the PSS algorithm is usually used as a preprocessing scheme for the AMT estimation, reducing background noise while introducing musical residual noise. In our experiments, the AMT was estimated from the noisy EL speech directly without the preenhancement process. This is due to the fact that the AMT from the preenhanced EL speech is very different from the one obtained from the original EL speech, even after decreasing the relative threshold offset to take into account the tonelike nature of the musical residual noise. This might be related to the particular characteristics of EL speech and the preenhanced processing using the PSS algorithm.

The adaptation rule is based on the following consideration: If the AMT is low, the subtraction parameters will be increased to reduce the noise. The introduced musical noise will be masked by the background noise remaining in the enhanced speech due to the high spectral floor. If the AMT is high, however, it is necessary to keep the subtraction parameters to their minimal values because residual noise will stay below the AMT and will be naturally masked and inaudible. The adaptation rules of the subtraction parameters are performed as follows:

$$\alpha = \alpha_{\max}\left[\frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}}\right] + \alpha_{\min}\left[\frac{T(\omega) - T(\omega)_{\min}}{T(\omega)_{\max} - T(\omega)_{\min}}\right] T(\omega)_{\min} \le T(\omega) \le T(\omega)_{\max}, \tag{7}$$

$$\beta = \beta_{\max}\left[\frac{T(\omega)_{\max} - T(\omega)}{T(\omega)_{\max} - T(\omega)_{\min}}\right] + \beta_{\min}\left[\frac{T(\omega) - T(\omega)_{\min}}{T(\omega)_{\max} - T(\omega)_{\min}}\right] T(\omega)_{\min} \le T(\omega) \le T(\omega)_{\max}, \tag{8}$$

where $\alpha_{\min}$, $\alpha_{\max}$, $\beta_{\min}$, $\beta_{\max}$, $T(\omega)_{\min}$, and $T(\omega)_{\max}$ are the minimal and maximal values of α, β, and $T(\omega)$ are updated from frame to frame. According to a number of experiments with different noise types and levels for selecting the appropriate values for these parameters, we choose the following values to obtain a good tradeoff between residual noise and speech distortion:

1) $\alpha_{\min}=1$ and $\alpha_{\max}=6$.
2) $\beta_{\min}=0$ and $\beta_{\max}=0.02$.

III. EXPERIMENTS

A. Subjects

Six male laryngectomees with the total removal of the larynx participated in the experiment. The participants had recovered from the fibrosis and edema resulting from radiation, and their neck tissue was supple enough so as to permit them to use EL effectively. All of the subjects were native speakers of Mandarin Chinese. The subjects with laryngectomies ranged from 48 to 70 years with a mean age of 58.68, and they had at least 2 years of experience using the device so that they were proficient at using EL for demonstration purposes.

Six listeners individually carried out the perceptual task in a soundproof room. Their ages varied from 20 to 32, with a mean age of 26.36. All of them were unfamiliar with EL speech, and none of them had hearing problems in both ears (pure-tone threshold better than 20 dB SPL across all frequencies). The listeners were reported to possess at least a college education, and they were able to correctly read and

comprehend the speech material used in the experiment. All the participants were monolingual Mandarin speakers.

B. Recordings

The recording procedure was carried out in a sound-proof room. Speech samples were collected by using a microphone mounted at a distance of 15 cm from the mouth, and amplified by using a multichannel conditioning amplifier (Brüel and Kjær, Model 2693). A locally made, hand-held EL was used (Model Hu Die 9201). The device had a built-in frequency range of 60 to 90 Hz, with an intensity range of 70 to 80 dB SPL. The EL speakers were instructed to use the pitch and intensity at a preset level throughout the recording. Recordings were taken at a sampling frequency of 20 kHz with 16-bits per sample. During the recording, speakers were provided with cards on which Chinese characters were printed representing the citation words. Five Chinese sentences, each of which was composed of six words, were used as the speech materials for acoustic and perceptual analyses. Instructions were given to the speakers before the recording took place. The speakers were instructed to read the speech materials three times at normal loudness and speaking rate.

White Gaussian noise and speech babble noise, taken from the Noisex-92 database designed for speech recognition in noisy environments, were chosen for the research of EL speech enhancement in noise in the experiments. Noise was added to the original EL speech signal with a varying SNR. All these sentences were processed by three enhancement algorithms. After that, these sentences were divided into six sets of sentences. Each set contained different sentences in randomized order with four enhancement conditions (three with enhancement algorithms and one with no algorithm), which was provided as the listening stimuli for the perceptual analyses.

C. Perceptual evaluation

In our perceptual study, listeners rated the acceptability of each sentence based on the criteria of the mean opinion score (MOS), which is a five-point scale (1: bad; 2: poor; 3: common; 4: good; 5: excellent). The listening tasks took place in a sound-proof room. The speech samples were presented to the listeners at a comfortable loudness level (65 dB SPL) via a high quality headphone. To eliminate the order effect, the order in which the speech samples were presented was randomized. A 4-sec pause was inserted before each citation word to allow the listeners to respond and to avoid rehearsal effect.

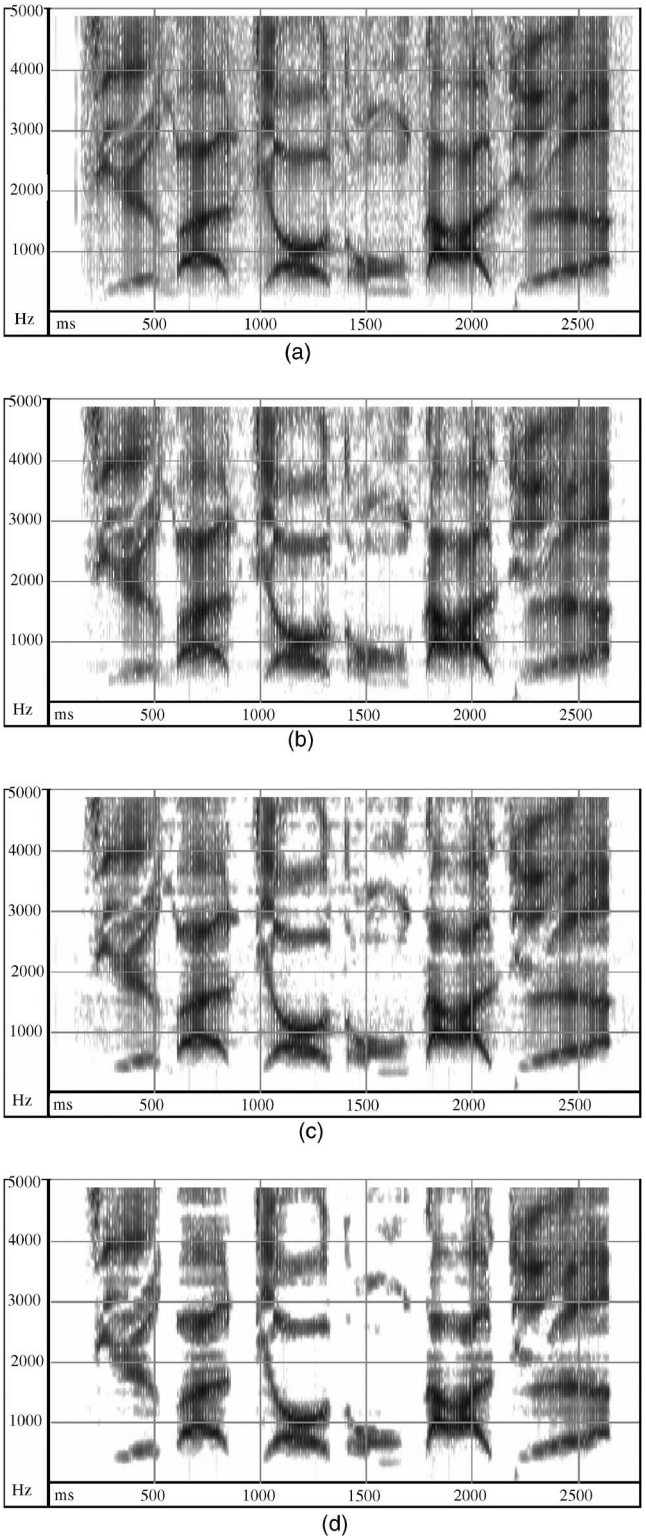


FIG. 3. Spectrograms of the phrase 'xi an jiao tong da xue': (a) original speech; (b) enhanced speech by the PSS algorithm; (c) by the AMT algorithm; and (d) by the SAMT algorithm.

IV. RESULTS

A. Acoustic analysis

The performance evaluation of the AMT and the SAMT algorithms was presented with a comparison with the PSS algorithm. In order to analyze the time-frequency distribution of the enhanced speech, we presented speech spectrograms that can give accurate information about residual noise and speech distortion. The speech material is a Chinese sentence "xi an jiao tong da xue" ('Xi'an Jiaotong University' in English).

Figure 3 shows the spectrograms of the original and enhanced EL speech without additive noise. Figure 3(a) shows that a certain amount of the radiated noise exists in the origi-

nal EL speech because of the continuous use of EL for phonation, especially during speech pauses. Figure 3(b) shows that the PSS algorithm is effective in reducing the radiated noise below 2 kHz as well as an amount of noise during speech pauses, but not to reduce the high-frequency noise. According to Fig. 3(c), the AMT algorithm reduces some noises above 2 kHz but keeps a few noises below 2 kHz. That is, the AMT achieves a better reduction of the high-frequency noise but a worse reduction of the low-frequency noise as compared to the PSS algorithm. Figure 3(d) shows that the SAMT algorithm not only reduces the radiated noise but also eliminates both the low- and the high-frequency noises completely.

Figures 4 and 5 show the spectrograms of the noisy and enhanced EL speech in the case of additive white noise and babble noise, respectively. In Fig. 4(a), the original speech is almost masked completely by white noise, especially in high-frequency regions. In Fig. 5(a), babble noise covers both the low- and high-frequency regions but more energy is concentrated in the low-frequency regions as compared to white noise. Figures 4(b) and 5(b) indicate that there are still many noises in the enhanced speech dealt with the PSS algorithm. The PSS algorithm reduces few low-frequency noises below 2 kHz, and even less high-frequency noises especially above 3 kHz. As shown in Figs. 4(c) and 5(c), the AMT algorithm does better than the PSS algorithm in eliminating the high-frequency noises. But it is still not good in reducing the low-frequency noises. Figures 4(d) and 5(d) associated with the SAMT algorithm show a much better performance in white noise reduction, in which the low- and high-frequency noises are eliminated completely.

B. Perceptual analysis

Perceptual analyses performed with the different enhancement algorithms are shown in Fig. 6. The score of the enhanced speech obtained from using the SAMT algorithm is the highest, followed by that from the AMT and the PSS algorithms. This is true for both the original speech and the noisy speech. With regard to the original speech, a repeated-measures analyses of variance (ANOVA) indicated a significant difference between original speech and enhanced speech $[F(3,1076)=16.512, p<0.05]$. Bonferroni *post hoc* test indicated that the score of the enhanced speech by the SAMT algorithm is significantly higher than those of the others ($p<0.05$). With regard to speech with additive noise, significant differences were found between original speech and enhanced speech in the acceptability [white noise: $F(3,1076)=6.839, p<0.05$; babble noise: $F(3,1076)=38.988, p<0.05$] as tested with a repeated-measures ANOVA. Bonferroni *post hoc* test indicated that the scores of enhanced speech by the AMT and the SAMT algorithms were significantly higher than those of the others ($p<0.05$). But no significant differences were found between the enhanced speech by the AMT algorithm and that by the SAMT algorithm, and between the enhanced speech by the PSS algorithm and the original speech.

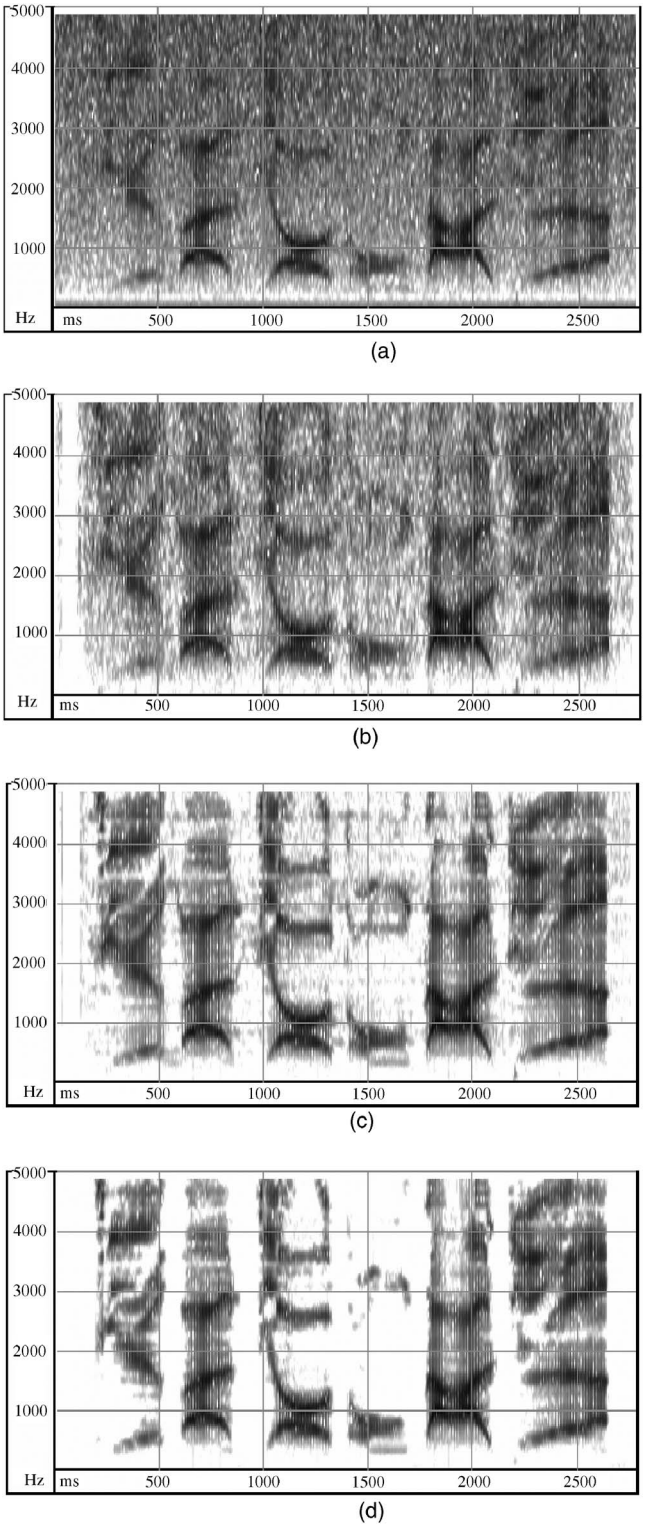


FIG. 4. Spectrograms of the phrase 'xi an jiao tong da xue': (a) noisy speech in the case of additive white noise at a SNR=0 dB; (b) enhanced speech by the PSS algorithm; (c) by the AMT algorithm; and (d) by the SAMT algorithm.

V. DISCUSSION

The results indicate that both the AMT algorithm and the SAMT algorithm are better suited for EL speech enhancement than the PSS algorithm, especially in the case of additive noise. Because the subtraction parameters are fixed and

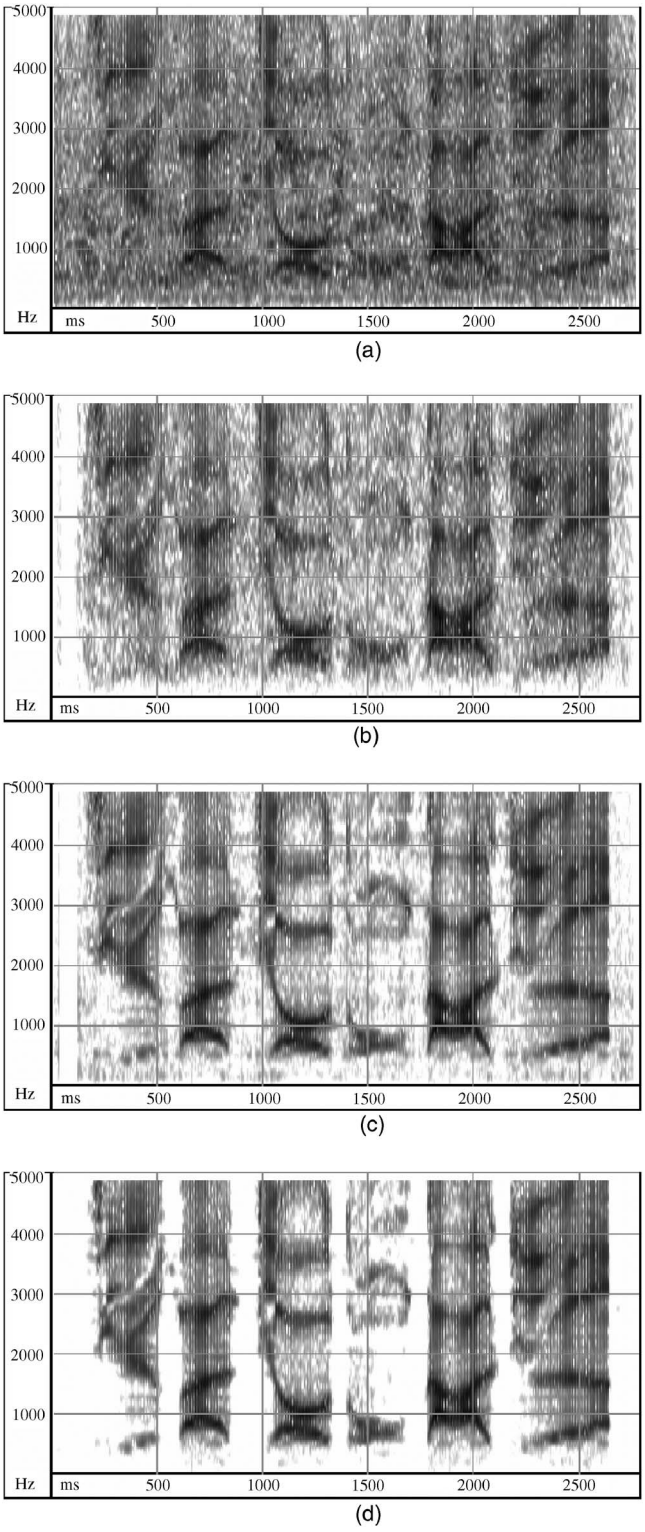


FIG. 5. Spectrograms of the phrase 'xi an jiao tong da xue': (a) noisy speech in the case of additive babble noise at a SNR=0 dB; (b) enhanced speech by the PSS algorithm; (c) by the AMT algorithm; and (d) by the SAMT algorithm.

unable to be adapted from frame to frame, the PSS algorithm cannot reduce the noise effectively, especially the high-frequency noise. These limitations will be worse for the enhancement of EL speech in the case of additive noise. With regard to the AMT and the SAMT algorithms, α and β can be adapted based on the AMT. Based on the frame-by-frame

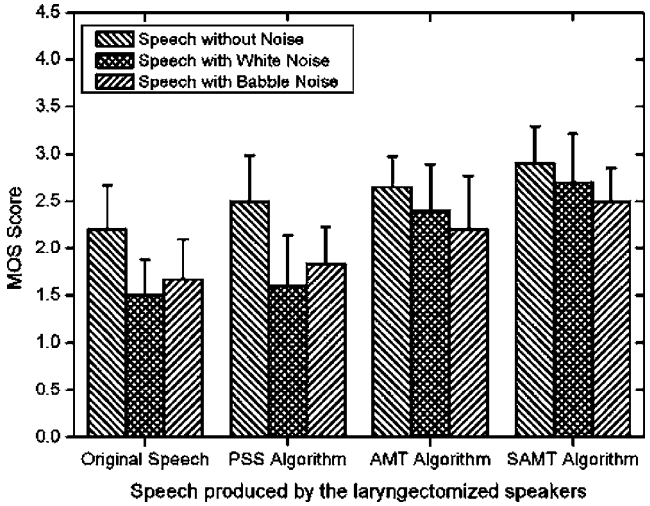


FIG. 6. Acceptability scores of the original and enhanced EL speech. The noisy speech in the case of additive noise has an input SNR of 0 dB.

adaptations of the subtraction parameters, these two algorithms can realize a good tradeoff between reducing noise, increasing intelligibility, and keeping the distortion acceptable to a human listener. Furthermore, the SAMT algorithm is superior to the AMT algorithm due to the supplementary scheme of the CCSS algorithm. This post-processing approach eliminates the cross-correlated parts of the noisy speech and compensates the deficit of the AMT algorithm in the reduction of low-frequency noise.

It is noted that a certain amount of residual noise is kept in the low-frequency regions of the enhanced speech by the AMT algorithm. This is based on the characteristics of the auditory masking model. According to the auditory masking theory, residual noise will be naturally masked and inaudible if the AMT is high. In this case, there is no need to reduce it in order to keep distortion as low as possible and the subtraction parameters are kept to their minimal values. If the AMT is low, residual noise will be annoying to the human listener and it is necessary to reduce it by increasing the subtraction parameters. The AMT value is higher in the low-frequency regions than that in the high-frequency regions, resulting in low values of subtraction parameters in the low-frequency regions and high values in the high-frequency regions based on the adaptation rules [see formulas (7) and (8)]. This leads to the results that some residual noises are kept in the low-frequency regions while noises in the high-frequency regions are eliminated almost completely. Due to the low-frequency deficit of EL (the output level below 550 Hz was about 30 dB SPL lower than that of normal speech) noted by Qi and Weinberg (1991), residual noise in the low-frequency regions would interfere in the intelligibility improvements. Therefore, a supplementary scheme of the CCSS approach is incorporated to reduce such noise as well as other correlated noise to further enhance EL speech. The results also indicate that the SAMT algorithm cannot only reduce the residual noise but also improve the low-frequency deficit of EL speech. The perceptual results confirm the acceptability improvements based on the SAMT algorithm,

which is consistent with the Qi and Weinberg report (1991) that the listeners preferred low-frequency-enhanced EL speech.

As the single channel subtractive-type speech enhanced methods, the AMT and the SAMT algorithms in this paper can be applied into the enhancement of EL speech in a practical situation. For example, an enhanced system embedded with these two algorithms can be developed. With the help of digital signal processing (DSP) technology, we can realize the enhancement function with a microprocessor and implant it into a telephone, microphone, or other electronic media. Different enhancement algorithms can be selected through the switch based on different noisy conditions (The SAMT algorithm is more powerful, but the AMT algorithm is more efficient). Along with the development of efficient enhancement methods, the quality of EL speech will be extensively improved for better perception.

VI. CONCLUSION

The present study investigated two enhancement algorithms of EL speech based on spectral subtraction: the AMT algorithm and the SAMT algorithm. Because these two algorithms took into account the auditory masking properties of the human ear to adapt the subtraction parameters in the enhancement process, a better effect of noise reduction was obtained and the perceptually annoying musical noise was efficiently reduced as compared to the PSS algorithm. Furthermore, the CCSS approach in the SAMT algorithm eliminates the residual noise in the low-frequency regions by the AMT algorithm. In addition, these two algorithms can effectively reduce both the radiated noise and the additive noise. As compared to the PSS algorithm in various noise types, the AMT and the SAMT algorithms show that the background noise is reduced efficiently while the distortion of EL speech remains acceptable.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (Grants No. 30070212 and 69925101). The authors express special appreciation to two anonymous reviewers for their comments and suggestions on the manuscript.

Arehart, H. K., Hansen, J. H. L., Gallant, S., and Kalstein, L. (**2003**). "Evaluation of an auditory masked threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners," Speech Commun. **40**, 575–592.

Barney, H. L., Haworth, F. E., and Dunn, H. K. (**1959**). "An experimental transistorized artificial larynx," Bell Syst. Tech. J. **38**, 1337–1356.

Berouti, M., Schwartz, R., and Makhoul, J. (**1979**). "Enhancement of speech corrupted by acoustic noise," *IEEE Proc. ICASSP-79* (IEEE, Washington, DC), pp. 208–211.

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. ASSP- **27**, 113–120.

Brandenburg, K., and Stoll, G. (**1994**). "ISO-MPEG-1 audio: A generic standard for coding of high quality digital audio," J. Audio Eng. Soc. **42**, 780–792.

Cohen, I., and Berdugo, B. (**2002**). "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Signal Process. Lett. **9**, 12–15.

Cole, D., Sridharan, S., Moody, M., and Geva, S. (**1997**). "Application of noise reduction techniques for alaryngeal speech enhancement," *IEEE Proc. TENCON-97* (IEEE, Brisbane), pp. 491–494.

Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. ASSP- **32**, 1109–1121.

Espy-Wilson, C. Y., Chari, V. R., and Huang, C. B. (**1996**). "Enhancement of alaryngeal speech by adaptive filtering," *IEEE Proc. ICSLP-96* (IEEE, Philadelphia), pp. 764–767.

Espy-Wilson, C. Y., Chari, V. R., MacAuslan, J., and Walsh, M. (**1998**). "Enhancement of electrolaryngeal speech by adaptive filtering," J. Speech Lang. Hear. Res. **41**, 1253–1264.

Hansen, J. H. L. (**1994**). "Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect," IEEE Trans. Speech Audio Process. **2**, 598–614.

Hansen, J. H. L., and Nandkumar, S. (**1995**). "Robust estimation of speech in noisy backgrounds based on aspects of the auditory process," J. Acoust. Soc. Am. **97**, 3833–3849.

Holly, S. C., Lernman, J., and Randolph, K. (**1983**). "A comparison of the intelligibility of esophageal, electrolarynx, and normal speech in quiet and in noise," J. Commun. Disord. **16**, 143–155.

Hu, Y., Bhatnagar, M., and Loizou, P. C. (**2001**). "A cross-correlation technique for enhancement speech corrupted with correlated noise," *IEEE Proc. ICASSP-01* (IEEE, Salt Lake City), pp. 673–676.

Johnston, J. D. (**1988**). "Transform coding of audio signals using perceptual noise criteria," IEEE J. Sel. Areas Commun. **6**, 314–323.

Knox, A. A., and Anneberg, M. (**1973**). "The effects of training in comprehension of electrolaryngeal speech," J. Commun. Disord. **6**, 110–120.

Lauder, E. (**1970**). "The laryngectomee and the artificial larynx–A second look," J. Speech Hear Disord. **35**, 62–65.

Lim, J. S., and Oppenheim, A. V. (**1979**). "Enhancement and bandwidth compression of noisy speech," Proc. IEEE **67**(12), 1586–1604.

Liu, H. J., Zhao, Q., Wan, M. X., and Wang, S. P. (**2006**). "Enhancement of electrolarynx speech based on auditory masking," IEEE Trans. Biomed. Eng. **53**(5), 865–874.

Lockwood, P., and Boudy, J. (**1992**). "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and projection, for robust recognition in cars," Speech Commun. **11**, 215–228.

McAulay, R. J., and Malpass, M. L. (**1980**). "Speech enhancement using a soft decision noise suppression filter," IEEE Trans. Acoust., Speech, Signal Process. ASSP- **28**, 137–145.

Nandkumar, S., and Hansen, J. H. L. (**1995**). "Dual-channel iterative speech enhancement with constraints on an auditory-based spectrum," IEEE Trans. Speech Audio Process. **3**, 22–34.

Niu, H. J., Wan, M. X., Wang, S. P., and Liu, H. J. (**2003**). "Enhancement of electrolarynx speech using adaptive noise cancelling based on independent component analysis," Med. Biol. Eng. Comput. **41**(6), 670–678.

Norton, R. L., and Bernstein, R. S. (**1993**). "Improved Laboratory Prototype electrolarynx (LAPEL): using inverse filtering of frequency response function of the human throat," Ann. Biomed. Eng. **21**, 163–174.

Painter, T., and Spanias, A. (**2000**). "Perceptual coding of digital audio," Proc. IEEE **88**, 451–515.

Pandey, P. C., Bhandarkar, S. M., Bachher, G. K., and Lehana, P. K. (**2002**). "Enhancement of alaryngeal speech using spectral subtraction," *IEEE DSP 2002* (IEEE, Aegean Island of Santorini), Vol. 2, pp. 59l–594.

Pratapwar, S. S., Pandey, P. C., and Lehana, P. K. (**2003**). "Reduction of background noise in alaryngeal speech using spectral subtraction with quantile based noise estimation," *7th World Multiconference on Systemics, Cybernetics and Informatics* (International Institute of Informatics and Systemics, Orlando), pp. 408–413.

Qi, Y., and Weinberg, B. (**1991**). "Low-frequency energy deficit in electrolaryngeal speech," J. Speech Hear. Res. **34**, 1250–1256.

Rothman, H. (**1982**). "Acoustic analysis of artificial electronic larynx speech," in *Electroacoustics Analysis and Enhancement of Alaryngeal Speech*, edited by A. Seikey (Charles Thomas, Springfield, IL), pp. 95–118.

Schroeder, M. R., Atal, B. S., and Hall, J. L. (**1979**). "Optimizing digital speech coders by exploiting masking properties of the human ear," J. Acoust. Soc. Am. **66**, 1647–1652.

Tsoukalas, D., Paraskevas, M., and Mourjopoulos, J. (**1993**). "Speech enhancement using psycho-acoustic criteria," *IEEE Proc. ICASSP-93* (IEEE, Minneapolis), pp. 359–362.

Tsoukalas, D. E., Mourjopoulos, J. N., and Kokkinakis, G. (**1997**). "Speech enhancement based on audible noise suppression," IEEE Trans. Speech

Audio Process. **5**, 479–514.
Usagawa, T., Iwata, M., and Ebata, M. (**1994**). "Speech parameter extraction in noisy environment using a masking model," *Proc. ICASSP-94* (IEEE, Adelaide), pp. 81–84.
Virag, N. (**1999**). "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech Audio Process. **7**, 126–137.
Weiss, M. S., Yeni-Komshian, G. H., and Heinz, J. M. (**1979**). "Acoustic and perceptual characteristics of speech produced with an electronic artificial larynx," J. Acoust. Soc. Am. **65**, 1298–1308.

Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance[a)]

Lisa Davidson[b)]
Department of Linguistics, New York University, 719 Broadway, 4th Floor, New York, New York 10003

(Received 31 October 2005; revised 24 April 2006; accepted 25 April 2006)

Ultrasound imaging of the tongue is increasingly common in speech production research. However, there has been little standardization regarding the quantification and statistical analysis of ultrasound data. In linguistic studies, researchers may want to determine whether the tongue shape for an articulation under two different conditions (e.g., consonants in word-final versus word-medial position) is the same or different. This paper demonstrates how the smoothing spline ANOVA (SS ANOVA) can be applied to the comparison of tongue curves [Gu, *Smoothing Spline ANOVA Models* (Springer, New York, 2002)]. The SS ANOVA is a technique for determining whether or not there are significant differences between the smoothing splines that are the best fits for two data sets being compared. If the interaction term of the SS ANOVA model is statistically significant, then the groups have different shapes. Since the interaction may be significant even if only a small section of the curves are different (i.e., the tongue root is the same, but the tip of one group is raised), Bayesian confidence intervals are used to determine which sections of the curves are statistically different. SS ANOVAs are illustrated with some data comparing obstruents produced in word-final and word-medial coda position. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2205133]

PACS number(s): 43.70.Jt [AL] Pages: 407–415

I. INTRODUCTION

Ultrasound imaging is becoming an increasingly popular technique for examining articulation in speech research. Previous research has shown that ultrasound imaging is a practical, low-cost, and noninvasive tool for acquiring articulatory data to examine tongue shapes corresponding to various sounds, answering phonological questions, conducting phonetic fieldwork, and use in speech rehabilitation (e.g., Bernhardt *et al.*, 2003; Bressmann *et al.*, 2005; Davidson, 2005; Gick, 2002; Stone, 2005; Stone *et al.*, 1992; Stone and Lundberg, 1996).

Ultrasound is an attractive technique for imaging articulation during speech because it provides an image of the length of the tongue. Other techniques for imaging the midsagittal contour of the length of tongue such as MRI and cinefluorography are also available. However, these methodologies are often prohibitively expensive or difficult to access. In most speech-related applications of ultrasound, researchers have focused on collecting data from the midsagittal contour of the tongue, although coronal slices have also been analyzed (Slud *et al.*, 2002). A sample image of a midsagittal tongue curve during the production of the fricative /z/ is shown in Fig. 1. In this and following ultrasound images, the tongue tip is on the right and the tongue root is on the left. The ability to image the entire contour of the tongue is a significant advantage of ultrasound over techniques like electromagnetic midsagittal articulography (EMMA) (Perkell *et al.*, 1992) or x-ray microbeam (Westbury, 1994), which only allow for the tracking of the flesh points to which the receivers are attached. Though a tongue surface can be partially reconstructed from fleshpoint data, there are two main shortcomings for fleshpoint tracking as compared to imaging techniques like ultrasound: (1) since the placement of receivers is limited by the gag reflex, it is difficult or impossible to acquire information about the shape or motion of the tongue root, and (2) there is always the possibility that an important shape of the tongue occurs between two receivers and cannot be accurately reconstructed.

While ultrasound has become important as a tool for both linguistic and clinical investigation, there has not been consensus regarding the quantification and statistical analysis of the data that are collected. Some methods that have been used so far include the overlay of a concentric grid with equally spaced radial lines on the tongue shape, which allows for measurements from a fixed point to the tongue surface on any of the lines in the grid (Bressmann *et al.*, 2005); a mean distance measure that averages the Euclidean distances between corresponding points on two curves being compared (Davidson, 2005); and principal components analysis (Slud *et al.*, 2002). Of these methods, the most common measurement technique for midsagittal tongue curves has been the concentric grid, which is implemented in several software packages for ultrasound imaging processing [e.g., University of Arizona's GLoSsatron (http://dingo.sbs.arizona.edu/~apilab/), Queen Margaret University College's Articulate Assistant (http://www.articulateinstruments.com/), the University of British Columbia's Ultrax (http://www.linguistics.ubc.ca/isrl/index.html), University of Toronto's Ultra-CATs (http://www.slp.utoronto.ca/English/Ultra-CATS.html); all websites last viewed on April 21, 2006]. The image in Fig. 2 demon-

[a)]Portions of this work were presented at the Ultrafest III workshop at the University of Arizona, April 2005 and the 50th Acoustical Society of America meeting in Minneapolis, MN, October 2005.

[b)]Electronic mail: lisa.davidson@nyu.edu

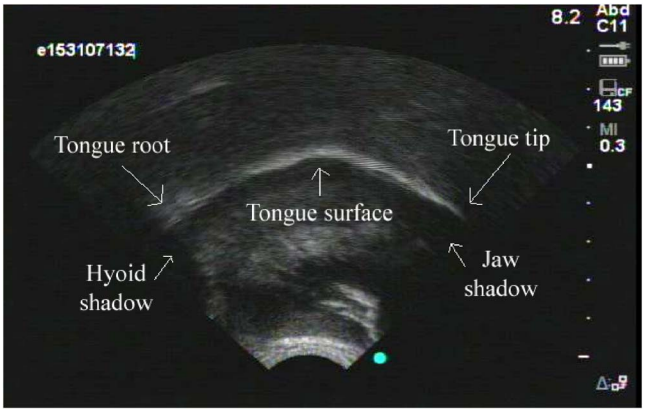


FIG. 1. (Color online) Midsagittal image of the frame corresponding to the midpoint of frication of the /z/ in the acoustic signal for "jazz dancer." The tongue tip is on the right and the tongue root is on the left.

strates a tongue curve representing the maximum constriction for the articulation of a /g/. Overlaid on top of the tongue curve is a grid with seven equally spaced radii. The origin of the radii is approximately at the center of the transducer. The ellipses indicate the points at which the radii intersect the tongue curve.

Using a radial grid, researchers can measure from the origin of the radii to the point at which each radius intersects the curve. For example, if a researcher were examining the difference in constriction degree for a velar stop like /g/ versus a velar fricative like /x/, measurements along one or more radii could be compared for multiple repetitions of the /g/ to those for the /x/, and then statistically analyzed with a *t* test. However, unless information from the entire tongue is recorded, it would be easy to miss taking a measurement at the most important location. For example, in the case of the /g/ in Fig. 2, the apex of the curve, marked with an X, is taken to be the point of maximum constriction. Since this point does not fall on a radius, the most relevant measurement is missed. Alternatively, the number of radii on the grid could be increased in order to make as many measurements as possible, but such a decision is an incomplete attempt to

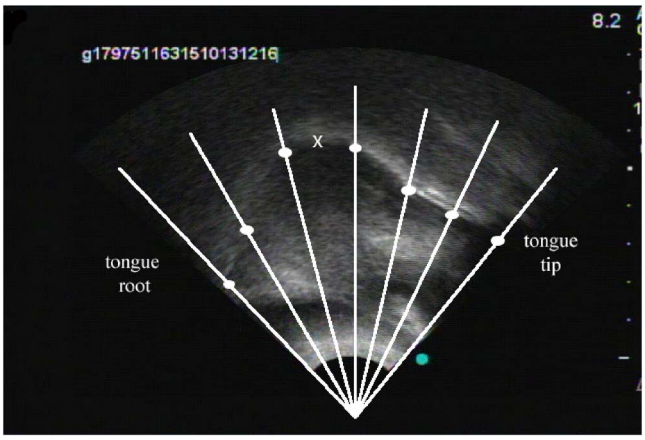


FIG. 2. (Color online) Midsagittal ultrasound image of the maximally raised position of the tongue dorsum for /g/ in "Baghdad" with a radial grid overlay. The white ellipses indicate where the radii intersect the tongue curve. The "X" indicates the location of maximum constriction along the tongue curve.

characterize the entire tongue surface, which is what the smoothing spline ANOVA described in this paper was explicitly developed to do. The other measurement techniques mentioned earlier are similarly dissatisfactory: either they are not suitable for individual comparisons (principal components), or there is no principled way of comparing individual sections of the tongue curve to determine where a difference lies (mean distance measures).

To address these issues, this paper introduces the smoothing spline ANOVA [SS ANOVA (Gu, 2002)] as a method for comparing tongue curve shapes. The SS ANOVA is a statistical method that allows for the holistic comparison of the entire tongue curve, whether it is obtained from ultrasound, MRI, or cinefluorography. This procedure has been used in other fields where similarities and differences of curve shapes must be assessed, such as plots of circadian rhythms in normal adults, patients with Cushing's syndrome, and patients with depression (Wang *et al.*, 2003). Because the mathematical details of both smoothing splines and the SS ANOVA have been well covered in both statistical and applied literature, this paper is intended primarily as a descriptive introduction of the technique for linguists or speech scientists who use ultrasound [or similar techniques, such as x-ray (e.g., Iskarous, 2005)] for speech research. References are provided for those desiring a more technical explanation of the procedures described in this paper.

To demonstrate the smoothing spline ANOVA for tongue curve comparison, it is illustrated with respect to the degree and location of maximum constriction of consonants in different word positions (e.g., *ba**g** dazzled* versus *Ba**g**hdad*). The data used in this paper to present the SS ANOVA come from an unpublished experiment, but this paper is not a report of the results of that study.

II. ULTRASOUND DATA COLLECTION

A. Data collection procedure

The stimuli consisted of three pairs of words and phrases containing the same consonant in different positions: *blac**k** top* versus *blac**k**top*, *ba**g** dazzled* versus *Ba**g**hdad*, and *ja**zz** dancer* versus *NA**S**DAQ*. These consonants were chosen because they are all lingual articulations that are easily imaged by the ultrasound. These words were produced by five monolingual native speakers of American English.

Midsagittal images of the tongue were recorded from a Sonosite Titan portable ultrasound machine using a 5–8 MHz Sonosite C-11 transducer with a 90° field of view and a depth of 8.2 cm. The incoming video signal from the ultrasound machine and an audio signal from an Audio Technica AT-813 microphone were synchronized and captured directly to a Dell computer using a Canopus ADVC-1394 capture card and Adobe Premiere 6.0. The Canopus card is designed to assure audio-video synchrony throughout the duration of the recording. The video frame rate is 29.97 Hz.

In order to compare images from different utterances, it is important to ensure that neither the speaker's head nor the transducer move during the experiment (Stone, 2005; Stone and Davis, 1995). Participants were seated in a sound-proof booth and their heads were stabilized using a moldable head

FIG. 3. Head and transducer stabilization setup. The speaker's head is encompassed by the moldable head stabilizer, which can be moved up and down on the Velcro strips against the wall of the soundproof booth. Another Velcro strap is pulled against the speaker's head for further stabilization. The transducer is stabilized with a microphone stand.

stabilizer (Comfort Company). The moldable head stabilizer is a rigid U-shaped foam brace designed to assist elderly people with low neck tone who have difficulty keeping their heads upright. The stabilizer is affixed to a wall in the soundproof booth with Velcro and is placed at the height of the participant's temples. Another piece of Velcro is then used to strap the speaker into the head stabilizer so that the head is entirely enclosed. Once the speaker is placed in the head stabilizer, one microphone stand to hold the transducer and another stand to hold the microphone are set up. The stand with the transducer is placed underneath the chin and the placement is adjusted until a satisfactory midsagittal tongue image is obtained. A picture of this setup is shown in Fig. 3.

A music stand was placed directly in front of the speaker at eye level. Eight pieces of paper each containing a randomization of the stimuli and fillers were placed on the music stand. The participant read the list, and then the experimenter turned the page. This resulted in eight repetitions of each phrase.

B. Edge extraction

After data collection, sections of the video files collected with Adobe Premiere containing the target phrases were transformed into JPEG stills. For the stops /k/ and /g/, the ultrasound frame with the most raised tongue body within the period of stop closure was chosen for comparison of tongue shapes for word-final versus word-medial codas. For the fricative /z/, the ultrasound frame roughly corresponding to the midpoint of the duration of frication on the acoustic record was chosen as the comparison frame. A sample image for the most raised tongue body for the /g/ in "Baghdad" is shown in Fig. 2 and the midpoint of the fricative for /z/ in "jazz dancer" is illustrated in Fig. 1. The decision to compare single frames as opposed to a sequence of frames was carried out both for theoretical reasons and for simplicity of presentation. First, one question that speech scientists may ask is whether the point of maximum constriction of a consonant differs with respect to some variable, such as word position, speech rate, or phonological environment (e.g., Browman and Goldstein, 1995; Kochetov, 2006). Second, in order to illustrate the SS ANOVA, the point of maximum constriction is used as a simple test case. However, the SS ANOVA has also been used to investigate comparisons along spatial and temporal dimensions, as illustrated by statistical methods developed to examine changes in the electroencephalograms (EEG) of epileptic patients (Guo *et al.*, 2003) or spatiotemporal changes in surface air temperature (Luo *et al.*, 1998).

For each repetition of the target phrases, the JPEG stills were loaded into EdgeTrak (version 1.0.0.4) for measurement (Li *et al.*, 2005). EdgeTrak is a computer program that automates the tracking of tongue contours by extracting (x,y) coordinates from the lower edge of the white curve in the ultrasound image. First, a few points on the tongue image are manually chosen, and then EdgeTrak uses an active contour model to determine the location of the tongue edge in the image. If the automatic tracking of the tongue edge does not produce satisfactory results, points can be manually added or subtracted to obtain the best fit. Sixty-four points were extracted for each tongue curve, which were then used for statistical analysis. A screenshot of the tongue curve extraction in EdgeTrak for the frame of the /g/ of "bag dazzled" is shown in Fig. 4.

III. SMOOTHING SPLINE ANOVA FOR COMPARING TONGUE SHAPES

A. Smoothing splines

The 64 points for each of the eight repetitions of /g/ for "bag dazzled" and "Baghdad" extracted with EdgeTrak are shown in Fig. 5. These repetitions are plotted in the statistical package S-Plus 2000 (the commercial version of the open-source R language for statistical computing). The first step is to fit the data using smoothing splines (Eubank, 1988; Green and Silverman, 1994; Wahba, 1990). Smoothing splines have also previously been employed in speech production research. For example, Ramsay *et al.* (1996) provides a technical introduction to the use of smoothing splines in a study of lip motion using OPTOTRAK, an optoelectronic tracking system that transduces the 3-D position of reflective markers. In what follows, a more intuitive introduction to smoothing splines is presented, focusing on how it applies to ultrasound data.

Smoothing splines are a type of natural cubic spline, which is a piecewise polynomial function that connects dis-

FIG. 4. Screenshot of the EdgeTrak extraction for the frame for /g/ shown in Fig. 1. The program is asked to provide 64 points to characterize the curve shape.

crete data points called knots. Smoothing splines include a smoothing parameter to find the best fit when the data tend to be noisy. More specifically, the function defining the smoothing spline contains two terms: one that attempts to fit the data and one that penalizes a fit which does not have the appropriate amount of smoothness. Although the penalty term does not allow the function to fit the data precisely, it ensures that the resulting spline has a suitable amount of smoothness. Natural cubic splines have the advantage that the shape of the data does not have to be known *a priori*.

The smoothing spline is estimated by minimizing the function in (1):

$$G(x) = \frac{1}{n}\sum_{\text{all } i} (y_i - f(x_i))^2 + \lambda \int_a^b (f''(u))^2 \, du, \qquad (1)$$

where n is the number of data points, and a and b are the x coordinates of the endpoint of the spline. The smoothing parameter λ is critical to the performance of the spline estimate. If λ is large, the curve will be smoother, whereas a small λ produces a wavier curve that attempts to fit each of the individual data points. The smoothing parameter is determined automatically using the generalized cross validation (GCV) method (technical details on GCV are discussed in Craven and Wahba, 1979; Ramsay *et al.*, 1996). The same function is used to estimate a spline whether the data contain the 64 points of one repetition or the 512 points of eight repetitions.

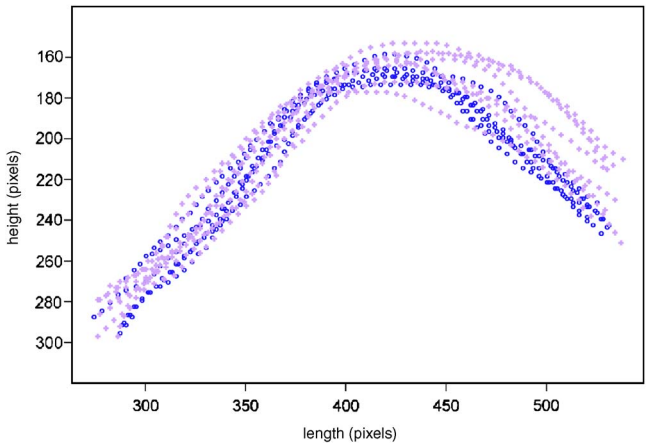


FIG. 5. (Color online) Raw data points from eight repetitions for comparison of the shapes for /g/ in "bag dazzled" and "Baghdad" for speaker TO. "bag dazzled" is represented by the dark blue "o" data points, and "Baghdad" by the pink '+' data points. The x axis is the length of the tongue, and the y axis is the height of the tongue. The scales correspond to the pixels of the original JPEGs, where 1 mm=2.63 pixels and the origin is in the top left corner (accounting for why the values on the y axis increase). Like the ultrasound images, the tongue tip is on the right and the tongue root is on the left.

An example of the smoothing splines corresponding to each data set from the eight repetitions of /g/ in "bag dazzled" and "Baghdad" for subject TO is shown in Fig. 6(a). In this figure, the axes are in pixels, where 1 mm =2.63 pixels. The vertical lines in the figure are a rough division of the tongue into three parts corresponding to the tongue anterior, the body/dorsum, and the root. This type of tentative division allows for the determination of statistical significance in the part of the tongue most relevant to the research question. For the purpose of the data discussed in this paper, the main region of interest for the coronal fricative /z/ is the rightmost third of the tongue corresponding to the anterior parts of the tongue, including the tip and blade. For the velar stops /k/ and /g/, the focus is on the middle third corresponding to the tongue body/dorsum. For now, the tongue is divided into three equal parts for lack of a better assumption about the most linguistically relevant way to determine such divisions. This issue will be discussed again in the general discussion.

Some differences between the consonantal articulations

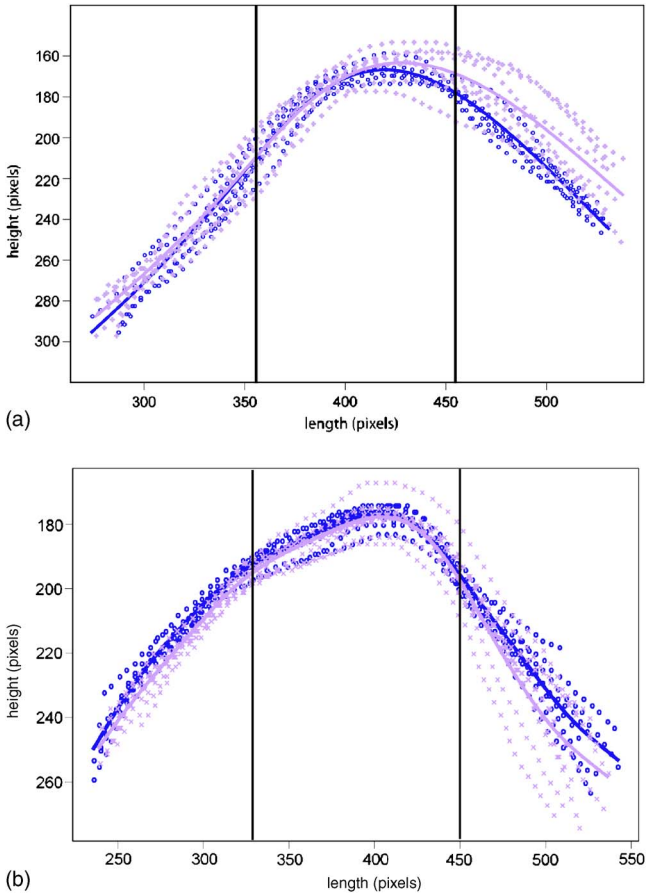


FIG. 6. (Color online) (a) Data points from eight repetitions and smoothing spline estimate (solid lines) for the /g/ in "bag dazzled" and "Baghdad" for speaker TO. "bag dazzled" is represented by the dark blue line and the "o" data points, and "Baghdad" by the pink line and '+' data points. (b) Data points and smoothing spline estimate for the /z/ in "jazz dancer" (dark blue) and "NASDAQ" (pink) for speaker RE.

can be seen impressionistically in Fig. 6. In Fig. 6(a), for example, the tongue blade and body for the /g/ of TO's "Baghdad" is somewhat higher and fronted. Figure 6(b) displays a comparison of the /z/ in "jazz dancer" and "NASDAQ" for speaker RE, with slight raising of the tongue anterior for the /z/ in "jazz dancer." Although no palate shape data were collected for this study, these differences likely correspond to differences in the degree and/or location of constriction for the consonant being produced. In the case of the /g/ of "Baghdad," the constriction location may be more fronted, whereas the /z/ in "jazz dancer" appears to have an increased constriction degree.

B. Smoothing spline ANOVA

The SS ANOVA has been used in applications that require a statistical technique to determine whether the shapes of multiple curves are significantly different from one another. In addition to the study of circadian rhythm mentioned in the Introduction (Wang *et al.*, 2003), SS ANOVAs have also been applied to studies in environmental science and epidemiology (Gu and Wahba, 1993a, b; Wahba *et al.*, 1995).

The SS ANOVA was implemented in S-Plus 2000 using the ASSIST library for fitting spline-based models (Wang and Ke, 2002). The SS ANOVA model is of the form in Eq. (2). Each component of f is estimated with a smoothing spline:

$$f = \mu + \beta x + \text{main group effect} + \text{smooth}(x) + \text{smooth}(x;\text{group}). \quad (2)$$

Unlike a standard ANOVA, the SS ANOVA does not return an F value. Instead, the smoothing parameters of the components smooth(x) and smooth(x;group) are compared to determine their relative contributions to the equation. In the ANOVA model, the main group effects correspond to the smoothing splines for each data set [for example, the dark blue data for "bag dazzled" versus pink data for "Baghdad" in Fig. 6(a), color online], smooth(x) is the single smoothing spline that would be the best fit for all of the data put together (not represented in these diagrams), and the interaction term smooth(x;group) is the smoothing spline representing the difference between a main effect spline and the smooth(x) spline.

The interaction term smooth(x;group) is examined to determine whether the curves representing each group are significantly different. If the two curves being compared have different shapes, then smooth(x;group) will be a significant component of f. Significance is determined by comparing the smoothing parameter value for the interaction term smooth(x;group) with the smoothing parameter value for smooth(x). If smooth(x) and smooth(x;group) are of the same order of magnitude, then it is likely that at least some regions along the two curves are significantly different. In this case, the order of magnitude refers to the nearest power of 10; thus, smoothing parameters with values of 8 and 30 would be considered to be within the same order of magnitude, since both are numerically close to 10^1. However, smoothing parameter values of 8 and 110 would be within different orders of magnitude, since 110 is nearest to 10^2. Furthermore, it should be emphasized that the order of magnitude criteria for the smoothing parameters is only a rough metric that does not guarantee that differences are not significant. In cases of extreme difference, such as values of 0.1 versus 10 000, it may be assumed that there are no significant differences among the curves. However, differences of 0.1 versus 10 may still contain a significant difference at some point along the curve. For the comparison of the /g/ of "bag dazzled" and "Baghdad," the smoothing parameters for smooth(x) and smooth(x;group) are 6.04 and 28.05, respectively. These values are within the same order of magnitude. Visual inspection suggests that the front third of the tongue shapes in Fig. 6(a) are significantly different from one another, but in order to confirm this, 95% Bayesian confidence intervals can be constructed to determine whether the curves are significantly different at any point in the comparison (Gu and Wahba, 1993b; Wahba, 1983).

C. Bayesian confidence intervals

The first step is to construct 95% Bayesian confidence intervals around the smoothing splines for the main effects curves themselves. This is illustrated in Fig. 7. When the

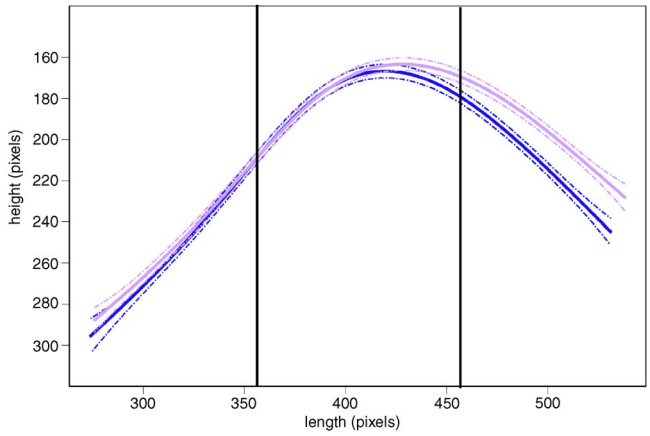


FIG. 7. (Color online) Smoothing spline estimate and 95% Bayesian confidence interval for comparison of the mean curves for /g/ in "bag dazzled" and "Baghdad" for subject TO. "bag dazzled" is represented by the dark blue line, and "Baghdad" by the pink line. The axes and scales are the same as in Fig. 6.

confidence intervals of the main effects curves overlap, the differences between two curves are not significant.

To better examine where significant differences are, Bayesian confidence intervals can also be constructed for the interaction curves. The interaction curves for each of the data sets being compared are a plot of the difference of the smoothing spline for each data set from the smoothing spline that is the best fit to all of the data [i.e., smooth(x)]. The interaction effects for the main effects curves shown in Fig. 7 are illustrated in Fig. 8. Though the mean interaction curves for each data set are mirror images, the confidence intervals for each one may be different, which is why both interaction curves are provided in the figures. If the confidence interval encompasses the zero on the y axis at any point along the interaction curve, there is no difference between the two curves being compared; the interaction at that point is not statistically significant. In Fig. 8, the Bayesian confidence

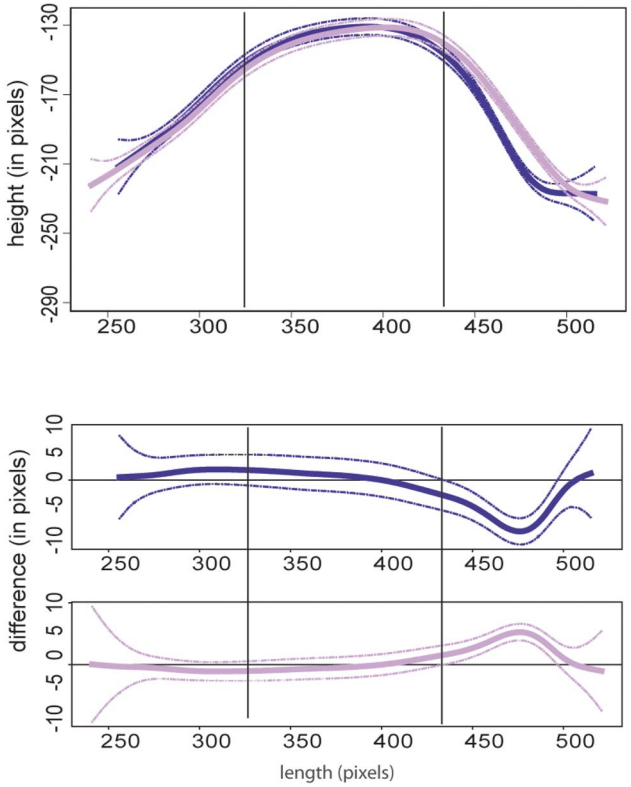


FIG. 9. (Color online) Smoothing splines for data sets (top) and interaction effects with Bayesian confidence intervals (bottom) for the shapes for /k/ in "black top" (dark blue) and "blacktop" (pink) for speaker RE.

intervals encompass zero for about two-thirds of the entire length of the tongue, starting at the tongue root. Thus, the front part of the tongue curves for "bag dazzled" and "Baghdad" are significantly different than one another.

Figure 9 contains the smoothing splines for the production of /k/ in "black top" versus "blacktop" for speaker RE. For this comparison, the smoothing parameter values for smooth(x) and smooth(x;group) were 0.88 and 0.44, respec-

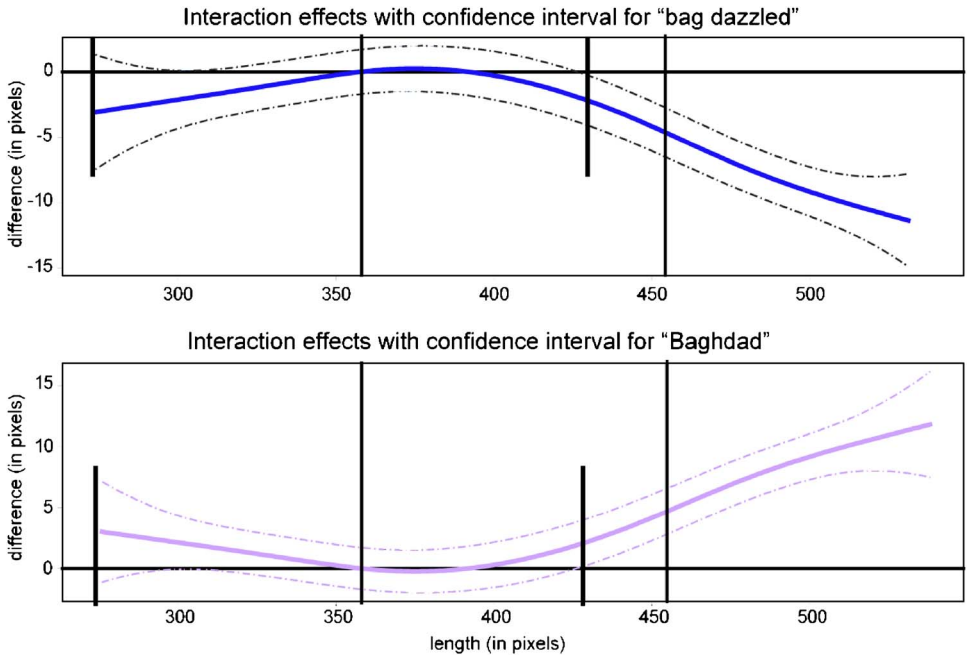


FIG. 8. (Color online) Interaction effects with Bayesian confidence intervals for the shapes for /g/ in "bag dazzled" and "Baghdad" for speaker TO. The splines representing the interaction effect are mirror images because they represent the difference of main effect spline (as shown in Fig. 7) from the spline that best fits all data for "bag dazzled" and "Baghdad." However, both images are shown because the confidence intervals can be different. The x axis is length, and the y axis is the difference between each data set and the spline that fits all data for "bag dazzled" and "Baghdad." When the confidence interval encompasses 0, the curves are not significantly different. The short, thick lines in each image demarcate the part of the interaction curve that is not significantly different.

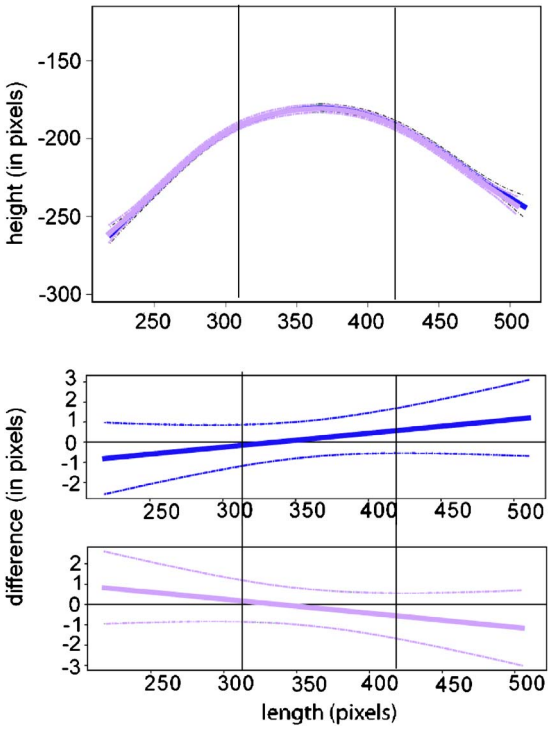


FIG. 10. (Color online) Smoothing splines for data sets (top) and interaction effects with Bayesian confidence intervals (bottom) for the shapes for /z/ in "jazz dancer" (dark blue) and "NASDAQ" (pink) for speaker SH.

tively. In this example, the confidence intervals for the interaction effects are different for "black top" (dark blue line) and "blacktop" (light/pink line) (color online). This is most evident at the ends of the curves, where ultrasound data are often less consistent since the imaging quality at the ends of the tongue curve may be slightly degraded, and therefore harder to accurately track. Such variability will be reflected in the Bayesian confidence intervals of the smoothing splines for both the main effects and the interaction. However, it is also clear in this figure that the confidence interval surrounding the interaction effect for "black top" (dark blue line) is wider than the interval for "blacktop" (light/pink line). This indicates greater variability for "blacktop."

In the example in Fig. 9, the section of the curve relevant to determining whether there is a difference in constriction for the two different types of /k/ is again the middle third. The interaction curves indicate that there is no significant difference anywhere in that region. There is a significant difference along most of the section corresponding to the anterior part of the tongue (the rightmost third), although at the very end of the curve the curves are again very close to one another. This is due to the increased variability at the tongue tip, as indicated by the widening confidence intervals.

Figure 10 demonstrates the production of /z/ in "jazz dancer" (black) and "NASDAQ" (gray) by speaker SH, in which there are no significant differences at all along any point in the curve. The smoothing parameter values for smooth(x) and smooth(x;group) were 7.92 and 2 608 893, respectively.

IV. GENERAL DISCUSSION

The smoothing spline ANOVA is a useful technique for providing a statistical analysis of differences among tongue shapes acquired by ultrasound imaging. When multiple repetitions of an utterance are collected, smoothing splines in conjunction with Bayesian confidence intervals are an appropriate method to account for the shapes that best fit the data and the variance in production. In the examples given above, the articulation corresponding to the most constricted position of a word-final consonant (e.g., *blac**k** top*) was compared to that of a word-medial consonant (e.g., *blac**k**top*). By looking at either the whole tongue curve or a particular region, depending on the researcher's interest, it can be determined whether the tongue shapes for a given articulation are the same or different when some context is varied. In the case of the /g/ in "bag dazzled" versus "Baghdad" for speaker TO (Fig. 7), a significant difference in the rightmost section of the tongue extending into the middle third of the tongue suggests a difference between the constrictions of /g/. In the example of /k/ for "black top" and "blacktop" for speaker RE (Fig. 9) and /z/ for "jazz dancer" and "NASDAQ" for speaker SH, however, there were no significant differences in the relevant regions.

One advantage of the SS ANOVA technique is that any changes in shape, rotation, or translation are taken into account in the statistical analysis. When the head and transducer are stabilized, it can be assumed that any changes not just in the tongue shape, but also in translation (shift on the x or y axis) or rotation of the tongue curve, are of interest to the question being researched. Translation changes, for example, may indicate a change in the backness dimension for a vowel, or may reflect the effects of coarticulation on the production of a consonant.

A few comments about the interpretability of the SS ANOVA should be mentioned. First, unlike methods such as electromagnetic midsagittal articulography (EMMA) (Perkell *et al.*, 1992) or cine-MRI (Stone *et al.*, 2001), ultrasound is not a point tracking technique. Although the smoothing splines representing the tongue shapes for articulation being compared may touch or cross in some spots, it is not the case that the location of contact occurs at the exact same point of the tongue. Thus, the fact that there will be no statistically significant difference between the curves at the point where tongue curves cross should be interpreted with care. As noted in the introduction, a point-tracking technique like EMMA is limited in that it can only provide information about tongue shape and motion for as many pellets as are placed on the tongue (usually around four), whereas the whole midsagittal or coronal contour of the tongue can be imaged by ultrasound.

Second, it is not immediately obvious how the tongue should be divided into linguistically relevant regions. While factor analysis and principle components analysis have been applied to the characterization of tongue configurations in vowel production, these methods are best suited to classifying the tongue shapes of related classes of sounds, not for examining differences in particular regions of interest (Harshman *et al.*, 1977; Hoole, 1999; Nix *et al.*, 1996; Stone and Lundberg, 1996). For example, Harshman *et al.* (1977) developed the PARAFAC ("parallel factors") algorithm in an effort to reduce the number of factors necessary to describe tongue shape. The measurements submitted to the algorithm

were based on tracings of midsagittal tongue curves from cinefluorograms which were divided into 18 sections individually determined for each speaker. The results of the PARAFAC analysis indicated that tongue shapes could generally be accounted for by two factors referred to as "front-raising" and "back-raising," which characterize the motion and shapes of the tongue blade and tongue dorsum, respectively. While this method is useful for classifying the overall tongue shape for particular articulations, it does not, for example, lend itself well to determining whether the constriction location and degree for an obstruent consonant in word-final position are statistically different from the same consonant in word-medial position.

In the examples presented in this paper, the tongue was partitioned into three equal sections that can be thought to roughly correspond to the tongue tip/blade, body/dorsum, and root. When examining the location of constriction for velar consonants like /k/ and /g/, the region of greatest interest was the middle third, or the dorsum of the tongue, since this is the section of the tongue that is most relevant to the formation of a velar constriction. However, it is possible—even likely—that the equal division of the tongue surface into three sections is neither the most anatomically nor linguistically accurate method for examining movements and constrictions of different parts of the tongue. One proposal by Iskarous *et al.* (2003) for segmenting the tongue uses conic arcs to model constriction location and constriction degree; perhaps this technique could be used in conjunction with the SS ANOVA to fully quantify tongue shape curves.

Third, related to the issue of linguistically relevant divisions is how to interpret a significant difference in a region of the tongue that is not obviously pertinent to the question being investigated. For example, if a researcher were studying a language that appeared to have a vowel distinction marked by advanced tongue root (ATR) (Ladefoged and Maddieson, 1996), it might be hypothesized that the only region of interest is the tongue root, which should be more advanced or retracted depending on the vowel being produced. However, since the SS ANOVA and the Bayesian confidence intervals for the interaction provide information about the entire tongue (that is, for example, a researcher cannot avoid the statistical comparison of the tongue blade even if it is not the region of interest), it is possible that significant differences will be revealed both in the tongue root and tongue blade region. Would the researcher want to assign any linguistic import to the distinction in the tongue blade? Or, if a difference were found only in the tongue blade region, would the researcher be forced to conclude that the vocalic distinction in question was not an ATR distinction? Such possibilities ought to be considered by researchers in advance so that they are prepared to interpret findings in which the SS ANOVA reveals an unexpected significant difference in some region of the tongue.

Finally, it is important to emphasize that the SS ANOVA is not appropriate for studies that would involve data collection over multiple sessions. It is extremely difficult to ensure that the transducer is placed in exactly the same place across more than one recording session, which results in a different slice of the tongue being imaged each time. This would rule out, for example, pretreatment/posttreatment studies that aim to use the SS ANOVA to quantify the effect of clinical intervention on an articulation of interest. However, the SS ANOVA could still be useful in clinical applications, such as the comparison of the tongue shapes collected within a single session corresponding to correctly produced velar stops with the disordered productions of alveolar stops as palatalized velar stops (Gibbon *et al.*, 1993).

In conclusion, the smoothing spline ANOVA is a promising method for speech researchers examining the tongue contour of an articulation at a moment in time, such as the most extreme articulation of a gesture of interest. In the future, development of methods that facilitate the analysis of changes over time, including an extension of the SS ANOVA to sequential frames of ultrasound data, will permit researchers to compare changes that span more than just the single frame representing the articulation of a sound being studied.

ACKNOWLEDGMENTS

The author thanks statistician Dr. Kyung Sin (Ph.D., UCSB) for her invaluable help in choosing and implementing the smoothing spline ANOVA for analyzing tongue curves. Thanks also to the members of the New York University Phonetics/Phonology Lab and the participants of Ultrafest III workshop at the University of Arizona, April 2005, for their comments on this work. Maureen Stone and Stefan Benus also provided invaluable input on the manuscript. This work was supported in part by the National Science Foundation CAREER Grant No. BCS-0449560.

Bernhardt, B., Gick, B., Bacsfalvi, P., and Ashdown, J. (**2003**). "Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners," Clin. Linguist. Phonetics **17**(3), 199–216.

Bressmann, T., Thind, P., Uy, C., Bollig, C., Gilbert, R., and Irish, J. (**2005**). "Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving, and symmetry: Data from 12 normal speakers and a partial glossectomee," Clin. Linguist. Phonetics **19**(6/7), 573–588.

Browman, C., and Goldstein, L. (**1995**). "Gestural syllable position effects in American English," in *Producing Speech: Contemporary Issues for Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (American Institute of Physics, New York).

Craven, P., and Wahba, G. (**1979**). "Smoothing noisy data with spline functions," Numer. Math. **31**, 377–403.

Davidson, L. (**2005**). "Addressing phonological questions with ultrasound," Clin. Linguist. Phonetics **19**(6/7), 619–633.

Eubank, R. (**1988**). *Spline Smoothing and Nonparametric Regression* (Dekker, New York).

Gibbon, F., Dent, H., and Hardcastle, W. (**1993**). "Diagnosis and therapy of abnormal alveolar stops in a speech-disordered child using electropalatography," Clin. Linguist. Phonetics **7**(4), 247–267.

Gick, B. (**2002**). "The use of ultrasound for linguistic phonetic fieldwork," J. Int. Phonetic Assoc. **32**(2), 113–122.

Green, P. J., and Silverman, B. W. (**1994**). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach* (Chapman and Hall, London).

Gu, C. (**2002**). *Smoothing Spline ANOVA Models* (Springer, New York).

Gu, C., and Wahba, G. (**1993a**). "Semiparametric analysis of variance with tensor product thin plate splines," J. R. Stat. Soc. Ser. B. Methodol. **55**, 353–368.

Gu, C., and Wahba, G. (**1993b**). "Smoothing spline ANOVA with component-wise Bayesian confidence intervals," J. Comput. Graph. Stat. **2**, 97–117.

Guo, W., Dai, M., Ombao, H., and von Sachs, R. (**2003**). "Smoothing spline ANOVA for time-dependent spectral analysis," J. Am. Stat. Assoc. **98**(463), 643–652.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.
Hoole, P. (**1999**). "On the lingual organization of the German vowel system," J. Acoust. Soc. Am. **106**, 1020–1032.
Iskarous, K. (**2005**). "Patterns of tongue movement," J. Phonetics **33**, 363–381.
Iskarous, K., Goldstein, L., Whalen, D., Tiede, M., and Rubin, P. (**2003**). "CASY: The Haskins Configurable Articulatory Synthesizer," in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Solé, D. Recasens, and J. Romero (Universitat Autónoma de Barcelona, Barcelona), pp. 185–188.
Kochetov, A. (**2006**). "Syllable position effects and gestural organization: Articulatory evidence from Russian," in *Papers in Laboratory Phonology VIII*, edited by L. Goldstein, D. Whalen, and C. Best (Mouton de Gruyter, Berlin).
Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford).
Li, M., Kambhamettu, C., and Stone, M. (**2005**). "Automatic contour tracking in ultrasound images," Clin. Linguist. Phonetics **19**(6/7), 545–554. EdgeTrak available at http://speech.maryland.edu/software.html. (website last viewed on 21 April 2006).
Luo, Z., Wahba, G., and Johnson, D. R. (**1998**). "Spatial-temporal analysis of temperature using smoothing spline ANOVA," J. Clim. **11**, 18–28.
Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). "Two cross-linguistic factors underlying tongue shapes for vowels," J. Acoust. Soc. Am. **99**, 3707–3717.
Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.
Ramsay, J. O., Munhall, K., Gracco, V., and Ostry, D. (**1996**). "Functional data analyses of lip motion," J. Acoust. Soc. Am. **99**, 3718–3727.
Slud, E., Stone, M., Smith, P., and Goldstein, M. (**2002**). "Principal components representation of the two-dimensional coronal tongue surface," Phonetica **59**, 108–133.
Stone, M. (**2005**). "A guide to analyzing tongue motion from ultrasound images," Clin. Linguist. Phonetics **19**(6/7), 455–502.
Stone, M., and Davis, E. P. (**1995**). "A head and transducer support system for making ultrasound images of tongue/jaw movement," J. Acoust. Soc. Am. **98**, 3107–3112.
Stone, M., and Lundberg, A. (**1996**). "Three-dimensional tongue surface shapes of English consonants and vowels," J. Acoust. Soc. Am. **99**, 3728–3737.
Stone, M., Faber, A., Rafael, L., and Shawker, T. (**1992**). "Cross-sectional tongue shape and linguopalatal contact patterns in [s], [esh], and [l]," J. Phonetics **20**(2), 253–270.
Stone, M., Davis, E. P., Douglas, A., Ness Aiver, M., Gullapalli, R., Levine, W. *et al.* (**2001**). "Modeling tongue surface contours from Cine-MRI images," J. Speech Lang. Hear. Res. **44**, 1026–1040.
Wahba, G. (**1983**). "Bayesian confidence intervals for the cross validated smoothing spline," J. R. Stat. Soc. Ser. B. Methodol. **45**(1), 133–150.
Wahba, G. (**1990**). *Spline Models for Observational Data* (Society of Industrial and Applied Mathematics, Philadephia).
Wahba, G., Wang, Y., Gu, C., Klein, R., and Klein, B. (**1995**). "Smoothing spline ANOVA for exponential families, with application to the Wisconsin epidemiological study of diabetic retinopathy," Ann. Stat. **23**(6), 1865–1895.
Wang, Y., and Ke, C. (**2002**). *ASSIST: A Suite of S-Plus Functions Implementing Spline Smoothing Techniques*. Available at http://www.pstat.ucsb.edu/faculty/yuedong/research (website last viewed on 21 April 2006).
Wang, Y., Ke, C., and Brown, M. (**2003**). "Shape-invariant modeling of circadian rhythms with random effects and smoothing spline ANOVA decompositions," Biometrics **59**, 804–812.
Westbury, J. (**1994**). *X-ray Microbeam Speech Production Database User's Handbook, Version 1* (Univ. of Wisconsin, Madison).

Some difference limens for the perception of breathiness[a)]

Rahul Shrivastav[b)] and Christine M. Sapienza
Department of Communication Science and Disorders, Dauer Hall, P.O. Box 117420, University of Florida, Gainesville, Florida 32611

(Received 7 March 2005; revised 11 April 2006; accepted 8 May 2006)

Perception of breathy voice quality appears to be cued by changes in the vowel spectrum. These changes are related to alterations in the intensity of aspiration noise and spectral slope of the harmonic energy [Shrivastav and Sapienza, J. Acoust. Soc. Am., **114** (4), 2217–2224 (2003)]. Ten young-adult listeners with normal hearing were tested using an adaptive listening task to determine the smallest change in signal-to-noise ratio that resulted in a change in breathiness. Six vowel continua, three female and three male, were generated using a Klatt synthesizer and served as stimuli. Results showed that listeners needed as much as 20-dB increase in aspiration noise to perceive a change in breathiness against a relatively normal voice. In contrast, listeners needed approximately an 11-dB increase in aspiration noise to discriminate breathiness against a severely breathy voice. The difference limens for breathiness were observed to vary across the six talkers. Voices having aspiration noise that was predominantly in the high frequencies had smaller difference limens. No significant differences for male and female voice were observed.
 [DOI: 10.1121/1.2208457]

PACS number(s): 43.71.Bp, 43.70.Gr, 43.71.Gv, 43.70.Dn [AL] Pages: 416–423

I. INTRODUCTION

Voice quality is a paralinguistic feature of speech that plays an important role in cueing information such as gender, age, and identity of the speaker, conveying appropriate emotion through speech, and influencing the overall intelligibility of speech. Following the ANSI (1960) definition of sound quality, voice quality may be defined as "the left over perception after pitch, loudness and phonetic category have been identified" (Titze, 1994, p. 252). The voice quality of some speakers may be described as being "breathy." Such voices are often characterized by "an audible escape of air resulting in thin and weak phonation" (Bassich and Ludlow, 1986, p. 133). Although breathy voice quality may often be heard in normal speakers, it is frequently encountered in dysphonic speakers who are unable to adequately close the glottis during voice production (Hammarberg *et al.*, 1980). For instance, the voice quality of patients with vocal fold paralysis or age-related vocal fold bowing is often described as breathy (Shindo *et al.*, 1996; Boone *et al.*, 2005). Additionally, in some languages such as Hmong and Gujarati, breathy voice quality also serves to contrast two phonological categories (Fischer-Jorgenson, 1967; Huffman *et al.*, 1987).

Although voice quality is essentially a perceptual construct resulting from specific changes to the speech acoustic signal (Kreiman and Gerratt, 2000), there is little consensus about the acoustic cues for the perception of breathiness. Past research correlating acoustic measures of voice with the perception of breathiness has been inconclusive due to inconsistent results across different experiments (Kreiman and Gerratt, 2000; Hillenbrand *et al.*, 1994). First, experiments differ in the measures found to correlate best with perceptual ratings of breathiness. For example, perceptual judgments of breathiness have been found to correlate with measures of spectral slope (de Krom, 1995; Klatt and Klatt, 1990), the intensity of the noise in the signal (Feijoo and Hernandez, 1990; Hirano *et al.*, 1988; Klatt and Klatt, 1990), as well as measures of frequency and amplitude perturbation (Feijoo and Hernandez, 1990; Martin *et al.*, 1995; Prosek *et al.*, 1987). Second, perceptual judgments of breathiness and specific acoustic measures have been observed to be highly inconsistent. For example, the correlation between aspiration noise and breathiness has been found to vary from 0.1 to 0.7 (Klatt and Klatt, 1990), that for pitch perturbation quotient and breathiness varied from 0.38 to 0.67 (Hirano *et al.*, 1988; Prosek *et al.*, 1987) and that for harmonic to noise ratio (HNR) and breathiness varied from −0.25 to −0.69 (de Krom, 1994; Martin *et al.*, 1995).

Such inconsistent findings may arise from a number of factors, including differences in the algorithms and procedures used to determine acoustic measures as well as differences in experimental design for obtaining perceptual data. In addition, Shrivastav and Sapienza (2003) attributed a part of these inconsistencies to the nature of the acoustic measures used to predict perceived breathiness in previous research. They argued that most acoustic measures such as the signal-to-noise ratio, spectral slope, or signal perturbation do not account for the nonlinear nature of the auditory transduction process. In contrast, a measure such as the "partial loudness of the harmonic energy" fares better at predicting perceptual judgments of breathiness because it accounts for some of the nonlinear processes in the auditory system. The partial loudness of the harmonic energy, which is calculated through the use of an auditory processing model as a signal processing front-end, was able to account for significantly

[a)]Part of this research was presented at the 145th Meeting of the Acoustical Society of America, Nashville, TN.
[b)]Electronic mail: rahul@csd.ufl.edu

 0001-4966/2006/120(1)/416/8/$22.50

greater variance in the perceptual ratings of breathiness than conventionally used acoustic measures (Shrivastav and Sapienza, 2003, Shrivastav, 2003).

The "partial loudness" calculated using a loudness model (Moore *et al.*, 1997) assumes that the aspiration noise in voices acts as an auditory masker for the harmonic energy produced by vocal fold vibration. The perception of breathiness, therefore, was found to be related to both aspiration noise and the spectral slope of the harmonic energy. An increase in the intensity of the aspiration noise would result in greater masking and a reduced partial loudness of the harmonic energy. Similarly, an increase in the spectral slope without any changes to the aspiration noise would also reduce the partial loudness of the harmonic energy. This is because greater spectral slope results in a lowering of the overall intensity of the harmonics, particularly in the higher frequencies. Finally, partial loudness may also be affected by changes in the spectral shape of the aspiration noise and the harmonic series. Two voices that have the same overall intensity of aspiration noise but which differ in their spectral shape may have different partial loudness. Such differences can arise because the auditory system is not equally sensitive at all frequencies.

These findings show that listeners' ability to discriminate two voices in terms of breathiness depends upon their sensitivity to small changes in the spectrum, such as those resulting from changes in the aspiration noise and spectral slope. However, few experiments have investigated the spectral change necessary for listeners to perceive a change in breathiness. Estimation of such difference limens (DL) is necessary to develop meaningful scales for voice quality. For example, this information can be used to understand and predict the acoustic changes that will or will not affect the perception of breathiness.

One experiment that provides some information regarding the DL for breathiness was reported by Kreiman and Gerratt (2005). They asked listeners to match the quality of a synthetic stimulus to that of a target by varying the aspiration noise-to-signal ratio (NSR). In this method-of-adjustment task, listeners were observed to show greater agreement in matching the most breathy voices, suggesting a smaller DL for these voices. Following this experiment, listeners were asked to make a "same-different" judgment for pairs of synthetically generated voices and the percent-correct score at different levels of aspiration noise were calculated. The data were then extrapolated to determine the NSR at which listeners could discriminate two voices with 75% accuracy. They found that listeners needed an average of 10.65-dB change in NSR to discriminate two voices identical in all respects except the intensity of aspiration noise. The differences across individual talkers were attributed to the variation in the source excitation patterns and voices with a low spectral slope were observed to have a greater DL for NSR.

However, this experiment was not specifically designed to determine the DL for intensity of aspiration noise. Thus, for example, these experiments did not specify how the DL for intensity of aspiration noise may change with respect to the intensity of noise in the standard. The goal of the present research was to determine the DL for breathiness for a series of voice stimuli that varied only in terms of their aspiration noise. Breathiness was manipulated by varying the level of the aspiration noise and, hence, the signal-to-noise ratios (SNRs) for the vowel stimuli and listeners were tested using a two-interval forced choice method. The SNR was varied by modifying the level of the aspiration noise in a Klatt-synthesizer (Klatt and Klatt, 1990), while maintaining all other synthesis parameters at a constant value. The DL for breathiness was determined at three different levels of SNR for each voice stimulus. The findings of this experiment will help understand how listeners judge breathy voice quality and help in the development of appropriate algorithms to predict the magnitude of breathiness from the vocal acoustic signal.

TABLE I. The parameters used to synthesize the six voice stimuli. The range of variation for AH is indicated in the table. AH was varied in 1-dB steps. All other parameters were kept constant.

	Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
F0	100	132	135	225	245	210
AV	60	60	60	60	60	60
OQ	50	35	60	75	20	50
SQ	100	150	400	200	200	300
TL	5	10	20	20	10	15
FL	10	5	15	5	0	0
AH	30–70	30–70	30–70	30–70	30–70	30–70
FNP	500	500	500	500	500	500
BNP	200	200	200	200	200	200
F1	800	650	1100	900	900	850
B1	100	110	125	175	145	300
DF1	0	0	20	0	10	35
DB1	0	0	40	0	25	50
F2	1200	1150	1550	1350	1350	1300
B2	125	135	150	200	200	275
F3	2850	2850	3800	3200	3100	2950
B3	175	165	175	150	225	350
F4	3500	3500	4000	3750	3600	3500
B4	350	350	350	350	350	4500
F5	4500	4500	4500	4500	4500	4500
B5	500	500	500	500	500	500

II. METHODS

A. Stimuli

Six instances of the vowel /a/ representing three male and three female, were synthesized using a Klatt Synthesizer (KLSyn; Sensimetrics Inc.). The LF model (Fant *et al.*, 1985) was used for generating the sound source. The three male talkers are henceforth referred to as M1, M2, and M3, and the three female talkers are referred to as F1, F2, and F3. Each vowel was 500 ms in duration. The initial parameters (fundamental frequency and first three formants) for synthesizing these voices were modeled after six voices arbitrarily chosen from a large database of dysphonic voices. Other synthesis parameters (formant bandwidths, open quotient, and spectral tilt) were then modified to achieve the most natural sounding stimuli as judged by two listeners in an informal listening test. The final parameters for each of these six talkers are shown in Table I. For each of these synthetic

vowels, a continuum of stimuli varying in the intensity of aspiration noise (AH) was created. Each continuum consisted of 41 stimuli, identical in all aspects except for AH, which was varied from 30 to 70 dB in 1-dB steps. The intensity of voicing (AV) and other synthesis parameters were maintained at a constant level to obtain a fixed intensity of the harmonics in each stimulus. Although other source parameters such as the open quotient and spectral tilt have been shown to affect the perception of breathiness (Klatt and Klatt, 1990), these were not manipulated in the present experiment. Therefore, a total of 246 stimuli were synthesized (6 talkers $\times$ 41 stimuli). To minimize listener responses to changes in overall intensity, all stimuli were scaled to have equal root mean square energy. Finally, the amplitude envelopes of all stimuli were modified to obtain a rise time and decay time of 20 ms. This was done to prevent the occurrence of an audible click at the onset and offset of the stimuli. All stimuli were generated with a sampling rate of 11.025 kHz and with 16-bit quantization.

B. Listeners

Ten young-adult females (mean age: 22 years) were recruited from the graduate program in Speech-Language Pathology at the University of Florida. All listeners were native speakers of American English and had no prior experience in making voice quality judgments. However, all listeners had studied about dysphonic voice quality in their coursework and were familiar with breathy voice quality. All listeners passed a hearing screening at 20 dB HL at octave frequencies between 250 and 4000 Hz. Listeners were paid for participating in the experiment.

C. Procedures

Each listener was tested in three 1-h sessions within a 2-week period. In each session, the DL for the six talkers was determined using a two-interval forced choice (2IFC) procedure. Listeners were presented a standard and a test stimulus and were asked to identify whether the two were same or different in terms of their breathiness. Listeners were instructed to avoid making responses based upon differences in overall loudness or the pitch of the two stimuli in each pair.[1] The 2IFC procedure was run with three different levels of the standard stimulus (AH values set at 30, 40, and 50 dB). Therefore, the DL for breathiness was calculated at three different levels of AH for each talker. Stimuli were presented binaurally at 75 dB SPL through an RP2 signal processor (Tucker-Davis Technology, Inc.) and TDH-39 headphones. The 2IFC task was controlled using ECos/Win (Avaaz Innovations, Inc.). The order of presentation of the six talkers was randomized across listeners and test sessions.

The 2IFC task proceeded as follows. Listeners heard pairs of stimuli ("standard" stimulus followed by the "test" stimulus) from the same talker. The test stimulus in the first pair always had a significantly greater level of AH than in the standard. Listeners were asked to respond whether the two members of each pair were the same or different in terms of their breathiness. The two stimuli in each pair were separated by a silent interval of 500 ms. If the listener responded the two stimuli in that pair to be "different," then the level of AH in the test stimulus was decreased. Conversely, if the listener indicated the two to have the "same" breathiness, then the level of AH in the test stimulus was increased. The DL for a particular voice was defined as the level of AH at which listeners judged the two voices to have the same breathiness 70.7% of the time (Levitt, 1971). The level of AH in the test stimulus was changed in 3-dB steps initially, but after the first two reversals in the listener's response the step size was reduced to 1 dB. The test proceeded until 12 reversals were obtained, and the DL was calculated by averaging the magnitude of change in the level of AH at the last eight reversals.

D. Acoustic analyses

Previous research has suggested that the DL for breathiness may vary across talkers and may depend upon the acoustic characteristics of that stimulus (Kreiman and Gerratt, 2005). To investigate possible interactions between the acoustic characteristic of voices and their DL for breathiness, the spectrum for the harmonic energy and the aspiration noise in each voice continuum was determined. Although each voice continua was generated using the same range of AH (30 to 70 dB) as input to the speech synthesizer, the resulting output spectra varied across the six voices. This is because the input aspiration noise was modified by a different set of filters (corresponding to the vowel formants) for each talker. The aspiration noise in each stimulus was isolated by generating each stimulus with the amplitude of voicing (AV) set to zero and the AH set to 60 dB. Similarly, a noise-free copy of each stimulus was generated by setting AH=0 and AV=60. Average fast Fourier transform (FFT) spectrum of the aspiration noise and the harmonic energy for each talker was obtained using 20-ms Hamming windows with a 10-ms overlap between successive windows. The harmonic energy and aspiration noise spectra for stimuli generated with AV=60 and AH=40 are shown in Fig. 1. The spectra for signals generated with AH=0 were characterized by measuring the difference in amplitude between the first and the second harmonic ($H1^*$-$H2^*$) and the difference in amplitude between the first harmonic and the third formant ($H1^*$-A3).[2] These measures estimate the spectral slope of the harmonic energy and have been shown to be correlated with differences in the voicing source characteristics (Hanson, 1997). The spectra for signals generated with AV=0 were compared by considering each spectrum as a probability distribution function and calculating the first four moments for each distribution. These moments were calculated because they have been successful in characterizing noise spectra in speech (Forrest *et al.*, 1988).

III. RESULTS

The average DL for all six talkers was found to be 20.74, 14.35, and 11.06 dB when the AH level for the standard stimulus was set to 30, 40, and 50 dB, respectively. A two-way analysis-of-variance (ANOVA) was used to determine if there were any differences in the DL across the six talkers and the three levels of AH for the standard stimulus.

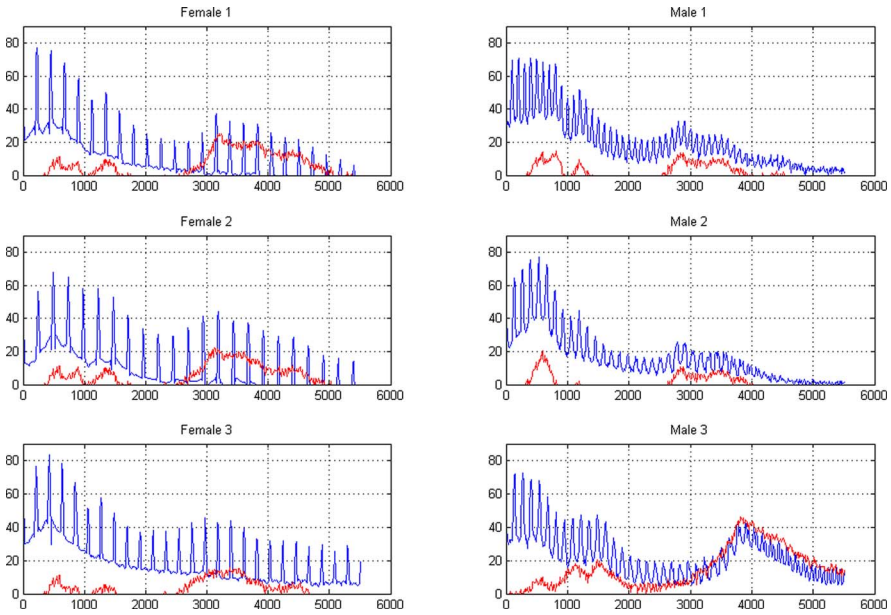


FIG. 1. (Color online) Harmonic energy and aspiration noise for stimuli generated with AV=60 and AH=40. These spectra show the stimuli at the 40-dB comparison level.

The six talkers and the three levels of AH in the standard were treated as independent variables and the DL served as the dependent variable. *Posthoc* tests using Scheffe correction were performed for each of the two independent variables.

A. Main effect of AH-level in the standard

The DL for breathiness was found to be significantly different at each of the three AH-levels in the standard stimulus ($F=171.639$; df=2; $p<0.001$). Listeners needed an average of 20.74-dB increase in AH to discriminate stimuli from a standard generated with AH set to 30 dB. In contrast, listeners only needed an 11.06-dB change in AH when the standard stimulus had AH values of 50 dB. In other words, listeners needed smaller changes in AH to discriminate breathiness in stimuli with high levels of noise, but needed a large change in AH if the standard stimulus had a high signal-to-noise ratio. *Posthoc* pairwise comparisons showed that the DL at each of the three AH intensities in the standard was significantly different from each other ($p<0.01$). These results are shown in Fig. 2.

B. Main effect of talker

Although all six talkers showed a decrease in DL as the level of AH in the standard stimulus increased, they were observed to have significantly different DL for each of the three standards ($F=812.351$; df=5; $p<0.001$). *Posthoc* pairwise comparisons showed that the six voices fell into three distinct groups (Group I consisting of talkers F3 and M2; Group II consisting of talkers F2 and M1; Group III consisting of talkers F1 and M3) that were significantly different from each other, suggesting that listeners' ability to discriminate breathiness in vowel stimuli is affected by factors other than the overall level of the AH alone. These differences may arise from differences in the spectral shape of the harmonic energy and/or that of the aspiration noise for the six talkers. Since the male and female talkers did not group together, these differences could not be attributed to changes in pitch across the six stimuli. The differences in DL for the six talkers are shown in Fig. 3.

C. Interaction between AH-level in the standard and talkers

A significant interaction was obtained between AH and talkers ($F=181.327$; df=10; $p<0.001$). This was further evident from the observation that the six talkers differed in the slope of their DL functions (change in DL with increasing AH levels in the standard stimulus). Talkers with the largest DL (F3 and M2) were observed to have a steeper slope as compared to those with smaller DL, further suggesting that listeners' ability to discriminate breathiness in two voices was affected by the nature of the harmonic and/or the aspiration noise spectra.

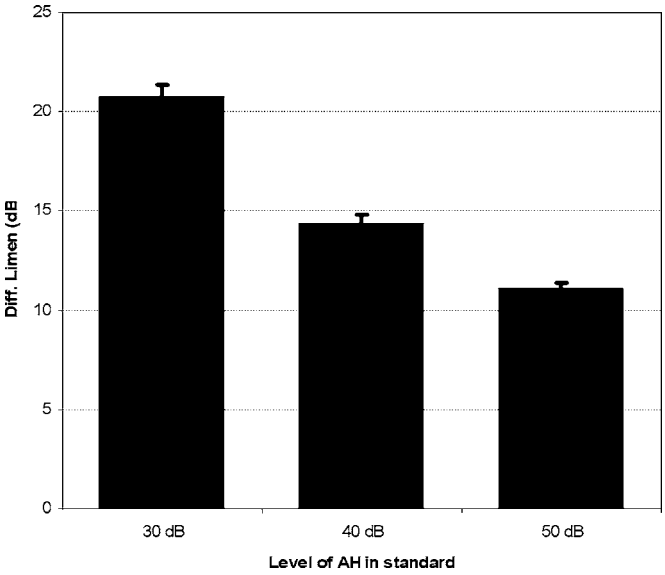


FIG. 2. Mean difference limen and standard deviation for all six talkers at the three levels of AH in the standard. The error bars show the standard error of the mean.

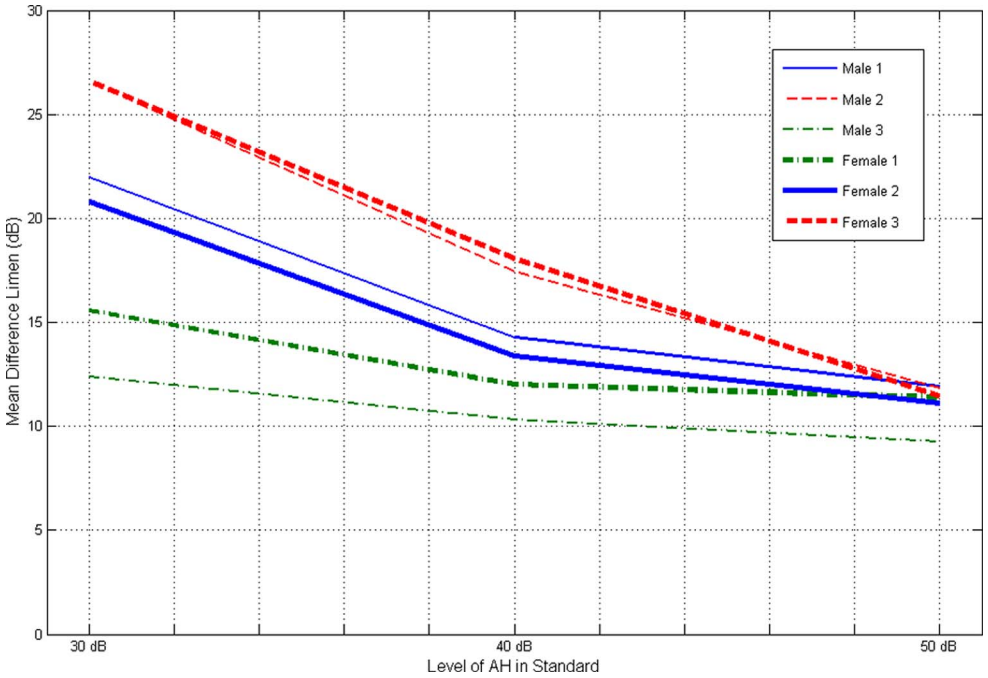


FIG. 3. (Color online) Difference limen functions for all six talkers. Talkers classified as "Group 1" are shown with dashed lines, "Group 2" shown with solid lines and "Group 3" shown with dash-dot lines. Female talkers are indicated with bolder lines.

D. Acoustic characteristics of the stimuli and their effects on DL

The results of the ANOVA showed that the six talkers could be separated into three groups, each consisting of two stimuli. Both talkers within a particular group were observed to have a similar DL function. To understand the potential reasons for differences across talkers, specific spectral characteristics of the harmonic energy and the aspiration noise for each talker were calculated and are shown in Table II. The spectral characteristics of the two talkers within each group were averaged and compared against the average for each of the other two groups. Although there are insufficient data to obtain statistically significant results, a comparison of the group averages suggests that differences in DL for the three groups of stimuli may be related to the spectral characteristics of the aspiration noise. Stimuli exhibiting the steepest DL function (Group III; talkers F3 and M2) were characterized by aspiration noise that had the lowest mean frequency, largest standard deviation, least negative skewness, and the lowest kurtosis. In contrast, stimuli with a less steep DL function (Group I; talkers F1 and M3) had higher mean, lower standard deviation, highest negative skewness, and highest kurtosis. Together these findings suggest that stimuli in which the aspiration noise had greater energy in the higher frequencies exhibit a less steep DL function. No clear trends between DL function and $H1^*$-$H2^*$ or $H1^*$-A3 were observed. The aspiration noise spectra for each of the six talkers are shown in Fig. 4.

IV. DISCUSSION

The perception of breathy voice quality has been shown to be cued by specific changes in the acoustic spectrum of the vowels. These changes may be brought about by the presence of aspiration noise, which tends to mask the higher harmonics in the vowel spectrum (Shrivastav and Sapienza,

TABLE II. Spectral characteristics for the harmonic energy and aspiration noise for the six talkers.

	Harmonic energy		Aspiration noise			
	$H1^*$-$H2^*$	$H1^*$-A3	Mean	SD	Skewness	Kurtosis
F1	3.54	38.34	3.48	0.668	−1.723	6.724
M3	1.60	34.06	3.949	0.242	−2.81	46.01
Group 1 average	**2.57**	**36.20**	**3.7145**	**0.455**	**−2.2665**	**26.367**
F2	−10.22	15.03	3.266	0.767	−1.588	4.44
M1	−1.67	33.76	2.286	1.238	−0.224	−1.284
Group 2 average	**−5.95**	**24.40**	**2.776**	**1.0025**	**−0.906**	**1.578**
F3	−5.41	31.25	3.046	0.977	−1.212	1.535
M2	−5.20	33.33	1.737	1.324	0.367	−1.652
Group 3 average	**−5.31**	**32.29**	**2.3915**	**1.1505**	**−0.4225**	**−0.0585**

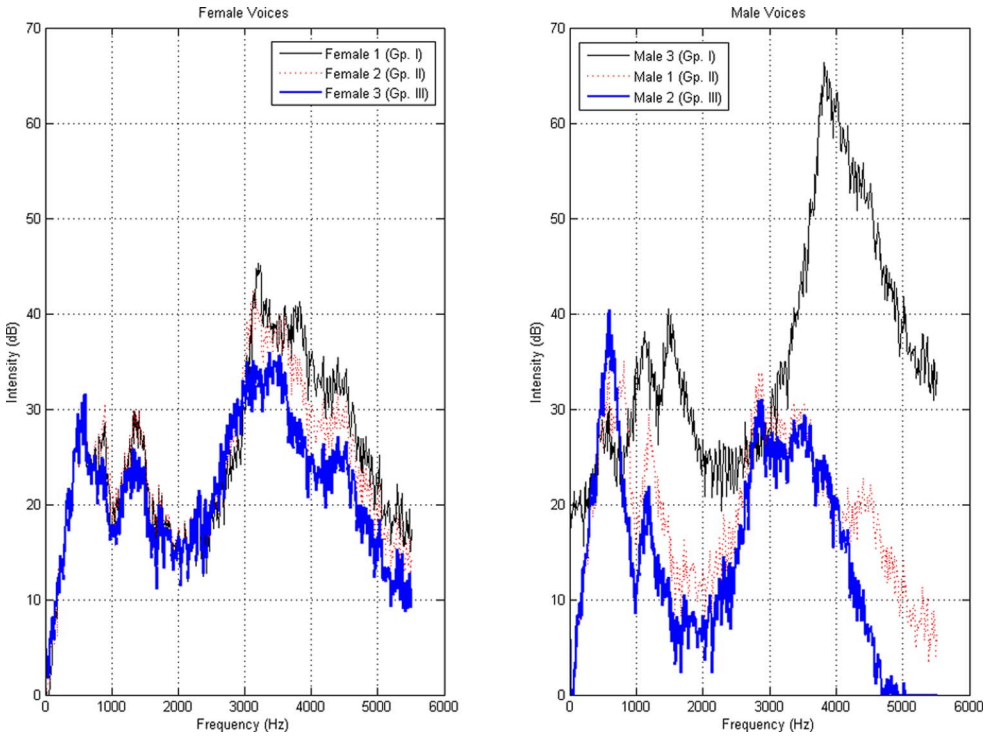


FIG. 4. (Color online) Aspiration noise spectra estimated by synthesizing stimuli with AV=0 and AH=60. Female talkers are shown in the left panel and male talkers in the right panel. Grouping of talkers is based on the ANOVA results. See text for details.

2003). The present experiment attempted to determine the smallest change in the acoustic spectrum of vowels that is necessary for listeners to perceive a change in the degree of breathiness. Results showed that the DL for breathiness in a particular voice decreases as the level of AH in the standard increases. In other words, listeners needed a smaller change to discriminate two voices with greater aspiration noise than those with less aspiration noise. However, the six talkers were found to differ in their DL, with some voices needing considerably greater change in aspiration noise to be perceived as being more (or less) breathy. *Post hoc* analysis of the vowel spectra suggest that these differences may be related to the differences in the aspiration noise spectra. Smaller DL were observed for voices where the aspiration noise was dominated by energy in the high frequencies, whereas voices with aspiration noise concentrated at relatively lower frequencies exhibited larger DL for breathiness.

The differences in DL across the six talkers may be explained on the basis of auditory masking. Shrivastav and Sapienza (2003) suggested that the perception of breathiness could be related to a change in the partial loudness of the harmonic energy in that voice. In voices with poor harmonic energy-to-aspiration noise ratio, a small change in aspiration noise can result in a reasonably large change in partial loudness and, hence, its breathiness. In contrast, those stimuli in which the intensity of the aspiration noise is considerably less than that of the harmonic energy require a much greater increase in aspiration noise for the same effect. Further, the intensity of the lower harmonics is significantly greater than that for the higher harmonics for all talkers (Fig. 1). Therefore, for aspiration noise with the same overall intensity, noise in the higher frequency would be more effective in masking some of the harmonics than noise with predominantly low-frequency energy.

In the present experiment breathiness was controlled by systematically manipulating the level of the aspiration noise. However, the DL for breathiness was significantly larger than those for intensity discrimination of pure tones or broadband stimuli reported previously (for example, Jesteadt *et al.*, 1977). The large DL observed in this experiment may suggest that discrimination of stimuli on the basis of breathiness is related to the masked thresholds for aspiration noise. Listeners may perceive a change in breathiness when the change in the level of AH is greater than its masked threshold. Thus, stimuli with a high SNR needed a greater change in AH level for a perceivable change in breathiness. This would also explain the relatively smaller DL for the 50-dB standard than for the 30-dB standard.

Discrimination of stimuli based on breathiness may also be viewed as resulting from changes in the spectral shape of the harmonic energy brought about as the harmonics are masked by the aspiration noise. The DL for breathiness is comparable to those found for spectral shape discrimination. For example, profile analysis experiments have found that listeners require approximately 10–15-dB change in the target component of a profile before listeners can consistently discriminate the profile from its standard (Drennan and Watson, 2001a, b; Mason *et al.*, 1984). A similar range of discrimination thresholds (5–20 dB) have been reported for broadband noises and speech-shaped noise presented for durations of 250 ms or greater (Farrar *et al.*, 1987; Narendran, 2004).

The DL for stimuli with the highest AH level (AH in standard = 50 dB) are comparable to the DL for NSR estimated by Kreiman and Gerratt (2005) using the paired comparison task. However, Kreiman and Gerratt (2005) observed that a listener's ability to match the noise intensity in a test stimulus to that of a target varied with the spectral slope of the harmonic energy. In contrast, the present experiment found the breathiness DL to be related to the aspiration noise spectra. These differences possibly result from differences in the talkers used in the two experiments and it is likely that neither the spectral slope of the harmonic energy nor the aspiration noise spectra alone are sufficient to explain differences across talkers. Rather, the differences in DL across talkers may be related to the nature of the aspiration noise relative to the harmonic energy as both of these changes may contribute to a change in the partial loudness of the harmonic energy. Voices with steeper spectral slope are characterized by a greater reduction in the intensity of higher harmonics than observed for voices with a less steep spectral slope. Therefore, in voices with steeper spectral slope, the same aspiration noise would result in a greater auditory masking and greater change in partial loudness of the harmonic energy.

All stimuli used in the present experiment were generated at a sampling rate of 11.025 kHz and consequently had little energy above 5 kHz. The use of a relatively low sampling rate may affect the extent to which the obtained DL reflects the DL for natural voice stimuli. An informal analysis of several natural voices that were perceived as "breathy" in a previous experiment (Shrivastav and Sapienza, 2003) showed that most had little energy over 5–6 kHz. However, few voices, particularly those classified as being "severely breathy," were observed to have significant energy at frequencies above 6 kHz. Therefore, the present findings may not correctly reflect the DL for breathiness in the most breathy voices. Unfortunately, most previous experiments that have systematically manipulated the level of aspiration noise for the study of breathiness (for example, Klatt and Klatt, 1990; Kreiman and Gerratt, 2005) have also used stimuli generated with similar bandwidths. Therefore, there is further need to obtain empirical data that shows how energy above 5 kHz may affect the DL for breathiness.

Despite this concern, the findings of this experiment have some important implications. First, these findings show that the perception of breathiness results from multiple acoustic changes in the speech signal. Changes in the magnitude of breathiness were observed to be related to the aspiration noise spectrum as well as the relative level of the aspiration noise. Models to explain the perception of voice quality need to account for such interactions between acoustic cues. Second, these findings highlight one reason why many of the measures proposed to quantify breathiness often do not show high correlation with perceptual data. Differences in DL across the three standards (varying in the level of AH) suggest that an equal increase in the AH level may not result in an equal increase in breathiness. Again, models to explain the perception of voice quality and measures proposed to quantify voice quality need to address such relationships.

V. CONCLUSIONS

Listeners needed as much as 20-dB increase in aspiration noise to discriminate breathiness in a voice with very little noise (i.e., in relatively "normal" voices). In contrast, listeners need approximately 11-dB increase in aspiration noise to discriminate breathiness in voices that have a high intensity of aspiration noise (i.e., for "moderate" or "severe" breathy voices). The DL for breathiness decreases as the intensity of the aspiration noise in the standard stimulus increases. The DL functions for the six talkers tested in the present experiment were found to vary considerably. In general, voices characterized by aspiration noise that is predominantly in higher frequency were observed to have smaller DL. These findings may be explained based on auditory masking of the harmonic energy by the aspiration noise. A model to explain the perception of dysphonic voice quality needs to account for interactions between multiple changes in the acoustic spectrum.

ACKNOWLEDGMENTS

The authors would like to thanks Maria Pinero for help in recruiting and testing listeners. The authors would also like to thank Anders Lofqvist and two anonymous reviewers for their comments.

[1]A pilot experiment showed that instructing listeners to ignore changes in loudness and pitch when making perceptual ratings of voice quality had the desired effect. Listeners gave the same stimulus the same average rating, despite random variation in its overall SPL. The overall intensity of the stimuli was varied by up to 6 dB and the fundamental frequency was varied by 30 Hz.

[2]The amplitudes of the first two harmonics were corrected for their proximity to the first formant frequency using the formula described by Hanson (1997). The amplitude of the third formant could not be appropriately corrected because this correction requires knowledge of the formant frequencies for a neutral vowel produced by the same speaker.

ANSI, S. (**1960**). *USA standard: Acoustical terminology (s1.1)* (American National Standards Institute, Inc., New York).

Bassich, C. J., and Ludlow, C. L. (**1986**). "The use of perceptual methods by new clinicians for assessing voice quality," J. Speech Hear Disord. **51**(2), 125–133.

Boone, D. R., McFarlane, S. C., and Von Berg, S. L. (**2005**). *The Voice and Voice Therapy* (Allyn & Bacon, Boston).

de Krom, G. (**1995**). "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments," J. Speech Hear. Res. **38**(4), 794–811.

Drennan, W. R., and Watson, C. S. (**2001a**). "Sources of variation in profile analysis. I. Individual differences and extended training," J. Acoust. Soc. Am. **110**(Pt. 1), 2491–2497.

Drennan, W. R., and Watson, C. S. (**2001b**). "Sources of variation in profile analysis. I. Component spacing, dynamic changes, and roving level," J. Acoust. Soc. Am. **110**(Pt. 1), 2498–2504.

Fant, G., Liljencrants, J., and Lin, Q. (1985). "A four parameter model of glottal flow," Speech Transmission Laboratory Quarterly Report, pp. 1–3.

Farrar, C. L., Reed, C. M., Ito, Y., Durlach, N. I., Delhorne, L. A., Zurek, P. M., and Braida, D. L. (**1987**). "Spectral-shape discrimination. I. Results from normal-hearing listeners for stationary broadband noises," J. Acoust. Soc. Am. **81**, 1085–1092.

Feijoo, S., and Hernandez, C. (**1990**). "Short-term stability measures for the evaluation of vocal quality," J. Speech Hear. Res. **33**(2), 324–334.

Fischer-Jorgenson, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels in Gujarati," Indian Linguist. **28**, 71–139.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., and Wedin, L. (**1980**). "Perceptual and acoustic correlates of abnormal voice qualities," Acta Oto-Laryngol. **90**(5-6), 441–451.
Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.
Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech Hear. Res. **37**(4), 769–778.
Hirano, M., Hibi, S., Yoshida, T., Hirade, Y., Kasuya, H., and Kikuchi, Y. (**1988**). "Acoustic analysis of pathological voice. Some results of clinical application," Acta Oto-Laryngol. **105**(5-6), 432–438.
Huffman, M. K. (**1987**). "Measures of phonation type in Hmong," J. Acoust. Soc. Am. **81**, 495–504.
Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169–177.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.
Kreiman, J., and Gerratt, B. (**2000**). "Measuring voice quality, in *Voice Quality Measurement*, 1st ed., edited by R. D. Kent and M. J. Bell (Singular, San Diego), pp. 73–101.
Kreiman, J., and Gerratt, B. (**2005**). "Perception of aperiodicity in pathological voice," J. Acoust. Soc. Am. **117**, 2201–2211.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Martin, D., Fitch, J., and Wolfe, V. (**1995**). "Pathologic voice type and the acoustic prediction of severity," J. Speech Hear. Res. **38**(4), 765–771.
Mason, C. R., Kidd, G., Jr., Hanna, T. E., and Green, D. M. (**1984**). "Profile analysis and level variation," Hear. Res. **13**(3), 269–275.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**(4), 224–239.
Narendran, M. M. (**2004**). "Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners," unpublished doctoral dissertation, Indiana University, Bloomington, IN.
Prosek, R. A., Montgomery, A. A., Walden, B. E., and Hawkins, D. B. (**1987**). "An evaluation of residue features as correlates of voice disorders," J. Commun. Disord. **20**(2), 105–117.
Shindo, M. L., Zaretsky, L. S., and Rice, D. H. (**1996**). "Autologous fat injection for unilateral vocal fold paralysis," Ann. Otol. Rhinol. Laryngol. **105**(8), 602–606.
Shrivastav, R. (**2003**). "The use of an auditory model in predicting perceptual ratings of breathy voice quality," J. Voice **17**(4), 502–512.
Shrivastav, R., and Sapienza, C. (**2003**). "Objective measures of breathy voice quality obtained using an auditory model," J. Acoust. Soc. Am. **114**, 2217–2224.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

Temporal properties in clear speech perception

Sheng Liu
Hearing and Speech Research Laboratory, Department of Biomedical Engineering, University of California, Irvine, Irvine, California 92697

Fan-Gang Zeng[a)]
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology-Head and Neck Surgery, University of California, Irvine, Irvine, California 92697

(Received 15 January 2005; revised 28 April 2006; accepted 4 May 2006)

Three experiments were conducted to study relative contributions of speaking rate, temporal envelope, and temporal fine structure to clear speech perception. Experiment I used uniform time scaling to match the speaking rate between clear and conversational speech. Experiment II decreased the speaking rate in conversational speech without processing artifacts by increasing silent gaps between phonetic segments. Experiment III created "auditory chimeras" by mixing the temporal envelope of clear speech with the fine structure of conversational speech, and vice versa. Speech intelligibility in normal-hearing listeners was measured over a wide range of signal-to-noise ratios to derive speech reception thresholds (SRT). The results showed that processing artifacts in uniform time scaling, particularly time compression, reduced speech intelligibility. Inserting gaps in conversational speech improved the SRT by 1.3 dB, but this improvement might be a result of increased short-term signal-to-noise ratios during level normalization. Data from auditory chimeras indicated that the temporal envelope cue contributed more to the clear speech advantage at high signal-to-noise ratios, whereas the temporal fine structure cue contributed more at low signal-to-noise ratios. Taken together, these results suggest that acoustic cues for the clear speech advantage are multiple and distributed. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2208427]

PACS number(s): 43.71.Es, 43.71.Gv, 43.71.Ky [ALF] Pages: 424–432

I. INTRODUCTION

When speech communication becomes difficult, a talker may adopt a different style of speech, "clear speech." This style differs from the everyday speech style, herein referred to as "conversational speech." Previous studies have demonstrated a significant intelligibility advantage for clear speech over conversational speech in both normal-hearing and hearing-impaired listeners across a wide range of listening conditions including quiet, noisy, and reverberant backgrounds (Chen, 1980; Picheny *et al.*, 1985; Payton *et al.*, 1994; Gagne *et al.*, 1995; Schum, 1996; Uchanski *et al.*, 1996; Helfer, 1997; Bradlow and Bent, 2002; Ferguson and Kewley-Port, 2002; Gagne *et al.*, 2002; Krause and Braida, 2002; Bradlow *et al.*, 2003; Liu *et al.*, 2004). Several acoustic differences have been identified between clear and conversational speech, including slower speaking rate, greater temporal modulation, enhanced fundamental frequency variation, expanded vowel space, and higher energy distribution at high frequencies for clear speech (Picheny *et al.*, 1986; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Krause and Braida, 2002;2004; Liu *et al.*, 2004). However, the exact acoustic cues that are responsible for the clear speech advantage remain largely elusive. The present study focuses on the role of temporal information in clear speech perception. Three temporal properties, including speaking rate, temporal envelope, and temporal fine structure, are examined.

Speaking rate is determined by both word and pause durations and is one of the most extensively studied temporal characteristics in speech perception. As early as the 1950s, Fairbanks and colleagues used magnetic tape recorders with different playback speeds to perform uniform time compression and expansion of speech sounds (Fairbanks *et al.*, 1954). Depending on the original speaking rate and the talker's gender, normal-hearing listeners could generally tolerate time-compressed and expanded speech for ratios up to two (Fairbanks *et al.*, 1957; Beasley *et al.*, 1972). However, elderly listeners and persons with certain central auditory processing disorders were found to have a particular difficulty in perceiving the time-compressed speech (Kurdziel *et al.*, 1976; Gordon-Salant and Fitzgibbons, 1995;1997).

Using an explicit pitch-tracking method, coupled with manipulations of the input and output sampling rates, more recent digital time-scaling algorithms could uniformly time compress and expand speech without changing voice pitch (Malah, 1979). Picheny *et al.* (1989) used Malah's algorithm to increase the clear speech rate to match the naturally produced conversational speech rate and to decrease the conversational speech rate to match the naturally produced clear speech rate. They found that uniform time scaling degraded speech intelligibility for both sped-up clear speech and slowed-down conversational speech, suggesting that digital artifacts contaminated the results.

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

 0001-4966/2006/120(1)/424/9/$22.50

Uchanski *et al.* (1996) used nonuniform time scaling to alter phonetic segment lengths to reflect previously measured segmental durational differences between clear and conversational speech. The nonuniform time-scaling method still produced lower intelligibility than unprocessed speech, but the degree of degradation was much less than uniform time scaling. Krause and Braida (2002; 2004) employed "natural" clear speech, training talkers to produce clear speech with the same speaking rate as conversational speech. The talkers were able to produce "fast" clear speech that had the same speaking rate as conversational speech. Perceptual results still showed significantly higher intelligibility for "fast" clear speech than same-rate conversational speech, but there was a global trend of decreasing intelligibility with increasing speaking rate for both clear and conversational speech.

Following this line of research, we employed two different techniques to further probe the role of speaking rate in clear speech perception. Experiment I employed newer signal-processing algorithms, which introduced fewer digital artifacts than algorithms in the 1980s, to uniformly time-scaled speech (Moulines and Laroche, 1995; Kawahara *et al.*, 1999). These newer time-scaling algorithms typically utilized sophisticated pitch extraction algorithms (e.g., pitch synchronous overlap add method or PSOLA) and avoided producing tonal noise in voiceless fricatives and degrading transitional portions in stop consonants. Experiment I is consistent with a recent trend in which classic speech studies are replicated using new digital signal-processing technology (Liu and Kewley-Port, 2004; Assmann and Katz, 2005).

Experiment II decreased speaking rate by inserting silent gaps between phonemes in the conversational speech. This experiment was partially motivated by recent studies which showed a high correlation between temporal processing and speech perception in special populations, including elderly listeners, cochlear-implant users, and persons with auditory neuropathy (Gordon-Salant and Fitzgibbons, 1997; Zeng *et al.*, 1999; Fu, 2002; Zeng *et al.*, 2005a). Inserting gaps between speech segments increased amplitude modulation and provided an extended time window, allowing the listener to process speech more efficiently. The present study differs from Uchanski's nonuniform time-scaling study in the following three ways. First, we did not attempt to match phoneme durations between clear and conversation speech. Instead, we inserted silent gaps proportionally in conversational speech so that the gap-inserted conversational speech had the same overall duration as the clear speech but was free of digital processing artifacts. Second, different from the 10% change in the overall duration in the Uchanski *et al.* (1996) study, we increased the average sentence duration (1.31 seconds) in the original conversational speech by 50% to match the average sentence duration (1.97 seconds) found in clear speech (Liu *et al.*, 2004). Finally, we used different speech materials (BKB sentences) than the nonsense sentences used in the Uchanski *et al.* study.

In addition to speaking rate, other temporal properties play a significant role in clear speech perception. Rosen (1992) divided temporal information into three categories according to the rate of wave fluctuations: envelope (2–50 Hz), periodicity (50–500 Hz), and fine structure (500–10 000 Hz). The temporal envelope cue from a limited number of spectral channels has been shown to be sufficient for speech recognition in quiet (Dudley, 1939; Houtgast and Steeneken, 1985; Van Tasell *et al.*, 1987; Drullman, 1995; Shannon *et al.*, 1995), but periodicity and fine structure are critical for speech recognition in noise, particularly when the noise is a competing voice (Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004; Kong *et al.*, 2005; Nie *et al.*, 2005; Zeng *et al.*, 2005b). It is possible that all three temporal cues are enhanced in clear speech (Bradlow *et al.*, 2003; Krause and Braida, 2004; Liu *et al.*, 2004); however, no study has directly assessed the relative contributions of these temporal cues to clear speech perception.

Experiment III used a novel processing scheme called "auditory chimera" to examine the relative contributions of temporal envelope and fine structure cues to clear speech perception (Smith *et al.*, 2002). The "chimera" scheme is reminiscent of previous cue-trading studies in the segmental domain, in which conflicting burst release and formant transition cues were combined in a single synthetic stimulus to examine their relative contribution to stop consonant recognition (Walley and Carrell, 1983; Dorman and Loizou, 1996). To synthesize a chimaeric sound, Smith *et al.* first divided two broadband signals into several sub-bands, then used the Hilbert transform to extract the temporal envelope and fine structure in each sub-band, and finally mixed one signal's temporal envelope with another signal's fine structure. Smith *et al.* tested the intelligibility of chimerized speech and found that the temporal envelope, rather than the temporal fine structure, made the most contributions to speech intelligibility. However, Smith *et al.* did not test speech recognition in noise, nor did they use any clear speech materials.

In summary, the present study conducted three experiments to evaluate the relative contributions of speaking rate, temporal envelope, and temporal fine structure to the clear speech advantage. Experiment I measured speech intelligibility as a function of signal-to-noise ratios using processed conversational speech that was uniformly stretched to match the duration of the clear speech, or by using processed clear speech that was uniformly compressed to match the duration of the conversational speech. Experiment II measured speech intelligibility using only processed conversational speech that was nonuniformly stretched by proportionally increasing silent gaps between phonetic segments to match the duration of the clear speech. Experiment III measured "chimerized" speech intelligibility using processed speech that contained either the clear speech envelope with conversational speech fine structure or the conversational speech envelope with clear speech fine structure. If a chimera containing the clear speech envelope produces the highest intelligibility, we would conclude that the envelope characteristics of clear speech are responsible for the clear speech advantage. If, in contrast, a chimera containing the clear speech fine structure cues is more intelligible, we would then reach a different conclusion that the fine structure characteristics of clear speech are responsible for the clear speech advantage.

II. EXPERIMENT I. UNIFORMLY TIME-SCALED SPEECH

A. Methods

1. Subjects

Ten normal-hearing listeners were recruited from the Undergraduate Social Science Subject Pool at the University of California, Irvine. Local Institutional Review Board approval was obtained for the experimental protocol. Informed consent was also obtained for each individual subject. None of the subjects reported any speech or hearing impairment. All subjects were native English speakers and received course credit for their participation. Five of the ten subjects were tested with original and processed fast clear speech, while the other five subjects were tested with original and processed slow conversational speech.

2. Stimuli

The original stimuli consisted of 144 sentences recorded in both clear and conversational speech styles. The sentences were modified from the original Bamford-Kowal-Bench (BKB) sentences used for British children (Bench and Bamford, 1979). A male adult talker recorded these sentences with a sampling rate of 16 KHz in a sound-treated room at the Phonetics Laboratory of the Department of Linguistics at Northwestern University (Bradlow *et al.*, 2003).

COOL EDIT PRO 2 (currently known as ADOBE AUDITION) was used to uniformly stretch or compress the original speech signal to change the speaking rate without changing the pitch. The processing algorithm was based on the pitch-synchronous overlap and add method (PSOLA) (Moulines and Laroche, 1995). First, the input waveform was decomposed into a stream of short-time signals based on pitch-synchronous marks. Second, the pitch-synchronous short-time signal was either eliminated or duplicated based on the predefined stretch factor. Third, the modified short-time signal was added to synthesize the stretched and compressed stimulus. The original pitch was preserved during processing and the duration of each voiced or silent segment in the speech was uniformly changed. Different from the earlier methods that changed sampling rate to perform time scaling (Malah, 1979), the newer algorithms used large units (i.e., pitch periods) to perform time scaling, reversed segments to avoid tonal noise in fricative consonants, and preserved the transitional properties in stop consonants. Presumably, these manipulations introduced minimal digital artifacts.

The speaking rate of clear speech was increased to match the speaking rate of conversational speech, and the speaking rate of conversational speech was decreased to match the rate of clear speech for each individual sentence. On average, the speaking rate was increased by 33% for the sped-up clear speech or decreased by 50% for the slowed-down conversational speech (the average duration for conversational speech was 1.31 s compared with 1.97 s for clear speech). Figure 1 shows spectrograms of the original clear speech (top left panel), the original conversational speech (top right), the processed slow conversational speech (left panel on the second row), and the processed fast clear speech (right panel on the second row). Note that the overall duration at the sentence level was matched, but the duration at the phonemic level was clearly not matched. Additionally, note the smeared harmonic structure and formant transitions in both types of processed speech.

All sentences were normalized to have the same overall root-mean-square (rms) level. The speech presentation level was fixed at 65 dBA. The noise level was varied to produce different signal-to-noise ratios. The speech signal was digitally mixed with a speech-spectrum-shaped noise to produce signal-to-noise ratios ranging from −15 to +10 dB.

3. Procedure

Normal-hearing subjects listened to the stimuli monaurally presented via Sennheiser HDA 200 headphones in an IAC double-walled, sound-treated booth. Sentences were presented only once for each subject over the course of the entire experiment. All subjects went through a practice session consisting of five sentences in quiet to become familiar with the test materials and procedures. To collect data, subjects were asked to type the sentences presented through the headphones. A MATLAB program recorded the subject's response and reminded the subject to double-check the spelling before accepting each answer. Speech recognition scores were automatically calculated by counting the number of correct keywords identified.

Experiment I had a total of 28 listening conditions, including the original and processed stimuli $(2)\times 2$ speech styles $\times 7$ signal-to-noise ratios from −15 to 10 dB in 5-dB steps and in quiet. Each condition had eight sentences containing three to four keywords each. The average percent-correct score from eight sentences was reported. In addition, the speech reception threshold (SRT) corresponding to the 50% correct score and the dynamic range (DR) corresponding to the dB difference between the signal-to-noise ratios producing 10% and 90% of the asymptotic performance was derived from the psychometric function (Zeng and Galvin, 1999; Liu *et al.*, 2004).

A mixed-design ANOVA was performed with speech style as a between-subjects factor and processing and signal-to-noise ratio as within-subjects factors. The processing factor examined the difference in performance between the original clear speech and the processed fast clear speech, as well as between the original conversational speech and processed slow conversational speech. A difference was significant at the 0.05 level.

B. Results and discussion

Figure 2 shows percent-correct scores as a function of signal-to-noise ratio obtained from the original clear (open circles with the solid line), original conversational (filled circles with the dotted line), uniformly time-scaled slow conversational (filled triangles with the dashed line), and uniformly time-scaled fast clear (open triangles with the dot-dashed line) speech. Table I shows three fitting parameters and two derived parameters for the perceptual data from experiment I. Several observations can be made from these data.

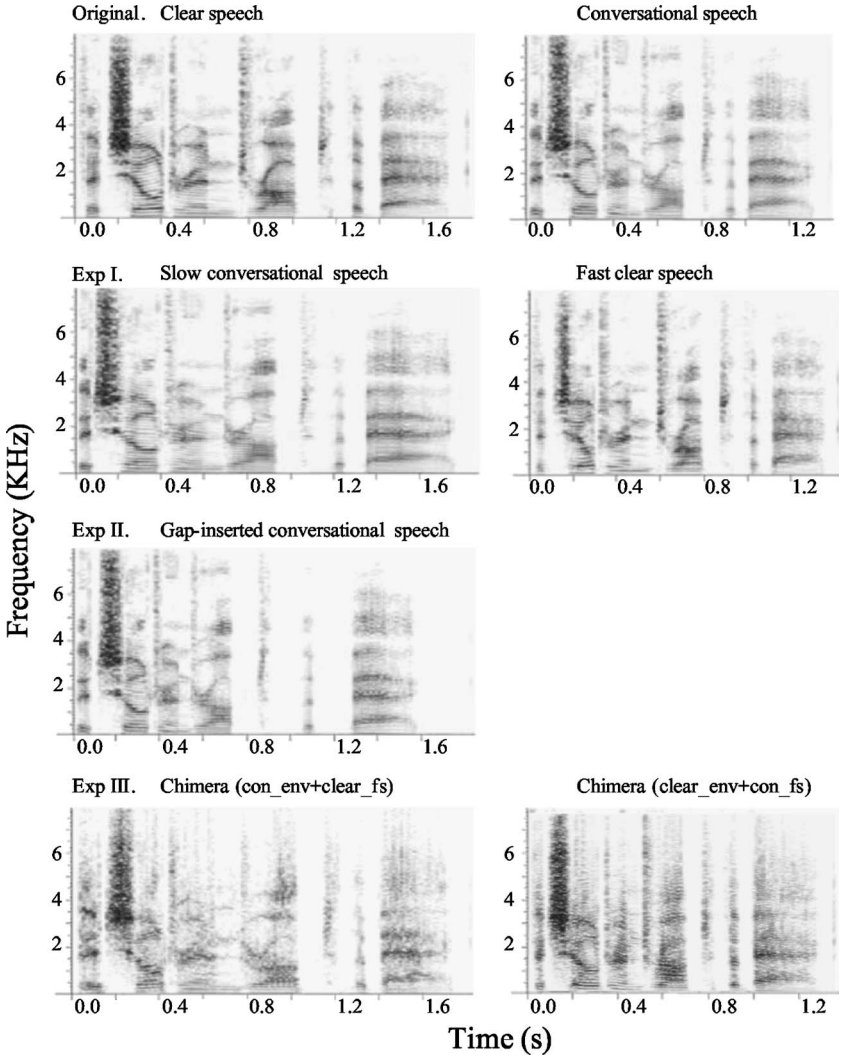


FIG. 1. Spectrograms for the sentence "The children dropped the bag" in seven stimulus conditions: clear speech (upper left), conversational speech (upper right), uniformly stretched conversational speech (second row, left), uniformly compressed clear speech (second row, right), gap-inserted conversational speech (third row), chimera containing conversational speech envelope and clear speech fine structure (bottom left), and chimera containing clear speech envelope and conversational speech fine structure (bottom right).

First, the original clear speech produced significantly better performance than the original conversational speech [$F(1,8)=15.0, p<0.05$]. The SRT was −8.8 dB for clear speech and −6.7 dB for conversational speech. Second, the processed slow conversational speech produced marginally better performance than the processed fast clear speech [$F(1,8)=4.9, p=0.06$]. The SRT was −5.9 dB for slow conversational speech and −4.0 dB for fast clear speech. Better performance with slow conversational speech suggests that either lowering speaking rate improved speech performance or time compression produced more processing artifacts than time expansion. We shall consider the latter in Sec. V. Third, the original clear speech produced significantly better performance than fast clear speech [$F(1,4)=28.4,\ p<0.05$)], but the original conversational speech produced similar performance to slow conversational speech [$F(1,4)=2.1, p>0.05$). No significant interactions were found. This result further implicates possibly more processing artifacts with time compression than time expansion. Finally, the fact that none of the processed speech produced better performance than the original speech suggests that digital processing artifacts are still a confounding factor in these newer signal-processing algorithms.

III. EXPERIMENT II. NONUNIFORMLY STRETCHED SPEECH

A. Methods

1. Subjects

Fifteen subjects were recruited to participate in this experiment using the same human subject protocol as experiment I. A within-subjects design was implemented, in which all subjects listened to the original clear, the original conversational, and the silent-gap-inserted conversational speech.

2. Stimuli

The same BKB sentences were used in this experiment as in the previous experiment. For the silent-gap-inserted conversational speech, the speaking rate was nonuniformly decreased by proportionally increasing silent gaps between phonetic segments in the conversational speech. To avoid the possibility that the silent interval between a vowel and a voiced stop consonant was inadvertently increased (Picheny *et al.*, 1986), silent gaps shorter than 10 ms were kept intact. No phonetic segments in the original conversational speech were altered; only the duration of the silent gap between these segments was proportionally increased by a predeter-

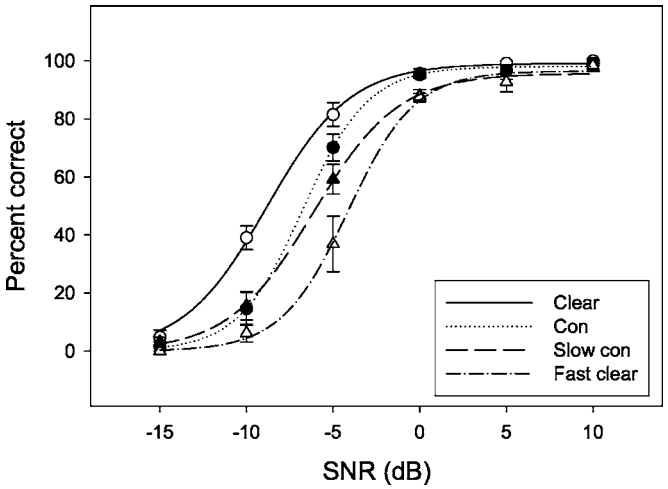


FIG. 2. Percent-correct scores as a function of signal-to-noise ratios for the original clear (open circles), original conversational (filled circles), uniformly time-scaled slow conversational (filled triangles), and fast clear speech (open triangles). Lines represent the best fitting sigmoidal psychometric functions for clear speech (solid line), conversational speech (dotted line), processed slow conversational speech (dashed line), and processed fast clear speech (dotted dashed line).

mined ratio to match the overall duration of the stretched conversational speech to that of the original clear speech. Finally, different from the 5-ms linear ramp used in the Uchanski *et al.* study, no additional ramping was used in the present study.

The left panel on the third row in Fig. 1 shows the spectrogram of the nonuniformly stretched conversational speech. Note that the gap-inserted conversational speech had the same duration as the original clear speech, but contained no apparent processing artifacts, such as smeared harmonic structure and formant transitions.

Each sentence was normalized to have the same overall root-mean-square (rms) level. Because increasing the silent intervals did not add any energy, the overall rms level in the processed speech had to be increased by an average of 1.8 dB to match the original speech overall rms level. The effect of this rms level normalization on speech intelligibility will be examined in Sec. V.

3. Procedure

The same protocol as experiment I was used in experiment II. Experiment II had a total of 15 listening conditions, including three stimulus types (original clear stimuli, original conversational stimuli, and the gap-inserted stimuli) presented at five signal-to-noise ratios (−15 to +5 dB in 5-dB steps). Each condition used eight sentences for each subject in the test. Different sentences were presented in each condition with the sentence presentation order being randomized. A within-subjects ANOVA was performed to examine the main effect of speech style and signal-to-noise ratio.

B. Results and discussion

Figure 3 shows percent-correct scores as a function of signal-to-noise ratio obtained from the original clear speech (open circles with the solid line), the original conversational speech (filled circles with the dotted line), and the gap-inserted conversational speech (filled triangles with the dashed line). A within-subjects ANOVA shows a significant main effect for both speech style $[F(2,28)=12.6, p<0.05]$ and signal-to-noise ratio $[F(4,56)=795.8, p<0.05]$. The interaction between speech style and signal-to-noise ratio was significant $[F(8,112)=2.4, p<0.05]$. The percent-correct scores at −5 dB SNR were 81.0%, 71.6%, and 62.0% for the original clear, gap-inserted conversational, and original conversational speech, respectively. The corresponding SRT values were −8.7, −7.5, and −6.2 dB (Table I). The result from experiment II appears to suggest that speaking rate accounts for roughly 50% of the clear speech advantage. We shall return to this point in Sec. V.

IV. EXPERIMENT III. CHIMERIC SPEECH

A. Methods

1. Subjects

Forty subjects were recruited to participate in experiment III using the same human subject protocol as in experiments I and II. The subjects were equally divided into four groups with each group being tested with the original clear speech, the original conversational speech, the clear speech

TABLE I. Comparison of parameters derived from the psychometric function in experiments I, II, and III. The asymptotic performance level "*S*," intercept "*a*," slope, speech-reception-threshold (SRT), and dynamic range (dB) were defined by Eqs. (1), (2), (3), and (4) in Liu *et al.* (2004).

Expt.	Stimuli	*S* (%)	*a* (dB)	Slope (%/dB)	SRT (dB)	DR (dB)
I	Clear	99.1	−8.8	10.2	−8.8	10.5
	Con	98.1	−6.7	13.0	−6.7	8.3
	Fast-clear	96.5	−4.1	12.8	−4.0	8.5
	Slow-con	95.7	−6.1	9.8	−5.9	10.7
II	Clear	100.0	−8.7	10.6	−8.7	10.3
	Con	100.0	−6.2	10.2	−6.2	10.7
	Gap-con	98.4	−7.6	9.6	−7.5	11.4
III	Clear	98.2	−8.7	13.0	−8.7	8.3
	Con	100.0	−5.3	9.4	−5.3	11.7
	Clear_env+con_fs	96.1	−4.4	15.0	−4.3	7.1
	Con_env+clear_fs	100.0	−5.0	8.2	−5.0	13.3

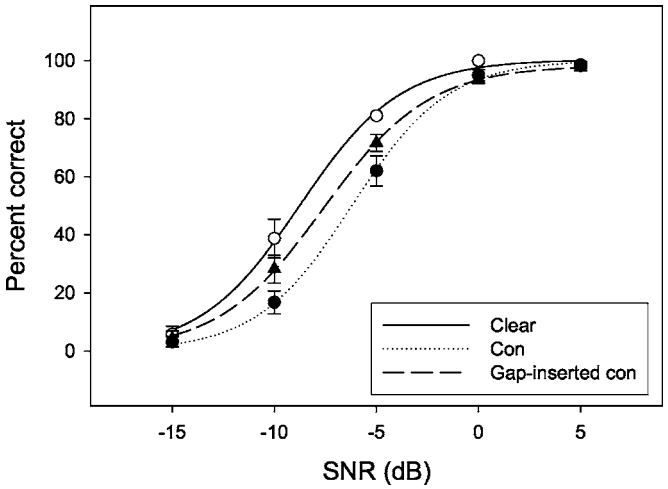


FIG. 3. Percent-correct scores as a function of signal-to-noise ratios for the original clear (open circles), gap-inserted conversational (filled triangles), and original conversational (filled circles) speech. Lines represent best-fit sigmoidal psychometric functions for clear speech (solid line), conversational speech (dotted line), and gap-inserted conversational speech (dashed line).

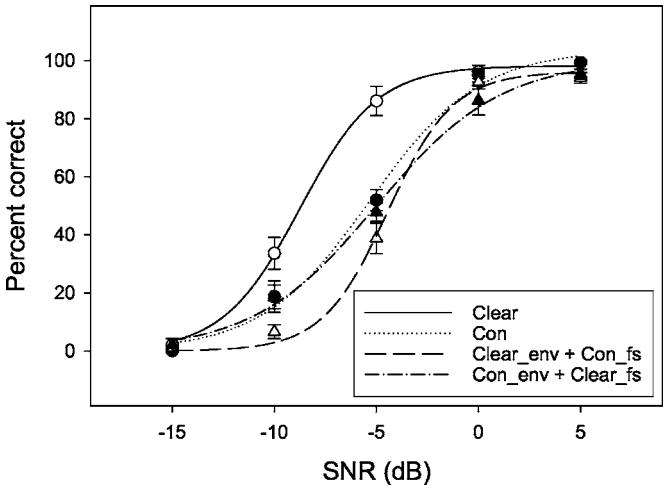


FIG. 4. Percen-correct scores as a function of signal-to-noise ratios for the original clear (open circles), original conversational (filled circles), chimera of clear speech envelope and conversational speech fine structure (open triangles), and chimera of clear speech fine structure and conversational speech envelope (filled triangles). Lines represent the best-fit sigmoidal psychometric function for clear speech (solid line), conversational speech (dotted line), chimera of clear speech envelope and conversational speech fine structure (dashed line), and chimera of clear speech fine structure and conversational speech envelope (dotted dashed line).

envelope and conversational speech fine-structure chimera, and the conversational speech envelope and clear speech fine-structure chimera, respectively.

2. Stimuli

The same BKB sentences were used in this experiment, which chimerized clear and conversational speech to create two types of new stimuli that contained either the clear speech envelope and conversational speech fine structure, or the conversational speech envelope and clear speech fine structure (Smith *et al.*, 2002). To create these stimuli, both the clear and conversational speech stimuli were spectrally divided into 16 logarithmically spaced filters spanning a frequency range of 80 to 8000 Hz (Greenwood, 1990). The number of bandlimited filters was chosen to avoid cochlear filtering with a low number of filters and filter ringing with a high number of filters (Zeng *et al.*, 2004). The bandpassed signal was then decomposed into its envelope and fine structure via the Hilbert transform. The bandlimited conversational speech envelope was nonuniformly stretched to align each segment in the original conversational speech to that in the original clear speech. The nonuniformly stretched conversational envelope was then used to amplitude modulate the clear speech fine structure. Similarly, nonuniform compression was used to match the clear speech envelope to the original conversational fine structure. Finally, the chimerized bandlimited signals were summed to form the chimerized speech.

The bottom-left panel in Fig. 1 shows the conversational speech envelope and clear speech fine structure chimera ("con_env+clear_fs"), and the bottom-right panel shows the clear speech envelope and conversational speech fine structure chimera ("clear_env+con_fs"). Because the temporal envelope was adjusted to match the duration between clear and conversational speech, the temporal fine structure determines both the overall sentence duration and individual phoneme duration in the chimera. For example, the "con_env +clear_fs" chimera (bottom left) has the same relative and overall durations as the original clear speech. Note also the slight spectral smearing in the chimerized speech. If, at a given SNR, the chimera containing the clear speech envelope produces the highest intelligibility, we would conclude that the envelope characteristics of clear speech underlie the superior intelligibility of clear speech. If, in contrast, the chimera containing the clear speech fine structure cues is more intelligible, we would reach a different conclusion that the fine structure characteristics of clear speech are responsible for its superior intelligibility.

3. Procedure

The experimental protocol used in experiments I and II was also used in experiment III. Experiment III had a total of 24 conditions, including 2 original speech stimuli and two chimeras (4 types of stimuli) $\times 5$ signal-to-noise ratios from -15 to 10 dB in 5-dB steps and in quiet (6 signal-to-noise ratios). Each condition used eight sentences for each subject in the test. A mixed ANOVA design was performed with stimulus type being the between-subjects factor and signal-to-noise ratio being the within-subjects factor.

B. Results and discussion

Figure 4 shows percent-correct scores as a function of signal-to-noise ratio for the original clear speech (open circles with the solid line), the original conversational speech (filled circles with the dotted line), the clear speech envelope and conversational speech fine structure chimera ("clear_env+con_fs," open triangles with the dashed line), and the conversational speech envelope and clear speech fine structure chimera ("con_env+clear_fs," filled triangles with the dot-dashed line). Table I shows fitted and derived parameters from the psychometric functions (experiment III). Both

stimulus type [$F(3,36)=17.8, p<0.05$] and signal-to-noise ratio [$F(5,180)=916.3, p<0.05$] were significant factors. The interaction between these two factors was also significant [$F(15,102)=8.9, p<0.05$]. Several observations can be made from the obtained data.

First, the original clear speech produced significantly better performance than all the other three stimuli, including the original conversational speech (a *posthoc* Bonferroni test, $p<0.05$). The SRT value was −8.7 dB for the original clear speech, as opposed to −5.3, −4.3, and −5.0 dB for the original conversational speech, the clear_env+con_fs, and the con_env+clear_fs chimera, respectively. The generally poorer performance with "auditory chimera" suggests the presence of processing artifacts.

Second, the significant interaction between stimulus type and signal-to-noise ratio occurred between the chimerized speech stimuli. At low SNRs (−10 and −5 dB), the con_env+clear_fs chimera produced an approximately 10-percentage-point better performance than the clear_env +con_fs chimera, implying a significant role of the fine structure cue in clear speech. At high SNRs (e.g., 0 dB), the reverse was true with the clear_env+con_fs chimera producing 6-percentage-point better performance than the con_env+clear_fs chimera, implying a significant role of the envelope cue in clear speech. Although confounded by processing artifacts, results from experiment III supported the hypothesis that the temporal envelope is critical for speech recognition in quiet and the temporal fine structure is critical for speech recognition in noise.

V. GENERAL DISCUSSION

A. Summary and comparison

Table I summarizes three fitting parameters and two derived parameters for the perceptual data from experiments I, II, and III [see Eqs. (1)–(4) in Liu *et al.*, 2004]. First, the original clear speech always produced the same or higher asymptotic performance (S), lower speech reception thresholds (SRT), and a steeper slope than the original conversational speech. The only exception was for experiment I, in which the conversational speech produced a steeper slope. Second, the SRT value was essentially identical for the original clear speech (−8.8, −8.7, and −8.7 dB) but varied greatly for the conversational speech (−6.7, −6.2, and −5.3 dB) in the present three experiments, which used the same materials and the same procedure but different subjects. For comparison, these SRT values were closely matched to the −8.5- and −6.3-dB SRT values found in the Liu *et al.* (2004) study, which used the same materials and the same procedure and an additional independent group of normal-hearing subjects. These SRT values suggest that acoustic cues in clear speech are less susceptible to individual variability than conversational speech.

We can also use the intelligibility difference in percentage points, which is equal to the product of the slope and the SRT difference between the conditions to quantify the clear speech and signal-processing effects. Except for the gap-inserted conversational speech producing 13-percentage-point higher intelligibility than the original conversation speech, all processed speech stimuli produced lower intelligibility than the original speech. The original clear speech produced intelligibility 29 percentage points higher than the uniformly stretched conversational speech, while the original conversational speech produced intelligibility 35 percentage points higher than the uniformly compressed clear speech. Similarly, Picheny *et al.* (1989) found a 30-percentage-point difference between the original and compressed clear speech and a 13-percentage-point difference between the original and the expanded conversational speech. These results suggest the presence of digital signal-processing artifacts as a confounding factor in the evaluation of the role of speaking rate in clear speech perception.

B. Signal-processing artifacts

To identify the source of processing artifacts, reversibility was tested in experiment I (Picheny *et al.*, 1989). We used the same COOL EDIT program to first compress the clear speech and then stretch the processed stimulus back to its original duration. The recovered clear speech had apparent audible processing artifacts and significantly lower intelligibility than the original clear speech. On the other hand, the recovered conversational speech, which underwent the expansion process first and the compression process second, had no audible artifacts and essentially the same intelligibility as the original conversational speech. The reversibility test revealed that the processing algorithm introduced more processing artifacts during compression than during expansion, and additionally that compression followed by expansion is irreversible while expansion followed by compression is reversible. Close examination of the "fast clear speech" spectrogram (right panel on the second row in Fig. 1) already shows a less accurate representation of formant transitions compared with the original clear speech. A processing artifact in time compression is a result of deleting segments that introduce discontinuities in fast changes, such as frequency transitions. Therefore, the compressed clear speech produced worse performance than the original clear speech, while the stretched conversational speech produced similar performance to the original conversational speech.

The chimerized speech may have introduced different types of processing artifacts than the uniformly time-scaled speech. The chimera method first extracted the bandlimited temporal envelope and fine structure from two sentences of different durations. The envelope had to be compressed (in clear speech) or stretched (in conversational speech) to match the duration of the fine structure, which remained intact. There were at least three sources of processing artifacts. The first artifact stemmed from alterations in modulation frequencies, which were introduced by digital resampling in temporal envelopes. The second artifact was introduced by the segment mismatch between one sentence's temporal envelope and another sentence's fine structure. The third artifact was due to the bandpass filtering in the analysis-synthesis process, which was generally irreversible. Clearly, these processing artifacts degraded performance and confounded the interpretation of the present results.

C. Speaking rate

While both uniform time scaling and chimerizing introduced processing artifacts, inserting silent gaps to decrease the conversational speech rate did not introduce any artifacts. At first glance, the results from experiment II seem to indicate that longer pauses between speech segments improved the perception of conversational speech by 1.3 dB in terms of the SRT measure. However, one may question whether this improvement is truly a result of the decreased speaking rate in the processed conversational speech.

Recall from Sec. II A 2 in experiment I that the average overall duration was 1.31 s for conversational speech and 1.97 s for clear speech, indicating that, on average, 0.66 s of silent gaps had to be inserted in the conversational speech to match the duration of the clear speech. As described in Sec. II A, a normalization procedure was employed to equalize the overall rms for all processed and original sentences. This normalization procedure increased the overall rms level by 1.8 dB for the gap-inserted conversational speech. Because the inserted silent gaps did not contribute to the overall rms level, the short-term rms level had to be increased proportionally by 1.8 dB for all phonetic segments. If we assume that the listener used a short-term window (tens to hundreds of milliseconds), instead of a 1- or 2-s window, to calculate the sentence-level rms level, then the effective short-term signal-to-noise ratio would be 1.8 dB higher than suggested by the overall rms level. Therefore, it is possible that the observed 1.3-dB improvement in SRT was a result of the rms level normalization employed at the sentence level. If this short-term rms level hypothesis holds true, then inserting silent gaps in conversational speech would not necessarily improve intelligibility.

Because longer silent gaps or pauses were consistently observed in clear speech, the above examination on the role of the short-term rms level brings about an important question: to what extent is the so-called clear speech advantage a result of the increased short-term signal-to-noise ratio? To answer this question, we removed all pauses in the original clear and original conversational speech and calculated their rms levels. For male talker materials used in the present study, we found that the pause-removed clear speech had a 0.2 dB higher overall rms level than the pause-removed conversational speech. Clearly, this 0.2-dB difference cannot account for the observed 3-dB clear speech advantage, suggesting that acoustic cues other than speaking rate contribute significantly to the clear speech advantage.

D. Temporal envelope and fine structure

Previous studies have emphasized the importance of the temporal envelope in speech recognition (Van Tasell *et al.*, 1987; Rosen, 1992; Drullman *et al.*, 1994; Shannon *et al.*, 1995), but recent results have suggested a complementary role of the temporal fine structure in speech recognition in noise (Nie *et al.*, 2005; Zeng *et al.*, 2005b). Although auditory chimera introduced digital processing artifacts (Smith *et al.*, 2002; Zeng *et al.*, 2004), results from experiment III suggest that this idea can be extended and applied to clear speech perception. At high signal-to-noise ratios, the chimera with the clear speech envelope and the conversational speech fine structure produced higher intelligibility than the chimera with conversational speech envelope and clear speech fine structure. On the other hand, the reverse was true at low signal-to-noise ratios. As far as clear speech perception is concerned, the present result suggests that the temporal envelope and fine structure contribute complementarily to the clear speech advantage. The temporal envelope contributes to the clear speech advantage in quiet, while the temporal fine structure contributes to the clear speech advantage in noise.

VI. CONCLUSIONS

The present study used three methods to evaluate temporal properties in clear speech perception. The three methods included (1) uniform time scaling to increase the clear speech rate or decrease the conversational speech rate; (2) nonuniform time scaling to decrease the conversational speech rate by increasing pauses between phonetic segments in conversational speech; and (3) "auditory chimera" with clear speech temporal envelope and conversational speech fine-structure or vice versa (Smith *et al.*, 2002). Based on acoustic analysis and perceptual data, we reached the following conclusions:

(1) Consistent with previous studies, the present study found a consistent clear speech advantage corresponding to a 2–3-dB difference in the speech reception threshold between clear and conversational speech.
(2) While both uniform time compression and stretching introduced processing artifacts, time compression was found to be more detrimental than time stretching in terms of processing reversibility and the degree of performance degradation.
(3) Increasing silent gaps in conversational speech decreased the speaking rate without introducing any processing artifacts. Perceptual results showed a 1.3-dB advantage in SRT for the gap-inserted conversational speech, accounting for roughly half of the overall clear speech advantage. Acoustic analysis indicated that this improvement in SRT might be a result of an increased short-term signal-to-noise ratio due to the rms level normalization at the sentence level.
(4) Although auditory chimera introduced digital processing artifacts, perceptual results from the chimerized clear and conversational speech suggested a complementary role of temporal envelope and fine structure in speech perception: the temporal envelope contributes more to the clear speech advantage at high signal-to-noise ratios, while the temporal fine structure contributes more at low signal-to-noise ratios.

ACKNOWLEDGMENTS

We thank Tiffany Chua, Elsa Del Rio, Paul Meneses, and Frank Z. Yu for stimulus processing and data collection. We thank Ann R. Bradlow for providing speech materials. Abby Copeland, Alex Francis, and two anonymous reviewers provided helpful comments on an earlier draft of this paper. This work was funded by the National Institutes of Health,

Department of Health and Human Services (2 RO1-DC02267).

Assmann, P. F., and Katz, W. F. (**2005**). "Synthesis fidelity and time-varying spectral change in vowels," J. Acoust. Soc. Am. **117**, 886–895.

Beasley, D. S., Schwimmer, S., and Rintelmann, W. F. (**1972**). "Intelligibility of time-compressed CNC monosyllables," J. Speech Hear. Res. **15**, 340–350.

Bench, J., and Bamford, J. (**1979**). *Speech-hearing Tests and the Spoken Language of Hearing-impaired Children* (Academic, London).

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities: sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Chen, F. (**1980**). *Acoustic Characteristics of Clear and Conversational Speech at Segmental Level* (Massachusetts Institute of Technology, Cambridge, MA).

Dorman, M. F., and Loizou, P. C. (**1996**). "Relative spectral change and formant transitions as cues to labial and alveolar place of articulation," J. Acoust. Soc. Am. **100**, 3825–3830.

Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Dudley, H. (**1939**). "The vocoder," Bell Lab. Rec. **17**, 122–126.

Fairbanks, G., Everitt, W. L., and Jerger, R. P. (**1954**). "Method for time or frequency compression-expansion of speech," IRE Trans. Audio **2**, 7–12.

Fairbanks, G., Guttman, N., and Miron, M. S. (**1957**). "Effects of time compression upon the comprehension of connected speech," J. Speech Hear Disord. **22**, 10–19.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Fu, Q. J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.

Gagne, J., Rochette, A., and Charest, M. (**2002**). "Auditory, visual, and audiovisual clear speech," Speech Commun. **37**, 213–230.

Gagne, J., Querengesser, C., Folkeard, P., Munhall, K., and Mastern, V. (**1995**). "Auditory, visual and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech," The Volta Review **97**, 33–51.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1995**). "Recognition of multiply degraded speech by young and elderly listeners," J. Speech Hear. Res. **38**, 1150–1156.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Helfer, K. S. (**1997**). "Auditory and auditory-visual perception of clear and conversational speech," J. Speech Lang. Hear. Res. **40**, 432–443.

Houtgast, T., and Steeneken, H. J. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1071–1077.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based *F*0 extraction: Possible role of a repetitive structure in sounds," Crystallogr. Rep. **27**, 187–207.

Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.

Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.

Kurdziel, S., Noffsinger, D., and Olsen, W. (**1976**). "Performance by cortical lesion patients on 40 and 60% time-compressed materials," J. Am. Aud Soc. **2**, 3–7.

Liu, C., and Kewley-Port, D. (**2004**). "Vowel formant discrimination for high-fidelity speech," J. Acoust. Soc. Am. **116**, 1224–1233.

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G. (**2004**). "Clear speech perception in acoustic and electric hearing," J. Acoust. Soc. Am. **116**, 2374–2383.

Malah, D. (**1979**). "Time-domain algorithms for harmonic bandwidth reduction and time scaling of speech signals," IEEE Trans. Acoust., Speech, Signal Process. **27**, 121–133.

Moulines, E., and Laroche, J. (**1995**). "Non-parametric techniques for pitch-scale and time-scale modification of speech," Crystallogr. Rep. **16**, 175–205.

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., Stickney, G., and Zeng, F. G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). "Speaking clearly for the hard of hearing. III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," J. Speech Hear. Res. **32**, 600–603.

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.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Schum, D. J. (**1996**). "Intelligibility of clear and conversational speech of young and elderly talkers," J. Am. Acad. Audiol **7**, 212–218.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Stickney, G. S., Zeng, F. G., Litovsky, R. Y., and Assmann, P. F. (**2004**). "Cochlear implant speech recognition with speech masker," J. Acoust. Soc. Am. **116**, 1081–1091.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.

Van Tasell, D. J., 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.

Walley, A. C., and Carrell, T. D. (**1983**). "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 1011–1022.

Zeng, F. G., and Galvin, J. J. (**1999**). "Amplitude mapping and phoneme recognition in cochlear implant listeners," Ear Hear. **20**, 60–74.

Zeng, F. G., Kong, Y. Y., Michalewski, H. J., and Starr, A. (**2005a**). "Perceptual consequences of disrupted auditory nerve activity," J. Neurophysiol. **93**, 3050–3063.

Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**1999**). "Temporal and speech processing deficits in auditory neuropathy," NeuroReport **10**, 3429–3435.

Zeng, F. G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y. Y., and Chen, H. (**2004**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351–1354.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005b**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.

Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis

Richard M. Stibbard[a] and Jeong-In Lee
20 Beeches Road, Cheltenham, Glos., GL53 8NQ, UK

(Received 5 February 2005; revised 14 April 2006; accepted 18 April 2006)

In a follow-up study to that of Bent and Bradlow (2003), carrier sentences containing familiar keywords were read aloud by five talkers (Korean high proficiency; Korean low proficiency; Saudi Arabian high proficiency; Saudi Arabian low proficiency; native English). The intelligibility of these keywords to 50 listeners in four first language groups (Korean, $n=10$; Saudi Arabian, $n=10$; native English, $n=10$; other mixed first languages, $n=20$) was measured in a word recognition test. In each case, the non-native listeners found the non-native low-proficiency talkers who did not share the same first language as the listeners the least intelligible, at statistically significant levels, while not finding the low-proficiency talker who shared their own first language similarly unintelligible. These findings indicate a mismatched interlanguage speech intelligibility *detriment* for low-proficiency non-native speakers and a potential intelligibility problem between mismatched first language low-proficiency speakers unfamiliar with each others' accents in English. There was no strong evidence to support either an intelligibility benefit for the high-proficiency non-native talkers to the listeners from a different first language background or to indicate that the native talkers were more intelligible than the high-proficiency non-native talkers to any of the listeners.
 [DOI: 10.1121/1.2203595]

PACS number(s): 43.71.Gv, 43.71.Hw, 43.71.Es [ALF] Pages: 433–442

I. INTRODUCTION

In the world as a whole, non-native speakers of English now outnumber native speakers (Crystal, 2003; Graddol, 1997). Many communications in English take place involving non-native speakers, and it is thus of importance to gain information on how well mutual intelligibility is maintained in these interactions. Research shows that the intelligibility of non-native speech is a complex matter which depends on many factors, including pronunciation (Anderson-Hsieh and Koehler, 1988; Bent and Bradlow, 2003; Bent, 2001; Gass and Varonis, 1984; Hinofitis and Bailey, 1981; Jenkins, 2000, 2002; Kranke and Christison, 1983; Munro and Derwing, 1999; Smith and Bisazza, 1982; Smith and Rafiqzad, 1979), grammar (Ensz, 1982), vocabulary (Politzer, 1976), discourse (Albrechtsen, *et al.* 1980), and familiarity with the topic (Gass and Varonis, 1984). It is therefore evident that intelligibility is not a matter of the phonetics of the utterance alone; listener variables too play an active role. Work on speech perception has therefore increasingly moved to focus on subject variables, particularly on the interaction between talker and listener variables (Bent and Bradlow, 2003; Flege, 1987; Flege, 1988; Flege and Fletcher, 1992; van Wijngaarden, 2001; van Wijngaarden *et al.*, 2002).

In a follow up to the Bent and Bradlow (2003) study of the mutual intelligibility of Korean, Chinese, and native English speakers speaking English, this paper reexamines their key findings regarding non-native talkers. These were (a) that, for the non-native listeners from the same first language (L1) background as the talkers, the intelligibility of the high-proficiency non-native talkers was better than or equal to that of the native talkers, a phenomenon which Bent and Bradlow term the "matched interlanguage speech intelligibility benefit and (b) that for listeners who did not share the talkers' L1, the high-proficiency non-native talkers were also judged to be either more intelligible or statistically no less intelligible than non-native talkers, termed the "mismatched interlanguage speech intelligibility benefit" (Bent and Bradlow, 2003, p. 1607).

Using the same type of spoken data, keywords embedded in carrier sentences, but using Korean and Saudi Arabian non-native speakers and modifying the experimental procedure with the intention of avoiding a possible familiarity effect, the present study found that

1. For the native English and mixed L1 groups of listeners, both low-proficiency non-native talkers were significantly less intelligible than the native English or high-proficiency non-native talkers;
2. For the Korean and Saudi listeners, in each case the low-proficiency non-native talker from the L1 background which did not match their own was the least intelligible at a statistically significant level. This indicates not a mismatched interlanguage intelligibility benefit, but a detriment for these low-proficiency non-native talkers in communication with listeners from another first language background;
3. For all listeners, the high-proficiency non-native talkers were either equally intelligible as or more intelligible than the native talkers, indicating that the native talkers were not found to be more intelligible than the non-native talkers, even by their fellow native listeners.

[a]Formerly of the Department of Culture, Media and Communication, University of Surrey, UK. Electronic mail: rmstibbard@yahoo.co.uk

A. Bent and Bradlow's study

The Bent and Bradlow (2003) paper "The interlanguage speech intelligibility benefit" studied the interaction between speaker and listener variables by investigating the intelligibility of native and non-native talkers' speech in English to native English and non-native listeners with the same and different L1 as the talkers. They used an experimental design in which stimulus sentences containing key words were embedded in white noise at a constant level. These stimulus sentences were read aloud by four non-native talkers (two Chinese and two Koreans), and by one native English talker. The non-native talkers were divided on the basis of native English listeners' judgments of their speech into one high-proficiency talker of each first language and one low-proficiency talker of each first language.

Bent and Bradlow measured the intelligibility of the embedded key words by playing the stimuli to four groups of listeners: Koreans ($n=10$), Chinese ($n=21$), native English ($n=21$), and mixed ($n=12$), the first languages of those in the mixed group being neither Korean nor Chinese nor English. They thus had a measure of the effect on speech intelligibility of the listeners' sharing or not sharing the talker's L1.

Bent and Bradlow make a claim for the existence of a "matched interlanguage speech intelligibility benefit." By this they mean that "for non-native listeners, intelligibility of a high-proficiency non-native talker (and in one case a low-proficiency talker) from the same native language background was equal to the intelligibility of the native talker" (Bent and Bradlow, 2003, p. 1607). This claim is well supported by the high-proficiency talker data. In both cases in the study, non-native listeners rated the high-proficiency non-native talker who shared their L1 significantly higher than the native talker. Bent and Bradlow make no claim that the "benefit" extends to low-proficiency talkers.

They also found a "mismatched interlanguage speech intelligibility benefit," by which they mean that "for non-native listeners, intelligibility of a high-proficiency non-native talker from a different native language background was greater than or equal to the intelligibility of the native talker" (*ibid*). In one case out of the two, that of the Chinese listeners rating the high-proficiency Korean talker, a non-native listener group gave a higher intelligibility rating to the mismatched L1 high-proficiency non-native talker than to the native talker. In the other case, that of the Korean listeners rating the high-proficiency Chinese talker, the intelligibility score was equal to that given to the native talker. The high-proficiency Korean talker was judged by all listeners except the native English listener group to be the most intelligible of all the talkers. The native English listeners judged the native English talker the most intelligible in the study, and this talker the second most intelligible. These results indicate that the high intelligibility scores received by this speaker were possibly due a generally high level of intelligibility rather than specifically to the effects of a mismatched interlanguage intelligibility benefit.

In the terms of Bent and Bradlow, an intelligibility benefit consists of receiving intelligibility scores greater than *or equal to* those received by the native talker. Taking the inclusive definition used by Bent and Bradlow, the existence of a matched and mismatched interlanguage intelligibility benefit is supported. However, it is possible to take issue with this use of the term "benefit" to describe these findings, as it might be argued that the word "benefit" should be used only to describe cases in which a talker received higher intelligibility scores than another talker, not those cases in which the scores were simply equal. If a benefit is taken in those terms, then their evidence for a mismatched interlanguage intelligibility benefit is not strong.

The intelligibility ratings given by the non-native listeners to the low-proficiency talkers in the Bent and Bradlow study are interesting. In each case, these listeners rated the talker who shared their language background higher than the one who did not, significantly so in the case of the Korean listeners and Chinese low-proficiency talker. This indicates not a mismatched interlanguage speech intelligibility benefit, but a detriment in cases involving low-proficiency talkers.

B. The present study

The present study was conducted to investigate whether evidence would be found to support the hypothesis of a matched or mismatched interlanguage intelligibility benefit. In the interests of continuity and control over variables, an experimental design similar to that of Bent and Bradlow (2003) was used, but with the modifications described below. In addition, the definition of the word "benefit" was made less inclusive to exclude cases of equal intelligibility scores and include only those cases where higher intelligibility scores were given.

Bent and Bradlow studied Chinese and Korean speakers; this study investigates Saudi Arabian and Korean speakers, in the interests of widening the number of languages studied. It is possible, as Bent and Bradlow speculate (2003, p. 1607), that the Korean and Chinese languages may share certain phonological similarities, which may mean that the case of a mismatched interlanguage intelligibility benefit which they observed was in fact due to similarities between these two particular languages leading to familiarity between the two groups with each others' accents in English. They give the example of a highly constrained syllable structure shared by the two languages which includes an absence of final consonant clusters. If this is so, then what they observed may simply have been an effect of this similarity rather than a genuine case of mismatched interlanguage intelligibility benefit. Also, in addition to possible phonological similarities, the geographical and cultural proximity of Korea and China may have the effect of further familiarizing the speakers with each others' first languages and accents in English.

The rationale behind the choice of Arabic and Korean for the present study was that they are in many respects phonologically rather distant from each other and may also be a less familiar pairing due to greater geographical and cultural distance. For these reasons, they may be a better test of the research question than the Korean/Chinese pairing.

A further modification made to the Bent and Bradlow study was to change the order in which the sentence stimuli

TABLE I. Profile of the pool of ten non-native speakers. The final column shows the speakers selected to act as talkers in this study: KH=Korean high proficiency; KL=Korean low proficiency; SH=Saudi high proficiency; SL=Saudi low proficiency.

Speaker	Age (yrs.)	Age of arrival (yrs.)	Length of residence (yrs.)	Time learning English (yrs.)	IELTS	TOEFL	Talker selected
Korean 1	28	27	1	16	n/a	236	
Korean 2	28	27	1	16	n/a	256	KH
Korean 3	31	26	5	20	6.0	n/a	
Korean 4	32	28	4	20	6.0	n/a	KL
Korean 5	30	26	4	30	6.0	n/a	
Saudi 1	20	18	2	16	n/a	n/a	
Saudi 2	18	16	2	14	n/a	n/a	
Saudi 3	41	33	8	27	7.0	n/a	SH
Saudi 4	32	29	3	13	6.0	n/a	SL
Saudi 5	31	30	1	15	n/a	237	

were played. In the Bent and Bradlow study, these were played to the listeners in a constant blocked format by talker; the non-native listeners heard all the sentences read by the high-proficiency matched first language talker, then the high-proficiency mismatched first language talker, then the native speaker of English, then the low-proficiency matched first language talker, and finally the low-proficiency mismatched first language talker. The native speaker listeners heard the blocks in the following order: Chinese high-proficiency; Korean high-proficiency; native speaker; Chinese low-proficiency; Korean low-proficiency. The reason for this ordering was to ensure consistency across the non-native listener groups with respect to the talker-listener native language match and mismatch, as well as to ensure that superior performance in listening to the non-native talkers could not be attributed to a practice effect due to hearing the native talker first.

However, it is our view that this could have familiarized the listeners to the voice of each talker in turn due to hearing the same talker repeatedly, and that this could have led listeners to mark similar results until the talker changed. It was thus decided to play all the stimuli in random order not in blocks by talker with the intention of avoiding such a possible familiarity effect.

In addition, the decision was made not to embed the stimuli in white noise as was done in the Bent and Bradlow study. Presumably, this was done in order to avoid a ceiling effect, although this is not stated. The present authors decided not to add white noise as it was felt that doing so would make the stimuli even more unrepresentative of real speech than is unavoidable in the use of isolated sentences, and because white noise could introduce other poorly controlled effects including the masking of particular speech sounds such as fricatives and low intensity sounds more than others. For these reasons, the stimuli were prepared using high quality recordings designed to reproduce natural speech as closely as possible within the experimental design. No ceiling effect was apparent in the data.

II. METHODOLOGY

A. Talkers

This section describes the procedure by which four non-native talkers (two Korean, two Saudi Arabian) were selected from a pool of ten (five Korean, five Saudi Arabian), to act as talkers for this experiment.

Table I shows a demographic profile of the pool of ten non-native speakers, including age, age of arrival in the UK, length of residence in the UK, time spent learning English, and their Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS) scores, where these were available.

The ten non-native speakers were recorded reading aloud all the 30 sentences in sentence lists one and two (Appendix A). From these recordings, a subset consisting of the following six sentences was used for talker selection: (1) "The mother stirs the tea;" (2) "The lady packed her bag;" (3) "The family like fish;" (4) "The girl held a mirror;" (5) "Mother made some curtains;" (6) "The man tied his scarf." The use of the plural form of the verb "like" with a singular group noun "family" (sentence 3) is common in British English and was not considered grammatically incorrect by the authors; none of the participants commented on this sentence.

These six sentences spoken by each of the ten speakers, a total of 60 recordings, were randomly mixed in order to minimize familiarity with the speaker's voice and were played to five native speaker listeners. The listeners were asked to rate the intelligibility of each sentence on a scale of 1 to 5 as follows: 1=heavy foreign accent; very difficult to understand; 2=heavy to moderate foreign accent; somewhat difficult to understand; 3=moderate foreign accent; almost difficult to understand; 4=slight foreign accent; never difficult to understand; 5=no foreign accent; very easy to understand.

Each sentence was marked out of five, so the total of six sentences were marked out of 30, which gave a maximum intelligibility score of 150 points for each speaker from the five native listeners. Table II shows a summary of the results of this evaluation, including total, mean, standard deviation,

TABLE II. Results of the native speaker evaluation used to select the four talkers used in the study from the pool of ten.

Speaker	Total	Mean	SD	Range	Talker selected
Korean 1	116	3.87	1.05	3	
Korean 2	139	4.63	0.63	2	KH
Korean 3	127	4.23	0.63	2	
Korean 4	92	3.07	0.41	3	KL
Korean 5	108	3.60	0.89	3	
Saudi 1	146	4.87	0.52	1	
Saudi 2	131	4.37	0.84	2	
Saudi 3	140	4.67	0.89	2	SH
Saudi 4	86	2.87	0.82	3	SL
Saudi 5	100	3.33	0.98	3	

and range. The final column of Table II indicates the speakers who were chosen as talkers. Talkers with closely matching scores at the higher and lower end of the rating scale were chosen, Korean 2 and Saudi 3 being selected as the high proficiency speakers (total scores 139 and 140, respectively), and Korean 4 and Saudi 4 as the low proficiency speakers (total scores 92 and 86, respectively, the lowest two scores). Saudi 1 received the highest total score, at 146, but was not used so as to maintain a more closely matched pair of high proficiency speakers.

B. Stimuli

As in the Bent and Bradlow study, carrier sentences were taken from the revised Bamford-Kowal-Bench Standard Sentence Test (Bent and Bradlow, 2003, p. 1602), originally devised by Bamford and Wilson (1979) and Bench and Bamford (1979). These were chosen on the basis of an independent evaluation of the familiarity of the contained key words to non-native speakers as explained in Bent and Bradlow (2003, p. 1602) and were chosen for consistency with the Bent and Bradlow study.

These sentences were then divided into two lists of 15 sentences covering 45 key words each. For the native English, the high-proficiency Korean, and the high-proficiency Saudi Arabian talker, the 15 sentences from Sentence List 1 were used. For the low-proficiency Korean and the low-proficiency Saudi Arabian talker, the 15 sentences from Sentence List 2 were used, in order to avoid a practice effect. The recordings made in the talker selection stage were used so as to avoid a talker practice effect due to repeated recordings of the same sentences.

The four non-native talkers and the native English talker read these sentences in the same order and were recorded in a quiet, carpeted room using a Sony MZ-R30 Minidisk recorder player with a Sony ECM-T150 microphone. Minidisk recordings produce a high quality recording which is indistinguishable to the human ear from the uncompressed wave form; thus despite the fact that they use a psychoacoustic compression algorithm (Tsutsui *et al.*, 1992) they are sufficient for perception experiments such as the present one.

These recordings were then digitized at a sampling rate of 48 kHz and 16-bit resolution and transferred to an IBM compatible personal computer using sound editing software (COOL EDIT 2000). Noise reduction was performed on all files to remove a small amount of background hiss using the noise reduction facility available in COOL EDIT. This software performs noise reduction by analyzing the frequencies of sound in a section of the sound file containing background noise only and lowering the amplitude of those frequencies throughout the rest of the sound file, thus increasing the signal-to-noise ratio. The noise reduction level was set at 60%, and the FFT size to 4096 points. These digital speech files were then edited and segmented into sentence length files and rms amplitude was normalized to the same level sentence by sentence. As delivery of the stimuli was to be by internet, these files were then converted to mp3 format so as to ensure fast download times. The two authors listened to the resulting sound files and judged subjectively that the sound quality was very clear with no audible background noise and that the speech signal was not audibly impaired with relation to the original.

The 75 sentences thus obtained were randomly mixed and used as the test stimuli.

C. Listeners

A total of 50 listeners participated in this study. All were students at the University of Surrey, UK. The listeners formed four groups by first language background: non-native speakers of English with Korean first language ($n=10$), non-native speakers of English with Saudi Arabian Arabic first language ($n=10$), native speakers of English ($n=10$), and non-native speakers of English with mixed first language other than Korean or Saudi Arabian Arabic ($n=20$). No individual served both as talker and listener.

The non-native mixed group were selected to represent as wide a range of other first languages as possible and consisted of: Chinese (Putonghua) ($n=3$), Chinese (Cantonese) ($n=2$), Albanian ($n=2$), Greek ($n=2$), Thai ($n=1$), Bangladeshi ($n=1$), Malaysian ($n=1$), Algerian ($n=1$), Jordanian ($n=1$), Finnish ($n=1$), Norwegian ($n=1$), Ethiopian ($n=1$), Polish ($n=1$), Nigerian ($n=1$), and Sri Lankan ($n=1$). All the participants were volunteers and there was no incentive to participate. Profile data for the three groups of non-native listeners is shown in Table III.

D. Data collection procedures

Data collection was via an internet web site which was developed for this study. This acted as the front-end to a program which collected and tabulated the results. The site homepage contained instructions and a sound test page which listeners could use to check the sound quality and adjust the volume control on their personal computer (PC). The format of the test was also introduced. Before starting the test, the second author gave verbal instructions to the listeners, which were the same as those on the homepage of the website.

The purpose of using a website was not to facilitate the gathering of data from remote sites, but to automate the data collection procedure and to obviate the need for data input from paper questionnaires. All the responses were gathered in a single information technology (IT) laboratory at the Uni-

TABLE III. Profile of the three non-native listener groups. The first value in each cell is the mean, with standard deviations following in parentheses. Not all respondents supplied a result for IELTS and TOEFL scores; in these cases, the number supplying this information is given.

	Listener group		
	Korean	Saudi Arabian	Mixed
No. of respondents	10	10	20
Age (years)	27.53 (4.64)	31.69 (3.04)	26.28 (4.34)
Length of learning English (years)	9.33 (5.50)	8.77 (5.04)	11.83 (6.74)
Time in UK (years)	1.83 (1.22)	2.81 (1.35)	1.86 (1.56)
TOEFL (computer)	235.66 (12.50) ($n=3$)	265 (7.07) ($n=2$)	295 (N/A) ($n=1$)
TOEFL (paper)	550 (N/A) ($n=1$)	528 (31.11) ($n=2$)	565.33 (16.62) ($n=3$)
IELTS	6.08 (0.38) ($n=6$)	5.83 (0.76) ($n=3$)	6.5 (0.81) ($n=7$)

versity of Surrey in the presence of the second author. No more than three listeners were present at any one time. The computers used were all situated in the same setting and equipped to the same specification.

Once the listener was satisfied with the sound quality and familiar with the test format, a "start" button started the test. Each question was displayed in turn, and at the same time the listener heard the appropriate stimulus. The listener's task was to listen to the sentence played and to enter the keywords of the stimulus sentence in the empty form fields provided. After filling out the keywords, the listener was prompted to click the "next" button to proceed to the next question. The program then stored the results on the server where the web site was hosted. There was no time limit to fill out the answer for each question, but listeners were allowed to listen to each stimulus sentence once only. This was achieved by tracking what had already been played in the database and removing this from the play list. The second author watched the procedure each time and was satisfied that no cheating (such as attempting to listen more than once to an item) took place.

Intelligibility scores were determined by a keyword-correct count. All the test results were double marked independently of the authors of this study by two assistants who discussed any differences of opinion until an agreement was reached. Words with added or deleted morphemes were counted as incorrect. However, obvious spelling errors were counted as correct.

As each talker said 15 sentences, and each sentence had three keywords, one talker could receive a score from a minimum of zero to a maximum of 45.

E. Word familiarity test

After completing the word recognition test, the listener was, again online, presented with a word familiarity test page. The 73 unique keywords from the sentence lists were presented to the listeners in written format, and the listeners were asked to rate their familiarity on a scale of 1 to 4, as follows: (1) = I don't know this word; (2) = I'm not sure about its meaning; (3) = I know its meaning but am not sure how to pronounce it; (4) = I know this word.

F. Listener profiles

After completing the word recognition test and the word familiarity test, the listeners were asked to complete a short questionnaire on their personal details, which included the information given in Table III, namely, age, years spent learning English, length of residence in the UK, TOEFL (computer and/or paper), and/or IELTS (as applicable), and first language.

The listener had to complete the whole test including the word familiarity test and personal details including first language for the results to be stored. If the answers were incomplete, the results were automatically discarded. However, not all listeners were able to supply English language test results, so missing results were allowed for these.

III. RESULTS

A. Word familiarity test

66% of the listeners gave the maximum rating of 4 to all of the words. 76% of the listeners gave a rating of 4 to at least 98% of the words. Only ten words were given average scores less than 3.8: "curtains" (3.6), "held" (3.6), "hurt" (3.7), "purse" (3.6), "sat" (3.7), "scarf" (3.5), "stirs" (3.3), "swept" (3.6), "tied" (3.7), and "worries" (3.7). The non-native listeners were thus all highly familiar with the target keywords. A correlation of the results of this test and the keyword intelligibility test reported below was highly significant ($p=0.000$), so it was taken that the use of these keywords was a valid measure of intelligibility.

B. Keyword intelligibility test

Table IV shows the intelligibility scores of the five talkers to each of the four listener groups as raw scores out of a maximum possible of 45, with standard deviations in parentheses.

To enable direct comparison with the data presented by Bent and Bradlow (2003), "rationalized" arcsine units (rau) are presented in Table V, as this is the way in which Bent and Bradlow presented their data, but all further analysis is performed on the raw data as there are no significant differences between the raw data and the rau converted data.

A repeated-measures analysis of variance (ANOVA) was conducted with talker (high-proficiency Korean, high-proficiency Saudi Arabian, native English, low-proficiency Korean, low-proficiency Saudi Arabian) as the within-subjects factor and listener (Korean, Saudi Arabian, mixed, native English) as the between-subjects factor. The result of this test showed that there were highly significant main effects of the talker [$F=87.573$, $P=0.000$] and listener group [$F=5.317$, $P=0.003$]. The interaction of talker and listener group was highly significant [$F=7.365$, $P=0.000$]. Figure 1 shows this interaction effect.

Post hoc pairwise comparisons of talker intelligibility were conducted within each listener group. Because of the large number of paired comparisons (ten for each listener

TABLE IV. Raw scores on word recognition test (maximum score possible 45) with standard deviations in parentheses.

	Talker				
Listener group	Korean high proficiency	Saudi high proficiency	Native English	Korean low proficiency	Saudi low proficiency
Korean	40.20 (2.53)	36.40 (2.72)	39.60 (2.22)	37.00 (3.33)	27.90 (4.61)
Saudi Arabian	39.30 (3.56)	39.10 (5.93)	39.10 (5.32)	32.70 (5.10)	35.20 (3.58)
Non-native Mixed L1	38.95 (4.26)	40.10 (3.45)	39.95 (3.17)	34.35 (4.91)	34.15 (5.44)
Native English	43.40 (3.03)	44.30 (0.95)	44.10 (2.85)	39.60 (2.80)	36.20 (2.30)

group), Bonferroni tests were used with the significance level set at $p<=0.005$ in order to compensate for the familywise error which would otherwise be associated with multiple pairwise comparisons. Table VI shows significant differences between the intelligibility scores given by the four listener groups to the various talkers.

As Table IV shows, the Korean listeners found the Saudi Arabian low-proficiency talker less intelligible than all the other talkers, with an intelligibility score of 27.9 out of 45. Table VI shows that this was significantly different from these listeners' judgement of all the other talkers, ($p<=0.001$). These listeners found no significant difference between any of the other talkers.

The Saudi Arabian listeners gave the Korean low-proficiency talker an intelligibility score of 32.7, a lower score than they gave to all the other talkers. This was significantly different ($p<=0.003$) from the scores they gave to all the other talkers except the Saudi low-proficiency talker. They found no significant difference between any of the other talkers.

The Korean listeners rated the Saudi high-proficiency talker lower, with a score of 36.4, than either the native English (39.6), Korean high proficiency (40.2), or Korean low-proficiency talkers (37.0), although none of these differences was at a statistically significant level. The Saudi listeners rated the Korean high-proficiency talker with a score (39.3) very slightly higher than that which they gave to the native English talker and the Saudi high-proficiency talker (each 39.1).

The non-native mixed L1 listeners found the two low-proficiency non-native talkers less intelligible than any of the other talkers (Korean low proficiency: 34.35; Saudi low proficiency: 34.15), which was a significant difference from their judgements of the native and high-proficiency talkers ($p<=0.004$). They found no significant difference between the high-proficiency non-native talkers and the native talkers and no significant difference between the two low-proficiency talkers.

The native English listeners also found the two low-proficiency talkers less intelligible than any of the other talkers (Korean low proficiency: 39.6; Saudi low proficiency: 36.2), which was significant at $p<=0.002$. They also found the Saudi low-proficiency talker significantly less intelligible than the Korean low-proficiency talker ($p<=0.004$). These listeners too found no significant difference between the high-proficiency non-native talkers and the native talkers. The results of the mixed L1 listeners' and native English listeners' judgements indicate that the low-proficiency talkers were significantly less intelligible than the other talkers and thus that the talker selection process and division into high and low-proficiency groups was successful.

Neither the English listeners nor the non-native listeners

TABLE V. Scores on word recognition test expressed as rationalized arcsine units (Studebaker, 1985) with standard deviations in parentheses.

	Talker				
Listener group	Korean high proficiency	Saudi high proficiency	Native English	Korean low proficiency	Saudi low proficiency
Korean	91.69 (8.35)	80.55 (7.08)	89.50 (6.85)	82.38 (9.03)	61.24 (9.73)
Saudi Arabian	89.83 (12.23)	91.54 (18.87)	91.07 (17.59)	71.96 (11.47)	77.71 (8.84)
Non-native mixed L1	89.03 (13.24)	92.59 (12.31)	91.93 (11.74)	76.07 (11.95)	75.64 (12.76)
Native English	106.79 (11.95)	110.31 (6.72)	112.42 (11.69)	89.67 (8.07)	79.85 (5.42)

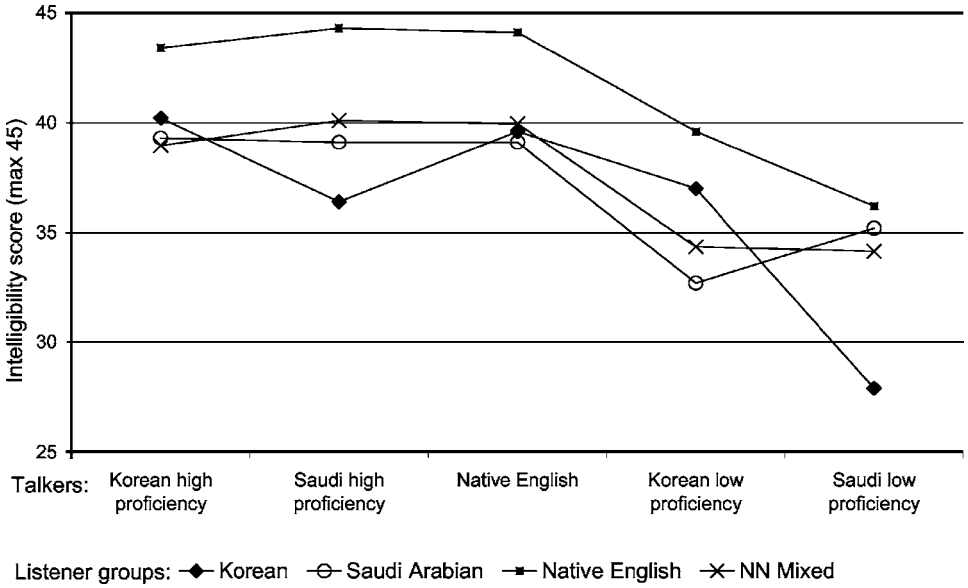


FIG. 1. Intelligibility scores as raw data (maximum possible score 45) for each talker as evaluated by each listener group.

rated the native English talker as more intelligible than the high-proficiency non-native talkers, indicating no intelligibility advantage for the native talkers over the high-proficiency non-native English talkers in any interaction.

To check the reliability of the keyword familiarity test, which was presented in written format, with the results of the intelligibility test, which was presented auditorily, a correlation analysis was performed keyword by keyword between the scores received by each keyword. The correlation result (Spearman's rho) was 0.635, $p=0.000$, indicating that there is a highly significant correlation between the two scores and that the format of the word familiarity test was a reliable predictor of the results in the intelligibility test.

To investigate the possibility that the differences in intelligibility might be due to the effect of varying speech rates, the durations of the stimulus sentences used in the keyword intelligibility test were measured. Table VII shows average sentence durations for each talker, with standard deviations in parentheses.

All pairwise comparisons among the five talkers were significant ($p=0.000$). However, for all listener groups, the average sentence durations for the five talkers did not significantly correlate with their intelligibility scores. Spearman rank correlations were: Korean listeners: rho=0.138, $p=0.239$; Saudi Arabian listeners: rho=0.085, $p=0.468$; Non-native mixed listeners: rho=0.181, $p=0.120$; native English listeners: rho=0.217, $p=0.061$.

Thus for example, the Saudi Arabian high-proficiency talker's sentences had the shortest mean duration but this speaker was judged significantly less intelligible than other talkers only by the Korean listeners. At the other extreme, the Saudi Arabian low-proficiency talker had the longest mean duration, but was again judged significantly less intelligible only by the Korean listeners. This indicates that speech effects are not directly related to intelligibility ratings.

TABLE VI. Multiple pairwise comparisons showing all significant differences in ratings received by talker pairs from each listener group. HP =high proficiency; LP=low proficiency. The first column shows the listener group, the second and third columns those talkers who were given significantly different intelligibility ratings by that listener group, and the last column the p value of that difference. All p values shown are significant at $p \Leftarrow 0.005$ (Bonferroni).

Listener group	Talker pairs		Sig.
Korean	Saudi LP	Saudi HP	0.000
	Saudi LP	English	0.000
	Saudi LP	Korean HP	0.000
	Saudi LP	Korean LP	0.001
Saudi	Korean LP	Korean HP	0.003
	Korean LP	English	0.000
	Korean LP	Saudi HP	0.002
Non-native Mixed L1	Saudi LP	Saudi HP	0.000
	Saudi LP	English	0.000
	Saudi LP	Korean HP	0.004
	Korean LP	Saudi HP	0.000
	Korean LP	English	0.000
	Korean LP	Korean HP	0.001
Native	Saudi LP	Saudi HP	0.000
	Saudi LP	English	0.000
	Saudi LP	Korean LP	0.004
	Saudi LP	Korean HP	0.000
	Korean LP	Korean HP	0.002
	Korean LP	English	0.000
	Korean LP	Saudi HP	0.001

TABLE VII. Average sentence durations for the five talkers in order of increasing duration. Means in milliseconds and standard deviations in parentheses are shown.

Talker	Average sentence duration (ms)
Saudi high proficiency	1808 (209)
Native English	1843 (203)
Korean low proficiency	1928 (217)
Korean high proficiency	2100 (242)
Saudi low proficiency	3013 (256)

IV. DISCUSSION

No evidence was found here that the talker/listener groups who did not share the same first language enjoyed a "mismatched interlanguage speech intelligibility benefit," taken here, in contrast to the Bent and Bradlow use of the term "benefit" as including equal intelligibility scores, to mean only cases in which higher intelligibility scores were given by the listeners. Rather, the most striking finding from this study is that speech intelligibility was clearly lowest in those cases involving low-proficiency non-native talkers and non-native listeners with a different first language, in every case at a statistically significant level.

Thus in both cases the non-native listeners found the low-proficiency non-native talker with whom they did not share the same first language the most difficult of all the talkers to understand. This provides evidence for a mismatched interlanguage speech intelligibility *detriment* in those cases involving the low-proficiency non-native talkers, and indicates that low-proficiency learners may find difficulty in being understood either by non-native listeners who do not share their first language or by native English listeners.

The case of the Korean listeners, who rated the Saudi high-proficiency talker lower than either the native English, Korean high-proficiency, or Korean low-proficiency talkers, provided further, although weaker, evidence against the existence of a mismatched interlanguage intelligibility benefit. This difference was not at a statistically significant level and the degree to which this isolated finding can be generalized from is limited, as the same phenomenon is not found in the case of the Saudi listeners rating the Korean high-proficiency talker.

There was limited evidence in favor of a matched interlanguage speech intelligibility benefit, as evidenced in higher intelligibility scores given by non-native listeners to non-native talkers with the same L1 background. The Korean listeners rated the Korean high-proficiency talker highest and the Korean low-proficiency talker higher than the Saudi low-proficiency talker. The Saudi listeners also rated the Saudi low-proficiency talker higher than the Korean low-proficiency talker, but this latter effect was not statistically significant.

This finding, which shows clear evidence of a talker/listener interaction effect, is in line with the large body of research indicating that non-native speech is likely to be more intelligible to other non-natives from the same language background due to the shared knowledge of the target and native languages and of the transfer effects in perception and production (Bent and Bradlow, 2003, p. 1607).

It thus seems likely that non-native learners who are exposed largely or entirely to the English of their fellow first language speakers will be able to communicate successfully with these speakers, as they can adjust their production and compensate in their perception on the basis of this shared knowledge, but may experience more problems once they leave this environment and attempt to communicate in English with other non-native speakers with different accents or native English speakers.

To take a well-known example, the substitution of [p^h] for English /f/ by Korean first language speakers (e.g., "coffee" pronounced as [$'k^h ɒp^h i$] is likely to be highly disruptive to intelligibility for listeners who are not aware of this feature of Koreans' pronunciation of English, whereas for other Korean listeners, familiar with this pronunciation and the reason behind it, it causes much less disruption to intelligibility. In a teaching context where learners are exposed overwhelmingly to local teachers who share their first language-influenced perception and production problems, this familiarity with the local accent may lull learners into a false sense of their intelligibility which may not be applicable in the wider world. Evidence from nonlaboratory data has shown also that learners are surprisingly unable to use contextual clues to disambiguate phonetically faulty utterances and instead rely heavily on the acoustic signal, often without success (Jenkins, 2000, pp. 80, 81).

It is thus important that the teaching of listening exposes learners to a range of accents beyond those of local speakers if international intelligibility is to be facilitated and that pronunciation teaching eradicates disruptive features of local accents such that learners are able to conform to an internationally intelligible norm.

There was also no strong evidence to support the hypotheses either that native listeners perform better at understanding others, or that native speakers are more intelligible than high-proficiency non-native speakers: Not even the native listeners found their fellow native speaker significantly more intelligible than the high-proficiency speakers. Thus arguments that native speaker speech is fundamentally more intelligible than non-native speech are not supported by this data. Equally, there was no evidence that the native talkers were any less intelligible than the high-proficiency non-native speakers (Munro and Derwing, 1999). More important than native or non-native status was the proficiency level of the non-native speakers.

V. LIMITATIONS OF THE STUDY

This study investigated the intelligibility between talkers and listeners from various first language backgrounds under controlled conditions using isolated sentences, rather than meaningful connected speech in a genuine communicative setting, in which context would have helped to disambiguate unclear speech.

The study focused primarily on subject variables, particularly on the role of the match or mismatch of talker and listeners, and excluded other possible factors that can affect speech intelligibility such as speaking rate or the listener's attitude to a particular foreign accent. Factors such as sociolinguistic variables, which might in real communication interact with the first language background variable, were not taken into consideration.

Both this study and that by Bent and Bradlow used Korean listeners to rate Korean and other non-native talkers and found that these listeners gave higher intelligibility scores to their fellow Koreans than to the mismatched L1 talkers. It is conceivable that there is some sociolinguistic factor at work

which is not addressed by the design of these studies. Further research is needed with other first language groups to investigate this issue.

A possible interfering factor is the use of two different word lists for the high-proficiency and low-proficiency talkers. It is conceivable that the words used in the lists were of inherently different difficulty, although the results of the keyword familiarity test indicate that this was not so.

The participants were all recruited from the student population at the University of Surrey. This may not be an accurate reflection of the various first language speakers represented. It is to be expected that these non-native talkers, who were studying outside their home countries, may be more used to communicating with speakers from other first languages and with native English speakers than are the wider populations of their respective countries. Similarly, the native English talkers who participated in this study may be more used to speaking with non-native speakers than is usual in the native English population at large. They are to this extent also more representative of the international community than those who do not have such contacts, but it is possible that the effects found in this study might be magnified if a similar experiment was carried out using more typical, less travelled, members of each community.

The use of two dimensions on the single rating scale (degree of foreign accent and intelligibility) may not have been a felicitous choice as it presupposes that the two are directly related. Research (e.g., Munro and Derwing, 1999) indicates that this assumption may not be a safe one and it is recommended that in future work these scales should not be conflated.

Both this study and that of Bent and Bradlow (2003) suffer from the very small sample of talkers and are thus highly dependent on the characteristics of particular individuals' speech. The degree to which it is possible to generalize from the findings of either study is thus very limited. Further research on intelligibility, both between non-native speakers and between native and non-native speakers, should be conducted using a wider variety of data types, including larger experimental studies and field studies in which contextual and sociolinguistic factors are allowed to play a role.

ACKNOWLEDGMENT

The authors would like to thank Joseph Juil Kim for the development of the website.

APPENDIX A: SENTENCE LISTS

Stimulus sentences were taken from the Bench-Kowal-Bamford sentence lists (Bench and Bamford, 1979). Keywords used in the study are underlined.

Sentence list 1

1) The book tells a story.
2) The thin dog was hungry.
3) He needed his holiday.
4) The milk came in a bottle.
5) Father looked at the book.
6) The cook cut some onions.
7) The lady goes to the shop.
8) The children dropped the bag.
9) She found her purse.
10) They washed in cold water.
11) The ball broke the window.
12) The table has three legs.
13) The shoes were very dirty.
14) The car hit a wall.
15) The girl caught a cold.

Sentence list 2

1) The old man worries.
2) She used her spoon.
3) The mother stirs the tea.
4) The clever girls are reading.
5) The floor looked clean.
6) The lady packed her bag.
7) He broke his leg.
8) The family like fish.
9) The cleaner swept the floor.
10) She hurt her hand.
11) The girl held a mirror.
12) He paid his bill.
13) Mother made some curtains.
14) They sat on a wooden bench.
15) The man tied his scarf.

Albrechtsen, D., Henriksen, B., and Færch, C. (**1980**). "Native speaker reactions to learners' spoken interlanguage," Lang. Learn. **30**, 365–396.

Anderson-Hsieh, J., and Koehler, K. (**1988**). "The effect of foreign accent and speaking rate on native speaker comprehension," Lang. Learn. **38**(4), 561–613.

Bamford, J., and Wilson, I. (**1979**). "Methodological considerations and practical aspects of the BKB sentence lists," in *Speech-hearing Tests and the Spoken Language of Hearing-Impaired Children*, edited by J. Bench and J. Bamford (Academic, London), pp. 148–187.

Bench, J., and Bamford, J. (**1979**) (eds.) *Speech-hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Bent, T. (**2001**). "Interlanguage benefit for non-native speaker intelligibility," J. Acoust. Soc. Am. **109**(5), Pt. 2, 2472.

Bent, T., and Bradlow, A. R. (**2003**). "The interlanguage speech intelligibility benefit," J. Acoust. Soc. Am. **114**(3), 1600–1610.

Crystal, D. (**2003**). *English as a Global Language* (Cambridge University Press, Cambridge).

Ensz, K. Y. (**1982**). "French attitudes toward typical speech errors of American speakers of French," The Modern Language Journal **66**, 133–139.

Flege, J. E. (**1987**). "The production and perception of speech sounds in a foreign language" in *Human Communication and its Discorders*, edited by H. Winitz, Vol. 3 (Ablex, Norwood, NJ).

Flege, J. E. (**1988**). "Factors affecting degree of perceived foreign accent in English sentences," J. Acoust. Soc. Am. **84**(1), 70–79.

Flege, J. E., and Fletcher, K. (**1992**). "Talker and listener effects on the perception of degree of perceived foreign accent," J. Acoust. Soc. Am. **91**(1), 370–389.

Gass, S., and Varonis, E. M. (**1984**). "The effect of familiarity on the comprehensibility of nonnative speech," Lang. Learn. **34**, 65–89.

Graddol, D. (**1997**). *The Future of English?: A Guide to Forecasting the Popularity of the English Language in the 21st Century* (The British Council, London).

Hinofitis, F. B., and Bailey, K. M., (**1981**). "American undergraduates' reactions to the communication skills of foreign teaching assistants" in *On TESOL '80: Building Bridges*, edited by J. Fisher, M. Clarke, and J. Schachter (TESOL, Washington, D.C), pp. 120–133.

Jenkins, J., (**2000**). *The Phonology of English as an Int. Language* (Oxford University Press, Oxford).

Jenkins, J., (**2002**). "A sociolinguistically based, empirically researched pronunciation syllabus for English as an Int. Language," Appl. Linguist.

23(1), 83–103.
Kranke, K., and Christison, M. A., (**1983**). "Recent language research and some language teaching principles," TESOL Quarterly **17**(4), 635–650.
Munro, M. J., and Derwing, T. M., (**1999**). "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," Lang. Learn. **49**, Suppl. 1, 285–310.
Politzer, R. L., (**1976**). "Linguistic accuracy and intelligibility" in *Proc. 4th Int. Congr. Appl. Linguistics* (Hochschul-Verlag, Stuttgart). pp. 505–513.
Smith, L. E., and Bisazza, J. A., (**1982**). "The comprehensibility of three varieties of English for college students in seven countries," Lang. Learn. **32**, 259–269.
Smith, L. E., and Rafiqzad, K., (**1979**). "English for cross-cultural communication: the question of intelligibility," TESOL Quarterly **13**(3), 371–380.
Studebaker, G. A. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res. **28**, 455–462.
Tsutsui, K., Suzuki, H., Shimoyoshi, O., Sonohara, M., Akagiri, K., and Heddle, R. M., (**1992**). *ATRAC: Adaptive transform acoustic coding for MiniDisc*. Reprinted from 93rd Audio Engineering Society Convention, San Francisco, Oct. 1–4, 1992. ⟨http://www.minidisc.org/aes_atrac.html⟩. Accessed 2 February 2005.
van Wijngaarden, S. J., (**2001**). "Intelligibility of native and non-native Dutch speech." Speech Commun. **35**, 103–113.
van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T., (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners." J. Acoust. Soc. Am. **111**, 1906–1916.

Speech feature extraction method using subband-based periodicity and nonperiodicity decomposition[a)]

Kentaro Ishizuka,[b)] Tomohiro Nakatani, and Yasuhiro Minami
NTT Communication Science Laboratories, NTT Corporation, Hikaridai 2-4, Seikacho, Sourakugun, Kyoto, 619-0237, Japan

Noboru Miyazaki
NTT Cyber Space Laboratories, NTT Corporation, Hikarino-oka 1-1, Yokosuka City, Kanagawa, 239-0847, Japan

(Received 1 November 2004; revised 24 April 2006; accepted 24 April 2006)

This paper proposes a speech feature extraction method that utilizes periodicity and nonperiodicity for robust automatic speech recognition. The method was motivated by the auditory comb filtering hypothesis proposed in speech perception research. The method divides input signals into subband signals, which it then decomposes into their periodic and nonperiodic components using comb filters independently designed in each subband. Both features are used as feature parameters. This representation exploits the robustness of periodicity measurements as regards noise while preserving the overall speech information content. In addition, periodicity is estimated independently in each subband, providing robustness as regards noise spectrum bias. The framework is similar to that of a previous study [Jackson *et al.*, Proc. of Eurospeech. (2003), pp. 2321–2324], which is based on cascade processing motivated by speech production. However, the proposed method differs in its design philosophy, which is based on parallel distributed processing motivated by speech perception. Continuous digit speech recognition experiments in the presence of noise confirmed that the proposed method performs better than conventional methods when the noise in the training and test data sets differs. [DOI: 10.1121/1.2205131]

PACS number(s): 43.72.Ar, 43.72.Ne, 43.71.An [DOS] Pages: 443–452

I. INTRODUCTION

The performance of automatic speech recognition (ASR) systems is worse in "real world," such as in a car on the street, or at a station, than in ideal environments where there is no noise or channel distortion. There are three major approaches for coping with the problem (Gong, 1995). The first is a signal preprocessing approach such as speech enhancement (e.g., Lim and Oppenheim, 1978; Koo *et al.*, 1989) and noise reduction techniques (e.g., Boll, 1979; Lockwood and Boudy, 1992; Sim *et al.*, 1998). The second is a model compensation approach such as parallel model combination (e.g., Varga and Moore, 1990; Gales and Young, 1993; Minami and Furui, 1995) and model adaptation methods (e.g., Lee *et al.*, 1991; Vaseghi and Milner, 1993; Bernard *et al.*, 2004). The third is an approach that employs robust speech features to deal with noise or channel distortion. Focusing solely on the third approach, this paper proposes a new method for extracting robust speech feature representation.

Mel frequency cepstral coefficients (MFCCs), which are based on findings related to the pitch perception characteristics of the human auditory system, have been widely used for ASR since Davis and Mermelstein (1980) showed that MFCCs provide better ASR performance than other features such as linear frequency cepstral coefficients (Bogert *et al.*, 1963) and linear prediction cepstral coefficients (Itakura, 1975). However, MFCCs are insufficiently robust in the real world (e.g., Gong, 1995; de Veth *et al.*, 2001; Singh *et al.*, 2002) because their feature representations are easily contaminated by noise or channel distortion. This vulnerability has made it necessary to develop more robust speech feature extraction methods.

Most robust speech feature extraction methods have been based on findings related to the human auditory system such as perceptually based linear prediction analysis (PLP) (Hermansky *et al.*, 1985; Hermansky 1990) and other auditory filter-based methods (e.g., Gao *et al.*, 1992; Aikawa *et al.*, 1996; Li *et al.*, 2001). In particular, several methods based on the characteristics of the auditory nerve response to periodic signals improve ASR robustness in noisy environments. Such methods include the generalized synchrony detector (GSD) (Seneff, 1988), the average localized synchrony detector (ALSD) (Ali *et al.*, 2002), the ensemble interval histogram (EIH) (Ghitza, 1988, 1994), and zero-crossing with peak amplitude (ZCPA) (Kim *et al.*, 1999).

GSD and ALSD focus on the characteristic frequencies (CFs) of the auditory filter, namely the frequencies to which the filters respond most sensitively. These methods measure the synchronicities of subband signals at the CFs and use these as speech features for ASR. The methods assume that

[a)]Portions of this work were presented in "Speech feature extraction method representing periodicity and aperiodicity in sub bands for robust speech recognition," Proceedings of the 29th International Conference on Acoustics, Speech and Signal Processing, Montreal, Canada, May 2004, and "Improvement in robustness of speech feature extraction method using sub-band based periodicity and aperiodicity decomposition," Proceedings of the 8th International Conference on Spoken Language Processing, Jeju, Korea, October 2004.

[b)]Electronic mail: ishizuka@cslab.kecl.ntt.co.jp

an auditory filter is a kind of frequency analyzer and adopt a strategy designed to improve robustness in noisy environments by enhancing the spectral peaks of the periodic signals around the CFs of the auditory filter. In terms of engineering, subband cross-correlation analysis (SBCOR) (Kajita and Itakura, 1995) uses the autocorrelation coefficients at the center frequencies of bandpass filters as speech features for ASR. While SBCOR indeed uses the autocorrelation function, it only focuses on a fixed coefficient corresponding to the center frequency of a bandpass filter. Therefore, we can consider that SBCOR employs the same strategy as that used by GSD or ALSD. However, because these methods focus on the synchronicities between the input signals and the CFs of the auditory filter, they are not good at handling the periodicity of signals that deviate from the center frequency of one of the bandpass filters.

Auditory nerve firings are not phase locked to their CFs but to lower-frequency components than the CFs (e.g., Rose *et al.*, 1971; Johnson, 1980; Greenberg, 2004). In the time domain, auditory nerve firings can represent frequencies that are different from their CFs. EIH and ZCPA are based on such phase-locking characteristics of the auditory nerves. These methods can handle any periodicity by generating a histogram of periodicities extracted from the subband signals, and then use the histogram as a speech feature for ASR.

The above methods can also substantially improve noise robustness by using the periodicity of the speech signal, because it is essentially difficult for the periodicity to be contaminated by interferer signals. Indeed, the human auditory system may use such speech signal characteristics. Psychoacoustic research has revealed that the human auditory system is very sensitive to the harmonicity that is related to the periodicity of sound (e.g., Darwin and Carlyon, 1995). However, speech signals consist not only of strict periodic signals such as steady parts of vowels and voiced consonants, but also of nonperiodic signals such as fluctuations intrinsically included in vowels, voiced consonants, stops, fricatives, and affricates. Because the above methods focus strongly on periodicity, they offer no advantage when it comes to representing the nonperiodic characteristics of speech. Therefore, they often degrade the ASR performance in clean (no-noise) or low-noise environments.

In terms of psychoacoustics, de Cheveigné (1997) used concurrent vowel identification experiments to suggest that the human auditory system can suppress harmonic interferes and perceive the residual target signal. This finding suggests that the human auditory system may process both the harmonic (periodic) feature, and the residue after canceling the harmonicity (nonperiodic) feature, which deviates from the dominant periodicity. Ishizuka and Aikawa (2002) also showed that very small fundamental frequency (F0) fluctuations of vowels improve human vowel identification in the presence of harmonic noise. Their results suggest that the quasi-periodicity of the speech sounds helps human speech perception in the presence of interferers, and support the importance of nonperiodic features that deviate from strict periodicity. In addition, in terms of engineering and speech production research, Jackson *et al.* (2003) showed that the ASR accuracy in noisy environments can be improved by using the decomposed periodic and nonperiodic features of speech signals. Their result indicates the efficiency of representing speech signals by their periodic and nonperiodic features for ASR.

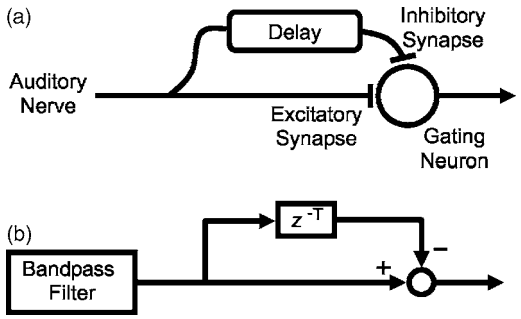


FIG. 1. (a) Neural cancellation filter (de Cheveigné, 1997). (b) An implementation of the neural cancellation filter using a delay unit and bandpass filter banks.

This paper proposes a feature extraction method that utilizes both periodic and nonperiodic features for each subband using auditory filter banks and comb filters. Henceforth, this proposed method is called Subband-based Periodicity and Aperiodicity DEcomposition (SPADE). The development of this decomposition method based on comb filters was motivated by the auditory comb filtering hypothesis (de Cheveigné, 1997), and its periodicity representation was motivated by the auditory nerve firing characteristics in the time domain. Section II describes the background to SPADE and focuses on the auditory comb filtering hypothesis and the advantages of its sound representation compared with the conventional speech features used for ASR. Section III describes the SPADE algorithm in detail. In Sec. IV, evaluation experiments undertaken with the noisy continuous digit speech recognition database show that SPADE can improve the ASR performance in the presence of noise in the real world.

II. BACKGROUND

A. Auditory comb filtering hypothesis

As described above, in terms of psychoacoustics, de Cheveigné (1993, 1997) found that the human auditory system can suppress harmonic interferers and perceive the residual target signal through concurrent vowel identification experiments. He proposed the auditory comb filtering hypothesis based on his findings. Although the existence of such mechanisms in the human auditory system has yet to be confirmed, this hypothesis suggests that the human auditory system may process both the periodic and nonperiodic features. In his hypothesis, he proposed a neural model of harmonic interference cancellation. Figure 1(a) shows the model: a neural cancellation filter. This model is implemented by a neuron with two synapses. One is an excitatory synapse fed by an input spike train, the other is an inhibitory synapse fed by a delayed input spike train. The neuron fires each time an input spike arrives, unless a delayed input spike arrives simultaneously from the delayed path. The length of the delay can be tuned adaptively to the rate of the input spikes, which have the dominant periodicity in the spikes, de

Cheveigné thought that such neural comb filtering mechanisms may be independently implemented corresponding to each auditory nerve fiber or group of fibers. This model provides a good explanation of the experimental results obtained for concurrent vowel identification (de Cheveigné, 1997).

Such mechanisms can be easily implemented using delay units and bandpass filter banks. One example is shown in Fig. 1(b). Note that the delay units should be tuned adaptively, and that the filter banks should have the same characteristics as auditory nerve fibers. SPADE is based on this implementation. This implementation realizes a rich representation of the input sound with the decomposed dominant periodicity and nonperiodicity, which is the residue after the suppression of the dominant periodicity. This sound representation has advantages compared with the conventional speech features used for ASR as explained below. In this paper, to simplify the formulation, we ignore such mechanisms inherently implemented in the human auditory system as the half-wave rectification and the sound compression.

B. Advantages of the sound representation based on the hypothesis

Key aspects of SPADE are the decomposition of the speech signal into its periodic and nonperiodic features, and the utilization of both features as speech features for ASR. The periodicity in a channel is calculated in a more adaptive manner than with GSD, ALSD, or SBCOR, which only focus on the fixed center frequency of the bandpass filter. Because SPADE employs an adaptive mechanism to search for the dominant periodicity within the search range, it can deal with frequencies that differ from the center frequency of the bandpass filter. The periodicity/nonperiodicity decomposition cannot be performed after the transformation from speech signals to the MFCC or PLP, because these methods inherently lose information about the periodicity and nonperiodicity of the input signals. Even if we apply some kind of statistical processing to these methods, it is very difficult to restore the information without using speech signal statistics. SPADE can exploit periodicity and nonperiodicity properties that the conventional speech features cannot deal with. The decomposition of speech signals into two properties and their subsequent use allow us to employ the robustness exhibited by periodic features without losing certain essential speech signal information included in the nonperiodic features. When the noise in the test data differs from that in the training data, the difference between the nonperiodic features of the two sets of data becomes large. However, the difference between the periodic features may remain small because the periodicity can be extracted from noisy speech more easily than the nonperiodicity. In contrast to SPADE, with the MFCC or PLP, the difference between the environmental sounds in the training and test data directly distorts all parameters, because the sound representation of the features considers the estimated spectrum of the whole sound signal rather than just speech signal characteristics. Therefore, SPADE is possibly more robust than the conventional features in noisy environments. On the other hand, if there is very little noise, SPADE can utilize both the periodic and nonperiodic to achieve a high ASR performance. Conventional robust features that utilize the periodic property cannot employ the nonperiodic property, which includes important information about speech signals. SPADE employs both periodic and nonperiodic properties to achieve high ASR performance under any conditions.

These key features of SPADE are similar to those of the method proposed by Jackson *et al.* (2003). However, because their study is based on a cascade processing approach motivated by human speech production mechanisms, the first step in their method depends strongly on the accuracy of a pitch-scaled harmonic filter (PSHF) (Jackson and Shadle, 2001). Therefore, any failure to decompose the harmonicity in the first step may have a large effect, and so they needed to use F0 values estimated from clean speech data to decompose the periodicity and nonperiodicity of noisy speech data in their evaluation experiment. By contrast, SPADE employs an independent periodicity estimation within each subband and a periodicity/nonperiodicity decomposition design based on a parallel distributed processing approach motivated by the human speech perception process. This is another key feature of SPADE. In general, it is difficult to estimate the fundamental frequencies from noisy speech signals. However, even in the presence of noise, some frequency regions where the speech signal energy is strong have high signal-to-noise ratios (SNRs). Based on this property of speech signals, SPADE estimates the periodicity independently in each subband, and thus can reliably estimate the periodicity in such high SNR regions and successfully decompose the speech signals into two properties. Because of the mechanism, unlike previous studies (Jackson and Shadle, 2001; Jackson *et al.*, 2003), SPADE is expected to offer such advantages as the ability to recover failed harmonicity estimations and to cope with an interferer whose energy is not distributed evenly in the frequency region. In summary, although the frameworks of SPADE and the system proposed by Jackson *et al.* (2003) are similar, these two approaches have different backgrounds and also differ in terms of design philosophy, which is mainly distinguished by the use of parallel or cascade processing.

III. METHOD

Figure 2 is a block diagram of SPADE. This method consists of six steps. In the first step, an input speech signal is divided into subband signals by bandpass filter banks for which this paper employs gammatone filter banks (Patterson and Moore, 1986). The center frequencies and bandwidths for each filter in the filter banks are decided in terms of the equivalent rectangular bandwidth (ERB) scale. The ERB is considered to be the critical bandwidth of an auditory filter and was measured as the spectral notch width of the notched masking noise, which can mask the pure tone positioned at the center frequency of the spectral notch (Patterson, 1976). ERB values can be calculated by using the approximation formula given below (Moore and Glasberg, 1983):

$$\mathrm{ERB} = 6.23f^2 + 93.39f + 28.52,$$

where ERB is the ERB value in Hz and f is the frequency in kHz. In our example, we use gammatone filter banks that

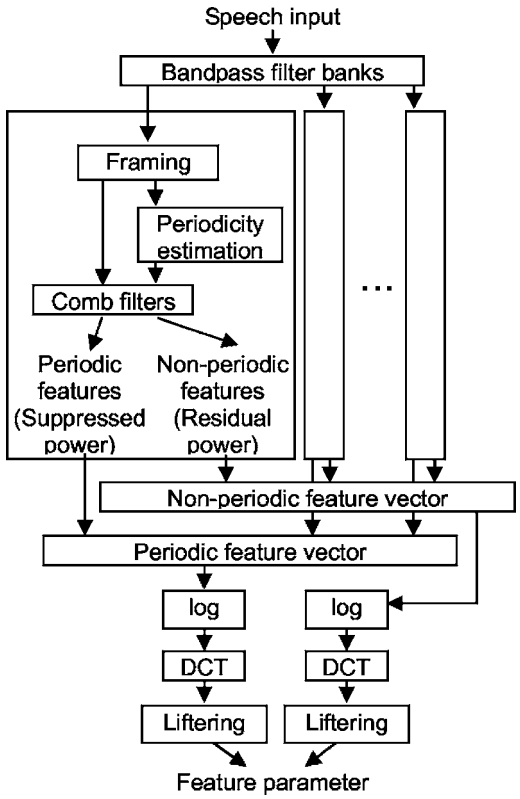


FIG. 2. Block diagram of the proposed speech feature extraction method "SPADE."

consist of 24 bandpass filters whose frequency characteristics are shown in Fig. 3.

In the second step, the output signal for each filter is analyzed using windows with a certain temporal length that are shifted a certain temporal length. In this paper, the frame length is fixed and the frame shift is synchronized for all subbands. In our example, we analyzed the filtered speech using 25-ms rectangular windows at 100 frames per second.

In the third step, the dominant periodicity is detected independently in each frame. To detect the periodicity, we use the autocorrelation method, which has previously been employed for F0 estimation (Rabiner, 1977; Hess, 1983). The method calculates the autocorrelation function of the signal in the frame and searches for the maximum peak of the function within a certain search range, e.g., from 50 to 400 Hz where the F0s of human speech exist. In this paper, the search range is fixed at 80 to 200 Hz, which is roughly a one-octave range covering the human F0 or its half

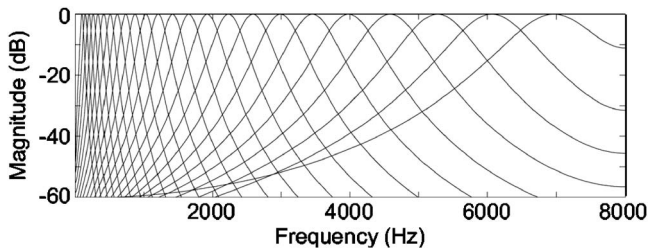


FIG. 3. An example of the frequency characteristics of 24-channel gammatone filter banks. The abscissa indicates the frequency. The ordinate indicates the magnitude.

or one-third values. The reason for the limited range employed with this method is given below. In fact, the F0s of human speech range from 50 to 400 Hz. However, any F0 estimation method sometimes estimates half, one-third, or double the actual F0 value. SPADE aims to decompose the periodic and nonperiodic features of the input sound by suppressing the dominant periodicity of the sound using comb filters described in the fourth step. In addition, SPADE utilizes only the power of the decomposed features described in the fifth step. Because the power of the periodic features is condensed in a harmonic component, if double the value of the actual F0 is estimated in this step, a comb filter designed based on the estimated F0 will fail to suppress a large part of the power of the dominant periodic feature. We consider that this failure yields the large difference between the estimated and actual values and affects the ASR performance. On the other hand, although the power of the residual nonperiodic feature is oversuppressed, a comb filter designed based on a half or one-third value estimation can suppress the whole power of the dominant periodic feature. Since the power of the nonperiodic feature may be broadly distributed in the spectra, we consider that oversuppression of the nonperiodic feature does not seriously affect the ASR performance. Therefore, SPADE gives priority to suppressing the dominant periodicity completely and concentrates the F0 search range at low F0 values. In addition, each subband is searched for in the same search range. The reason for the identical range for all subbands is given below. Even in the higher order bands, which do not cover the frequency of the search range, the dominant periodic signal has higher harmonic components whose spaces correspond to the F0. To suppress these periodic components by using comb filters, SPADE employs an identical F0 search range for each subband.

In the fourth step, the signal in the frame is comb filtered based on the periodicity detected in the third step. The use of suppression type comb filters for the decomposition is motivated by the auditory comb filter hypothesis (de Cheveigné, 1997). The characteristic of the comb filter is given by $H(z)$ in the z-transformation form, where n indicates the period with the maximum value detected in the third step:

$$H(z) = 1 - z^{-n}.$$

This comb filter is designed to suppress the whole power of the dominant periodicity of the input speech as described above.

In the fifth step, the power suppressed by the comb filtering and the power of the residual signal in the frame after the comb filtering are calculated as the sum of the square of the signals. The power suppressed by the comb filtering is calculated as the difference between the signal powers before and after the comb filtering. After this step, the power suppressed by the comb filtering is considered to be a periodic feature, and the power of the residual signal is considered to be a nonperiodic feature. Note that the term "nonperiodic" does not strictly mean signals that are completely without periodicity such as white noise. Because natural voiced speech signals are quasi-periodic signals rather than strictly periodic signals, such speech signals include quasi-periodic fluctuation components. Therefore, most of the fluctuation

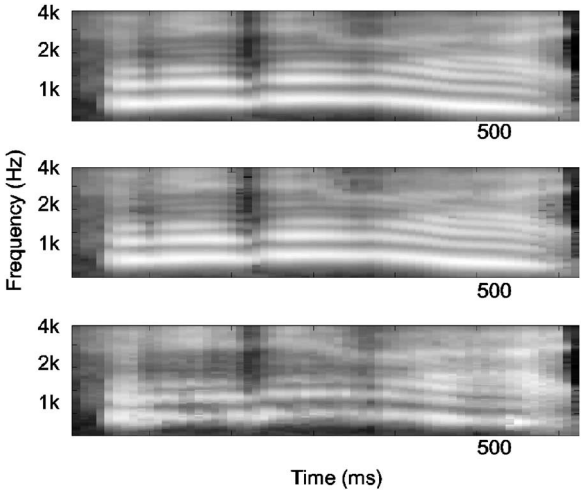


FIG. 4. Examples of the excitation patterns, i.e., the output powers from gammatone filter banks (top), and the periodic (middle) and aperiodic features (bottom) of the Japanese word "soredewa" read by a female speaker. White indicates the highest power and black the lowest power. The periodic feature well represents the power of the stable periodicity, whereas the aperiodic feature shows the power changing part, e.g., sound onset and format transition as white.

component remains in the nonperiodic features after the dominant periodicity has been suppressed. After this feature decomposition, the powers across the subbands at the same frame shifting point are combined for each feature and considered to be vectors. Figure 4 shows examples of the output power patterns from gammatone filter banks, which are similar to auditory excitation patterns (Moore and Glasberg, 1983), and the vectors of its periodic and nonperiodic features obtained after analyzing a speech sentence with 48-channel gammatone filter banks and a 25-ms frame at 100 frames per second. As seen in Fig. 4, the periodic features show strong power responses, especially at the stable part of the periodicity. In contrast, the nonperiodic features show weak power responses at the stable part of the periodicity, whereas they show stronger power responses around the onsets and offsets of voiced speech or formant transition parts, which can be considered the fluctuations of sounds that deviate from the dominant periodicity, rather than the stable part of the periodicity.

In the final step, each power vector calculated in the fifth step is logarithmic transformed, and its cepstral coefficients are calculated from the logarithmic transformed power vector using the discrete cosine transform (DCT) as shown below, where N is the number of gammatone filters, m_j is the power vector for filter index number j, and c_i is the ith cepstral coefficient.

$$c_i = \sqrt{\frac{N}{2}} \sum_{j=1}^{N} \log(m_j) \cos\left(\frac{\pi i}{N} (j - 0.5) \right).$$

The formula includes both the logarithmic transformation and DCT, and is the same as that used with the conventional MFCC calculation (e.g., Huang *et al.*, 2001). These cepstral coefficients are calculated for each feature, and only certain low order coefficients, e.g., from the first to twelfth coefficients, are used as the speech features for ASR. Then, both cepstral coefficients are combined as the feature parameters, that is, if the coefficients from the first to twelfth order are used, then the total number of feature parameters is 24 (12 periodic and 12 nonperiodic feature coefficients).

IV. EXPERIMENTS

A. Aurora-2J

We conducted evaluation experiments with the AURORA-2J database (Nakamura *et al.*, 2003, 2005). AURORA-2J is the Japanese version of AURORA-2 (Hirsh and Pearce, 2000; Pearce and Hirsh, 2000), which is a benchmark English speech data set for continuous digit recognition in the presence of noise. That is, AURORA-2J is a database for evaluating the performance of Japanese continuous digit speech recognizers in the presence of environmental interferer sounds. The recognition task was continuous telephone band-limited digit speech recognition, and the number of vocabulary was 11 (from one to nine, and zero, which has two kinds of reading). The environmental interferer sounds are the same as those included in AURORA-2 and are recorded in real world. AURORA-2J includes two kinds of training data sets and three kinds of test data sets (test sets A, B, and C). One of the training sets is called a clean-speech training data set, which includes speech data spoken in a clean (noiseless) environment. The other is called a multicondition-speech training data set, which includes the clean speech data mixed with subway, babble, car, and exhibition sounds, which are the same environmental sounds as those in test set A at SNRs of 5, 10, 15, 20, and infinity (no noise condition) dB. Each training set includes 8440 continuous digit utterances spoken by 110 Japanese speakers (55 speakers, 4220 utterances for each sex). These utterances are the same for the two training sets, and only the noise conditions are different. Henceforth, the term "clean-speech training" and "multicondition-speech training" mean that the recognizer is trained using a clean-speech and a multicondition-speech training data set, respectively. The test data sets are convoluted with different telephone channel characteristics and mixed with different environmental sounds related to the training data sets. The differences are shown in Table I. Test set A includes speech data mixed with subway, babble, car, and exhibition sounds, which are the same environmental sounds as those in the multicondition-speech training data set. Test set B includes speech data mixed with restaurant, street, airport, and railway station sounds. Test set C includes speech data, whose channel characteristics (MIRS) differ from those of the other test sets and training data sets (G712), mixed with subway and street sounds. The SNRs are −5, 0, 5, 10, 15, 20, and infinity (no noise condition) dB. There are 1001 continuous digit speech utterances spoken by 104 speakers (52 speakers of each sex) for each combination of environmental sound and SNR. Consequently, test sets A and B include 28 028 utterances (1001 utterances, four environmental sounds, seven SNRs) for each, and test set C includes 14 014 utterances (1001 utterances, two environmental sounds, seven SNRs). The multicondition-speech training data set includes the same environmental sounds as those in test set A and half of test set C. Therefore, with

TABLE I. Noise and channel conditions for three test sets in AURORA-2J. The noise condition in test set A is closed only with multicondition-speech training. With clean-speech training, the noise condition is always open. Half of the noise types in test set C are closed only with multicondition-speech training, and the other half are open.

Training condition	Test set	Kind of distortion	
		Noise	Channel
Clean-speech training	A	Open	Closed
	B	Open	Closed
	C	Open	Open
Multicondition-speech training	A	Closed	Closed
	B	Open	Closed
	C	Open/Closed	Open

multicondition-speech training, the noise conditions are "closed" for test set A and half of test set C. On the other hand, the noise conditions are "open" for test set B and the other half of test set C with multicondition-speech training. With clean-speech condition training, the noise conditions are "open" for all test sets. Because of the difference between the channel characteristics of the training data sets and test set C, the channel conditions of test set C are "open" under both training conditions.

AURORA-2J provides the baseline recognition performance achieved by a conventional MFCC-based speech recognizer. This MFCC-based recognizer uses 23 mel-scale bandpass filter banks, 25-ms Hamming windows at 100 frames per second, 12-order MFCCs and a log power, and their deltas and accelerations (that is, 39 dimensions in total) as the feature parameters. The recognizer also uses 16-state 20-Gaussian mixture hidden Markov models (HMMs) as a pattern classifier for each digit, and three-state 36-Gaussian mixture HMMs for silence (nonspeech segment). In this experiment, the HMMs were trained using HMM Toolkit (HTK) version 3.1 developed by the Speech Vision and Robotics Research Group of Cambridge University Engineering Department. In accordance with AURORA-2J, the performance of a speech recognizer should be measured in terms of average word accuracies between SNRs of 0 to 20 dB. This paper mainly employs this criterion. The average word accuracies achieved by the baseline MFCC-based method were 46.17% with clean-speech training and 85.93% with multicondition-speech training.

B. The effect of the number of feature parameters

We first measured the robustness by comparing the word accuracies obtained with SPADE and the MFCC. SPADE used 24-channel gammatone filter banks as bandpass filter banks, 25-ms rectangular windows at 100 frames per second, and 12-order coefficients for each feature, giving 24 coefficients in total. A log power was also used as a feature parameter. In addition, we employed the deltas and accelerations of these features; therefore the feature parameters had a total of 75 dimensions. The pattern classifier consisted of 16-state 24-Gaussian mixture HMMs for each digit. The HMMs were trained using HTK version 3.1. The evaluation

TABLE II. Experimental evaluation results obtained with AURORA-2J. This table shows the word accuracies achieved by the baseline MFCC, 12- and 24-order MFCCs (12-/24-order MFCC), SPDC, SNDC, GTCC, and SPADE for each test set and each type of training.

Feature	Test set A	B	C	Overall
Clean-speech training (% Accuracy)				
Baseline MFCC	46.51	43.98	49.90	46.17
12-order MFCC	46.89	42.33	51.63	46.02
24-order MFCC	45.45	43.20	46.71	44.80
SPDC	46.63	41.00	48.80	44.81
SNDC	39.51	37.25	42.97	39.30
GTCC	43.75	38.25	49.52	42.70
SPADE	**54.45**	**52.34**	**58.01**	**54.32**
Multicondition-speech training (% Accuracy)				
Baseline MFCC	91.53	80.39	85.83	85.93
12-order MFCC	**92.04**	79.33	**85.95**	85.74
24-order MFCC	91.19	82.95	84.83	**86.62**
SPDC	90.64	80.99	84.80	85.61
SNDC	85.43	73.64	74.28	78.48
GTCC	**92.04**	79.71	85.27	85.76
SPADE	89.92	**83.30**	84.94	86.28

category established in AURORA-2J was 3 (this means the HMM topology is changed), that is, the feature extraction process and the number of Gaussian mixtures for the HMMs were changed from the baseline result. To measure the effect of the difference in the number of Gaussian mixtures for the HMMs, we evaluated the performance of a 12-order MFCC using 16-state 24-Gaussian mixture HMMs for each digit. In addition, to measure the effect of the number of feature dimensions, we evaluated the performance of a 24-order MFCC (75 dimensions in total) using 16-state 24-Gaussian mixture HMMs. In this case, the number of feature dimensions is indeed the same for the MFCC and SPADE; however, it should be noted that the MFCC includes more detailed information about the spectral shape because it uses higher order coefficients than SPADE.

Table II compares the average word accuracies achieved by the AURORA-2J baseline MFCC, the 12-order MFCC with 24-Gaussian mixture HMMs, the 24-order MFCC with 24-Gaussian mixture HMMs, and SPADE. Figure 5 shows the mean word accuracies and their error bars obtained from a 12-order MFCC with 24-Gaussian mixture HMMs, a 24-order MFCC with 24-Gaussian mixture HMMs, and SPADE with clean-speech training. The error bars were calculated from the word accuracies for each noise and channel condition. The performance of the 12-order MFCC with 24-Gaussian mixture HMMs rather than 20-Gaussian mixture HMMs (baseline) did not improve with either type of training with respect to the average word accuracies. The performance of the 24-order MFCC was slightly worse than the 12-order MFCC. SPADE performs significantly better than the 12- and 24-order MFCC with clean-speech training. However, with multicondition-speech training, SPADE was only able to achieve a slight improvement for test set B (open noise condition). In addition, by comparison with the

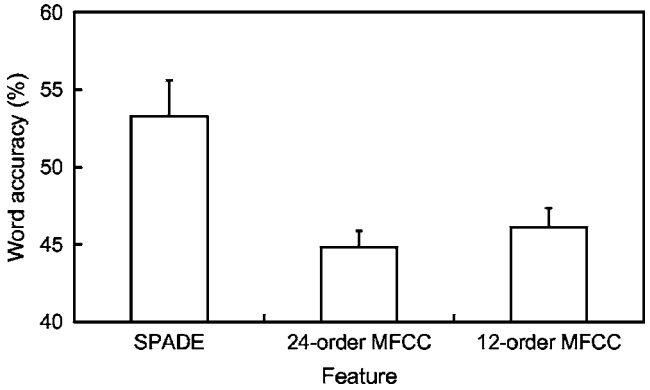


FIG. 5. Experimental evaluation results obtained with AURORA-2J under clean-speech training. The word accuracies and their error bars achieved by a 12-order MFCC with 24-Gaussian mixture HMMs, a 24-order MFCC with 24-Gaussian mixture HMMs, and SPADE for each test set and each type of training.

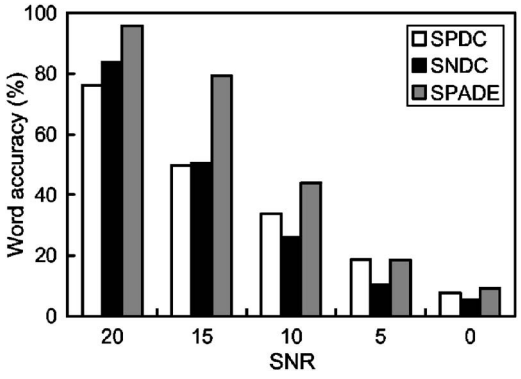


FIG. 6. Experimental evaluation results obtained with AURORA-2J under clean-speech training. The word accuracies achieved by SPDC, SNDC, and SPADE at each SNR in the presence of railway station sound.

24-order MFCC, SPADE could not improve the average word accuracy with multicondition-speech training.

C. The effect of using both periodic and nonperiodic features

To evaluate the effect of using both periodic and nonperiodic features, we evaluated the performance when using coefficients based on either periodic or nonperiodic features calculated in SPADE. Henceforth, the use of coefficients based only on periodic features is called "SPDC," and the use of coefficients based only on nonperiodic features is called "SNDC." Both SADC and SNDC also use 24-channel gammatone filter banks, 25-ms rectangular windows at 100 frames per second, 12-order coefficients and a log power, and their deltas and accelerations (39 dimensions in total). The pattern classifier consisted of 16-state 24-Gaussian mixture HMMs for each digit.

Table II shows the average word accuracies achieved by SPDC, SNDC, and SPADE. There was no improvement in the SPDC performance with respect to the average word accuracies with either type of training. The SNDC performance also deteriorated with respect to the average word accuracies with both types of training. Figure 6 shows the word accuracies for SPDC, SNDC, and SPADE at each SNR in the presence of railway station sound. As shown in Fig. 6, SNDC achieves higher/lower accuracies than SPDC when the SNRs are high/low.

D. The effect of filter banks

We evaluated the ASR performance when using the gammatone filter banks without the periodicity and nonperiodicity decomposition in order to evaluate their effects separately. The difference between this method and the MFCC method lies in the bandpass filter banks, namely, this method uses gammatone filter banks rather than mel-scale filter banks. Henceforth, this method is called "GTCC" (cepstral coefficients using gammatone filter banks). By comparing GTCC with SPADE we can measure the effect of decomposition. The feature parameter conditions for GTCC are the same as those for SPDC or SNDC for the experiments described above. In addition, to evaluate the effect of the number of filter banks, we compared the performance of SPADE with 24-, 36- and 48-channel gammatone filter banks under clean-speech training conditions.

Table II shows the average word accuracies achieved by GTCC and SPADE. There was no improvement in the GTCC performance with respect to the average word accuracy with either type of training. The GTCC performance only improved for test set A with multicondition-speech training. SPADE performed better than GTCC. Figure 7 shows the performance obtained from SPADE with each of the above numbers of channels for the filter banks. There is no significant difference between the average word accuracies obtained by SPADE with 24-, 36-, and 48-channel gammatone filter banks.

E. The effect of cepstral mean normalization

Cepstral mean normalization (CMN; Atal, 1974) is a technique that is widely used for reducing the effects of differences in channel characteristics. It is expected that these effects will also be reduced by applying CMN to SPADE. This section presents a method for applying CMN to SPADE and examines its effectiveness with an evaluation experiment.

The method we used for applying CMN to SPADE is as follows. The speech feature representation of SPADE is simi-

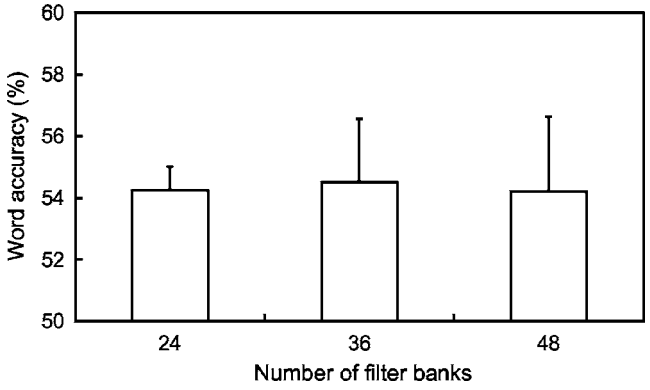


FIG. 7. Experimental evaluation results obtained with AURORA-2J. The word accuracies achieved by SPADE with 24-, 36-, and 48-channel gammatone filter banks.

TABLE III. Experimental evaluation results obtained with AURORA-2J. This table shows the word accuracies achieved by the 12- and 24-order MFCCs with CMN, and SPADE with CMN for each test set and each type of training.

Feature	Test Set A	B	C	Overall
Clean-speech training (% Accuracy)				
12-order MFCC+CMN	45.63	46.23	50.99	46.94
24-order MFCC+CMN	51.32	52.24	55.94	52.61
SPADE+CMN	**58.02**	**58.68**	**61.98**	**59.08**
Multicondition-speech training (% Accuracy)				
12-order MFCC+CMN	**92.31**	86.65	**91.20**	89.82
24-order MFCC+CMN	91.25	87.78	90.08	89.63
SPADE+CMN	91.72	**87.87**	90.55	**89.94**

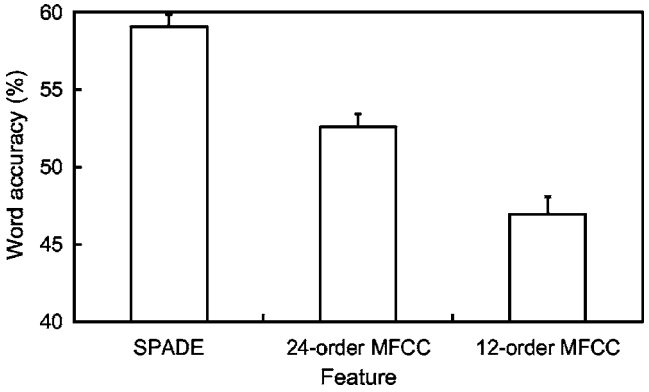


FIG. 8. Experimental evaluation results obtained with AURORA-2J under clean-speech training. The word accuracies and their error bars achieved by a 12-order MFCC with CMN, a 24-order MFCCs with CMN, and SPADE with CMN for each test set and each type of training.

lar to that of MFCC except for periodicity/nonperiodicity decomposition, therefore CMN can be applied to SPADE in the same way that it is applied to MFCC. The coefficient of speech features calculated by SPADE is given as $C(n,t)$, where n is the index number of dimensions and t is the discrete time stamp of the temporal frame. The mean cepstral coefficient $M(n)$ is calculated as below, where T is the number of frames in one speech segment, e.g., one utterance:

$$M(n) = \frac{1}{T}\sum_{t=1}^{T} C(n,t).$$

The normalized coefficients $N(n,t)$ are calculated as below:

$$N(n,t) = C(n,t) - M(n).$$

$N(n,t)$ is calculated for all n and t. Both the training and test data are normalized.

We experimentally evaluated the effect of applying CMN to SPADE with AURORA-2J under both clean- and multicondition-speech training conditions. Table III and Fig. 8 show the results, which indicate that CMN can improve the performance, especially for test set C with multicondition-speech training. It should be noted that applying CMN to SPADE improves its performance under both open (test set C) and closed channel conditions (test sets A and B). We conducted an experiment to compare the above results with those obtained with the 12- and 24-order MFCCs using CMN. Table III shows the outcome. The results indicate that SPADE with CMN achieves better average word accuracies than the 12- and 24-order MFCCs under both conditions. With clean-speech training, SPADE with CMN performs significantly better than MFCCs with CMN as shown in Fig. 8.

F. Discussion

By comparing Tables I–III, we can conclude that the following:

(i) SPADE is effective in improving ASR performance under open noise conditions (i.e., all test sets with clean-speech training, and test set B with multicondition-speech training).

(ii) SPADE does not improve ASR performance under closed noise conditions (i.e., test sets A and C with multicondition-speech training).

(iii) SPADE can improve the ASR performance regardless of whether the channel distortion is added to the signal with clean-speech training.

The reason for the improved SPADE performance under open noise conditions can be explained as follows. It is possible that the periodic features of the training data include largely information about voiced speech in the training data. When the noise in the test and training data sets differs, the difference between the nonperiodic features of the training and test data becomes large, whereas the difference between the periodic features remains small. With the MFCC, the difference between the interferers in the training and test data directly distorted all the parameters and thus degraded the word accuracies.

The limitation of SPADE is that it does not improve the ASR performance under closed noise conditions. Under such conditions, the MFCCs of the training and test data are contaminated by the same kinds of noise, therefore the difference between the training and test data distributions becomes small. This effect substitutes for the advantageous feature of SPADE and improves the ASR performance even without periodicity/nonperiodicity decomposition. As a result, SPADE's performance did not improve with multicondition-speech training.

With clean-speech training, SPADE improved the average word accuracy without the need for any noise reduction techniques. These results suggest that the performance of this method will improve further when it is combined with certain noise reduction methods. A comparison of the baseline MFCC, GTCC, and SPADE results indicates that the improvements were not caused by the difference in the bandpass filter banks, but by the decomposition of the periodicity and nonperiodicity. A comparison of the results that we obtained with SPADE, the 12-order MFCC, and the 24-order MFCC with 24-Gaussian mixture HMMs confirmed that the improvements were not due to the number of Gaussian mixtures or the number of parameters. The fact that the 24-order MFCC result was slightly worse than the 12-order MFCC

result can be attributed to the use of too large a number of feature parameters. In practice, when the number of features is too large, the performance is sometimes worse than with a small number of features. Nonetheless, SPADE performs better than the 24-order MFCC. This result indicates that the sound representation provided by SPADE contains more information than the representation provided by the MFCC. A comparison of the SPADE and SPDC/SNDC results shows that there is no advantage in employing either periodic or nonperiodic features for ASR in such noisy environments where there are various environmental interferer sounds at various SNRs. This result provides evidence for the effectiveness of exploiting both features. These experimental results confirm that the feature representation provided by SPADE is effective for robust ASR in real noisy environments under open noise conditions.

In addition, applying CMN to SPADE improves the performance, and effectively absorbs the influence of the differences between the channels of the training and test data. CMN also improves the performance under closed channel conditions. SPADE with CMN performs better than MFCCs with CMN under open noise conditions.

V. CONCLUSION

This paper proposed a speech feature extraction method that utilizes periodic and nonperiodic features for robust ASR. The method, which is called "SPADE," uses gammatone filter banks and comb filters to divide speech signals into the above two features. An evaluation experiment with the AURORA-2J noisy continuous digit speech database showed that SPADE provides better performance in the presence of noise than the conventional MFCC-based feature extraction method, especially under open noise conditions, and confirmed the effectiveness of the subband-based periodicity/nonperiodicity decomposition and the utilization of both of the features. The results indicate that such an enhancement in sound representation can improve the robustness of an ASR system.

On the other hand, a possible limitation of SPADE is that it cannot improve the performance under closed noise conditions. Applying CMN to SPADE can absorb the channel distortion effect and improve the performance. An evaluation experiment using AURORA-2J showed that the performance achieved by SPADE with CMN was better than that achieved by MFCCs with CMN.

ACKNOWLEDGMENTS

The authors thank Dr. Alain de Cheveigné (CNRS, France), Professor Kiyoaki Aikawa (Tokyo University of Technology), Professor Yoshinao Shiraki (Shonan Institute of Technology), and three anonymous reviewers for their fruitful criticisms and comments about this research. The authors also thank members of the Spoken Dialogue System Group at NTT Cyber Space Laboratories for their useful suggestions as regards the evaluation experiments in relation to the commercial use of ASR for real applications. Lastly, the authors thank Professor Biing-Huang Juang (Georgia Institute of Technology), Professor Fumitada Itakura (Meijo University), and Dr. Erik McDermott and members of the Signal Processing Research Group at NTT Communication Science Laboratories for fruitful discussions about this research. This research employed the noisy speech recognition evaluation environment AURORA-2J produced and distributed by the IPSJ SIG-SLP Noisy Speech Recognition Working Group. AURORA-2J utilizes HMM Toolkit (HTK) developed by the Speech Vision and Robotics Research Group of Cambridge University Engineering Department.

Aikawa, K., Singer, H., Kawahara, H., and Tohkura, Y. (**1996**). "Cepstral representation of speech motivated by time-frequency masking: An application to speech recognition," J. Acoust. Soc. Am. **100**, 603–614.

Ali, A. M., Spiegel, J. V., and Mueller, P. (**2002**). "Robust auditory-based speech processing using the average localized synchrony detection," IEEE Trans. Speech Audio Process. **10**, 279–292.

Atal, B. S. (**1974**). "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Am. **55**, 1304–1312.

Bernard, A., Gong, Y., and Cui, X. (**2004**). "Can back-ends be more robust than front-ends? Investigation over the Aurora-2 database," *Proc. of the 29th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. **1**, pp. 1025–1028.

Bogert, B., Healy, M., and Tukey, J. (**1963**). "The quefrency analysis of time series for echoes," in *Proc. Symp. on Time Series Analysis*, edited by M. Rosenblatt (Wiley, New York), pp. 209–243.

Boll, S. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**, 113–210.

Darwin, C. J., and Carlyon, R.P. (**1995**). "Auditory grouping," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 387–424.

Davis, S. B., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-28**, 357–366.

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 Veth, J., Cranen, B., and Boves, L. (**2001**). "Acoustic features and distance measure to reduce vulnerability of ASR performance due to the presence of a communication channel and/or background noise," in *Robustness in Language and Speech Technology*, edited by J.-C. Junqua and G. van Noord (Kluwer Academic, Dordrecht, Netherlands), pp. 9–45.

Gales, M., and Young, S. (**1993**). "HMM recognition in noise using parallel model combination," *Proc. of the 3rd European Conference on Speech Communication and Technology (Eurospeech)*, pp. 837–840.

Gao, Y., Huang, T., Chen, S., and Haton, J.-P. (**1992**). "Auditory model based speech processing," *Proc. of the 2nd International Conference on Spoken Language Processing (ICSLP)*, pp. 73–76.

Ghitza, O. (**1988**). "Temporal non-place information in the auditory nerve firing patterns as a front-end for speech recognition in a noisy environment," J. Phonetics **16**, 109–124.

Ghitza, O. (**1994**). "Auditory models and human performance in tasks related to speech coding and speech recognition," IEEE Trans. Speech Audio Process. **2**, 115–132.

Gong, Y. (**1995**). "Speech recognition in noisy environments: A survey," Speech Commun. **16**, 261–291.

Greenberg, S. (**2004**). "Speech processing in the auditory system: An overview," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).

Hermansky, H. (**1990**). "Perceptual Linear Predictive (PLP) analysis of speech," J. Acoust. Soc. Am. **87**, 1738–1752.

Hermansky, H., Hanson, B., and Wakita, H. (**1985**). "Low-dimensional representation of vowels based on all-pole modeling in the psychophysical domain," Speech Commun. **4**, 181–187.

Hess, W. (**1983**). "Short-term analysis pitch determination," in *Pitch Determination of Speech Signals* (Springer-Verlag, New York).

Hirsh, H. G., and Pearce, D. (**2000**). "The AURORA experimental frame-

work for the performance evaluation of speech recognition systems under noisy conditions," *Proc. of the ISCA Tutorial and Research Workshop on Automatic Speech Recognition (ISCA ITRW ASR)*, pp. 181–188.
Huang, X., Acero, A., and Hon, H. (**2001**). "Speech signal representation," in *Spoken Language Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Ishizuka, K., and Aikawa, K. (**2002**). "Effect of F0 fluctuation and amplitude modulation of natural vowels on vowel identification in noisy environments," *Proc. of the 7th International Conference on Spoken Language Processing (ICSLP)*, pp. 1633–1636.
Itakura, F. (**1975**). "Minimum prediction residual principle applied to speech recognition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-23**, 67–72.
Jackson, P. J. B., and Shadle, C. H. (**2001**). "Pitch-scaled estimation of simultaneous voiced and turbulence-noise components in speech," IEEE Trans. Speech Audio Process. **9**, 713–726.
Jackson, P. J. B., Moreno, D. M., Russell, M. J., and Hernando, J. (**2003**). "Covariation and weighting of harmonically decomposed streams for ASR," *Proc. of 8th European Conference on Speech Communication and Technology (Eurospeech)*, pp. 2321–2324.
Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.
Kajita, S., and Itakura, F. (**1995**). "Robust feature extraction using SBCOR analysis," *Proc. of the 20th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 421–424.
Kim, D. S., Lee, S. Y., and Kil, R. M. (**1999**). "Auditory processing of speech signals for robust speech recognition in real-world noisy environments," IEEE Trans. Speech Audio Process. **7**, 55–69.
Koo, B., Gibson, J., and Gray, S. (**1989**). "Filtering of colored noise for speech enhancement and coding," *Proc. of the 14th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 349–352.
Lee, C.-H., Lin, C.-H., and Juang, B.-H. (**1991**). "A study on speaker adaptation of the parameters of continuous density hidden Markov models," IEEE Trans. Signal Process. **39**, 806–814.
Li, Q., Soong, F. K., and Siohan, O. (**2001**). "An auditory system-based feature for robust speech recognition," *Proc. of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pp. 619–621.
Lim, J., and Oppenheim, A. (**1978**). "All-pole modeling of degraded speech," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-26**, 197–210.
Lockwood, P., and Boudy, J. (**1992**). "Experiments with a Nonlinear Spectral Subtractor (NSS), hidden Markov models and the projection, for robust speech recognition in cars," Speech Commun. **11**, 215–228.
Minami, Y., and Furui, S. (**1995**). "A maximum likelihood procedure for a universal adaptation method," *Proc. of the 20th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 129–132.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formula for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Nakamura, S., Yamamoto, K., Takeda, K., Kuroiwa, S., Kitaoka, N., Yamada, T., Mizumachi, M., Nishiura, T., Fujimoto, M., Saso, A., and Endo, T. (**2003**). "Data collection and evaluation of AURORA-2 Japanese corpus," *Proc. of the 8th IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 619–623.
Nakamura, S., Takeda, K., Yamamoto, K., Yamada, T., Kuroiwa, S., Kitaoka, N., Nishiura, T., Sasou, A., Mizumachi, M., Miyajima, C., Fujimoto, M., and Endo, T. (**2005**). "AURORA-2J: An evaluation framework for Japanese noisy speech recognition," IEICE Trans. Inf. Syst. **E88-D**, 535–544.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 123–177.
Pearce, D., and Hirsh, H. G. (**2000**). "The AURORA experimental framework for the performance evaluation of speech recognition systems under noise conditions," *Proc. of the 6th International Conference on Spoken Language Processing (ICSLP)*, Vol. **4**, pp. 29–32.
Rabiner, L. R. (**1977**). "On the use of autocorrelation analysis for pitch detection," IEEE Trans. Acoust., Speech, Signal Process. **25**, 24–33.
Rose, J. E., Hind, J. E., Anderson, D. J., and Brugge, J. F. (**1971**). "Some effects of the stimulus intensity on response of auditory nerve fibers in the squirrel monkey," J. Neurophysiol. **34**, 685–699.
Seneff, S. (**1988**). "A joint synchrony/mean-rate model of auditory speech processing," J. Phonetics **16**, 55–76.
Sim, B. L., Tong, Y. C., Chang, J. S., and Tan, C. T. (**1998**). "A parametric formulation of the generalized spectral subtraction method," IEEE Trans. Speech Audio Process. **6**, 328–337.
Singh, R., Stern, R. M., and Raj, B. (**2002**). "Signal and feature compensation methods for robust speech recognition," in *Noise Reduction in Speech Applications*, edited by G. M. Davis (CRC, Boca Raton, FL), pp. 219–244.
Varga, A., and Moore, R. (**1990**). "Hidden Markov model decomposition of speech and noise," *Proc. of the 15th International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, pp. 845–848.
Vaseghi, S., and Milner, B. (**1993**). "Noisy speech recognition based on HMMs, Wiener filters and re-evaluation of most likely candidates," *Proc. of the 18th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 437–440.

Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech

Johan Sundberg and Maria Nordenberg
KTH Music Acoustics, Department of Speech Music Hearing, KTH, and School of Computer Science and Communication, KTH, Stockholm, SE-100 44 Sweden

(Received 12 January 2006; revised 26 April 2006; accepted 5 May 2006)

The overall slope of long-term-average spectrum (LTAS) decreases if vocal loudness increases. Therefore, changes of vocal loudness also affects the α measure, defined as the ratio of spectrum intensity above and below 1000 Hz. The effect on α of loudness variation was analyzed in 15 male and 16 female voices reading a text at different degrees of vocal loudness. The mean range of equivalent sound level (L_{eq}) amounted to about 28 dB and the mean range of α to 19.0 and 11.7 dB for the female and male subjects. The L_{eq} vs. α relationship could be approximated with a quadratic function, or by a linear equation, if softest phonation was excluded. Using such equations α was computed for all values of L_{eq} observed for each subject and compared with observed values. The maximum and the mean absolute errors were 2.4 dB and between 0.1 and 0.6 dB. When softest phonation was disregarded and linear equations were used, the maximum error was less than 2 dB and the mean absolute errors were between 0.2 and 0.7 dB. The strong correlation between L_{eq} and α indicates that for a voice L_{eq} can be used for predicting α.
 [DOI: 10.1121/1.2208451]

PACS number(s): 43.72.Ar, 43.70.Aj [BHS] Pages: 453–457

I. INTRODUCTION

Long-term average spectrum (LTAS) is a classical method in voice research. It provides an overview of the mean spectral characteristics of a voice. It also has the advantage of being an analysis that is quick and easy to perform, and it is part of several voice analysis programs.

LTAS analysis was explored by Frøkjaer-Jensen and Prytz (1976). In this study they also examined voice quality changes associated with voice therapy of patients suffering from recurrent nerve paralysis. They proposed the now commonly used LTAS measure α of spectral balance, defined as the ratio between the sound energy above and below 1000 Hz:

$$\alpha = I_{HF}/I_{LF}, \quad (1)$$

where I_{HF} is the intensity for frequencies >1000 Hz and I_{LF} is the intensity for frequencies ≤ 1000 Hz. Thus, α, often expressed in dB, increases if the high-frequency content of a voice increases.

As illustrated in Fig. 1, an increase of vocal loudness affects an LTAS curve in such a way that the levels at frequencies between about 1500 and 3000 Hz increase more than the levels at lower frequencies (Nordenberg and Sundberg, 2004). Therefore, α can be expected to vary with vocal loudness.

Frøkjaer-Jensen and Prytz (1976) observed this type of variation in α as a function of instantaneous overall SPL during 20 s of running speech, as shown Fig. 2. They also found that after therapy α increased by an average of 3 or 4 dB. The SPL variation of the healthy voices amounted to 25 dB or more, while the variation for some pathological voices varied by no more than about 10 dB. The concomitant α variation was about 30 dB and slightly narrower in the pathological voices.

The balance between high- and low-frequency partials is also affected by type of phonation. Thus, Kitzing (1986) derived α data from LTAS analyses of ten healthy voices faking breathy, pressed, soft, and normal/sonorous voice qualities. He found that the α belonged to a set of potent criteria for differentiating voice quality.

Löfqvist (1986) performed LTAS analysis on healthy and pathological voices and found that the α ratio, although robust, failed to discriminate the two types of voices. Kitzing and Åkerlund (1993) analyzed LTAS of dysphonic voices before and after therapy and found a small effect on the α ratio. They also pointed out the need to analyze more thoroughly the influence of voice intensity on the LTAS shape.

The aim of the present investigation was to analyze how the α measure is affected by vocal loudness variation.

II. METHOD

Acoustic signals were recorded from 15 male (mean age =29 years, range [23,35] years) and 16 female (mean age=28 years, range [21,40], years) speakers. All speakers reported that they had healthy voices at the time of the recording. Each speaker was asked to read the same Swedish text repeatedly. This took at least 40 s, which is sufficient to produce an LTAS that is independent of the text (Fritzell *et al.*, 1974). The subjects read the text with party noise presented over headphones at five different sound levels (see Table I). To check the subjects' consistency, one of these sound levels, the one at 85 dB, was presented twice. The subjects were

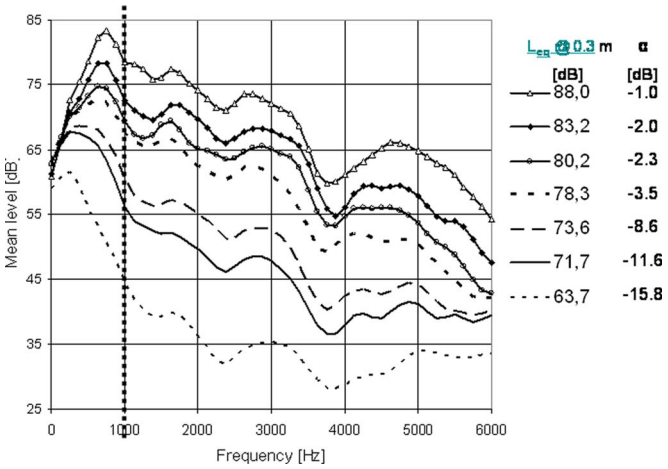


FIG. 1. LTAS curves observed when a female subject read the same text at different degrees of vocal loudness.

asked to make their voices heard in the presence of this noise. Then, they read the same text without headphones as softly as possible without whispering.

The audio signal was picked up at 0.3 m distance by a TCM 110 AV-JEFE head-mounted electret microphone and recorded on two channels of a Sony TCD-D10 DAT recorder. The amplification of the channels was adjusted to differ by 20 dB in order to secure a good signal-to-noise ratio also at extreme degrees of vocal loudness. For each subject the microphone was calibrated by holding it next to a sound level meter (B&K, type 2215, precision sound level meter) and recording, on both channels, a sustained vowel sound produced by the experimenter at two levels. The sound pressure levels observed on the sound level meter were announced on the tape. The same microphone gain was then used for the entire recording.

The recordings were digitized and transferred to computer files and analyzed using the Soundswell Signal Workstation 4.0 (HiTech Development, Solna, Sweden). The equivalent sound level (L_{eq}), defined as the logarithm of sound energy averaged over time, was measured by means of the Histogram module. Long-term-average spectra (LTAS)

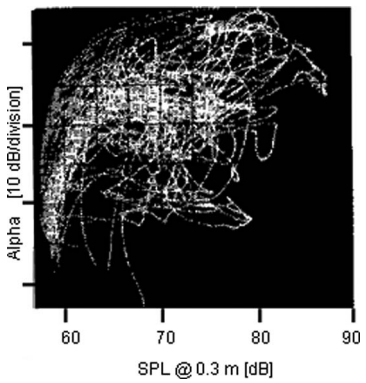


FIG. 2. Example of the variation of α as function of instantaneous overall SPL in running speech according to Frøkjaer-Jensen and Prytz (1976). The y axis, originally inverted by mistake (Frøkjaer-Jensen, personal communication), has been corrected by turning their graph upside-down.

TABLE I. Order and L_{eq} of the party noise presented over headphones to the subjects during the recording.

Reading no.	L_{eq} of party noise (dB)
1	85
2	95
3	85
4	81
5	88
6	77
7	No added noise

were obtained from the Line Spectrum module, using a Hanning window, 400 Hz analysis bandwidth and a 0–6000-Hz frequency range.

Each LTAS was copied into an Excel file and each of its level values was transformed to linear intensity amplitude. Then all intensity amplitudes for frequencies ≤ 1000 Hz were summed. Likewise, all intensity amplitudes in the frequency range 1000 Hz $< f \leq 6000$ Hz were summed. These sums were expressed in dB and the difference between the sum for the high frequency and that of the low frequency, thus corresponding to the α ratio expressed in dB, was determined.

Voiceless consonants were included in the LTAS analysis and this increases the LTAS level above about 5000. The effect of voiceless consonants on the α ratio was examined for some randomly selected spectra and was found to be less than 0.1 dB.

III. RESULTS

On average, an increase of the party noise level by 1 dB caused the female and male subjects to increase their L_{eq} by 0.81 and 0.76 dB, respectively (see Fig. 3). The subjects' max L_{eq} varied between 100.3 and 81.8 dB for the females and 100.1 and 83.3 dB for the males. The L_{eq} difference between loudest and softest reading was on average 28.1 dB (SD 5.2 dB) and 28.4 dB (SD 4.7 dB) for the female and the male subjects, respectively. The corresponding variation of α was 19.0 dB (SD 3.6 dB) and 11.7 dB (SD 2.3 dB).

All subjects were exposed twice to the party noise L_{eq} of 85 dB. The second time the female and male subjects produced an L_{eq} which was, on average, 2.3 dB (SD 1.6 dB) and

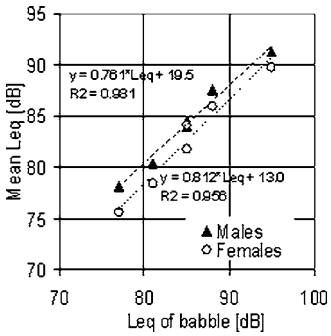


FIG. 3. Variation of female and male subjects' mean L_{eq} (open and filled circles) in response to the indicated variation of the party noise level.

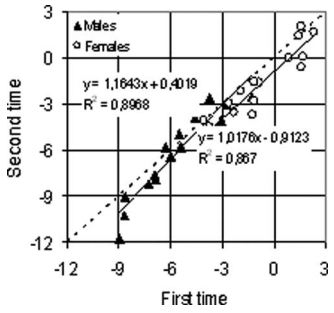


FIG. 4. Replicability of the α measure observed when the subjects read the same text two times with the same level of the party noise which they heard in headphones. The dashed line represents the case of no difference.

0.5 dB (SD 1.6 dB) lower than the first time. The second time the female and male subjects' α values were, on average, 0.9 dB (SD 0.9 dB) and 0.5 dB (SD 1.0 dB) lower, respectively (see Fig. 4).

Figure 5 shows typical examples of the relationship between α and L_{eq} for two female and two male subjects. For subjects F16 and M2 the relationship could be approximated by a linear function. The slope and intercepts, however, differed substantially between these subjects. The data pertaining to the other two subjects F10 and M14 could be better approximated by a quadratic function, mainly due to very low values for these subjects' lowest L_{eq}. Also in this case, the equations for these two subjects differed considerably.

Quadratic approximations were computed for each subject. As shown in Table II the intersubject variation of the factors and the intercept of the equations were considerable for both the female and the male subjects. The coefficient of determination $R2$, on the other hand, was quite high.

Using the quadratic approximation for each subject, α values were calculated for each of the observed L_{eq} values and the results were compared with the observed α values. The maximum and average error of these predictions, i.e., differences between observed and predicted values, are shown in Fig. 6. The greatest error amounted to 2.4 dB. The average absolute error was generally less than 0.5 dB.

As illustrated in Fig. 5, the need for a quadratic approximation was often caused by the lowest L_{eq} value, produced when the subjects were instructed to speak as softly as they could. For degrees of vocal loudness typically occurring in speech, the data points could be approximated by linear equations. The mean slope, intercept, and determination coefficient for these equations are listed in Table III. On average the female and male subjects increased their α by 0.44 and 0.33 dB for a 1-dB increase of L_{eq}. Using the linear approximations for each subject α values were calculated for each reading of each subject, however excluding the softest reading. The maximum and average error of these predictions are shown in Fig. 7. The greatest error was less than 2 dB. The average absolute error varied within the range of 0.3 to 0.7 dB for the female subjects and 0.2 to 0.4 dB for the male subjects.

TABLE II. Means (M) and standard deviations (SD) of A, B, and C and of the squared correlation ($R2$) for the quadratic approximation $\alpha=A(L_{eq})^2+B(L_{eq})+C$ of the α versus L_{eq} relationship for the female and the male subjects.

		A	B	C	R2
Female	M	−0.018	3.34	−157.4	0.993
Female	SD	0.007	1.06	43.2	0.013
Male	M	−0.005	1.23	−73.0	0.979
Male	SD	0.010	1.60	61.9	0.026

IV. DISCUSSION

The effect on spectral balance of vocal loudness variation is well documented in the literature (Klatt and Klatt, 1990). A linear relationship between the level of the first formant, or the overall SPL of a vowel, and the level of the singer's formant was observed by Cleveland and Sundberg (1985) and quantified by Sjølander and Sundberg (2004) in a study of vowels sung by five professional baritone singers. In the latter study the levels of the singer's formant and the first formant were compared. The results showed that on average the difference between them decreased by 0.76 dB per dB increase of SPL. In the present investigation of running speech we found that α increased by, on average, 0.33 dB per dB increase of L_{eq} for the male subjects.

We found that the relationship between α and L_{eq} can be approximated by a linear or quadratic function. The same does not appear to be true for the relationship between α and SPL. For instance, the results obtained by Frøkjaer-Jensen and Prytz (1976) for running speech showed that α increased quickly for low SPL values and very slowly for high SPL values. Ternström and collaborators (2006) and Ternström (1993) analyzed what they called the spectral balance as a function of the SPL of different vowels produced in running speech at widely varying degrees of vocal loudness. Their measure of spectral balance, similar to α, was defined as the level difference between the 2–6-kHz range and the 0–1-kHz

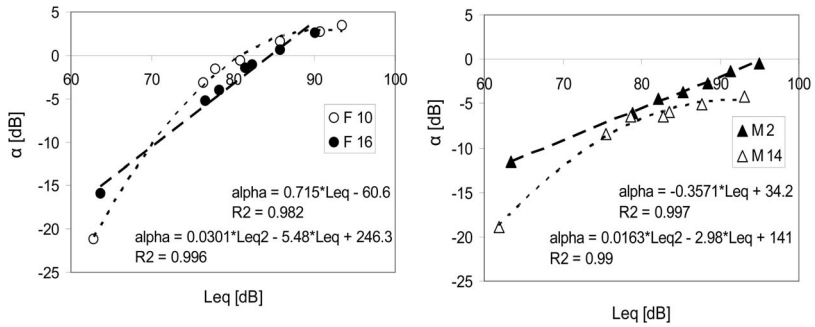


FIG. 5. Typical examples of the relationship between α and L_{eq} for female subjects F 10 and F 16 (left panel, open and filled symbols) and male subjects M 2 and M 14 (right panel, filled and open symbols). The curves show the best linear or polynomial approximations of the data.

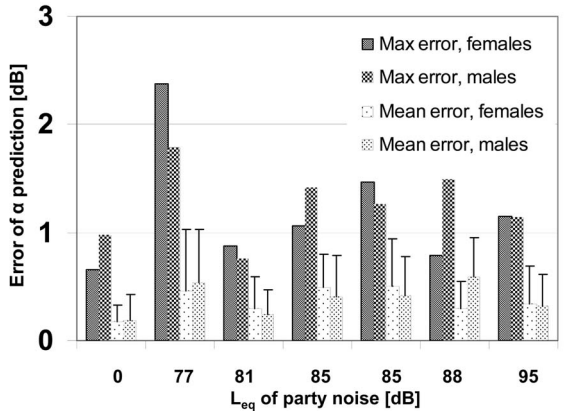


FIG. 6. Maximum error and average error observed for the female and male subjects when their α values were calculated from the L_{eq} values, approximating the individual subject's α vs. L_{eq} relationships by a quadratic function. The bars represent one SD.

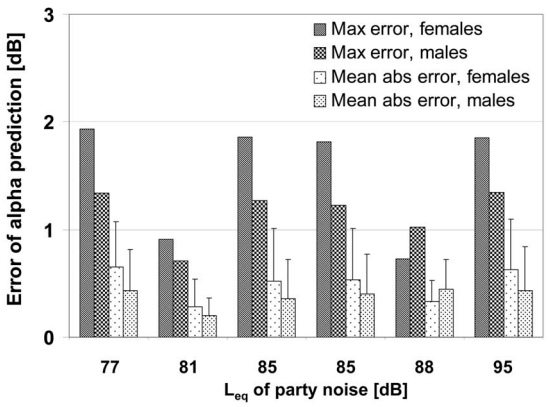


FIG. 7. Maximum error and average error observed for the female and male subjects when their α values were calculated from the L_{eq} values. In these calculations their α vs. L_{eq} relationships were approximated by a linear equation. In constructing these equations each subject's softest reading was excluded. The bars represent one SD.

range of the vowel spectrum. They found that, for a given vowel, the balance tended to remain rather constant at extreme degrees of vocal loudness while it increased for intermediate degrees. The reason for this discrepancy may be that, while L_{eq} is an average over time, SPL is computed over a short time window. Hence, the SPL of speech sounds is highly affected not only by vocal loudness but also by the frequency distance between the strongest spectrum partial and the first formant (Gramming and Sundberg, 1988; Titze, 1994). Moreover, the intensity level above 1000 Hz is strongly influenced by the frequencies of formants 1 and 2. Also of relevance may be that our subjects were not instructed to use their maximum degree of vocal loudness.

Our results show that the effect of vocal loudness variation on α is predictable provided that L_{eq} rather than SPL is used for quantifying vocal loudness. By and large, α was found to increase by 0.44 and 0.33 dB per dB increase of L_{eq} for female and male subjects. However, the relationship showed a strong interindividual variation. Therefore reasonably accurate predictions probably need to be based on several recordings at well-separated degrees of vocal loudness. Then, the relationship between α and L_{eq} can be determined and approximated by a polynomial trendline. This trendline can be used for predicting α for new L_{eq} values. This procedure would allow for informative comparisons of voice characteristics, e.g., before and after treatment.

The high correlation between L_{eq} and α implies that most of the variation of α can be explained by variation in L_{eq}. This suggests that, conversely, α can be used for estimating L_{eq}. This could be valuable, since vocal loudness has a strong influence on voice source characteristics and hence on the radiated sound. However, as the intersubject variation of the relationship between L_{eq} and α was substantial, accurate prediction for a given voice would require that the prediction be based on the L_{eq}-to-α relationship for the individual voice analyzed.

TABLE III. Means (M) and standard deviations (SD) of the slope and intercept (Icpt) and of the squared correlation ($R2$) for the linear approximation of α as a function of L_{eq} for the female and the male subjects.

		Slope	Icpt	$R2$
Female	M	0.442	38.0	0.911
Female	SD	0.124	10.7	0.118
Male	M	0.326	33.3	0.856
Male	SD	0.141	12.1	0.144

V. CONCLUSIONS

The α measure, defined as the ratio of spectrum intensity in the frequency ranges $1000 < f \leq 6000$ Hz and $0 < f \leq 1000$ Hz, and frequently expressed in dB, is strongly influenced by variations of vocal loudness. Here we have shown that α is a function of the equivalent sound level L_{eq}. This function can be approximated by a second-order polynomial, or by a linear equation, if softest possible phonation is excluded. For a group of 16 females and 15 males α increased by 0.44 and 0.33 dB per dB increase of L_{eq}. The relationship between α and L_{eq} varied substantially between subjects, so prediction of α for given L_{eq} values must be based on observations of the individual voice. The maximum error of such predictions was found to be 2.4 dB in our subject group, while the mean absolute error varied between 0.2 to 0.6 dB. If the softest possible phonation was excluded, the α vs. L_{eq} relationship could be approximated by a linear function, producing prediction errors $\leq$1.9 and 1.3 dB for the female and male subjects, respectively, the mean absolute errors to being 0.7 and 0.4 dB, respectively. The results indicate that for an individual voice α can be predicted from L_{eq}, provided that this relationship is known for that voice.

ACKNOWLEDGMENTS

The recordings used in this investigation were made at the Department of Speech Music Hearing, KTH, for coauthor MN's thesis work in Logopedics at Lund University.

Cleveland, T., and Sundberg, J. (**1985**). "Acoustic analyses of three male voices of different quality," in SMAC 83. *Proceedings of the Stockholm Internat Music Acoustics Conf.*, Vol. 1, edited by A. Askenfelt, S. Felicetti, E. Jansson, and J. Sundberg, Stockholm: R. Sw. Acad. Music, **46**(1), 143–156.

Fritzell, B., Hallén, O., and Sundberg, J. (**1974**). "Evaluation of Teflon injection procedures for paralytic dysphonia," Folia Phoniatr. **26**, 414–421.
Frøkjaer-Jensen, B., and Prytz, S. (**1976**). "Registration of voice quality," Brüel & Kjaer Technical Review **3**, 3–17.
Gramming, P., and Sundberg, J. (**1988**). "Spectrum factors relevant to phonetogram measurement," J. Acoust. Soc. Am. **83**, 2352–2360.
Kitzing, P. (**1986**). "LTAS criteria pertinent to the measurement of voice quality," J. Phonetics **14**, 477–482.
Kitzing, P., and Åkerlund, L. (**1993**). "Long-time average spectrograms of dysphonic voices before and after therapy," Folia Phoniatr. **45**, 53–61.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.
Löfqvist, A. (**1986**). "The long-term-average spectrum as a tool in voice research," J. Phonetics **14**, 471–475.
Nordenberg, M., and Sundberg, J. (**2004**). "Effect on LTAS of vocal loudness variation," Logopedics Phoniatrics Vocology. **29**, 183–191.
Sjølander, P., and Sundberg, J. (**2004**). "Spectrum effects of subglottal pressure variation in professional baritone singers," J. Acoust. Soc. Am. **115**, 1270–1273.
Ternström, S. (**1993**). "Long-time average spectrum characteristics of different choirs in different rooms," Voice **2**, 55–77.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

Pitch-based monaural segregation of reverberant speech

Nicoleta Roman[a)]
Department of Computer Science and Engineering, The Ohio State University, Columbus, Ohio 43210

DeLiang Wang[b)]
Department of Computer Science and Engineering & Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210

(Received 13 April 2005; revised 20 January 2006; accepted 23 March 2006)

In everyday listening, both background noise and reverberation degrade the speech signal. Psychoacoustic evidence suggests that human speech perception under reverberant conditions relies mostly on monaural processing. While speech segregation based on periodicity has achieved considerable progress in handling additive noise, little research in monaural segregation has been devoted to reverberant scenarios. Reverberation smears the harmonic structure of speech signals, and our evaluations using a pitch-based segregation algorithm show that an increase in the room reverberation time causes degraded performance due to weakened periodicity in the target signal. We propose a two-stage monaural separation system that combines the inverse filtering of the room impulse response corresponding to target location and a pitch-based speech segregation method. As a result of the first stage, the harmonicity of a signal arriving from target direction is partially restored while signals arriving from other directions are further smeared, and this leads to improved segregation. A systematic evaluation of the system shows that the proposed system results in considerable signal-to-noise ratio gains across different conditions. Potential applications of this system include robust automatic speech recognition and hearing aid design.
 [DOI: 10.1121/1.2204590]

PACS number(s): 43.72.Dv [DOS] Pages: 458–469

I. INTRODUCTION

In a natural environment, a desired speech signal often occurs simultaneously with other interfering sounds such as echoes and background noise. While the human auditory system excels at speech segregation from such complex mixtures, simulating this perceptual ability computationally remains a great challenge. In this paper, we study the monaural separation of reverberant speech. Our monaural study is motivated by the following two considerations. First, an effective one-microphone solution to sound separation is highly desirable in many applications including automatic speech recognition and speaker recognition in real environments, audio information retrieval, and hearing prosthesis. Second, although binaural listening improves the intelligibility of target speech under anechoic conditions (Bronkhorst, 2000), this binaural advantage is largely diminished by reverberation (Plomp, 1976; Culling *et al.*, 2003); this underscores the dominant role of monaural hearing in realistic conditions.

Various techniques have been proposed for monaural speech enhancement including spectral subtraction (e.g., Martin, 2001), Kalman filtering (e.g., Ma *et al.*, 2004), subspace analysis (e.g., Ephraim and Trees, 1995), and autoregressive modeling (e.g., Balan *et al.*, 1999). However, these methods make strong assumptions about the interference and thus have difficulty in dealing with a general acoustic background. Another line of research is the blind separation of signals using independent component analysis (ICA). While standard ICA techniques perform well when the number of microphones is greater than or equal to the number of observed signals such techniques do not function in monaural conditions. Some recent sparse representations attempt to relax this assumption (e.g., Zibulevsky *et al.*, 2001). For example, by exploiting *a priori* sets of time-domain basis functions learned using ICA, Jang *et al.* (2003) attempted to separate two source signals from a single channel but the performance is limited.

Inspired by the human listening ability, research has been devoted to build speech separation systems that incorporate known principles of auditory perception. According to Bregman (1990), the auditory system performs sound separation by employing various cues including pitch, onset time, spectral continuity, and location in a process known as auditory scene analysis (ASA). This ASA account has inspired a series of computational ASA (CASA) systems that have significantly advanced the state-of-the-art performance in monaural separation (e.g., Weintraub, 1985; Cooke, 1993; Brown and Cooke, 1994; Wang and Brown, 1999; Hu and Wang, 2004) as well as in binaural separation (e.g., Roman *et al.*, 2003; Palomaki *et al.*, 2004). Generally, CASA systems follow two stages: segmentation (analysis) and grouping (synthesis). In segmentation, the acoustic input is decomposed into sensory segments, each of which originates from a single source. In grouping, the segments that likely come from the same source are put together. A recent overview of both monaural and binaural CASA approaches can be found in Brown and Wang (2005). Compared with speech enhance-

[a)]Electronic mail: roman.45@osu.edu
[b)]Electronic mail: dwang@cse.ohio-state.edu

 0001-4966/2006/120(1)/458/12/$22.50

ment techniques described above, CASA systems make few assumptions about the acoustic properties of the interference and the environment.

CASA research, however, has been largely limited to anechoic conditions, and few systems have been designed to operate on reverberant input. A notable exception is the binaural system proposed by Palomaki *et al.* (2004) which includes an inhibition mechanism that emphasizes the onset portions of the signal and groups them according to common location. Evaluations in reverberant conditions have also been reported for a series of two-microphone algorithms that combine pitch information with binaural cues or signal-processing techniques (Luo and Denbigh, 1994; Shamsoddini and Denbigh, 2001; Barros *et al.*, 2002).

At the core of many CASA systems is a time-frequency (T-F) mask. Specifically, the T-F units in the acoustic mixture are selectively weighted in order to enhance the desired signal. The weights can be binary or real (Srinivasan *et al.*, 2004). The binary T-F masks are motivated by the masking phenomenon in human audition, in which a weaker signal is masked by a stronger one in the same critical band (Moore, 2003). Additionally, from the speech segregation perspective, the notion of an *ideal binary mask* has been proposed as the computational goal of CASA (Wang, 2005). Such a mask can be constructed from *a priori* knowledge about target and interference; specifically a value of 1 in the mask indicates that the target is stronger than the interference and 0 indicates otherwise. Speech reconstructed from the ideal binary mask has been shown to be highly intelligible even when extracted from multisource mixtures and also to produce large improvements in robust speech recognition and human speech intelligibility (Cooke *et al.*, 2001; Roman *et al.*, 2003; Brungart *et al.*, 2006).

Perceptually, one of the most effective cues for speech segregation is the fundamental frequency (F0) (Darwin and Carlyon, 1995). Accordingly, much work has been devoted to build computational systems that exploit the F0 of a desired source to segregate its harmonics from the interference (for a review see Brown and Wang, 2005). In particular, the system proposed by Hu and Wang (2004) employs differential strategies to segregate resolved and unresolved harmonics. More specifically, periodicities detected in the response of a cochlear filterbank are used at low frequencies to segregate resolved harmonics. In the high-frequency range, however, the cochlear filters have wider bandwidths and a number of harmonics interact within the same filter, causing amplitude modulation (AM). In this case, their system exploits periodicities in the response envelope to group unresolved harmonics. In this paper, we propose a pitch-based speech segregation method that follows the same principles while simplifying the calculations required for extracting periodicities. The method shows good performance when tested with a variety of noise intrusions under anechoic conditions. However, when F0 varies with time in a reverberant environment, reflected waves with different F0s arrive simultaneously with the direct sound. This multipath situation causes smearing of harmonic structure (Darwin and Hukin, 2000). Due to weakened harmonicity, the performance of pitch-based segregation degrades in reverberant conditions.

One method for removing the reverberation effect is to pass the reverberant signal through a filter that inverts the reverberation process and hence reconstructs the original signal. However, because a typical room impulse response is not minimum phase, perfect one-microphone reconstruction requires a noncausal infinite impulse response filter with a large delay (Neely and Allen, 1979). In addition, one needs to have *a priori* knowledge of the room impulse response, which is often impractical. Several methods have been proposed to estimate the inverse filter in unknown acoustical conditions (Furuya and Kaneda, 1997; Gillespie *et al.*, 2001; Nakatani and Miyoshi, 2003). In particular, the system developed by Gillespie *et al.* (2001) estimates the inverse filter from an array of microphones using an adaptive gradient-descent algorithm that maximizes the kurtosis of linear prediction (LP) residuals. The inverse filter results in reduction of perceived reverberation as well as enhanced harmonicity. In this paper, we employ a one-microphone adaptation of this method proposed by Wu (2003; Wu and Wang, 2006).

The dereverberation algorithms described above are designed to enhance a single reverberant source. Here, we investigate the effect of inverse filtering as preprocessing for a pitch-based speech segregation system in order to improve its robustness in reverberant environments. The key idea is to estimate a filter that inverts the room impulse response corresponding to the target source. The effect of applying this inverse filter on the reverberant mixture is twofold: It improves the harmonic structure of the target signal while smearing those signals originating at other locations. Using a signal-to-noise ratio (SNR) evaluation, we show that the inverse filtering stage improves the separation performance of our pitch-based system. To our knowledge, this is the first study that addresses monaural speech segregation with room reverberation.

The rest of the paper is organized as follows. The next section defines the problem domain and presents a model overview. Section III gives a detailed description of the dereverberation stage. Section IV gives a detailed description of the pitch-based segregation stage. Section V presents systematic results on pitch-based segregation both in reverberant and inverse-filtered conditions. We also make a comparison with the spectral subtraction method. Section VI concludes the paper.

II. MODEL OVERVIEW

The speech received at one ear in a reverberant enclosure undergoes both convolutive and additive distortions:

$$y(t) = h(t) * s(t) + n(t), \tag{1}$$

where "$*$" indicates convolution. $s(t)$ is the clean (anechoic) target speech signal to be recovered, $h(t)$ models the acoustic transfer function from target speaker location to the ear, and $n(t)$ is the reverberant background noise which usually contains interfering sources at other locations. As explained in the Introduction, the problem of monaural speech segregation has been studied extensively in the additive condition by employing the periodicity of target speech. However, room reverberation poses an additional challenge by smearing the

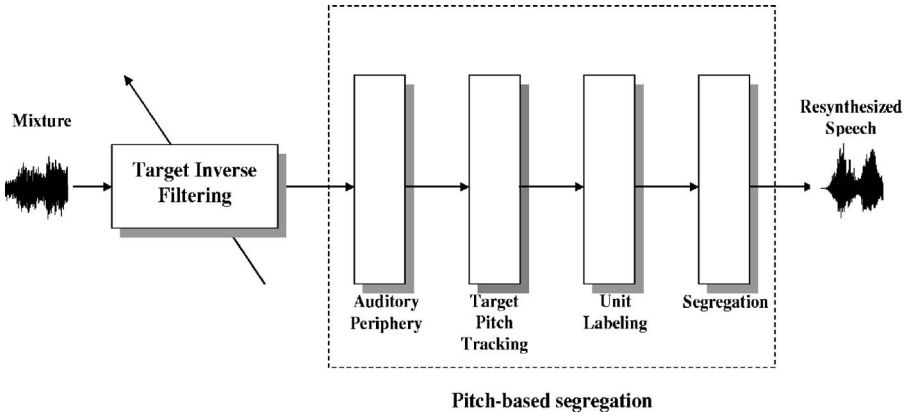


FIG. 1. Schematic diagram of the proposed two-stage model.

spectrum and weakening the harmonic structure. Consequently, we propose a two-stage speech segregation model: (1) inverse filtering with respect to the target location in order to enhance the periodicity of the target signal; (2) pitch-based speech segregation. Figure 1 illustrates the architecture of the proposed model.

The input to our model is a monaural mixture of two or more sound sources in a small reverberant room (6 m $\times$4 m$\times$3 m). The receiver—the left ear of a Knowles Electronic Manikin for Auditory Research (KEMAR) (Burkhard and Sachs, 1975)—is fixed at (2.5 m, 2.5 m, and 2 m) while the acoustic sources are located at a distance of 1.5 m from the receiver. The impulse response corresponding to the acoustic transfer function from a source to the receiver is simulated using a room acoustic model. Specifically, the simulated reflections from the walls are given by the image reverberation model (Allen and Berkley, 1979) and are convolved with the measured head related impulse responses (HRIR) of the KEMAR (Gardner and Martin, 1994). This represents a realistic input signal at the ear. Specific room reverberation times are obtained by varying the absorption characteristics of room boundaries (Palomaki *et al.*, 2004). Note that two different positions in the room produce impulse responses that differ greatly. The original clean signals are upsampled at the HRIR sampling frequency of 44 kHz and then convolved with the corresponding room impulse responses. Finally, the resulting reverberant signals are added together and resampled at 16 kHz.

In the first stage, a finite impulse response filter is estimated that inverts the target room impulse response. Adaptive filtering strategies for estimating this filter are sensitive to background noise (Haykin, 2002). For simplicity, we perform this estimation during an initial training stage using reverberant speech from the target location in the absence of background noise. We employ the inverse-filtering method by Gillespie *et al.* (2001), which uses a relatively small amount of training data. During testing, the inverse filter is applied to a mixture signal consisting of a reverberant target signal and interfering signals. The result is then fed to the next stage. We emphasize that this initial training is not utterance dependent; that is, the utterances used in training and testing can be totally different.

In the second stage, we employ a pitch-based segregation system to separate the inverse-filtered target signal. The signal is analyzed using a gammatone filterbank (Patterson *et al.*, 1988) in consecutive time frames to produce a T-F decomposition, where a basic T-F unit refers to the response of a particular filter channel in a particular time frame. Our system computes a correlogram which is a standard technique for periodicity extraction (Licklider, 1951; Slaney and Lyon, 1993). Specifically, autocorrelation is computed at the output of a particular channel and the set of the autocorrelations for all channels forms the correlogram. In the high-frequency range, we use response envelopes and extract AM rates. The system then groups those T-F units where the underlying target is stronger than the combined interference by comparing the extracted periodicities with an estimated target pitch. Labeling at the T-F unit level is a local decision and therefore prone to noise. Following Bregman's conceptual model, previous CASA systems employ an initial segmentation stage followed by a grouping stage in which segments likely to originate from the same source are grouped together (see, e.g., Wang and Brown, 1999). To enhance the robustness, we further perform segmentation. The result of this process is a binary T-F mask corresponding to the target stream.

Finally, a speech wave form is resynthesized from the resulting binary mask using a method described by Weintraub (1985; see also Brown and Cooke, 1994). The signal is reconstructed from the output of the gammatone filterbank. To remove across-channel differences, the output of the filter is time reversed, passed through the gammatone filter, and reversed again. The mask is employed to retain the acoustic energy from the mixture that corresponds to one's in the mask and nullifies the others.

III. TARGET INVERSE FILTERING

As described in the Introduction, inverse filtering is a standard strategy used for deriving the anechoic signal. We employ the method proposed by Gillespie *et al.* (2001) which attempts to blindly estimate the inverse filter from single-source reverberant speech. Their method was originally proposed for multi-microphone situations and has subsequently been extended to monaural recordings by Wu and Wang (2006). Based on the observation that peaks in the LP residual of speech are weakened by reverberation, an adaptive algorithm estimates the inverse filter by maximizing the kurtosis of the inverse-filtered LP residual of reverberant speech $\tilde{z}(t)$

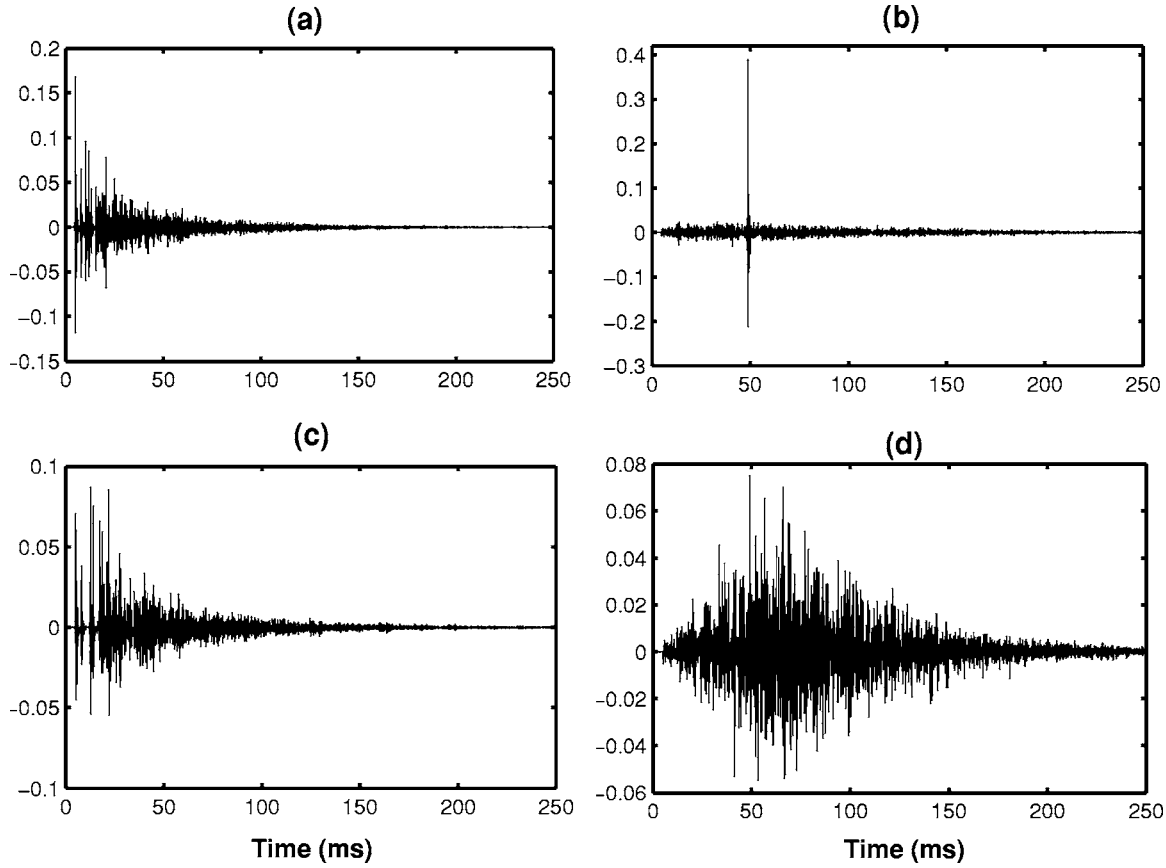


FIG. 2. Effects of inverse filtering on room impulse responses. (a) A room impulse response for a target source presented in the median plane. (b) The effect of convolving the impulse response in (a) with an estimated inverse filter. (c) A room impulse response for one interfering source at 45° azimuth. (d) The effect of convolving the impulse response in (c) with the estimated inverse filter.

$$\tilde{z}(t) = \mathbf{q}\mathbf{y}_r^T(t), \tag{2}$$

where $\mathbf{y}_r(t) = [y_r(t-L+1), \ldots, y_r(t-1), y_r(t)]$ and $y_r(t)$ is the LP residual of the reverberant speech from the target source, and $\mathbf{q}$ is an inverse filter of length L. The inverse filter is derived by maximizing the kurtosis of $\tilde{z}(t)$, which is defined as

$$J = \frac{E[\tilde{z}^4(t)]}{E^2[\tilde{z}^2(t)]} - 3. \tag{3}$$

The gradient of the kurtosis with respect to the inverse filter $\mathbf{q}$ can be approximated as follows (Gillespie *et al.*, 2001):

$$\frac{\partial J}{\partial \mathbf{q}} \approx \left\{ \frac{4(E[\tilde{z}^2(t)]\tilde{z}^3(t) - E[\tilde{z}^4(t)]\tilde{z}(t))}{E^3[\tilde{z}^2(t)]} \right\} \mathbf{y}_r(t). \tag{4}$$

Consequently, the optimization process in the time domain is given by the following update equation:

$$\hat{\mathbf{q}}(t+1) = \hat{\mathbf{q}}(t) + \mu f(t) \hat{\mathbf{y}}_r(t), \tag{5}$$

where $\hat{\mathbf{q}}(t)$ is the estimate of the inverse filter at time t, μ denotes the update rate, and $f(t)$ denotes the term inside the braces of Eq. (4).

However, a direct time-domain implementation of the above update equation is not desirable since it results in very slow convergence or no convergence at all under noisy conditions (Haykin, 2002). In this paper, we use the fast-block LMS (least mean square) implementation for one microphone signals described by Wu and Wang (2006). This method shows good convergence when applied to one-microphone reverberant signals for a range of reverberation times. The signal is processed block by block using a size L for both filter length and block length with the following update equations:

$$\mathbf{Q}'(n+1) = \mathbf{Q}(n) + \frac{\mu}{M} \sum_{m=1}^{M} \mathbf{F}(m) \mathbf{Y}_r^*(m), \tag{6}$$

$$\mathbf{Q}(n+1) = \frac{\mathbf{Q}'(n+1)}{|\mathbf{Q}'(n+1)|}, \tag{7}$$

where $\mathbf{F}(m)$ and $\mathbf{Y}_r(m)$ represent the fast Fourier transform (FFT) of $f(t)$ and $\mathbf{y}_r(t)$ for the m*th* block, and $\mathbf{Q}(n)$ represents the estimate for the FFT of inverse filter $\mathbf{q}$ at iteration n. M represents the number of blocks and the superscript * indicates the complex conjugation. Equation (7) ensures that the estimate of the inverse filter is normalized.

The system is trained on reverberant speech from the target source sampled at 16 kHz and presented alone. We employ a training corpus consisting of ten speech signals from the TIMIT database: five female utterances and five male utterances. An inverse filter of length $L = 1024$ is adapted for 500 iterations on the training data.

Figure 2 shows the outcome of convolving an estimated inverse filter with both the target impulse response as well as

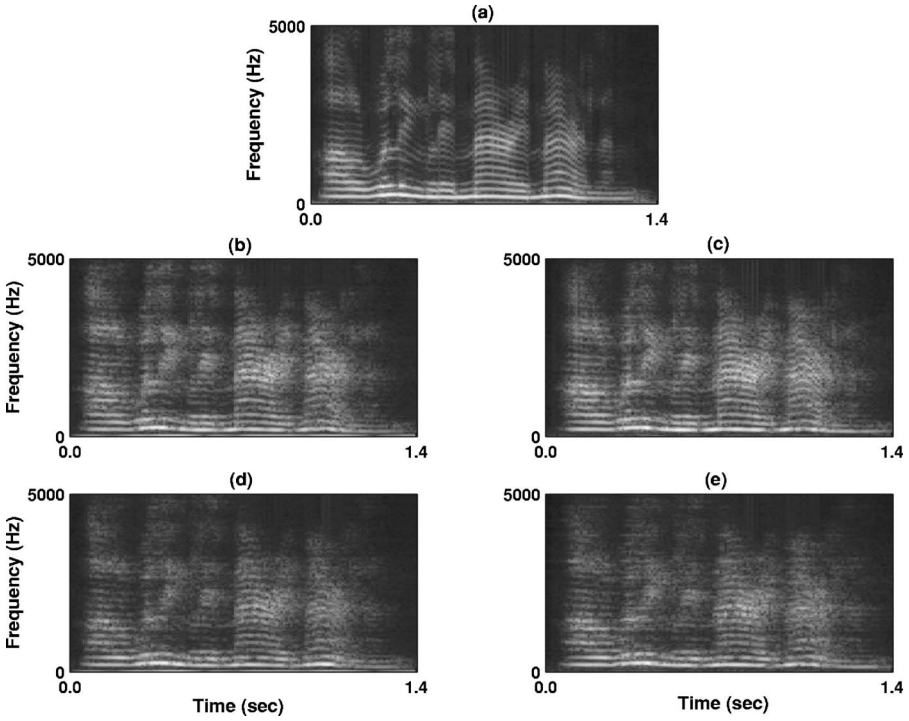


FIG. 3. Effects of reverberation and target inverse filtering on the harmonic structure of a voiced utterance. (a) Spectrogram of the anechoic signal. (b) Spectrogram of the reverberant signal corresponding to the impulse response in Fig. 2(a). (c) Spectrogram of the inverse-filtered signal corresponding to the equalized impulse response in Fig. 2(b). (d) Spectrogram of the reverberant signal corresponding to the room impulse response in Fig. 2(c). (e) Spectrogram of the inverse filtered signal corresponding to the impulse response in Fig. 2(d).

the impulse response at a different source location. The room reverberation time T_{60} is 0.35 s (T_{60} is the time required for the sound level to drop by 60 dB following the sound offset). The two source azimuths are 0° (target) and 45°. As can be seen in Fig. 2(b), the equalized response for the target source is far more impulselike compared to the room impulse response in Fig. 2(a). On the other hand, the impulse response corresponding to the interfering source is further smeared by the inverse filtering process, as seen in Fig. 2(d). Figure 3 illustrates the effect of reverberation as well as that of inverse filtering on the harmonic structure of a voiced utterance. The filters in Fig. 2 are convolved with an anechoic signal to generate the signals in Fig. 3. For a constant pitch contour, reverberation produces elongated tails but preserves the harmonicity. However, once the pitch varies reverberation smears the harmonic structure. For a given change in pitch frequency, higher harmonics vary their frequencies more rapidly compared to lower ones. Consequently, higher harmonics are more susceptible to reverberation as can be seen in Fig. 3(b). Figure 3(c) shows that an inverse filter is able to recover some of the harmonic components in the signal; for example, the harmonic series starting at about 1.0 s is more visible in Fig. 3(c) than in Fig. 3(b). To exemplify the smearing effect on the spectrum of an interfering source, we show the convolution of the same utterance with the filters corresponding to Figs. 2(c) and 2(d) and the results are given in Figs. 3(d) and 3(e), respectively.

Finally, the target inverse filter is applied on the reverberant mixture and the resulting signal feeds to the second stage of our model described below.

IV. PITCH-BASED SPEECH SEGREGATION

The proposed pitch-based segregation system uses a given target pitch track to group harmonically related components from the target source. Our system follows the segmentation and grouping steps of Hu and Wang (2004). However, we simplify their algorithm by extracting periodicities directly from the correlogram. Also, compared to the sinusoidal modeling scheme for computing AM rates in Hu and Wang (2004), our simplified method is more robust to intrusions in the high frequency range. A detailed description of our model is given below.

A. Auditory periphery and feature extraction

The signal is filtered through a bank of 128 fourth-order gammatone filters with center frequencies between 80 and 5000 Hz (Patterson *et al.*, 1988). In addition, envelopes are extracted for channels with center frequencies higher than 800 Hz. A Teager energy operator is applied to the signal to extract its envelope (Rouat *et al.*, 1997). This is defined as $E(n)=x^2(n)-x(n+1)x(n-1)$ for a signal $x(n)$, where n denotes the sampling step. Then, the signals are low-pass filtered at 800 Hz using a third-order Butterworth filter and high-pass filtered at 64 Hz.

The correlogram $A(c,j,\tau)$ for channel c, time-frame j, and lag τ is computed by the following autocorrelation using a window of 20 ms ($K=320$):

$$A(c,j,\tau)=\frac{\sum_{k=0}^{K} g(c,j-k)g(c,j-k-\tau)}{\sqrt{\sum_{k=0}^{K} g^2(c,j-k)}\sqrt{\sum_{k=0}^{K} g^2(c,j-k-\tau)}}, \quad (8)$$

where g is the gammatone filter output and the correlogram is updated every 10 ms. The range for τ corresponding to the plausible pitch range of 80 to 500 Hz is from 32 to 200. At high frequencies, the autocorrelation based on response envelopes reveals the amplitude modulation rate that coincides with the F0 for one periodic source. Hence,

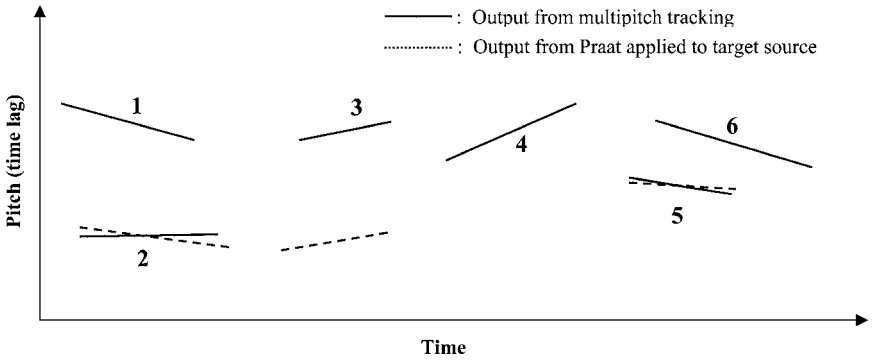


FIG. 4. Illustration of sequential organization. Solid lines illustrate a set of pitch contours from a multipitch tracking algorithm, each denoted by a number. Dashed lines show a set of pitch contours from Praat applied to the target signal before mixing. Note that these contours are drawn here for purposes of explanation, i.e., they are not actually produced from the algorithms. A comparison between the sets results in the selection of contours 2 and 5 as estimated target pitch contours.

an additional envelope correlogram $A_E(c,j,\tau)$ is computed for channels in the high-frequency range (>800 Hz) by replacing the filter output g in Eq. (8) with its extracted envelope. This correlogram representation of the acoustic signal has been successfully used in Wu *et al.* (2003) for multipitch analysis.

Finally, the cross-channel correlation between normalized autocorrelations in adjacent channels is computed in each T-F unit as

$$C(c,j) = \sum_{\tau=0}^{N-1} A(c,j,\tau)A(c+1,j,\tau), \tag{9}$$

where $N=200$ corresponds to the minimum pitch frequency of 80 Hz. Since adjacent channels activated by the same source tend to have similar autocorrelation responses, the cross-channel correlation has been used in previous segmentation studies (see, e.g., Wang and Brown, 1999). Similarly, envelope cross-channel correlation $C_E(c,j)$ is computed for channels in the high-frequency range (>800 Hz) to capture common amplitude modulation.

B. Unit labeling

A pitch-based segregation system requires a robust pitch detection algorithm. We employ the multipitch tracking (estimation) algorithm proposed by Wu *et al.* (2003) that gives good performance for a variety of intrusions. The system combines correlogram-based peak and channel selection within a statistical framework in order to form multiple tracks that correspond to different harmonic sources. When the interference is also a harmonic source, their system produces two pitch tracks each of which consists of a set of continuous pitch contours which do not overlap with each other, but the two sets may overlap in time; a pitch contour is a consecutive set of pitch points. The multipitch tracking system, however, does not address the issue of whether a particular pitch contour belongs to the target source or the interference. Assigning individual pitch contours to either the target or the interference is the issue of sequential organization (Bregman, 1990), and a challenging computational task which has been little addressed in previous CASA studies (Brown and Cooke, 1994; Hu and Wang, 2004). A recent study by Shao and Wang (2006) uses trained speaker models to address the sequential organization problem in the specific context of cochannel speech (two-speaker mixtures). In this paper, we do not attempt to address this problem and instead assume an "ideal" assignment for the two pitch tracks, i.e., an "ideal" binary decision for each of the contours in the contour union of the two tracks (as each track generally contains multiple contours). For this, an estimated pitch track from the target signal is extracted using Praat (Boersma and Weenink, 2002) and then used for the sole purpose of assigning whether an individual pitch contour corresponds to the target pitch track. This is explained in Fig. 4, which illustrates a set of pitch contours from the multipitch tracking algorithm of Wu *et al.* (2003) and the corresponding target pitch contours from Praat. The contours from the mixture data are marked as solid lines with numerical labels, while the target pitch contours from Praat are marked as dashed lines. In this situation, a comparison between the two sets results in the selection of contours 2 and 5 as estimated target pitch contours, which are used to group individual T-F units that belong to the target as described below. See Wu *et al.* (2003) for extensive treatment of multipitch tracking for noisy speech.

The labeling of an individual T-F unit is carried out by comparing the estimated target pitch with the periodicity of the correlogram. The correlogram has the well-known property that it exhibits a peak at the signal period as well as the multiples of the period. Note that an autocorrelation response is quasiperiodic due to the bandpass nature of a filter channel and the number of peaks in the correlogram increases with increasing center frequency of the channel. For a particular T-F unit, we should select the peak that best captures the periodicity of the underlying signal. In the low-frequency range, the system selects the peak for which the corresponding time lag l is the closest to the estimated target pitch lag p in $A(c,j,\tau)$. Statistics collected in individual channels show that the distribution of selected time lags is sharply centered around the target pitch lag and its variance decreases with increased center frequency. Hence, a T-F unit is discarded if the distance between the two lags $|p-l|$ exceeds a threshold θ_L. We have found empirically that a value of $\theta_L = 0.15(F_s/F_c)$ results in a good performance, where F_s is the sampling frequency and F_c is the center frequency of channel c. Finally, the peak height indicates the strength of the target signal in the mixture. The unit is thus labeled 1 if $A(c,j,l)$ is close to the maximum of $A(c,j,\tau)$ in the plausible pitch range

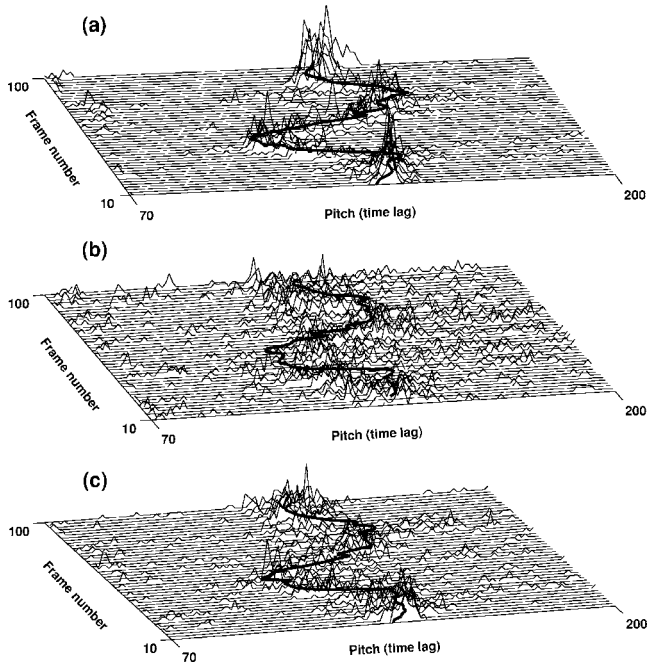


FIG. 5. Histograms of selected peaks in the high-frequency range (>800 Hz) for a male utterance. (a) Results for the anechoic signal. (b) Results for the reverberant signal. (c) Results for the inverse-filtered signal. The solid lines are the corresponding pitch tracks.

$$\frac{A(c,j,l)}{\max\limits_{\tau\in[32,200]} A(c,j,\tau)} > \theta_P, \tag{10}$$

where θ_P is fixed to 0.85. The unit is labeled 0 otherwise.

In the high-frequency range, we adapt the peak selection method of Wu *et al.* (2003). First, the envelope correlogram $A_E(c,j,\tau)$ of a periodic signal exhibits a peak both at the pitch lag and at the double of the pitch lag. Thus, the system selects all the peaks that satisfy the following condition: A peak with time lag l must have a corresponding peak that falls within the 5% interval around the double of l. If no peaks are selected, the T-F unit is labeled 0. Second, to deal with the situation where the pitch lag corresponding to the interference is half that of the target pitch, our system selects the first peak that is higher than half of the maximum peak in $A_E(c,j,\tau)$ for $\tau\in[32,200]$. Finally, the T-F unit is labeled 1 if the distance between the time lag corresponding to the selected peak and the estimated target pitch lag does not exceed a threshold of $\Delta=15$. The unit is labeled 0 otherwise. All the above parameters were optimized by using a small training set and found to generalize well over a test set.

The distortions on harmonic structure due to room reverberation are generally more severe in the high-frequency range. Figure 5 illustrates the effect of reverberation as well as inverse filtering in frequency channels above 800 Hz for a single male utterance. The filters in Figs. 2(a) and 2(b) are used to generate the reverberant signal and the inverse-filtered signal, respectively. At each time frame, we display the histogram of time lags corresponding to selected peaks. As can be seen from the figure, inverse filtering results in sharper peak distributions and improved harmonicity in comparison with the reverberant condition. The corresponding

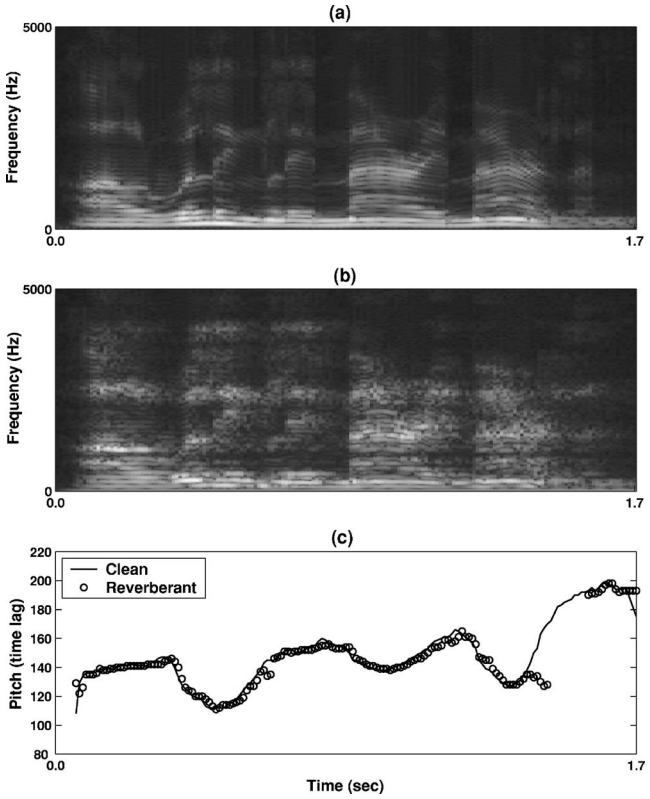


FIG. 6. Comparison of pitch tracking in anechoic and reverberant conditions for a male voiced utterance. (a) Spectrogram of the anechoic signal. (b) Spectrogram of the reverberant signal corresponding to the impulse response in Fig. 2(a). (c) Pitch tracking results. The solid line indicates the anechoic pitch track. The 'o' track indicates the reverberant track.

pitch tracks are extracted using Praat for each separate condition. To illustrate the effect of inverse filtering on the harmonic structure of the signals originating at the target location, we apply the T-F labeling described above to both the reverberant as well as the inverse-filtered male utterance. The signals are then reconstructed from the resulting T-F masks using the resynthesis method described in Sec. II. The reconstructed signal retains 79% of the target energy in the inverse-filtered condition compared to only 58% in the reverberant condition. As a reference, the corresponding labeling in the anechoic condition retains 94% of the target energy.

C. Segregation

The final segregation of the acoustic mixture into a target and a background stream is based on combined segmentation and grouping. A segment is a contiguous region of T-F units, each of which should be dominated by the same sound source. The main objective of the final segregation is to improve on the T-F unit labeling described above using segment-level features. The following steps follow the general segregation strategy in the Hu and Wang model (2004).

In the first step, segments are formed using temporal continuity and cross-channel correlation. Specifically, neighboring T-F units are iteratively merged into segments if their corresponding cross-channel correlation $C(c,j)$ exceeds a threshold $\theta_C=0.985$. The segments formed in this step are primarily located in the low-frequency range. A segment agrees with the target pitch at a given time frame if more

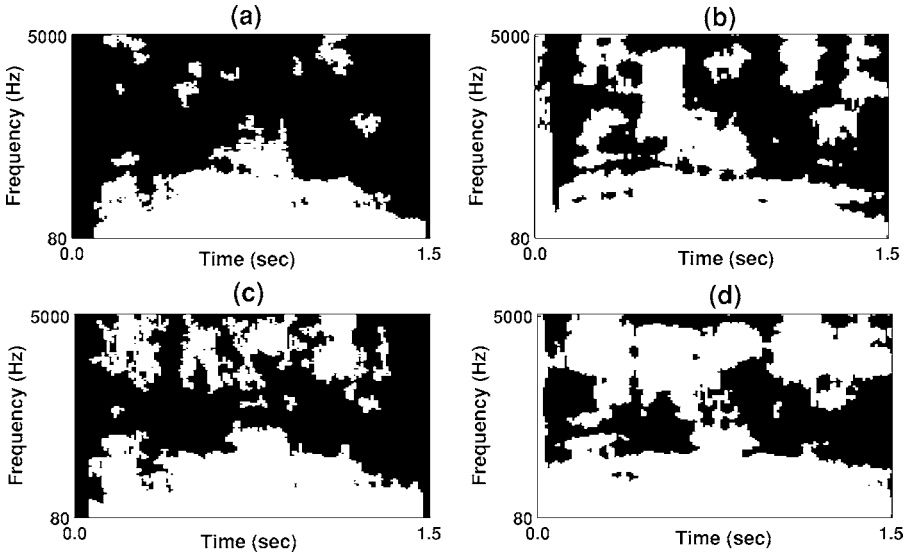


FIG. 7. Binary mask estimation for a mixture of target male utterance and interference female speech in reverberant and inverse-filtered conditions. (a) The estimated binary mask on the reverberant mixture. (b) The ideal binary mask for the reverberant condition. (c) The estimated binary mask on the filtered mixture. (d) The ideal binary mask for the inverse-filtered condition. The white regions indicate T-F units that equal 1 and the black regions indicate T-F units that equal 0.

than half of its T-F units are labeled 1. A segment that agrees with the target pitch for more than half of its length is grouped into the target stream; otherwise it goes to the background stream.

The second step primarily deals with potentially missing segments in the high-frequency range. Segments are formed by iteratively merging T-F units that are labeled 1 but not selected in the first step for which the envelope cross-channel correlation $C_E(c,j)$ exceeds the threshold θ_C. Segments shorter than 50 ms are removed. All these segments are grouped to the target stream.

The final step performs an adjustment of the target stream so that all T-F units in a segment bear the same label and no segments shorter than 50 ms are grouped. Furthermore, the target stream is iteratively expanded to include neighboring units that do not belong to either stream but are labeled 1.

With the T-F units belonging to the target stream labeled 1 and the other units labeled 0, the segregated target speech wave form is then resynthesized from the resulting binary T-F mask for systematic performance evaluation, to be discussed in the next section.

V. RESULTS

Two types of ASA cues that can potentially help a listener to segregate one talker in noisy conditions are location and pitch. Darwin and Hukin (2000) compared the effects of reverberation on spatial, prosodic, and vocal-tract size cues for a sequential organization task where the listener's ability to track a particular voice over time is examined. They found that while location cues are seriously impaired by reverberation, the F0 contour and vocal-tract length are more resistant cues. In our experiments, we have also observed that pitch tracking is robust to moderate levels of reverberation. To illustrate this, Fig. 6 compares the results of the pitch tracking algorithm of Wu *et al.* (2003) on a single male utterance in anechoic and reverberant conditions where T_{60}=0.35 s. The only distortions observed in the reverberant pitch track compared to the anechoic one are elongated tails and some deletions in the time frames where pitch changes rapidly.

Culling *et al.* (2003) have shown that while listeners are able to exploit the information conveyed by the F0 contour to separate a desired talker, the smearing of individual harmonics caused by reverberation degrades their separation capability. However, compared to location cues, the pitch cue degrades gradually with increasing reverberation and remains effective for speech separation (Culling *et al.*, 2003). In addition, as illustrated in Fig. 5, inverse filtering with respect to target location enhances signal harmonicity. We therefore assess the performance of two viable pitch-based strategies: (1) segregating the reverberant target from the reverberant mixture and (2) segregating the inverse-filtered target from the inverse-filtered mixture. Consequently, the speech segregation system described in Sec. IV is applied separately on the reverberant mixture and the inverse-filtered mixture.

To conduct a systematic SNR evaluation, a segregated signal is reconstructed from a binary mask following the method described in Sec. II. Given our computational objective of identifying T-F regions where the target is stronger than the interference, we use the signal reconstructed from the ideal binary mask as the ground truth to compute the output SNR (see Hu and Wang, 2004)

$$\mathrm{SNR}_{\mathrm{OUT}} = 10 \log_{10} \frac{\sum_t s_{\mathrm{IBM}}^2(t)}{\sum_t [s_{\mathrm{IBM}}(t) - s_E(t)]^2}, \tag{11}$$

where $s_{\mathrm{IBM}}(t)$ represents the target signal reconstructed using the ideal binary mask and $s_E(t)$ the estimated target reconstructed from the binary mask produced by our model. The input SNR is computed in the standard way as the ratio of target signal energy to noise signal energy expressed in decibels. Note that the target signal refers to the reverberant target signal in the reverberant condition and to the inverse-filtered signal in the inverse-filtered condition.

Figure 7 shows the binary masks produced by our system for a mixture of target male speech presented at 0° and interference female speech at 45°. Reverberant signals as

TABLE I. Output SNR results for target speech mixed with a female interference at three input SNR levels and different reverberation times.

Reverberation time (s)	−5 dB	0 dB	5 dB
Anechoic	8.78	11.61	13.93
T_{60}=0.05	7.25	8.54	10.65
T_{60}=0.10	7.35	8.16	9.46
T_{60}=0.15	6.37	7.09	8.24
T_{60}=0.20	5.59	6.52	7.39
T_{60}=0.25	4.74	6.06	6.79
T_{60}=0.30	4.47	5.57	6.22
T_{60}=0.35	4.55	5.36	6.13

well as inverse-filtered signals for both target and interference are produced by convolving the original anechoic utterances with the filters from Fig. 2. The signals are mixed to give an overall 0 dB input SNR in both conditions. The figure also displays the ideal binary masks. The results show an improved segregation capacity in the high frequency range in the inverse-filtered case [Fig. 7(c)] as compared to the reverberant case [Fig. 7(a)].

We perform the SNR evaluations using as target a set of ten voiced male sentences collected by Cooke (1993) for the purpose of evaluating voiced speech segregation systems. The following five noise intrusions are used: white noise; babble noise; a male utterance; music; and a female utterance. These intrusions represent typical acoustical interferences occurring in real environments. In all cases, the target is fixed at 0°. The babble noise is obtained by presenting natural speech utterances from the TIMIT database at the following eight separate directions around the target source: ±20°; ±45°; ±60°; and ±135°. For the other intrusions, the interfering source is located at 45°, unless otherwise specified. Also, the reverberation time for the experiments described below equals 0.35 s, unless otherwise specified. This reverberation time falls in the typical range for living rooms and office environments. When comparing the results between the two segregation strategies the target signal in each case is scaled to yield the desired input SNR. Each value in the following tables resents the average output SNR of one particular intrusion mixed with the ten target sentences.

We first analyze how pitch-based speech segregation is affected by reverberation. Table I shows the performance of our pitch-based segregation system applied directly on reverberant mixtures when T_{60} increases from 0.05 to 0.35 s. The mixtures are obtained using the female speech utterance as interference and three levels of input SNR: −5; 0; and 5 dB. The ideal pitch contours, not estimated ones, are used here for testing purposes. As expected, the system performance degrades gradually with increasing reverberation. Individual harmonics are increasingly smeared and this results in a gradual loss in energy, especially in the high-frequency range as illustrated also in Fig. 7. The decrease in output SNR for T_{60}=0.35 s compared to the anechoic condition ranges from 4.23 dB at −5 dB input SNR to 7.80 dB at 5 dB input SNR. Overall, however, the segregation algorithm provides consistent gains, showing the robustness of the pitch cue. Observe that a sizeable gain of 9.55 dB is obtained for the −5 dB input SNR even when T_{60}=0.35 s.

Now we analyze how the inverse-filtering stage impacts the overall performance. The results in Table II are given for both the reverberant case (Reverb) and inverse-filtered case (Inverse) at three input SNR levels: −5; 0; and 5 dB. The results are obtained using estimated pitch tracks as explained in Sec. IV B. The performance depends on input SNR and type of interference. A maximum improvement of 12.46 dB is obtained for the female interference at −5 dB input SNR. The proposed system (Inverse) has an average gain of 10.11 dB at −5 dB, 6.45 dB at 0 dB, and 2.55 dB at 5 dB. When compared to the reverberant condition a 2–3 dB improvement is observed for the male and female intrusions at all input SNR conditions. Almost no improvement is observed for white noise or babble noise. Moreover, inverse filtering decreases the system performance in the case of white noise at low SNRs because of the over-grouping of T-F units in the high-frequency range. For comparison, results using the ideal pitch tracks are presented in Table III. The improvement obtained by using ideal pitch tracks is small and shows that the pitch estimation method is accurate. We note that the variation in the output SNR values across different target sentences is relatively small—the standard deviation ranges from 1 to 2 dB—in both reverberant and inverse-filtered conditions.

As seen in the results presented above, the major advantage of the inverse-filtering stage occurs for a harmonic interference. In all the cases presented above the interfering source is located at 45°, and the inverse filtering stage further smears its harmonic structure. However, if the interfering source is located at a location near the target source the inverse filter will dereverberate the interference also. Table IV

TABLE II. Output SNR results using estimated pitch tracks for target speech mixed with different noise types at three input SNR levels and T_{60}=0.35 s. Target is at 0° and interference at 45°.

	−5 dB		0 dB		5 dB	
Input SNR	Reverb	Inverse	Reverb	Inverse	Reverb	Inverse
White noise	5.75	4.92	6.22	5.87	6.37	7.39
Babble noise	2.50	2.81	4.76	5.27	5.95	6.94
Male	0.67	4.54	3.96	6.68	5.76	7.76
Music	3.27	5.82	5.58	6.72	6.24	7.70
Female	4.87	7.46	5.51	7.70	6.13	7.95
Average	3.41	5.11	5.21	6.45	6.03	7.55

TABLE III. Output SNR results using ideal pitch tracks for target speech mixed with different noise types at three input SNR levels and T_{60}=0.35 s. Target is at 0° and interference at 45°.

	−5 dB		0 dB		5 dB	
Input SNR	Reverb	Inverse	Reverb	Inverse	Reverb	Inverse
White noise	5.94	5.38	6.19	6.10	6.37	7.56
Babble noise	3.25	4.23	5.14	5.71	5.95	7.40
Male	1.90	5.08	4.49	6.96	5.76	7.80
Music	3.89	6.25	5.73	6.93	6.24	7.80
Female	4.55	7.23	5.36	7.71	6.13	8.30
Average	3.90	5.63	5.38	6.68	6.09	7.77

shows SNR results for both white noise and female speech intrusions when the interference location is fixed at 0°, the same as the target location. As expected, in the white noise case, the results are similar to the ones presented in Table III. However, the relative improvement in output SNR obtained using inverse filtering is reduced to the range of 0.5–1 dB. This shows that smearing the harmonic structure of the interfering source plays an important role in boosting the segregation performance in the inverse-filtered condition.

As mentioned in Sec. I, this paper is the first study on monaural segregation of reverberant speech. As a result, it is difficult to quantitatively compare with existing systems. In an attempt to put our performance in perspective, we show a comparison with the spectral subtraction method, which is a standard speech enhancement technique (O'Shaughnessy, 2000). To apply spectral subtraction in practice requires robust estimation of interference spectrum. To put spectral subtraction in a favorable light, the average noise power spectrum is computed *a priori* within the silent periods of the target signal for each reverberant mixture. This average is used as the estimate of intrusion and is subtracted from the mixture. The SNR results are given in Table V, where the reverberant target signal is used as ground truth for the spectral subtraction algorithm and the inverse-filtered target signal is used as ground truth for our algorithm. As shown in the table, the spectral subtraction method performs significantly worse than our system, especially at low levels of input SNR. This is because of its well-known deficiency in dealing with nonstationary interferences. At 5 dB input SNR the spectral subtraction outperforms our system when the interference is white noise, babble noise, or music. In those cases of high-input SNR and relatively steady intrusion, the spectral subtraction algorithm tends to subtract little intrusion but it also introduces little distortion to the target signal. By comparison, our system focuses on target extraction that attempts to reconstruct the target signal on the basis of periodicity. Target components made inharmonic by reverberation are removed by our algorithm, thus introducing more target signal loss. It is worth noting that the ceiling performance of our algorithm without any interference is 8.89 dB output SNR.

VI. DISCUSSION

In natural settings, reverberation alters many of the acoustical properties of a sound source reaching our ears, including smearing its harmonic and temporal structures. Despite these alterations, moderate reverberant speech remains highly intelligible for normal-hearing listeners (Nabelek and Robinson, 1982). When multiple sound sources are active, however, reverberation adds another level of complexity to the acoustic scene. Not only does each interfering source constitute an additional masker for the desired source, but also does reverberation blur many of the cues that aid in source segregation. The recent results of Culling *et al.* (2003) suggest that reverberation degrades human ability to exploit differences in F0 between competing voices, producing a 5 dB increase in speech reception threshold for normally intonated sentences in monaural conditions.

We have investigated pitch-based monaural segregation in room reverberation and report the first systematic results on this challenging problem. We observe that pitch detection is relatively robust in moderate reverberation. However, the segregation capacity is reduced due to the smearing of the harmonic structure, resulting in gradual degradation in performance as the room reverberation time increases. As seen in Table I, compared to anechoic conditions there is an average decrement of 5.33 dB output SNR for a two-talker situation with T_{60}=0.35 s. This decrement is, however, consistent with the 5 dB increase in speech reception threshold reported by Culling *et al.* (2003).

TABLE IV. Output SNR results using ideal pitch tracks for target speech mixed with two types of noise at three input SNR levels and T_{60}=0.35 s. Target and interference are both located at 0°.

	−5 dB		0 dB		5 dB	
Input SNR	Reverb	Inverse	Reverb	Inverse	Reverb	Inverse
White noise	6.37	6.76	6.30	6.82	6.21	7.28
Female	4.82	5.51	5.74	6.65	6.28	7.57

TABLE V. Comparison between the proposed algorithm and spectral subtraction (SS). Results are obtained for target speech mixed with different noise types at three input SNR levels and T_{60}=0.35 s. Target is at 0° and interference at 45°.

	−5 dB		0 dB		5 dB	
Input SNR	SS	Proposed	SS	Proposed	SS	Proposed
White noise	2.40	3.36	6.54	4.93	10.47	6.48
Babble noise	−2.76	2.74	1.98	4.66	6.65	6.42
Male	−4.05	4.11	0.77	6.17	5.59	7.24
Music	−1.37	4.45	3.22	6.01	7.68	7.07
Female	−3.31	5.40	1.46	6.71	6.19	7.56
Average	−1.81	4.01	2.79	5.69	7.31	6.95

To reduce the smearing effects on the target speech, we have proposed a preprocessing stage which equalizes the room impulse response corresponding to target location. This preprocessing results in both improved harmonicity for signals arriving from the target direction and smearing of competing sources at other directions. We have found that this effect provides a better input signal for pitch-based segregation. The extensive evaluations show that our system yields substantial SNR gains across a variety of noise conditions. Our previous study shows a strong correlation between SNR gains measured against the ideal binary mask and improvements in automatic speech recognition and speech intelligibility scores (Roman *et al.*, 2003). Hence we expect similar improvements for the SNR gains achieved in the present study, although further evaluation is required to substantiate this projection.

The improvement in speech segregation obtained in the inverse-filtering case is limited by the accuracy of the estimated inverse filter. In our study, we have employed an algorithm that estimates the inverse filter directly from reverberant speech data. When the room impulse response is known, better inverse-filtering methods exist, e.g., the linear least square equalizer by Gillespie and Atlas (2002). This type of preprocessing leads to increased target signal fidelity and thus produces large improvements in speech segregation. In terms of applications to real-world scenarios our inverse-filtering faces several drawbacks. First, the adaptation of the inverse filter requires data on the order of a few seconds and thus any fast change in the environment (e.g., head movements and walking) will have an adverse impact on the inverse-filtering stage. Second, this stage needs to perform filter adaptation in the presence of no or weak interference. On the other hand, our pitch-based segregation stage can be applied without such limitations. Hence, whenever the adaptation of the inverse filter is infeasible, one can still apply our pitch-based segregation algorithm directly on the reverberant mixture.

Speech segregation in high input SNR conditions presents a challenge to our system. We employ a figure-ground segregation strategy that attempts to reconstruct the target signal by grouping harmonic components. Consequently, inharmonic target components are removed by our approach even in the absence of interference. While this problem is common in both anechoic and reverberant conditions, it worsens in reverberation due to the smearing of harmonicity. To address this issue probably requires examining the inharmonicity induced by reverberation and distinguishing such inharmonicity from that caused by additive noise. This is a topic of further investigation.

In the segregation stage, our system utilizes only pitch cues and thus is limited to the segregation of voiced speech. Other ASA cues such as onsets, offsets, and acoustic-phonetic properties of speech are also important for monaural separation (Bregman, 1990). Recent research has shown that these cues can be used to separate unvoiced speech (Hu and Wang, 2003; 2005). Future work will need to address unvoiced separation in reverberant conditions. Another limitation, already mentioned in Sec. IV B, concerns sequential grouping. Like previous studies, our system avoids this issue by assuming an "ideal" assignment of estimated pitch contours. Although some progress has been made on sequential grouping of cochannel speech (e.g., Shao and Wang, 2006), the general problem of sequential organization remains a considerable challenge in CASA.

ACKNOWLEDGMENTS

This research was supported in part by an AFOSR grant (FA9550-04-1-0117) and an NSF grant (IIS-0081058).

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

Barros, A. K., Rutkowski, T., Itakura, F., and Ohnishi, N. (**2002**). "Estimation of speech embedded in a reverberant and noisy environment by independent component analysis and wavelets," IEEE Trans. Neural Netw. **13**, 888–893.

Balan, R., Jourjine, A., and Rosca, J. (**1999**). "AR processes and sources can be reconstructed from degenerate mixtures," *Proc. 1st Int. Workshop on Independent Component Analysis and Signal Separation*, pp. 467–472.

Boersma, P., and Weenink, D. (**2002**). *Praat: doing Phonetics by Computer*, Version 4.0.26 (http://www.fon.hum.uva.nl/praat).

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT Press, Cambridge, MA).

Bronkhorst, A. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acustica **86**, 117–128.

Brown, G. J., and Cooke, M. (**1994**). "Computational auditory scene analysis," Comput. Speech Lang. **8**, 297–336.

Brown, G. J., and Wang, D. L. (**2005**). "Separation of speech by computational auditory scene analysis," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, eds. (Springer, New York), pp. 371–402.

Brungart, D., Chang, P., Simpson, B., and Wang, D. L. (**2006**). "Isolating the energetic component of speech-on-speech masking with an ideal binary time-frequency mask," (unpublished).

Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for

acoustic research," J. Acoust. Soc. Am. **58**, 214–222.
Cooke, M. P. (**1993**). *Modeling Auditory Processing and Organization* (Cambridge University Press, Cambridge, U.K).
Cooke, M. P., Green, P., Josifovski, L., and Vizinho, A. (**2001**). "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Commun. **34**, 267–285.
Culling, J. F., Hodder, K. I., and Toh, C. Y. (**2003**). "Effects of reverberation on perceptual segregation of competing voices," J. Acoust. Soc. Am. **114**, 2871–2876.
Darwin, C. J., and Carlyon, R. P. (**1995**). "Auditory grouping," in *The Handbook of Perception and Cognition*, vol. 6, B. C. J. Moore, ed. (Academic, London), pp. 387–424.
Darwin, C. J., and Hukin, R. W. (**2000**). "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention," J. Acoust. Soc. Am. **108**, 335–342.
Ephraim, Y., and Trees, H. L. (**1995**). "A signal subspace approach for speech enhancement," IEEE Trans. Speech Audio Process. **3**, 251–266.
Furuya, K., and Kaneda, Y. (**1997**). "Two-channel blind deconvolution for non-minimum phase impulse responses," *Proc. ICASSP*, pp. 1315–1318.
Gardner, W. G., and Martin, K. D. (**1994**). "HRTF measurements of a KEMAR dummy-head microphone," MIT Media Lab Perceptual Computing Technical Report #280.
Gillespie, B. W., and Atlas, L. E. (**2002**). "Acoustic diversity for improved speech recognition in reverberant environments," *Proc. ICASSP*, pp. 557–560.
Gillespie, B. W., Malvar, H. S., and Florencio, D. A. F. (**2001**). "Speech dereverberation via maximum-kurtosis subband adaptive filtering," *Proc. ICASSP*, vol. 6, pp. 3701–3704.
Haykin, S. (**2002**). *Adaptive Filter Theory*, 4th ed. (Prentice-Hall, Upper Saddle River, NJ).
Hu, G., and Wang, D. L. (**2003**). "Separation of stop consonants," *Proc. ICASSP*, vol. 2, pp. 749–752.
Hu, G., and Wang, D. L. (**2004**). "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw. **15**, 1135–1150.
Hu, G., and Wang, D. L. (**2005**). "Separation of fricatives and affricates," *Proc. ICASSP* vol. 1, pp. 1101–1104.
Jang, G.-J., Lee, T.-W., and Oh, Y.-H. (**2003**). "Single channel signal separation using time-domain basis functions" IEEE Signal Process. Lett. **10**(6), 168–171.
Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.
Luo, H. Y., and Denbigh, P. N. (**1994**). "A speech separation system that is robust to reverberation," *Proc. ISSIPNN*, pp. 339–342.
Ma, N., Bouchard, M., and Goubran, R. (**2004**). "Perceptual Kalman filtering for speech enhancement in colored noise," *Proc. ICASSP*, vol. 1, pp. 717–720.
Martin, R. (**2001**). "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Process. **9**, 504–512.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego, CA).
Nabelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.
Nakatani, T., and Miyoshi, M. (**2003**). "Blind dereverberation of single channel speech signal based on harmonic structure," *Proc. ICASSP*, pp. 92–95.
Neely, S. T., and Allen, J. B. (**1979**). "Invertibility of a room impulse response," J. Acoust. Soc. Am. **66**, 165–169.
O' Shaughnessy, D. (**2000**). *Speech Communications: Human and Machine*, 2nd ed. (Piscataway, IEEE Press, NJ).
Palomaki, K. J., Brown, G. J., and Wang, D. L. (**2004**). "A binaural processor for missing data speech recognition in the presence of noise and small-room reverberation," Speech Commun. **43**, 361–378.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Price, P. (**1988**). "APU Report 2341: An efficient auditory filterbank based on the gammatone function," Applied Psychology Unit, Cambridge.
Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of a single competing sound source (speech or noise)," Acustica **34**, 200–211.
Roman, N., Wang, D. L., and Brown, G. J. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.
Rouat, J., Liu, Y. C., and Morissette, D. (**1997**). "A pitch determination and voice/unvoiced decision algorithm for noisy speech," Speech Commun. **21**, 191–207.
Shao, Y., and Wang, D. L. (**2006**). "Model-based sequential organization in cochannel speech," IEEE Trans. Audio, Speech, Lang. Proc. **14**, 289–298.
Shamsoddini, A., and Denbigh, P. N. (**2001**). "A sound segregation algorithm for reverberant conditions," Speech Commun. **33**, 179–196.
Slaney, M., and Lyon, R. F. (**1993**). "On the importance of time—A temporal representation of sound," in *Visual Representations of Speech Signals*, M. P. Cooke, S. Beet, and M. Crawford, eds. (Wiley, New York), pp. 95–116.
Srinivasan, S., Roman, N., and Wang, D. L. (**2004**). "On binary and ratio time-frequency masks for robust speech recognition," *Proc. ICSLP*, pp. 2541–2544.
Wang, D. L. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, P. Divenyi, ed. (Kluwer Academic, Norwell, MA), pp. 181–197.
Wang, D. L., and Brown, G. J. (**1999**). "Separation of speech from interfering sounds based on oscillatory correlation," IEEE Trans. Neural Netw. **10**, 684–697.
Weintraub, M. (**1985**). "A theory and computational model of auditory monaural sound separation," Ph.D. dissertation, Stanford University Department of Electrical Engineering.
Wu, M. (**2003**). "Pitch tracking and speech enhancement in noisy and reverberant environments," PhD thesis, The Ohio State University, Department of Computer and Information Science.
Wu, M., and Wang, D. L. (**2006**). "A two-stage algorithm for one-microphone reverberant speech enhancement," IEEE Trans. Audio, Speech, Lang. Proc. **10**, 774–784.
Wu, M., Wang, D. L., and Brown, G. J. (**2003**) "A multipitch tracking algorithm for noisy speech," IEEE Trans. Speech Audio Process. **11**, 229–241.
Zibulevsky, M., Pearlmutter, B. A., Bofill, P., and Kisilev, P. (**2001**). "Blind source separation by sparse decomposition," in *Independent Component Analysis: Principles and Practice*, S. J. Roberts and R. M. Everson, eds. (Cambridge University Press, Cambridge).

An effective cluster-based model for robust speech detection and speech recognition in noisy environments

J. M. Górriz,[a] J. Ramírez, and J. C. Segura
Department of Signal Theory, University of Granada, Spain

C. G. Puntonet
Department of Computer Architecture and Technology, University of Granada, Spain

(Received 29 December 2005; revised 3 May 2006; accepted 5 May 2006)

This paper shows an accurate speech detection algorithm for improving the performance of speech recognition systems working in noisy environments. The proposed method is based on a hard decision clustering approach where a set of prototypes is used to characterize the noisy channel. Detecting the presence of speech is enabled by a decision rule formulated in terms of an averaged distance between the observation vector and a cluster-based noise model. The algorithm benefits from using contextual information, a strategy that considers not only a single speech frame but also a neighborhood of data in order to smooth the decision function and improve speech detection robustness. The proposed scheme exhibits reduced computational cost making it adequate for real time applications, i.e., automated speech recognition systems. An exhaustive analysis is conducted on the AURORA 2 and AURORA 3 databases in order to assess the performance of the algorithm and to compare it to existing standard voice activity detection (VAD) methods. The results show significant improvements in detection accuracy and speech recognition rate over standard VADs such as ITU-T G.729, ETSI GSM AMR, and ETSI AFE for distributed speech recognition and a representative set of recently reported VAD algorithms. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2208450]

PACS number(s): 43.72.Ne, 43.72.Dv [EJS] Pages: 470–481

I. INTRODUCTION

The emerging wireless communication systems require increasing levels of performance and speech processing systems working in noise adverse environments. These systems often benefit from using voice activity detectors (VADs) which are frequently used in such application scenarios for different purposes. Speech/nonspeech detection is an unsolved problem in speech processing and affects numerous applications including robust speech recognition,[1,2] discontinuous transmission,[3,4] estimation and detection of speech signals,[5,6] real-time speech transmission on the Internet[7] or combined noise reduction and echo cancelation schemes in the context of telephony.[8] The speech/nonspeech classification task is not as trivial as it appears, and most of the VAD algorithms fail when the level of background noise increases. During the last decade, numerous researchers have developed different strategies for detecting speech on a noisy signal[9–13] and have evaluated the influence of the VAD effectiveness on the performance of speech processing systems.[14] Most of them have focused on the development of robust algorithms with special attention on the derivation and study of noise robust features and decision rules.[12,15–17] The different approaches include those based on energy thresholds,[15] pitch detection,[18] spectrum analysis,[17] zero-crossing rate,[4] periodicity measures[19] or combinations of different features.[3,4,20]

The speech/pause discrimination may be described as an unsupervised learning problem. Clustering is an appropriate solution for this case where the data set is divided into groups which are related "in some sense." Despite the simplicity of clustering algorithms, there is an increasing interest in the use of clustering methods in pattern recognition,[21] image processing[22] and information retrieval.[23,24] Clustering has a rich history in other disciplines[25,26] such as machine learning, biology, psychiatry, psychology, archaeology, geology, geography, and marketing. Cluster analysis, also called data segmentation has a variety of goals. All of these are related to grouping or segmenting a collection of objects into subsets or "clusters" such that those within each cluster are more closely related to one another than objects assigned to different clusters. Cluster analysis is also used to form descriptive statistics to ascertain whether or not the data consists of a set of distinct subgroups, each group representing objects with substantially different properties.

The paper is organized as follows. Section II introduces the necessary background information on clustering analysis. Section III shows the feature extraction process and a description of the proposed long term information *C*-means (LTCM) VAD algorithm is given in Sec. IV. Section V discusses some remarks about the proposed method. A complete experimental evaluation is conducted in Sec. VI in order to compare the proposed method with a representative set of VAD methods and to assess its performance for robust speech recognition applications. Finally, we state some conclusions and acknowledgments in the last part of the paper.

[a]URL: http://www.ugr.es/~gorriz; Electronic mail: gorriz@ugr.es

 0001-4966/2006/120(1)/470/12/$22.50

TABLE I. Hard C-means pseudocode.

(1). Initialize a C-partition randomly or based on some prior knowledge. Calculate the cluster prototype matrix $\mathbf{M}=[\mathbf{m}_1,\ldots,\mathbf{m}_C]$
(2) Assign each object in the data set to the nearest cluster P_i[a].
(3). Recalculate the cluster prototype matrix based on the current partition
(4). Repeat steps (2)–(3) until there is no change for each cluster.

[a]That is, $\mathbf{x}_j \in P_i$ if $\|\mathbf{x}_j-\mathbf{m}_i\|<\|\mathbf{x}_j-\mathbf{m}_{i'}\|$ for $j=1,\ldots,N, i\neq i'$, and $i'=1,\ldots,C$

II. HARD PARTITIONAL CLUSTERING BASIS

Partitional clustering algorithms partition data into certain number of clusters, in such a way that, patterns in the same cluster should be "similar" to each other unlike patterns in different clusters. Given a set of input patterns $\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_j,\ldots,\mathbf{x}_N\}$, where $\mathbf{x}_j=(x_{j1},\ldots,x_{ji},\ldots,x_{jK})\in\mathbb{R}^K$ and each measure x_{jk} is said to be a feature, hard partitional clustering attempts to seek a C-partition of $\mathbf{X}$, $P=\{P_1,\ldots,P_C\}, C\leq N$, such that

(i) $P_i\neq\emptyset, i=1,\ldots,C$;
(ii) $\cup_{i=1}^{C}P_i=\mathbf{X}$;
(iii) $P_i\cap P_{i'}=\emptyset; i,i'=1,\ldots,C$ and $i\neq i'$.

The "similarity" measure is established in terms of a criterion function. The sum of squares error function is one of the most widely used criteria and is defined as

$$J(\mathbf{\Gamma},\mathbf{M})=\sum_{i=1}^{C}\sum_{j=1}^{N}\gamma_{ij}\|\mathbf{x}_j-\mathbf{m}_i\|^2, \tag{1}$$

where $\mathbf{\Gamma}=\gamma_{ij}$ is a partition matrix,

$$\gamma_{ij}=\begin{pmatrix}1 & \text{if } \mathbf{x}_j\in P_i\\ 0 & \text{otherwise}\end{pmatrix}$$

with $\Sigma_{i=1}^{C}\gamma_{ij}=1, \forall j, \mathbf{M}=[\mathbf{m}_1,\ldots,\mathbf{m}_C]$ is the cluster prototype or centroid (means) matrix with $\mathbf{m}_i=1/N_i\Sigma_{j=1}^{N}\gamma_{ij}\mathbf{x}_j$, the sample mean for the ith cluster and N_i the number of objects in the ith cluster. The optimal partition resulting of the minimization of the latter criterion can be found by enumerating all possibilities. It is unfeasible due to costly computation and heuristic algorithms have been developed for this optimization instead.

Hard C-means clustering is the best-known heuristic squared error-based clustering algorithm.[27] The number of cluster centers (prototypes) C is *a priori* known and the C-means iteratively moves the centers to minimize the total cluster variance. Given an initial set of centers the hard C-means algorithm alternates two steps:[28]

(i) for each cluster we identify the subset of training points (its cluster) that is closer to it than any other center;
(ii) the means of each feature for the data points in each cluster are computed, and this mean vector becomes the new center for that cluster.

In Table I we show a more detailed description of the C-means algorithm.

III. FEATURE EXTRACTION INCLUDING CONTEXTUAL INFORMATION

Let $x(n)$ be a discrete time signal. Denote by $\mathbf{y}_{n'}$ a frame containing the samples

$$\mathbf{y}_{n'}=\{x(i+n'\cdot D)\},\quad i=0,\ldots,L-1,\quad n=i+n'\cdot D, \tag{2}$$

where D is the window shift, L is the number of samples in each frame and n' selects a certain data window. Consider the set of $2\cdot m+1$ frames $\{\mathbf{y}_{l-m},\ldots,\mathbf{y}_l,\ldots,\mathbf{y}_{l+m}\}$ centered on frame $\mathbf{y}_l$, and denote by $Y(s,n')$, $n'=l-m,\ldots,l,\ldots,l+m$ its discrete Fourier transform (DFT), respectively,

$$Y_{n'}(\omega_s)\equiv Y(s,n')=\sum_{i=0}^{N_{\mathrm{FFT}}-1}x(i+n'\cdot D)\cdot\exp(-\mathbf{j}\cdot i\cdot\omega_s), \tag{3}$$

where $\omega_s=2_{\pi\cdot s}/N_{\mathrm{FFT}}$, $0\leq s\leq N_{\mathrm{FFT}}-1$, N_{FFT} is DFT resolution (if $N_{\mathrm{FFT}}>L$ then the DFT is padded with zeros) and $\mathbf{j}$ denotes the imaginary unit. The averaged energies for each n'th frame, $E(k,n')$, in K subbands ($k=1,2,\ldots,K$), are computed by means of

$$E(k,n')=\left(\frac{2K}{N_{\mathrm{FFT}}}\sum_{s=s_k}^{s_{k+1}-1}|Y(s,n')|^2\right)$$

$$s_k=\left\lfloor\frac{N_{\mathrm{FFT}}}{2K}(k-1)\right\rfloor,\quad k=1,2,\ldots,K, \tag{4}$$

where an equally spaced subband assignment is used and $\lfloor\cdot\rfloor$ denotes the "floor" function. Hence, the signal energy is averaged over K subbands obtaining a suitable representation of the input signal for VAD,[29] the observation vector at each frame n', defined as

$$\mathbf{E}(n')=(E(1,n'),\ldots,E(K,n'))^T\in\mathbb{R}^K. \tag{5}$$

The VAD decision rule is formulated over a sliding window consisting of $2m+l$ observation (feature) vectors around the frame for which the decision is being made (l), as we will show in the following sections. This strategy, known as "long term information,"[30] provides very good results using several approaches for VAD, however it imposes an m-frame delay on the algorithm that, for several applications including robust speech recognition, is not a serious implementation obstacle.

In the following section we show the way we apply **C**-means to modeling the noise subspace and to find a soft decision rule for VAD.

IV. HARD C-MEANS FOR VAD

In the LTCM VAD algorithm, the clustering method described in Sec. II is applied to a set of initial pause frames in order to characterize the noise subspace, that is, the generic feature vector described in Sec. II is defined in terms of energy observation vectors as we show in the following: each observation vector in Eq. (5) is uniquely labeled, by the integer $j\in\{1,\ldots,N\}$, and uniquely assigned (hard decision-

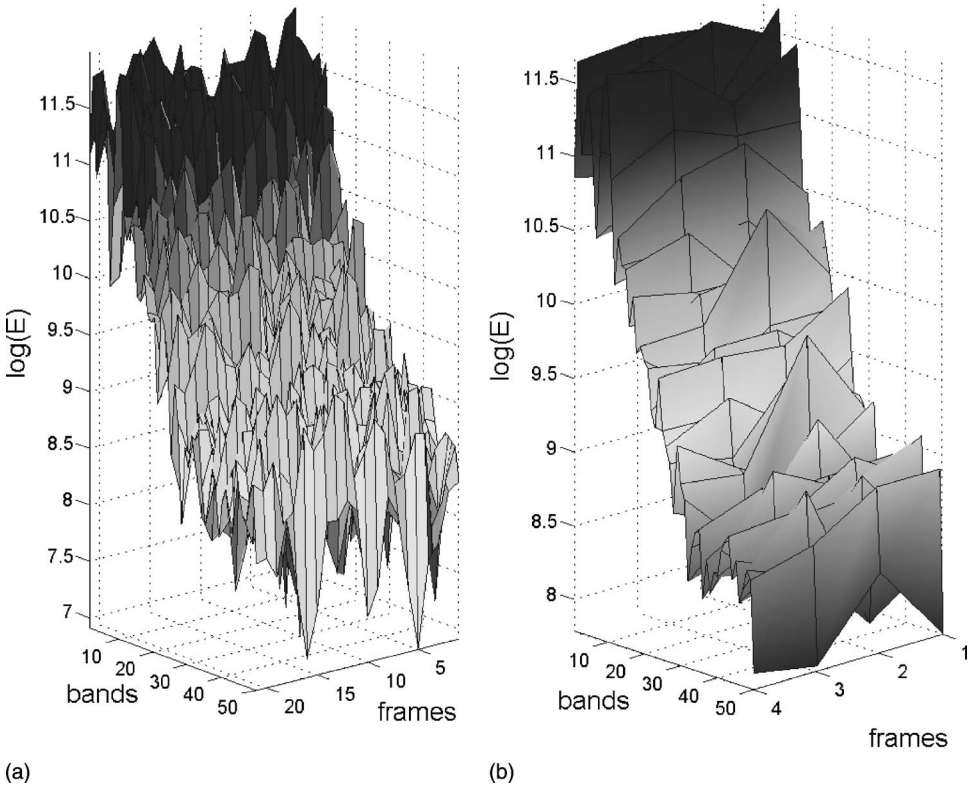


FIG. 1. (a) 20 noise log-energy frames, computed using $N_{\mathrm{FFT}}=256$ and averaged over 50 subbands. (b) Clustering approach to the latter set of frames using hard decision C-means ($C=4$ prototypes).

based clustering) to a prespecified number of prototypes $C<N$, labeled by an integer $i\in\{1,\ldots,C\}$. Thus, we are selecting the generic feature vector as $\mathbf{x}_j\equiv\mathbf{E}_j$.

The similarity measure to be minimized in terms of energy vectors is based on the squared Euclidean distance:

$$d(\mathbf{E}_j,\mathbf{E}_{j'})=\sum_{k=1}^{K}(E(k,j)-E(k,j'))^2=\|\mathbf{E}_j-\mathbf{E}_{j'}\|^2 \tag{6}$$

and can be equivalently defined as[28]

$$J(\mathcal{C})=\tfrac{1}{2}\sum_{i=1}^{C}\sum_{\mathcal{C}(j)=i}\sum_{\mathcal{C}(j')=i}\|\mathbf{E}_j-\mathbf{E}'_j\|^2=\tfrac{1}{2}\sum_{i=1}^{C}\sum_{\mathcal{C}(j)=i}\|\mathbf{E}_j-\mathbf{E}_i\|^2, \tag{7}$$

where $\mathcal{C}(j)=i$ denotes a many-to-one mapping, that assigns the jth observation to the ith prototype and

$$\bar{\mathbf{E}}_i=(\bar{E}(1,i),\ldots,\bar{E}(K,i))^T=\mathrm{mean}(\mathbf{E}_j),$$
$$\forall j,\quad \mathcal{C}(j)=i,\quad i=1,\ldots,C \tag{8}$$

is the mean vector associated with the ith prototype (the sample mean for the ith prototype $\mathbf{m}_i$ defined in Sec. II). Thus, the loss function is minimized by assigning N observations to C prototypes in such a way that within each prototype the average dissimilarity of the observations is minimized. Once convergence is reached, N K-dimensional pause frames are efficiently modeled by C K-dimensional noise prototype vectors denoted by $\bar{\mathbf{E}}_i^{\mathrm{opt}}$, $\quad i=1,\ldots,C$. We call this set of clusters C-partition or noise prototypes since, in this work, the word cluster is assigned to different classes of *labeled data*, that is $\mathbf{K}$ is fixed to 2, i.e., we define two clusters: "noise" and "speech" and the cluster "noise" consists of C prototypes. In Fig. 1 we observed how the complex nature of noise can be simplified (smoothed) using a this clustering approach. The clustering approach speeds the decision function in a significant way since the dimension of feature vectors is reduced substantially ($N\rightarrow C$).

Soft decision function for VAD

In order to classify the second data class (energy vectors of speech frames) we use a basic sequential algorithm scheme, related to Kohonen's leaning vector quantization (LVQ),[31] using a multiple observation (MO) window centered at frame l, as shown in Sec. II. For this purpose let us consider the same dissimilarity measure, a threshold of dissimilarity γ and the maximum clusters allowed $\mathbf{K}=2$.

Let $\hat{\mathbf{E}}(l)$ be the decision feature vector at frame l that is defined on the MO window as follows:

$$\hat{\mathbf{E}}(l)=\max\{\mathbf{E}(j)\},\quad j=l-m,\ldots,l+m. \tag{9}$$

The selection of this envelope feature vector, describing not only a single instantaneous frame but also a $(2m+1)$ entire neighborhood, is useful as it detects the presence of voice beforehand (pause-speech transition) and holds the detection flag, smoothing the VAD decision (as a hangover based algorithm in speech-pause transition[16,17]), as shown in Fig. 2.

Finally, the presence of the second "cluster" (speech frame) is detected if the following ratio holds:

$$\eta(l)=\log\left(1/K\sum_{k=1}^{K}\frac{\hat{E}(k,l)}{<\bar{\mathbf{E}}_i>}\right)>\gamma, \tag{10}$$

where $<\bar{\mathbf{E}}_i>=1/C\Sigma_{i=1}^{C}\bar{\mathbf{E}}_i=1/C\Sigma_{i=1}^{C}\Sigma_{j=1}^{N}\gamma_{ij}\mathbf{E}_j$ is the averaged noise prototype center and γ is the decision threshold.

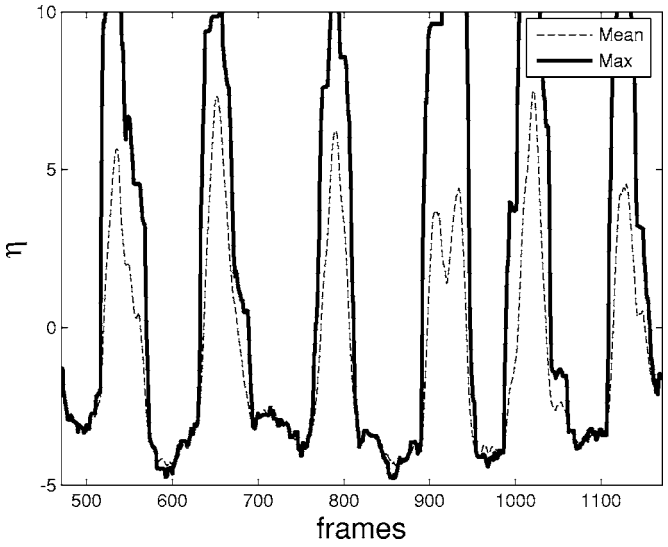


FIG. 2. Decision function in Eq. (10) for two different criteria: energy envelope [Eq. (9)] and energy average.

In order to adapt the operation of the proposed VAD to nonstationary and noise environments, the set of noise prototypes are updated according to the VAD decision during nonspeech periods [not satisfying Eq. (10)] in a competitive manner (only the closer noise prototype is moved towards the current feature vector):

$$\bar{\mathbf{E}}_{i'} = \arg_{\min}(\|\bar{\mathbf{E}}_i - \hat{\mathbf{E}}(l)\|^2) \quad i = 1, \ldots, C$$

$$\Rightarrow \bar{\mathbf{E}}_{i'}^{\text{new}} = \alpha \cdot \bar{\mathbf{E}}_{i'}^{\text{old}} + (1-\alpha) \cdot \hat{\mathbf{E}}(l), \tag{11}$$

where α is a normalized constant. Its value is close to one for a soft decision function (i.e., we selected in simulation $\alpha = 0.99$), that is, uncorrected classified speech frames contributing to the false alarm rate will not affect the noise space model significantly.

V. SOME REMARKS ON THE LTCM VAD ALGORITHM

The main advantage of the proposed algorithm is its ability to deal with on line applications such as DSR systems. The above-mentioned scheme is optimum in computational cost. First, we apply a batch hard C-means to a set of initial pause frames once, obtaining a fair description of the noise subspace and then, using Eq. (11), we move the nearest prototype to the previously detected as silence current frame. Any other on-line approach would be possible but it would be necessary to update the entire set of prototypes for each detected pause frame. In addition, the proposed VAD algorithm belongs to the class of VADs which model noise and apply a distance criterion to detect the presence of speech, i.e., Ref. 17.

A. Selection of an adaptive threshold

In speech recognition experiments (Sec. VI), the selection of the threshold is based on the results obtained in detection experiments [working points in receiving operating curves (ROC) for all conditions]. The working point (selected threshold) should correspond with the best tradeoff between the hit rate and false alarm rate, then the threshold is adaptively chosen depending on the noisy condition.

The VAD makes the speech/nonspeech detection by comparing the unbiased LTCM VAD decision to an adaptive threshold,[32] that is the detection threshold is adapted to the observed noise energy E. It is assumed that the system will work under different noisy conditions characterized by the energy of the background noise. Optimal thresholds (working points) γ_0 and γ_1 can be determined for the system working in the cleanest and noisiest conditions. These thresholds define a linear VAD calibration curve that is used during the initialization period for selecting an adequate threshold as a function of the noise energy E:

$$\gamma = \begin{cases} \gamma_0, & E \le E_0, \\ \dfrac{\gamma_0 - \gamma_1}{E_0 - E_1} + \gamma_0 - \dfrac{\gamma_0 - \gamma_1}{1 - E_1/E_2}, & E_0 < E < E_1, \\ \gamma_1, & E \ge E_1, \end{cases} \tag{12}$$

where E_0 and E_1 are the energies of the background noise for the cleanest and noisiest conditions that can be determined examining the speech databases being used. A high speech/nonspeech discrimination is ensured with this model since silence detection is improved at high and medium SNR levels while maintaining high precision detecting speech periods under high noise conditions.

The algorithm described so far is presented as pseudocode in the following:

(1) Initialize noise model:
  (a) Select N feature vectors $\{\mathbf{E}_j\}$, $j = 1, \ldots, N$.
  (b) Compute threshold γ.

(2) Apply C-means clustering to feature vectors, extracting C noise prototype centers

$$\{\bar{\mathbf{E}}_i\}, \quad i = 1, \ldots, C$$

(3) for l=init to end
  (a) Compute $\hat{\mathbf{E}}(l)$ over the MO window
  (b) if $\eta(l) > \gamma$ [Eq. (10)] than VAD = 1 else VAD = 0 and update noise prototype centers $\{\bar{\mathbf{E}}_i\}$, $i = 1, \ldots, C$ [Eq. (11)].

B. Decision variable distributions

In this section we study the distributions of the decision variable as a function of the long-term window length (m) in order to clarify the motivations for the algorithm proposed. A hand-labeled version of the Spanish SpeechDat-Car (SDC) (Ref. 33) database was used in the analysis. This database contains recordings from close-talking and distant microphones at different driving conditions: (a) stopped car, motor running, (b) town traffic, low speed, rough road, and (c) high speed, good road. The most unfavorable noise environment (i.e., high speed, good road) was selected and recordings from the distant microphone were considered. Thus, the m-order divergence measure between speech and silences was measured during speech and nonspeech periods, and the histogram and probability distributions were built. The 8 kHz input signal was decomposed into overlapping frames

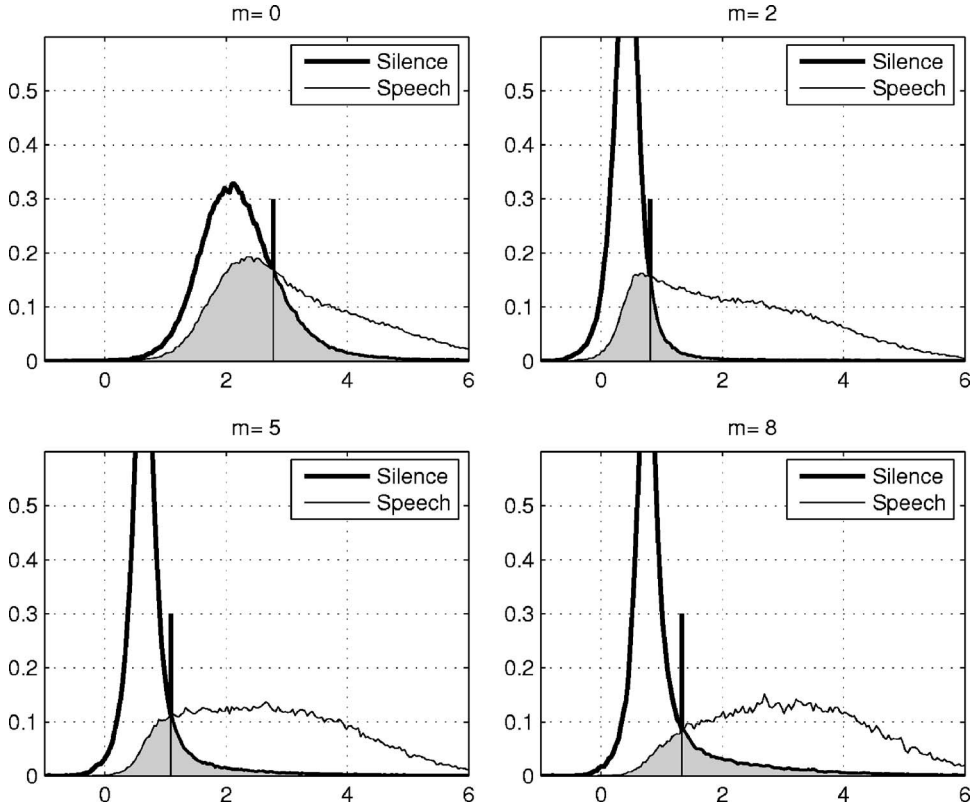


FIG. 3. Speech/nonSpeech distributions and error probabilities of the optimum Bayes classifier for $m=0,2,5$, and 8.

with a 10 ms window shift. Figure 3 shows the distributions of speech and noise for $m=0,2,5$, and 8. It is derived from this that speech and noise distributions are better separated when increasing the order of the long-term window. The noise is highly confined and exhibits a reduced variance, thus leading to high nonspeech hit rates. This fact can be corroborated by calculating the classification error of speech and noise for an optimal Bayes classifier. Figure 4 shows the misclassification errors as a function of the window length m. The speech classification error is approximately divided by three from 32% to 10% when the order of the VAD is increased from 0 to 8 frames. This is motivated by the separation of the distributions that takes place when m is increased as shown in Fig. 3. On the other hand, the increased speech detection robustness is only prejudiced by a moderate increase in the speech detection error. According to Fig. 4, the optimal value of the order of the VAD would be $m=8$. This analysis corroborates the fact that using long-term speech features[32] results beneficial for VAD since they reduce misclassification errors substantially.

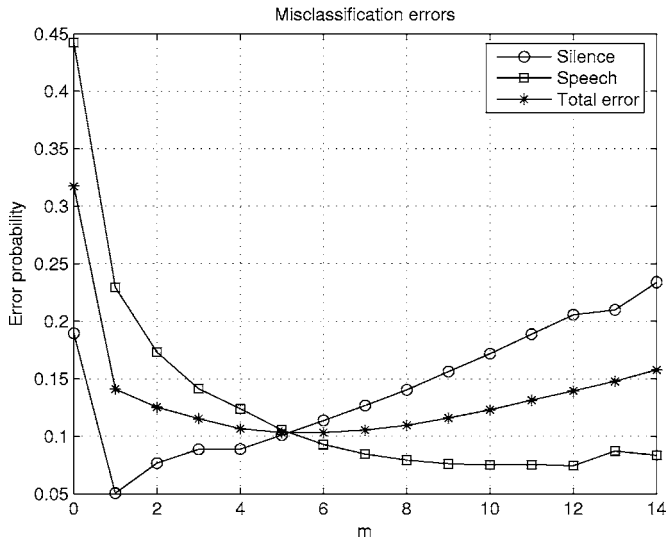


FIG. 4. Probability of error as a function of m.

VI. EXPERIMENTAL RESULTS

Several experiments are commonly carried out in order to assess the performance of VAD algorithms. The analysis is normally focused on the determination of the error probabilities in different noise scenarios and SNR values,[17,34] and the influence of the VAD decision on speech processing systems.[1,14] The experimental framework and the objective performance tests conducted to evaluate the proposed algorithm are described in this section.

A VAD achieves silence compression in modern mobile telecommunication systems reducing the average bit rate by using the discontinuous transmission (DTX) mode. The International Telecommunication Union (ITU) adopted a toll-quality speech coding algorithm known as G.729 to work in combination with a VAD module in DTX mode.[4] The ETSI AMR (Adaptive Multi-Rate) speech coder[3] developed by the Special Mobile Group (SMG) for the GSM system specifies two options for the VAD to be used within the digital cellular telecommunications system. In option 1, the signal is passed through a filterbank and the level of signal in each band is calculated. A measure of the SNR is used to make the VAD decision together with the output of a pitch detector, a tone detector and the correlated complex signal analysis module. An enhanced version of the original VAD is the AMR option

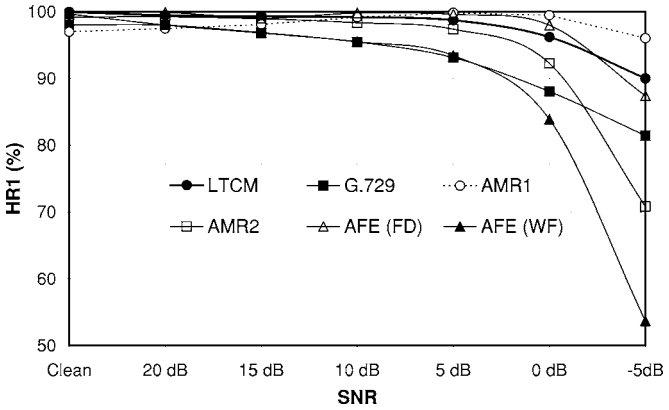


FIG. 5. Speech hit rates (HR1) of standard VADs as a function of the SNR for the AURORA 2 database.

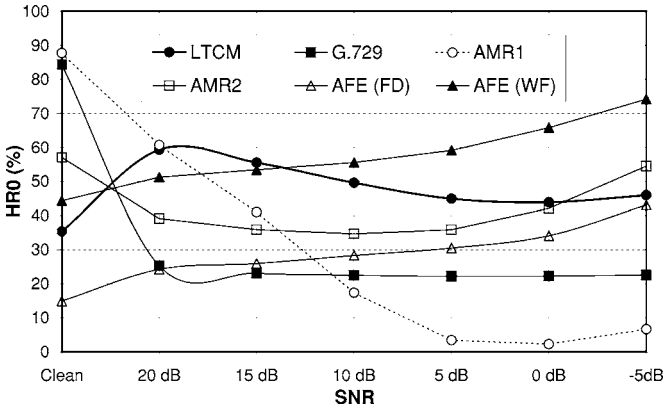


FIG. 7. Nonspeech hit rates (HR0) of standard VADs as a function of the SNR for the AURORA 2 database.

2 VAD which uses parameters of the speech encoder being more robust against environmental noise than AMR1 and G.729. Recently, a new standard incorporating noise suppression methods has been approved by the ETSI for feature extraction and distributed speech recognition (DSR). The so-called advanced front-end (AFE) (Ref. 36) incorporates an energy-based VAD (WF AFE VAD) for estimating the noise spectrum in Wiener filtering speech enhancement, and a different VAD for nonspeech frame dropping (FD AFE VAD).

Recently reported VADs are based on the selection of discriminative speech features, noise estimation and classification methods. Sohn *et al.* showed a decision rule derived from the generalized likelihood ratio test by assuming that the noise statistics are known *a priori*.[12] An interesting approach is the endpoint detection algorithm proposed by Li,[16] which uses optimal FIR filters for edge detection. Other methods track the power spectrum envelope of the signal[17] or use energy thresholds for discriminating between speech and noise.[15]

A. Evaluation under different noise environments

First, the proposed VAD was evaluated in terms of the ability to discriminate between speech and nonspeech in different noise scenarios and at different SNR levels. The AURORA 2 database[35] is an adequate database for this analysis since it is built on the clean Tldigits database that consists of sequences of up to seven connected digits spoken by American English talkers as source speech, and a selection of eight different real-world noises that have been artificially added to the speech at SNRs of 20 dB, 15 dB, 10 dB, 5 dB, 0 dB, and −5 dB. These noisy signals have been recorded at different places (suburban train, crowd of people (babble), car, exhibition hall, restaurant, street, airport, and train station), and were selected to represent the most probable application scenarios for telecommunication terminals. In the discrimination analysis, the clean Tldigits database was used to manually label each utterance as speech or nonspeech on a frame by frame basis for reference. Detection performance is then assessed in terms of the speech pause hit-rate (HR0) and the speech hit-rate (HR1) defined as the fraction of all actual pause or speech frames that are correctly detected as pause or speech frames, respectively,

$$HR1 = \frac{N_{1,1}}{N_1^{\text{ref}}}, \quad HR0 = \frac{N_{0,0}}{N_0^{\text{ref}}}, \tag{13}$$

where N_1^{ref} and N_0^{ref} are the number of real nonspeech and speech frames in the whole database and $N_{1,1}$ and $N_{0,0}$ are the number of real speech and nonspeech frames correctly classified, respectively.

Figures 5–8 provide comparative results of this analysis and compare the proposed VAD to standardized algorithms including the ITU-T G.729,[4] ETSI AMR,[3] and ETSI AFE

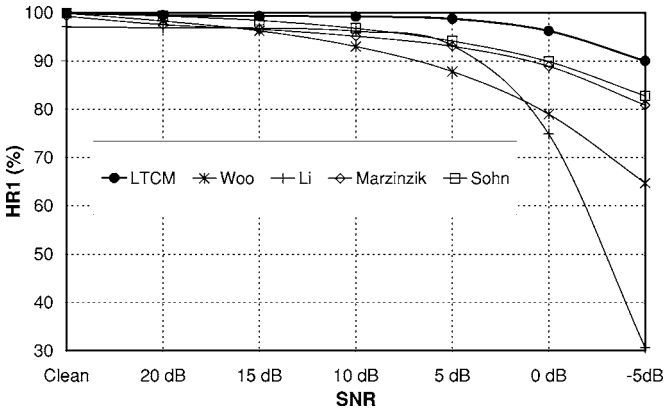


FIG. 6. Speech hit rates (HR1) of other VADs as a function of the SNR for the AURORA 2 database.

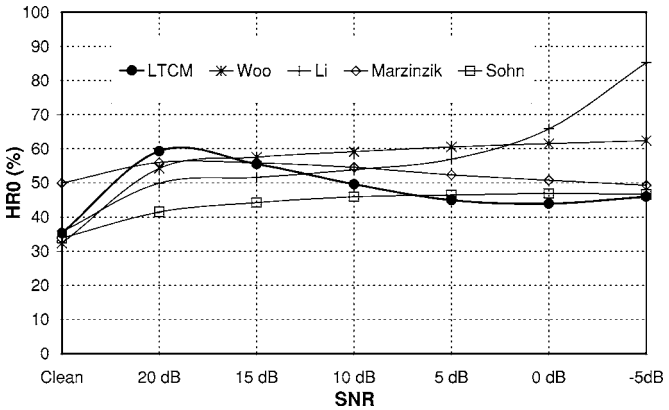


FIG. 8. Nonspeech hit rates (HR0) of other VADs as a function of the SNR for the AURORA 2 database.

TABLE II. Average speech/nonpeech hit rates for SNRs between clean conditions and −5 dB. Comparison to (a) standardized VADs and (b) other VAD methods.

			(a)			
	G.729	AMR1	AMR2	AFE (WF)	AFE (FD)	**LTCM**
HR0 (%)	31.77	31.31	42.77	57.68	28.74	**47.81**
HR1(%)	93.00	98.18	93.76	88.72	97.70	**97.57**
			(b)			
		Sohn	Woo	Li	Marzinzik	**LTCM**
HR0 (%)		43.66	55.40	57.03	52.69	**47.81**
HR1 (%)		94.46	88.41	83.65	93.04	**97.57**

(Ref. 36) in terms of the nonspeech hit-rate (HR0, Fig. 7) and speech hit-rate (HR1, Fig. 5) for clean conditions and SNR levels ranging from 20 to −5 dB. Note that results for the two VADs defined in the AFE DSR standard[36] for estimating the noise spectrum in the Wiener filtering (WF) stage and nonspeech frame dropping (FD) are provided. The results shown in these figures are averaged values for the entire set of noises.

It can be derived from Figures 7 and 5 that (i) ITU-T G.729 VAD suffers poor speech detection accuracy with the increasing noise level while nonspeech detection is good in clean conditions (85%) and poor (20%) in noisy conditions, (ii) ETSI AMR1 yields an extreme conservative behavior with high speech detection accuracy for the whole range of SNR levels but very poor nonspeech detection results at increasing noise levels. Although AMR1 seems to be well suited for speech detection at unfavorable noise conditions, its extremely conservative behavior degrades its nonspeech detection accuracy being HR0 less than 10% below 10 dB, making it less useful in a practical speech processing system, (iii) ETSI AMR2 leads to considerable improvements over G.729 and AMR1 yielding better nonspeech detection accuracy while still suffering fast degradation of the speech detection ability at unfavorable noisy conditions, (iv) The VAD used in the AFE standard for estimating the noise spectrum in the Wiener filtering stage is based in the full energy band and yields a poor speech detection performance with a fast decay of the speech hit rate at low SNR values. On the other hand, the VAD used in the AFE for frame dropping achieves a high accuracy in speech detection but moderate results in nonspeech detection, and (v) LTCM yields the best compromise among the different VADs tested. It obtains a good behavior in detecting nonspeech periods as well as exhibiting a slow decay in performance at unfavorable noise conditions in speech detection (90% at −5 dB).

Figures 6 and 8 compare the proposed VAD to a representative set of recently published VAD method.[12,15–17] It is worthwhile clarifying that the AURORA 2 database consists of recordings with very short nonspeech periods between digits and, consequently, it is more important to classify speech correctly than nonspeech in a speech recognition system. This is the reason to define a VAD method with a high speech hit rate even in very noisy conditions. Table II summarizes the advantages provided by LTCM VAD over the different VAD methods in terms of the average speech/nonspeech hit rates (over the entire range of SNR values). Thus, the proposed method with a 97.57% mean HR1 and a 47.81% mean HR0 yields the best trade-off in speech/nonspeech detection.

B. Receiver operating characteristic (ROC) curves

An additional test was conducted to compare speech detection performance by means of the ROC curves, a frequently used methodology in communications based on the hit and error detection probabilities,[17,29,37] that completely describes the VAD error rate. The AURORA subset of the Spanish SDC database[33] was used in this analysis. This database contains 4914 recordings using close-talking and distant microphones from more than 160 speakers. As in the whole SDC database, the files are categorized into three noisy conditions: quiet, low noise, and high noise conditions, which represent different driving conditions and average SNR values of 12 dB, 9 dB, and 5 dB. Thus, recordings from the close-talking microphone are used in the analysis to label speech/pause frames for reference, while recordings from the distant microphone are used for the evaluation of different VADs in terms of their ROC curves. The speech pause hit rate (HR0) and the false alarm rate (FAR0=100 -HR1) were determined in each noise condition for the proposed VAD and the G.729, AMR1, AMR2, and AFE VADs, which were used as a reference. For the calculation of the false-alarm rate as well as the hit rate, the "real" speech frames and "real" speech pauses were determined using the hand-labeled database on the close-talking microphone.

The sensitivity of the proposed method to the number of clusters used to model the noise space was studied. It was found experimentally that the behavior of the algorithm is almost independent of C, using a number of subbands $K=10$. Figure 9 shows that the accuracy of the algorithm (noise detection rate versus false alarm rate) in speech-pause discrimination is not affected by the number of prototypes selected as long as $C \geq 2$, thus the benefits of the clustering approach are evident. Note that the objective of the VAD is to work as close as possible to the upper left corner in this figure where speech and silence is classified with no errors. The effect of the number of subbands used in the algorithm is plotted in Fig. 10. The use of a complete energy average

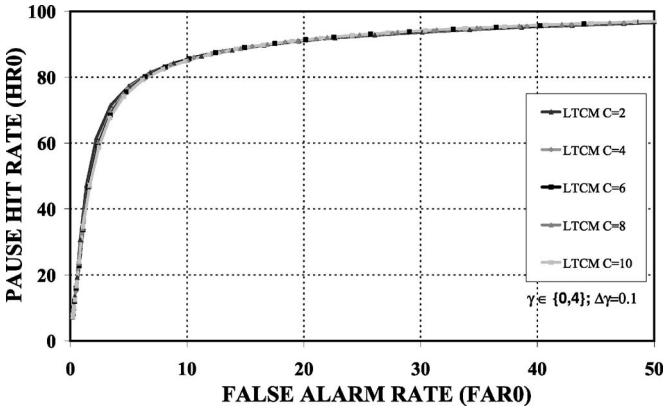


FIG. 9. ROC curves in high noisy conditions for different number of noise prototypes. The DFT was computed with $N_{\mathrm{FFT}}=256$, $K=10$ log-energy subbands were used to build features vectors and the MO-window contained $2\cdot m+1$ frames ($m=10$).

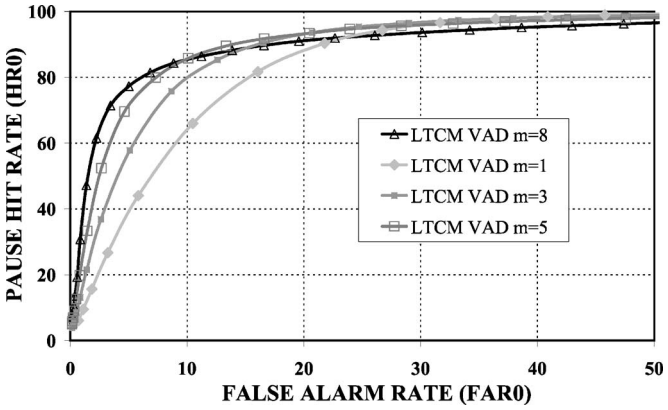


FIG. 11. Selection of the number of m (high, high speed, good road, 5 dB average SNR, $K=32$, $C=2$).

($K=1$) or raw data ($K=100$) reduces the effectiveness of the clustering procedure making its accuracy equivalent to other proposed VADs.

Figure 11 shows the speech pause hit rate (HR0) as a function of the false alarm rate (FAR0=100-HR1) of the proposed LTCM VAD for different values of the decision threshold and different values of the number of observations m. It is shown how increasing the number of observations (m) leads to better speech/nonspeech discrimination with a shift-up and to the left of the ROC curve in the ROC space. This enables the VAD to work closer to the "ideal" working point (HR0=100%, FAR0=0%) where both speech and nonspeech are classified ideally with no errors. These results are consistent with our preliminary experiments and the results shown in Figs. 3 and 4 that expected a minimum error rate for m close to eight frames.

Figure 12 shows the ROC curves of the proposed VAD and other reference VAD algorithms[12,15–17] for recordings from the distant microphone in high noisy conditions. The working points of the ITU-T G.729, ETSI AMR, and ETSI AFE VADs are also included. The results show improvements in detection accuracy over standardized VADs and over a representative set of VAD algorithms.[12,15–17] Among all the VAD examined, our VAD yields the lowest false alarm rate for a fixed nonspeech hit rate and also, the highest nonspeech hit rate for a given false alarm rate. The benefits are especially important over ITU-T G.729,[4] which is used along with a speech codec for discontinuous transmission, and over the[16] algorithm, that is based on an optimum linear filter for edge detection. The proposed VAD also improves Marzinzik[17] VAD that tracks the power spectral envelopes, and the Sohn[12] VAD, that formulates the decision rule by means of a statistical likelihood ratio test (LRT) defined on the power spectrum of the noisy signal.

It is worthwhile mentioning that the experiments described above yield a first measure of the performance of the VAD. Other measures of VAD performance that have been reported are the clipping errors.[38] These measures provide valuable information about the performance of the VAD and can be used for optimizing its operation. Our analysis does not distinguish between the frames that are being classified and assesses the hit rates and false alarm rates for a first performance evaluation of the proposed VAD. On the other hand, the speech recognition experiments conducted later on the AURORA databases will be a direct measure of the quality of the VAD and the application it was designed for. Clipping errors are evaluated indirectly by the speech recognition system since there is a high probability of a deletion error occurring when part of the word is lost after frame dropping.

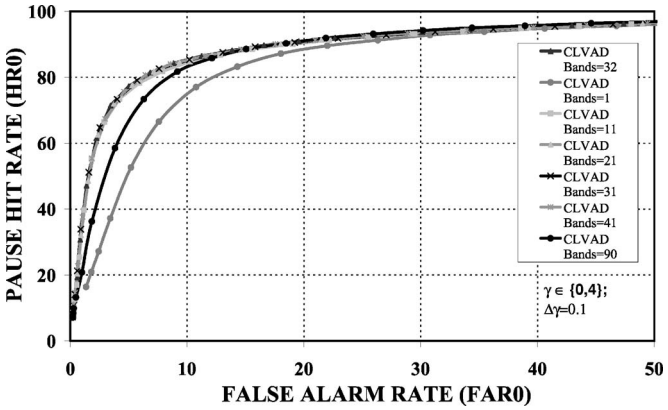


FIG. 10. ROC curves in high noisy conditions for different number of subbands. $N_{\mathrm{FFT}}=256$; $C=10$ and $m=10$.

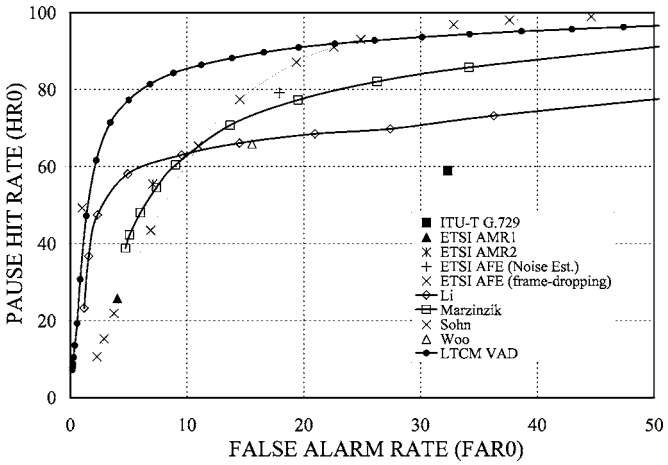


FIG. 12. ROC curves for comparison to standardized and other VAD methods (high, high speed, good road, 5 dB average SNR, $K=32$, $C=2$).

C. Assessment of the VAD on an ASR system

Although the discrimination analysis or the ROC analysis presented in the preceding section are effective to evaluate a given speech/nonspeech discrimination algorithm, the influence of the VAD in a speech recognition system was also studied. Many authors claim that VADs are well compared by evaluating speech recognition performance[15] since nonefficient speech/nonspeech discrimination is an important performance degradation source for speech recognition systems working in noisy environments.[1] There are two clear motivations for that: (i) noise parameters such as its spectrum are updated during nonspeech periods being the speech enhancement system strongly influenced by the quality of the noise estimation, and (ii) frame dropping, a frequently used technique in speech recognition to reduce the number of insertion errors caused by the acoustic noise, is based on the VAD decision and speech misclassification errors lead to loss of speech, thus causing irrecoverable deletion errors.

The reference framework (Base) is the distributed speech recognition (DSR) front-end[39] proposed by the ETSI STQ working group for the evaluation of noise robust DSR feature extraction algorithms. The recognition system is based on the HTK (Hidden Markov Model Toolkit) software package.[40] The task consists in recognizing connected digits which are modeled as whole word HMMs (Hidden Markov Models) with the following parameters: 16 states per word, simple left-to-right models, mixture of 3 Gaussians per state and only the variances of all acoustic coefficients (no full covariance matrix), while speech pause models consist of three states with a mixture of six Gaussians per state. The 39-parameter feature vector consists of 12 cepstral coefficients (without the zero-order cepstral coefficient), the logarithmic frame energy plus the corresponding derivatives (Δ) and acceleration ($\Delta\Delta$) coefficients.

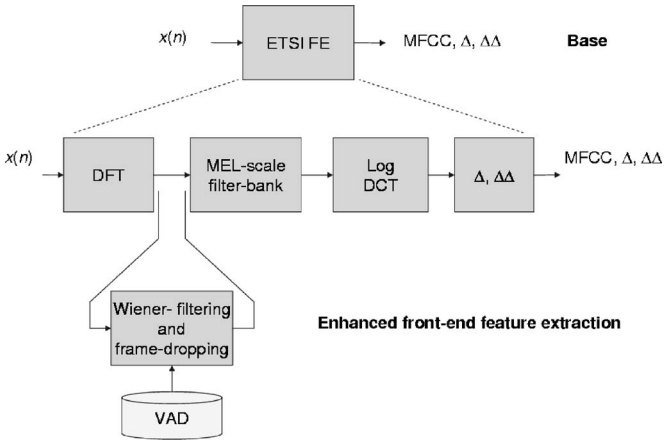


FIG. 13. Speech recognition experiments. Front-end feature extraction.

Two training modes are defined for the experiments conducted on the AURORA 2 database: (i) training on clean data only (Clean Training), and (ii) training on clean and noisy data (Multi-Condition Training). For the AURORA 3 SpeechDat-Car databases, the so-called well-matched (WM), medium-mismatch (MM) and high-mismatch (HM) conditions are used. AURORA 3 databases contain recordings from the close-talking and distant microphones. In WM condition, both close-talking and hands-free microphones are used for training and testing. In MM condition, both training and testing are performed using the hands-free microphone recordings. In HM condition, training is done using close-

TABLE III. Average word accuracy for the AUR0RA 2 database. (a) Clean training. (b) Multicondition training.

(a)

		Base + WF					Base + WF + FD				
	Base	G.729	AMR1	AMR2	AFE	**LTCM**	G.729	AMR1	AMR2	AFE	**LTCM**
Clean	99.03	98.81	98.80	98.81	98.77	98.88	98.41	97.87	98.63	98.78	99.18
20 dB	94.19	87.70	97.09	97.23	97.68	97.46	83.46	96.83	96.72	97.82	98.05
15 dB	85.41	75.23	92.05	94.61	95.19	95.14	71.76	92.03	93.76	95.28	96.10
10 dB	66.19	59.01	74.24	87.50	87.29	88.71	59.05	71.65	86.36	88.67	90.71
5 dB	39.28	40.30	44.29	71.01	66.05	72.48	43.52	40.66	70.97	71.55	75.82
0 dB	17.38	23.43	23.82	41.28	30.31	42.91	27.63	23.88	44.58	41.78	47.01
−5 dB	8.65	13.05	12.09	13.65	4.97	15.34	14.94	14.05	18.87	16.23	19.88
Average	60.49	57.13	66.30	78.33	75.30	79.34	57.08	65.01	78.48	79.02	**81.54**

(b)

		Base + WF					Base + WF + FD				
	Base	G.729	AMRl	AMR2	AFE	**LTCM**	G.729	AMRl	AMR2	AFE	**LTCM**
Clean	98.48	98.16	98.30	98.51	97.86	98.45	97.50	96.67	98.12	98.39	98.78
20 dB	97.39	93.96	97.04	97.86	97.60	97.93	96.05	96.90	97.57	97.98	98.41
15 dB	96.34	89.51	95.18	96.97	96.56	97.06	94.82	95.52	96.58	96.94	97.61
10 dB	93.88	81.69	91.90	94.43	93.98	94.64	91.23	91.76	93.80	93.63	95.39
5 dB	85.70	68.44	80.77	87.27	86.41	87.54	81.14	80.24	85.72	85.32	88.40
0 dB	59.02	42.58	53.29	65.45	64.63	66.23	54.50	53.36	62.81	63.89	66.92
−5 dB	24.47	18.54	23.47	30.31	28.78	31.21	23.73	23.29	27.92	30.80	32.91
Average	86.47	75.24	83.64	88.40	87.84	**88.68**	83.55	83.56	87.29	87.55	**89.35**

TABLE IV. Average word acccuracy for clean and multicondition AURORA 2 training/testing experiments. Comparison to (a) standard VADs and (b) recently presented VAD methods.

	(a)					
	G.729	AMR1	AMR2	AFE	**LTCM**	Hand-labeling
Base + WF	66.19	74.97	83.37	81.57	**84.01**	84.69
Base + WF+FD	70.32	74.29	82.89	83.29	**85.44**	86.86
	(b)					
	Woo	Li	Marzinzik	Sohn	**LTCM**	Hand-labeling
Base + WF	83.64	77.43	84.02	83.89	**84.01**	84.69
Base + WF+ FD	81.09	82.11	85.23	83.80	**85.44**	86.86

talking microphone material from all driving conditions while testing is done using hands-free microphone material taken for low noise and high noise driving conditions. Finally, recognition performance is assessed in terms of the word accuracy (WAcc) which takes into account the number of substitution errors (S), deletion errors (D), and insertion errors (I),

$$\mathrm{WAcc}(\%) = \frac{N - D - S - I}{N} \times 100\,\%\,, \tag{14}$$

where N is the total number of words in the testing database.

The influence of the VAD decision on the performance of different feature extraction schemes was studied. The first approach (shown in Fig. 13) incorporates Wiener filtering (WF) to the Base system as noise suppression method. The second feature extraction algorithm that was evaluated uses Wiener filtering and nonspeech frame dropping. The algorithm has been implemented as described for the first stage of the Wiener filtering noise reduction system present in the advanced front-end AFE DSR standard.[36] The same feature extraction scheme was used for training and testing and no other mismatch reduction techniques already present in the AFE standard (wave form processing or blind equalization) have been considered since they are not affected by the VAD decision and can mask the impact of the VAD on the overall system performance.

Table III shows the AURORA 2 recognition results as a function of the SNR for speech recognition experiments based on the G.729, AMR, AFE, and LTCM VAD algorithms. These results were averaged over the three test sets of the AURORA 2 recognition experiments. Notice that, particularly, for the recognition experiments based on the AFE VADs, we have used the same configuration used in the standard[36] with different VADs for WF and FD. Only exact speech periods are kept in the FD stage and consequently, all the frames classified by the VAD as nonspeech are discarded. FD has impact on the training of silence models since less nonspeech frames are available for training. However, if FD is effective enough, few nonspeech periods will be handled by the recognizer in testing and consequently, the silence models will have little influence on the speech recognition performance. As a conclusion, the proposed VAD outperforms the standard G.729, AMR1, AMR2, and AFE VADs when used for WF and also, when the VAD is used for removing nonspeech frames. Note that the VAD decision is used in the WF stage for estimating the noise spectrum during nonspeech periods, and a good estimation of the SNR is critical for an efficient application of the noise reduction algorithm. In this way, the energy-based WF AFE VAD suffers fast performance degradation in speech detection as shown in Fig. 5, thus leading to numerous recognition errors and the corresponding increase of the word error rate, as shown in Table III. On the other hand, FD is strongly influenced by the performance of the VAD and an efficient VAD for robust speech recognition needs a compromise between speech and nonspeech detection accuracy. When the VAD suffers a rapid performance degradation under severe noise conditions it loses too many speech frames and leads to numerous deletion errors; if the VAD does not correctly identify nonspeech periods it causes numerous insertion errors and the corresponding FD performance degradation. The best recognition performance is obtained when the proposed LTCM VAD is used for WF and FD. Note that FD yields better results for the speech recognition system trained on clean speech. This is motivated by the fact that models trained using clean speech do not adequately model noise processes, and normally cause insertion errors during nonspeech periods. Thus, removing efficiently speech pauses will lead to a significant reduction of this error source. On the other hand, noise is well modeled when models are trained using noisy speech and the speech recognition system tends itself to reduce the number of insertion errors in multicondition training as shown in Table III, part (a).

Table IV, part (a), compares the word accuracies aver-

TABLE V. Average word accuracy (%) for the Spanish, SDC database.

	Base	Woo	Li	Marzinzik	Sohn	G729	AMRl	AMR2	AFE	**LTCM**
WM	92.94	95.35	91.82	94.29	96.07	88.62	94.65	95.67	95.28	96.41
MM	83.31	89.30	77.45	89.81	91.64	72.84	80.59	90.91	90.23	91.61
HM	51.55	83.64	78.52	79.43	84.03	65.50	62.41	85.77	77.53	86.20
Avg.	**75.93**	89.43	82.60	87.84	90.58	75.65	74.33	90.78	87.68	**91.41**

aged for clean and multicondition training modes to the upper bound that could be achieved when the recognition system benefits from using the hand-labeled database. These results show that the performance of the proposed algorithm is very close to that of the reference database. In all the test sets, the proposed VAD algorithm outperforms standard VADs obtaining the best results followed by AFE, AMR2, AMR1, and G.729. Table IV, part (b), extends this comparison to other recently presented VAD methods.[12,15–17]

Table V shows the recognition performance for the Spanish SpeechDat-Car database when WF and FD are performed on the base system.[39] Again, the VAD outperforms all the algorithms used for reference yielding relevant improvements in speech recognition. Note that, these particular databases used in the AURORA 3 experiments have longer nonspeech periods than the AURORA 2 database and then, the effectiveness of the VAD results more important for the speech recognition system. This fact can be clearly shown when comparing the performance of the proposed VAD to Marzinzik[17] VAD. The word accuracies of both VADs are quite similar for the AURORA 2 task. However, the proposed VAD yields a significant performance improvement over Marzinzik[17] VAD for the AURORA 3 database.

VII. CONCLUSION

A new algorithm for improving speech detection and speech recognition robustness in noisy environments is shown. The proposed LTCM VAD is based on noise modeling using hard *C*-means clustering and employs long-term speech information for the formulation of a soft decision rule based on an averaged energy ratio. The VAD performs an advanced detection of beginnings and delayed detection of word endings which, in part, avoids having to include additional hangover schemes or noise reduction blocks. It was found that increasing the length of the long-term window yields to a reduction of the class distributions and leads to a significant reduction of the classification error. An exhaustive analysis conducted on the AURORA database showed the effectiveness of this approach. The proposed LTCM VAD outperformed recently reported VAD methods including Sohn's VAD, that defines a likelihood ratio test on a single observation, and the standardized ITU-T G.729, ETSI AMR for the GSM system and ETSI AFE VADs for distributed speech recognition. On the other hand, it also improved the recognition rate when the VAD is used for noise spectrum estimation, noise reduction and frame dropping in a noise robust ASR system.

ACKNOWLEDGMENTS

This work has received research funding from the EU 6th Framework Programme, under Contract No. IST-2002-507943 (HIWIRE, Human Input that Works in Real Environments) and SESIBONN and SR3-VoIP projects (TEC2004-06096-C03-00, TEC2004-03829/TCM) from the Spanish government. The views expressed here are those of the authors only. The Community is not liable for any use that may be made of the information contained therein.

[1] L. Karray and A. Martin, "Towards improving speech detection robustness for speech recognition in adverse environments," Speech Commun. **43**, 261–276 (2003).
[2] J. Ramírez, J. C. Segura, M. C. Benítez, A. de la Torre, and A. Rubio, "A new adaptive long-term spectral estimation voice activity detector," Proceedings of EUROSPEECH 2003, Geneva, Switzerland, 2003, pp. 3041–3044.
[3] ETSI, "Voice activity detector (VAD) tor Adaptive Multi-Rate (AMR) speech traffic channels," ETSI EN 301 708 Recommendation, 1999.
[4] ITU, "A silence compression scheme for G.729 optimized for terminals conforming to recommendation V.70," ITU-T/Recommendation G.729-Annex B, 1996.
[5] L. Krasny, "Soft-decision speech signal estimation," J. Acoust. Soc. Am. **108**, 2575 (2000).
[6] P. S. Veneklassen and J. P. Christoff, "Speech detection in noise," J. Acoust. Soc. Am. **32**, 1502 (1960).
[7] A. Sangwan, M. C. Chiranth, H. S. Jamadagni, R. Sah, R. V. Prasad, and V. Gaurav, "VAD techniques for real-time speech transmission on the Internet," IEEE International Conference on High-Speed Networks and Multimedia Communications, 2002, pp. 46–50.
[8] F. Basbug, K. Swaminathan, and S. Nandkumar, "Noise reduction and echo cancellation front-end for speech codecs," IEEE Trans. Speech Audio Process. **11**, 1–13 (2003).
[9] Y. D. Cho and A. Kondoz, "Analysis and improvement of a statistical model-based voice activity detector," IEEE Signal Process. Lett. **8**, 276–278 (2001).
[10] S. Gazor and W. Zhang, "A soft voice activity detector based on a Laplacian-Gaussian model," IEEE Trans. Speech Audio Process. **11**, 498–505 (2003).
[11] L. Armani, M. Matassoni, M. Omologo, and P. Svaizer, "Use of a CSP-based voice activity detector for distant-talking ASR," Proceedings of EUROSPEECH 2003, Geneva, Switzerland, 2003, pp. 501–504.
[12] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," IEEE Signal Process. Lett. **16**, 1–3 (1999).
[13] I. Potamitis and E. Fishier, "Speech activity detection and enhancement of a moving speaker based on the wideband generalized likelihood ratio and microphone arrays," J. Acoust. Soc. Am. **116**, 2406–2415 (2004).
[14] R. L. Bouquin-Jeannes and G. Faucon, "Study of a voice activity detector and its influence on a noise reduction system," Speech Commun. **16**, 245–254 (1995).
[15] K. Woo, T. Yang, K. Park, and C. Lee, "Robust voice activity detection algorithm for estimating noise spectrum," Electron. Lett. **36**, 180–181 (2000).
[16] Q. Li, J. Zheng, A. Tsai, and Q. Zhou, "Robust endpoint detection and energy normalization for real-time speech and speaker recognition," IEEE Trans. Speech Audio Process. **10**, 146–157 (2002).
[17] M. Marzinzik, and B. Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics," IEEE Trans. Speech Audio Process. **10**, 341–351 (2002).
[18] R. Chengalvarayan, "Robust energy normalization using speech/nonspeech discriminator for German connected digit recognition," Proceedings of EUROSPEECH 1999, Budapest, Hungary, 1999, pp. 61–64.
[19] R. Tucker, "Voice activity detection using a periodicity measure," IEE Proc.-Commun. **139**, 377–380 (1992).
[20] S. G. Tanyer and H. Özer, "Voice activity detection in nonstationary noise," IEEE Trans. Speech Audio Process. **8**, 478–482 (2000).
[21] M. R. Anderberg, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, *Cluster Analysis for Applications* (Academic, New York, 1973).
[22] A. Jain and P. Flynn, "Image segmentation using clustering," in *In Advances in Image Understanding. A Festschrift for Azriel Rosenfeld*, edited by N. Ahuja and K. Bowyer, (IEEE, 1996), pp. 65–83.
[23] E. Rasmussen, "Clustering algorithms," in *Information Retrieval: Data Structures and Algorithms*, edited by W. B. Frakes and R. Baeza-Yates (Prentice-Hall, Upper Saddle River, NJ, 1992), pp. 419–442.
[24] G. Salton, "Developments in automatic text retrieval," Science **109**, 974–980 (1991).
[25] A. Jain and R. Dubes, *Algorithms for Clustering Data*, Prentice-Hall advanced reference series (Prentice-Hall, Upper Saddle River, NJ, 1988).
[26] D. Fisher, "Knowledge acquisition via incremental conceptual clustering," Mach. Learn. **2**, 139–172 (1987).
[27] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Berkeley, 1967).

[28]T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference, and Prediction Series*, Springer Series in Statistics, 1st ed. (Springer, New York, 2001).
[29]J. Ramírez, J. C. Segura, C. Benítez, A. de la Torre, and A. Rubio, "An effective subband OSF-based VAD with noise reduction for robust speech recognition," IEEE Trans. Speech Audio Process. **13**, 1119–1129 (2005).
[30]J. M. Górriz, J. Ramírez, J. C. Segura, and C. G. Puntonet, "Improved MO-LRT VAD based on bispectra Gaussian model," Electron. Lett. **41**, 877–879 (2005).
[31]T. Kohonen, *Self Organizing and Associative Memory*, 3rd ed. (Springer-Verlag, Berlin, 1989).
[32]J. Ramírez, J. C. Segura, M. C. Benítez, A. de la Torre, and A. Rubio, "Efficient voice activity detection algorithms using long-term speech information," Speech Commun. **42**, 271–287 (2004).
[33]A. Moreno, L. Borge, D. Christoph, R. Gael, C. Khalid, E. Stephan, and A. Jeffrey, "SpeechDat-Car: A Large Speech Database for Automotive Environments," Proceedings of the II LRCE Conference, 2000.
[34]F. Beritelli, S. Casale, G. Rugeri, and S. Serrano, "Performance evaluation and comparison of G.729/AMR/Fuzzy voice activity detectors," IEEE Signal Process. Lett. **9**, 85–88 (2002).
[35]H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noise conditions," ISCA ITRW ASR2000 Automatic Speech Recognition: Challenges for the Next Millennium, Paris, France, 2000.
[36]ETSI, "Speech processing, transmission and quality aspects (STQ); distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithms," ETSI ES 202 050 Recommendation, 2002.
[37]J. M. Górriz, J. Ramírez, C. G. Puntonet, and J. C. Segura, "Generalized LRT-based Voice Activity Detector," IEEE Signal Process. Lett. (to be published).
[38]A. Benyassine, E. Shlomot, H. Su, D. Massaloux, C. Lamblin, and J. Petit, "ITU-T Recommendation G.729 Annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications," IEEE Commun. Mag. **35**, 64–73 (1997).
[39]ETSI, "Speech processing, transmission and quality aspects (stq); distributed speech recognition; front-end feature extraction algorithm; compression algorithms," ETSI ES 201 108 Recommendation, 2000.
[40]S. Young, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, *The HTK Book* (Cambridge University Press, Cambridge, 1997).

The violin bridge as filter

George Bissinger[a)]
Physics Department, East Carolina University, Greenville, North Carolina 27858

(Received 10 March 2006; revised 1 May 2006; accepted 2 May 2006)

The violin bridge filter role was investigated using modal and acoustic measurements on 12 quality-rated violins combined with systematic bridge rocking frequency f_{rock} and wing mass decrements Δm on four bridges for two other violins. No isolated bridge resonances were observed; bridge motions were complex (including a "squat" mode near 0.8 kHz) except for low frequency rigid body pivot motions, all more or less resembling rocking motion at higher frequencies. A conspicuous broad peak near 2.3 kHz in bridge driving point mobility (labeled BH) was seen for good and bad violins. Similar structure was seen in averaged bridge, bridge feet, corpus mobilities and averaged radiativity. No correlation between violin quality and BH driving point, averaged corpus mobility magnitude, or radiativity was found. Increasing averaged-over-f_{rock} Δm(g) from 0 to 0.12 generally increased radiativity across the spectrum. Decreasing averaged-over-Δm f_{rock} from 3.6 to 2.6 kHz produced consistent decreases in radiativity between 3 and 4.2 kHz, but only few-percent decreases in BH frequency. The lowest f_{rock} values were accompanied by significantly reduced radiation from the Helmholtz A0 mode near 280 Hz; this, combined with reduced high frequency output, created overall radiativity profiles quite similar to "bad" violins among the quality-rated violins. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2207576]

PACS number(s): 43.75.De [NHF] Pages: 482–491

I. INTRODUCTION

> "*It is difficult to imagine the reason of this; how it is that a little piece of maple, which merely serves to keep the strings off the finger-board, should have such a powerful effect on the tone of the instrument to which it is not fastened in any way, being merely kept in place by the pressure of the four strings*"
>
> (Heron-Allen, 1884).

The bridge—a seemingly minor, ~0.002 kg substructure on top of a ~0.4 kg violin—has been considered a vital ingredient of good violin tone for centuries. In his remarkable summary book of everything known about violins and their construction up to 1884, Heron-Allen then proceeded to provide an answer. *"The first explanation of this influence must be sought for in the fact that it is the principal channel by which the vibration of the strings pass, to the belly…, and to the back…."*[1] Obviously the prominent role of the bridge as the energy "gatekeeper" has been recognized for a long time. It is the first of the two primary, independent components of violin sound. The second filter component—the subsequent conversion of corpus vibrational energy to acoustic energy—has now been addressed quite generally from the behavior of the violin normal mode radiation efficiency over the audible range, the ratio of radiation damping to total damping, and an effective critical frequency for the violin.[2,3]

Numerous experimental studies of bridge motion have been published over the years,[4,5] but one aspect of the bridge has dominated discussion over the last 30 years or so, a broad peak in driving point mobility (input admittance) first observed as an impedance minimum by Reinicke,[6] the so-called "bridge-hill" near 3 kHz thought to be linked to the first in-plane rocking mode of the clamped-foot bridge near that frequency. Little work has been done however on the bridge's actual dynamic mechanical behaviors in relationship to corpus vibrations and subsequent radiation from the violin.

Jansson and co-workers have been leaders in experimental analysis of the bridge and its effects on violin response for many years,[7–13] stating early on that there was a correlation between the quality of a violin and the prominence of the bridge-hill. Of late their attention has turned to quantifying the effect of bridge modifications on the driving point mobility spectrum. In an especially revealing experiment they replaced the standard bridge (with heart and hip cutouts) by a solid bridge, which increased the rocking mode frequency from ~3 to ~8 kHz. Surprisingly the bridge-hill remained near 2.5 kHz.[12] (In a less declaratory experiment Reinicke in 1973 observed no significant movement in the impedance minimum when wedges were inserted into the side slots to stiffen the bridge.[6]) For the first time an experiment clearly showed that the bridge was not the predominant influence on the "bridge-hill," hence future references will be to the BH peak.

Following this experiment Beldie modeled the generalized effects of localized top plate stiffness in addition to bridge stiffness, concluding that material properties of the top plate where the bridge feet rest dominated measured BH behavior, with relatively minor contributions from the bridge itself.[14] Subsequently in an extensive experimental series Durup and Jansson investigated the effect of cutting rectangular segment *f*-holes into rectangular spruce plates, concluding that without *f*-holes no bridge-hill was seen.[13]

Modes of the various substructures do influence overall structural response, to some extent in proportion to their fraction of the overall mass, as they are subsumed into the

[a)]Electronic mail: bissingerg@gcu.edu

 0001-4966/2006/120(1)/482/10/$22.50

overall response of the violin. The bridge however falls into a special category because it is the string-to-corpus energy conduit/filter. In this work modal analysis and radiativity measurements on quality rated violins in an anechoic chamber were combined with systematic bridge waist and wing mass trimming experiments to investigate the effect of this filter on violin radiativity. This comprehensive approach was designed to provide experimental answers to questions such as: (1) does the bridge substructure demonstrate recognizable in-plane normal mode resonances while on the playable violin, (2) are enhanced bridge motions related to enhanced corpus vibrations and radiativity, (3) is there a relationship between BH magnitude and violin quality, (4) is the BH frequency f_{BH} sensitive to the rocking frequency f_{rock} of the bridge, (5) are there general radiativity trends arising from bridge waist (stiffness) or wing mass trims?

II. EXPERIMENT

A. VIOCADEAS measurements

The VIOCADEAS zero-mass-loading calibrated experimental measurements have been described elsewhere in considerable detail[15] (and references therein). Here only essentials will be presented. Simultaneous vibration and radiation measurements were performed in an anechoic chamber on violins hung by two thin elastics in an approximate "free-free" condition. The frequency range encompassed the BH hill near 2.3 kHz and the critical frequency for "good" violins.[2] Violin quality ratings for the 12-violin database were all by an outstanding professional violinist[15] following Weinreich's general 3-category scheme[16] of student (our bad category—rating 1–3), decent professional instrument (good category—rating 4–7), and fine solo instrument (excellent category—rating 8–10). The good violin data shown in plots were from three violins rated 7, while the bad were rated 2,3,3.

The calibrated measurements incorporated 9 points on the bridge proper, plus multiple points along a line on the corpus directly in front of the violin bridge from which the motion of each bridge foot was extracted. Earlier uncalibrated "free-free" vibration measurements on 20 violin bridges with a zero-mass-loading microphone were used to help categorize in- and out-of-plane bridge normal modes.[17]

Force hammer impacts at the driving point on the bridge G-corner in its plane were directed parallel ($F_{\|}$) or approximately perpendicular ($F_{\perp}$) to the plane of the violin. Tested violins were playable although chin and shoulder rests were removed. No damping was applied to strings at tension (A =440 Hz). The scanning laser picked up motion along the beam direction and generated mobility (velocity/force: complex) transfer functions $Y(\omega)$ for approximately 500 points over the corpus (top-ribs-back); bridge (9 points), tailpiece and neck-fingerboard measurements were in two perpendicular directions. Simultaneously a rotating 13-microphone array collected pressure data and generated radiativity (pressure/force; complex) transfer functions $R(\omega)$ at 266 points over a sphere. Bridge mode classifications as normal or complex were made using the mode animation capability in the modal analysis program.

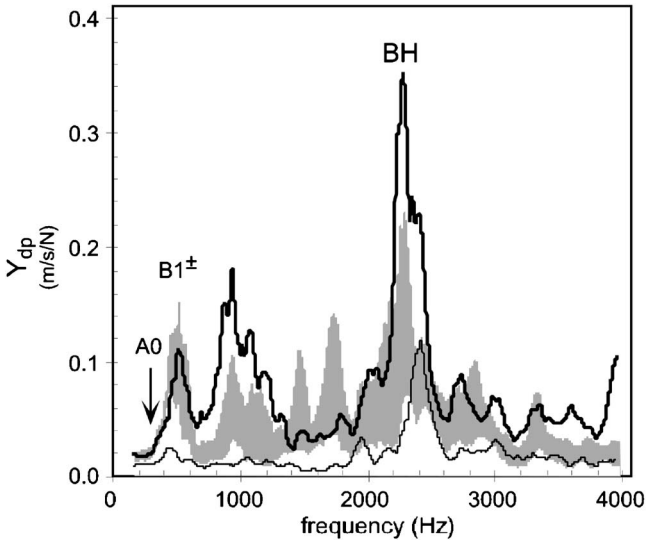


FIG. 1. Min-max driving point mobility magnitude for good (shaded) and bad violins (curves); A0,B1±,BH noted.

The proposed linkage of a prominent BH driving point mobility (input admittance) Y_{dp} with good sound quality was examined through a plot of minimum-maximum Y_{dp} ranges for good and bad violins (Fig. 1; low-lying strongly radiating "signature modes" labeled: A0, the compliant wall Helmholtz radiator (~280 Hz) and B1±, the first corpus bending modes nominally between 450 and 550 Hz). Driving point spectra did not show strong A0 excitation. The most prominent structure was the BH peak, and the largest BH peak was for a bad violin, with variability between bad violins being considerably larger than for good violins.

The 250-Hz (or other) band averages of driving point mobility Y_{dp}, averaged-over-bridge mobility $\langle Y_{brg}\rangle$, averaged-bridge-foot mobility $\langle Y_{brgft}\rangle$, averaged-over-corpus mobility $\langle Y_{corpus}\rangle$ and averaged-over-sphere radiativity $\langle R\rangle$ ($\langle\ \rangle$ denotes rms average) were computed directly from the spectra, not by averaging individual normal mode properties over these bands as was done in previous work. This greatly speeded data analysis, bypassed fitting and mode overlap problems and statistical fluctuations associated with the small numbers of modes in a band while giving comparable results.

As expected modal analysis results showed that vibrational energy transfer from strings to violin corpus was predominantly through the bridge, with bridge impedance ranging from 0.01× to 0.1× the nut or tailpiece impedance. Henceforth we assume the only important path for string energy to reach the corpus will be through the bridge.

B. Bridge Waist - Wing Mass Trims

The effects of bridge waist and wing mass trimming on violin radiativity were examined in a systematic way at the Oberlin Violin Acoustics Workshops in 2004 and 2005. The VIOCADEAS support fixture and force hammer excitation apparatus were removed as a unit from the anechoic chamber and mounted temporarily on a base that incorporated 5 evenly-spaced calibrated microphones in a semicircular array (r=0.3 m), at 30° intervals from 15° to 165°, oriented per-

pendicular to the top plate in the plane of the bridge. Eight Sonex 15 cm foam wedge absorbers were placed underneath the violin to reduce nearest-surface floor reflections. Low mass (0.5 g) accelerometers were placed at each bridge foot in front of the bridge. An 8-channel data acquisition system was used to collect accelerances from both bridge feet and radiativity transfer functions from the 5 microphones. The averaged 5-microphone radiativity will henceforth be labeled as a partial radiativity $\langle R_{\text{part}} \rangle$ to distinguish it from the averaged-over-sphere $\langle R \rangle$.

In the Oberlin 2004 experiment, two violins, an Andreas Guarneri (1660) and a Gregg Alf (2003), were fitted with four bridges each. No changes other than the sequenced bridge modifications were made to either instrument. The rocking mode frequency f_{rock} was measured off-violin using a separate apparatus incorporating a piezo-film contacting the bridge top along with a bridge-foot clamping vise. The bridge blank tops were first trimmed down to give the proper string height for each violin, and then each waist was trimmed to give $f_{\text{rock}}=3.6$ kHz (nominal). Three bridges then had their wing mass decremented by $\Delta m=0.04$, 0.08, and 0.12 g, all from the same location in the bridge wings, and the waist was again trimmed slightly to drop f_{rock} back to 3.6 kHz. One bridge in each set of four had no wing-mass decrement and was labeled $\Delta m=0$. Finally, waist thicknesses only were trimmed in successive stages to get $f_{\text{rock}}=3.4$, 3.2, 3.0, 2.8 kHz. Altogether 20 $\langle R_{\text{part}} \rangle$ measurements were made for each violin during the sequential modification process.

Mass removal from the waist was much more effective in changing f_{rock} than from the wings: $\Delta f/\Delta m \approx 24 \pm 14$ kHz/g vs. 0.75 ± 0.03 kHz/g, convincing support for simplified bridge models separating the bridge into a top mass and waist spring—irrespective of boundary conditions for the feet. Other experimental systematics from waist trimming ($2.8 \leq f_{\text{rock}} \leq 3.6$ kHz) were values for: (1) f_{rock} changes versus waist thickness x, $\Delta f/\Delta x \approx 180$ Hz/mm, (2) bridge mass changes for waist trims, $\Delta m/\Delta x \approx 0.009$ g/mm, (3) waist trims from 16.3–17.8 mm to 11.8–13.1 mm, 4.7 mm on average, (4) mass changes associated with waist trimming from 0.014 to 0.068 g, 0.043 g on average. The 40 separate bridge modifications/measurements in the 2004 two-violin experiment were made as rapidly as possible, precluding qualitative judgments. The 2005 experiment, where f_{rock} was dropped from 3.4 to 3.0 to 2.6 kHz but $\Delta m=0$, used only one bridge on one violin, but the instrument was played for a small group of listeners after each bridge trim for qualitative evaluation.

Using Oberlin 2004 data a 4×5 data "R matrix" can be created for each frequency or frequency band from $\langle R_{\text{part}} \rangle$ spectra, but it is completely impractical to present these matrices for any more than a few important bands for each violin to demonstrate complexities accompanying simultaneous f_{rock} and Δm changes. To examine any generalized f_{rock} or Δm trends the R matrices were reduced to a single row or column by averaging over f_{rock} for one bridge (Δm =constant) to scrutinize general $\langle R_{\text{part}} \rangle$ vs Δm effects, and by averaging over Δm for four different bridges (f_{rock} =constant) to scrutinize general $\langle R_{\text{part}} \rangle$ vs f_{rock} effects.

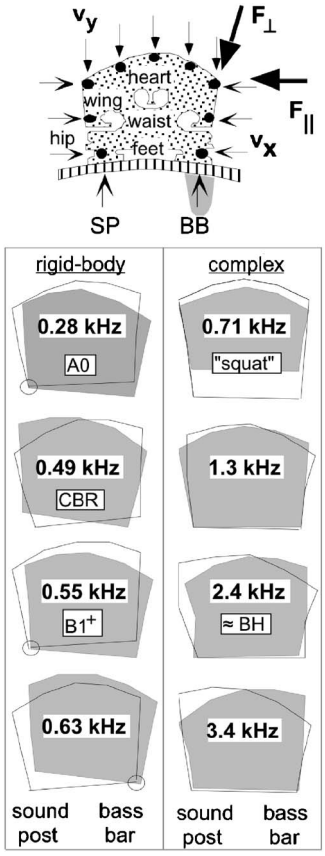


FIG. 2. (Top) bridge shape, force-response positions and directions (soundpost SP and bassbar BB from corpus); (bottom) motion extremes up to 4 kHz (A0, CBR, B1+, "squat," $f \approx f_{\text{BH}}$ labeled; ○ — rigid body SP or BB pivot.

III. RESULTS

Experimental results are presented in two main sections: (1) modal-acoustic analysis for dynamic and radiative behaviors of the violin or any particular substructure to understand energy transmission through the violin, and (2) systematic waist and wing mass trims to examine bridge filter effects on radiativity.

A. Modal analysis of bridge vibrations

Figure 2 shows the position and measured mobility direction(s) for each bridge point, including the driving point. Since the bridge feet were in intimate contact with the top plate, it was assumed that corpus motion measured immediately adjacent to the bridge feet could be used as a direct measure of bridge foot motion in the vertical direction; separate horizontal motion measurements were made on the sides of the bridge feet. The energy introduced at the bridge travels through an energy chain from driving point → bridge → bridge feet → corpus → radiation. Superimposing the mobility responses of each link it is possible to see if characteristic structures in driving point mobility spectra "migrate" from bridge to corpus to radiativity. Near f_{BH} modal average radiation efficiencies are nearing 1.[2] Since f_{BH} is also close to the ear's sensitivity maximum, such peaks should generally be audible.

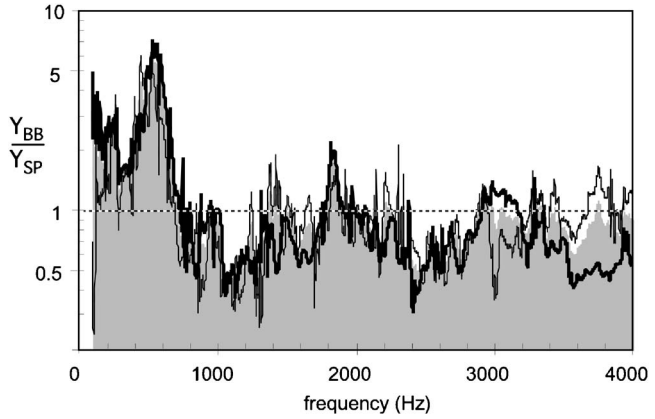


FIG. 3. Bassbar-soundpost bridge foot mobility ratio for $F_{\|}$ driving point excitation. (Shaded — 12-violin average; lines: thick - good, thin — bad).

1. Bridge motions on the violin

The observed bridge motions for all twelve violins were basically the same: (a) at frequencies <0.7 kHz the bridge generally pivoted as a rigid body around the soundpost or bass bar foot (exception—the CBR mode where both feet were active) with little deformation, (b) above ~ 0.7 kHz, to 4 kHz, motions were complex. No normal mode bridge motions—strong, 0–180° bridge motions unaccompanied by other substructure peaks—were seen. A sampling of observed bridge motions is presented in Fig. 2 for one violin.

In all 12 violins bass bar foot motion predominated below 600 Hz; above 600 Hz generally the sound post side was more active. Note however that relatively little foot motion was seen above 1 kHz (Fig. 2). The ratio of bass bar to soundpost mobilities shows this trend for the 12-violin average, as well as just the good or bad 3-violin subsets (Fig. 3).

2. Independent bridge modes?

Is the bridge capable of maintaining a vibration mode independent of the corpus and strings between which it resides? If so, there would be a peak in bridge mobility that does not coincide with any corpus or string mode peaks. Such a possibility was eliminated straightforwardly up to 4 kHz by examination of composite plots of Y_{dp}, $\langle Y_{brg}\rangle$, $\langle Y_{brgft}\rangle$ $\langle Y_{corpus}\rangle$, and $\langle R\rangle$; bridge mobility peaked *only* when corpus mobility peaked. This observation, consistent with recent simulations,[14,18] is unsurprising given the low mass of the bridge compared to the vibrating (varying boundary condition) corpus on which it rides, or even in comparison to the strings, which *in toto* have a mass similar to the bridge. Including tailpiece and neck-fingerboard averages in such plots also shows that only the corpus radiates significantly (cf. Fig. 1, Ref. 3).

One interesting complex mode where the bridge appeared to have a "squat" behavior in animation (see Fig. 2) was observed for 9 of the 12 violins tested. It fell far below the in-plane bridge rocking mode frequency, not above as would be expected for an in-plane mode. The overall motion and frequency placement were consistent with the first *out-of-plane* bending mode where the strings and top plate create boundary condition constraints on transverse motion at opposite ends of the bridge.

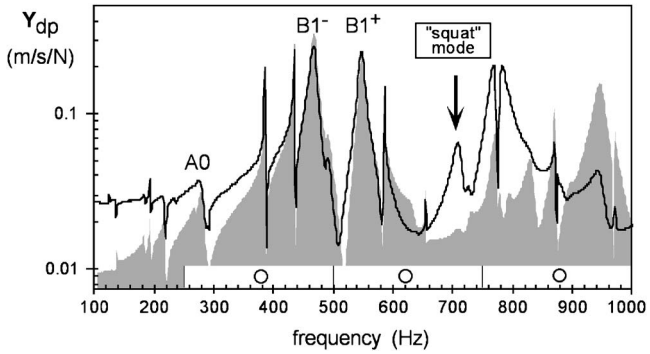


FIG. 4. $F_{\|}$ (shaded curve) and $F_{\perp}$ (solid line) driving point excitation shows "squat" mode near 700 Hz. (Major radiators A0, B1⁻, and B1⁺ labeled; note 250 Hz bands and band centers (○).

The motion suggests being more strongly excited by $F_{\perp}$ than $F_{\|}$ strikes, as was confirmed by experiment (Fig. 4). Similar response differences near this frequency were seen earlier by Trott.[19]

3. Energy chain

To illustrate how the driving point BH structure (Fig. 1) appears throughout the energy chain a sequence of 250-Hz band averages for good and bad violins is presented in Fig. 5 for the driving point Y_{dp}, bridge ($\langle Y_{brg}\rangle$), bridge feet ($\langle Y_{brgft}\rangle$), corpus ($\langle Y_{corpus}\rangle$), and averaged-over-sphere radiativity $\langle R\rangle$. A distinct BH peak near 2.3 kHz is evident throughout the chain, in the bad as well as good, being somewhat larger for the bad in Y_{dp} (although error bars overlap) and Y_{brg}. This result disagrees markedly with the quality association current in the literature.

4. BH mobility and radiativity versus quality

Overall the bad violin radiativity curve in Fig. 5(e) has a larger maximum-minimum range and falls off more on either side of the BH structure near 2.3 kHz. Can some audible difference in sound be ascribed to the difference between our good-bad broad-band excitation radiativity curves, even though the general string driving force is sawtooth-harmonic in character? For discussion and comparison purposes we use a simplified sound characterization scheme based on that of Dunnwald[20] with additional contributions from Meinel,[21] where relatively strong radiation in individual bands is associated with a certain general character to the sound: 190–650 Hz — "sonorous, full sound" (this band includes the first corpus bending modes B1⁻ and B1⁺; Dunnwald also notes the importance of a relatively strong A0 near 280 Hz); 650–1300 Hz — "nasal, boxy"; 1300–4200 Hz — "brilliant, clear"; 4200–6400 Hz — "harsh."

The band-by-band ratio of radiativities shown in Fig. 6 is useful for examining good-bad differences. In this ratio any common driving force cancels for broad-band or harmonic excitation. A significant difference between the quality classes appears only for a few individual 250 Hz bands, viz., 375, 1625, and 3375 Hz bands. Ratios in the BH region (2125–2625 Hz) do not differ significantly. The ~ 3.4 kHz region falls where the lower effective critical frequency of

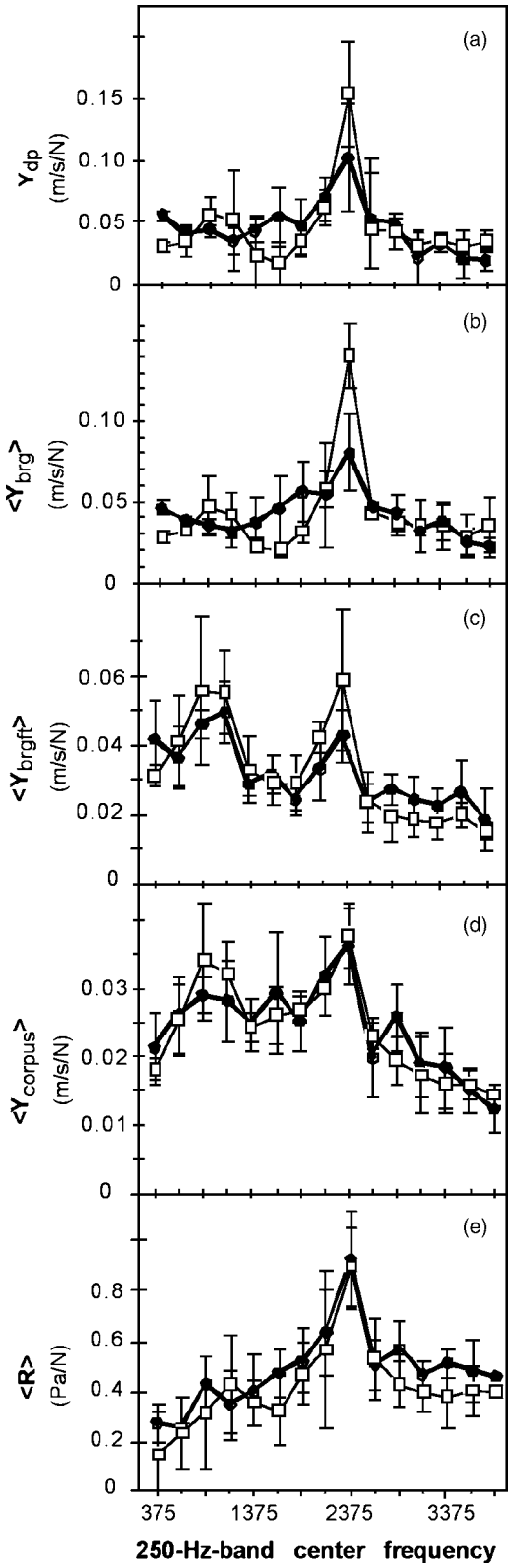


FIG. 5. Driving point (a), bridge (b), bridge feet (c), corpus (d), radiativity (e) for good (thick line, ●) and bad (thin line, □) violins. (1 s.d. errors shown).

these good violins can enhance radiativity,[2] a relative enhancement actually reinforced by holding/playing the violin.[3]

An important question concerning possible significant good-bad differences in band-average radiation efficiency near f_{BH} can be answered directly from the data in Fig. 5:

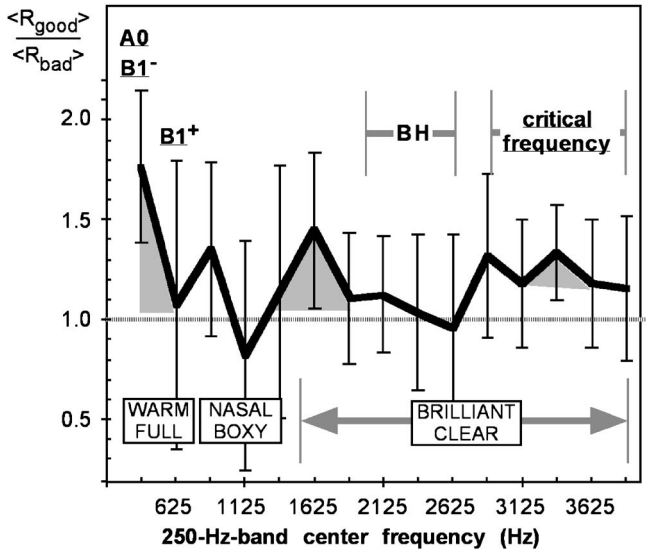


FIG. 6. Good-bad $\langle R \rangle$ ratio (1 s.d. errors). (Noted: BH, critical frequency, boxed quality descriptors, signature modes A0, B1$^-$, and B1$^+$; shading denotes significant difference from 1).

$\langle Y_{corpus} \rangle$ and $\langle R \rangle$ magnitudes are within error the same for good and bad violins, hence the BH radiation efficiency — computed from the $\langle R^2 \rangle / \langle Y^2_{corpus} \rangle$ ratio[22] — does not vary significantly between good and bad violins.

To look for trends versus violin quality Y_{dp}, $\langle Y_{brg} \rangle$, $\langle Y_{brgft} \rangle$, $\langle Y_{corpus} \rangle$, and $\langle R \rangle$ BH magnitudes were plotted for the 12 violins in Fig. 7. The 7-quality violins were not significantly different from the 2/3-quality violins, e.g., at the extremes the three 7-quality good violin magnitudes covered a range that usually overlapped the 2/3 violin values. While one might argue that these are poor statistics, a better argument might be that even in this small sample there are already exceptions — both inter- and intra-violin quality class — to any presumed correlation between violin quality and a large BH peak. The data shown in Figs. 5–7 that pertain to

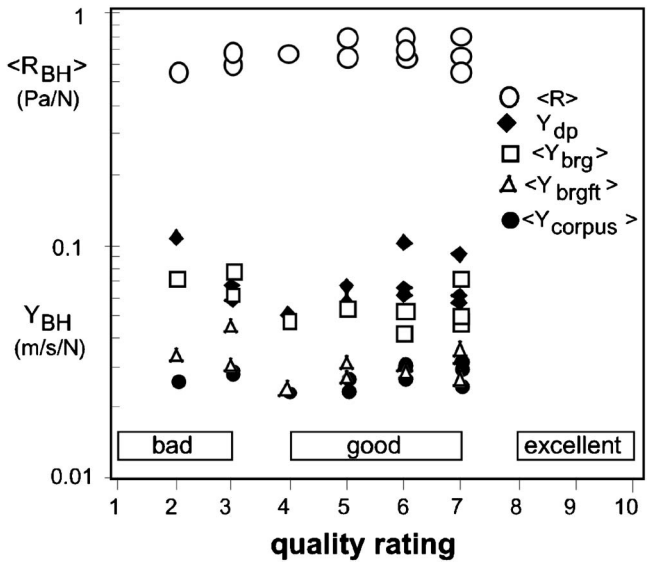


FIG. 7. Driving point(♦), average bridge(□), bridge feet(△), and corpus (●) BH mobility magnitudes and average BH radiativity(○) for 12 violins vs violin quality.

the presumed relationship between BH magnitude and violin quality can be summarized briefly as showing no experimental correlation between the driving point or corpus mobility, or the radiativity, or the radiation efficiency that is significantly different between good and bad violins. Rather, the good-bad differences so far appear to lie in the strengths of the other acoustically important regions *relative* to BH.

B. Bridge waist and wing mass trims

Waist-wing trimmings entailed removal of ≤ 0.12 g from the bridge; no effect was seen on corpus mode frequencies when proper string tension was maintained, hence mode shapes and radiation efficiencies should remain the same. Under these conditions $\langle R_{\text{part}} \rangle$ measurements were quite reliable for investigating intra-violin radiativity changes due to systematic f_{rock} and Δm variations. They were also reliable for inter-violin mode comparisons below ~ 0.7 kHz, since radiativity is close to isotropic when $\lambda >$ violin dimensions.

Above 1 kHz inter-violin broadband comparisons of $\langle R_{\text{part}} \rangle$ become more reliable than individual spectra because coincidental microphone placements at minima for one mode should tend to average out with equally coincidental placements at another mode's maximum within each band. Because the cavity mode A0 lying at $f \approx 280$ Hz is such an important radiator, and the *only* strongly radiating mode below about 450 Hz, it was isolated for special attention using the average over ± 10 Hz around the peak. The lowest band-average now covers 300–500 Hz, with band center at 400 Hz (all higher bands as before).

The R matrix from the Oberlin 2004 experiment provides a systematic framework in which to analyze possible filter effects from bridge waist or wing mass trims (all other violin properties held constant) on radiativity. Filter effects related to f_{rock} variations were the most straightforward to interpret since only one bridge waist was trimmed. Wing mass trims however were over four different bridges, i.e., four different pieces of wood, a possible complication.

1. f_{rock} and the BH frequency

Trimming the waist primarily changed the rocking mode frequency f_{rock}, with little effect on the bridge mass (3.6 to 2.8 kHz, max. $\Delta m = 0.068$ g; average $\Delta m = 0.045$ g). When the bridge was in place on the violin the BH centroid frequency f_{BH} (nominally 2.3 kHz) had a range $\Delta f_{\text{BH}} = 31 \pm 15$ Hz (average over all bridges for both violins) as f_{rock} varied 0.8 kHz, giving $\Delta f_{\text{BH}}/\Delta f_{\text{rock}} \approx 0.04$. Clearly f_{BH} was insensitive to f_{rock} variations, (although the Guarneri was somewhat more sensitive), in agreement with the general conclusions of Ref. 7. A more recent model of Woodhouse[18] incorporating the dynamic properties of the violin underneath a simplified bridge placed on top of a rectangular box without f-holes displayed higher f_{BH}values (~ 2700 to ~ 3500 Hz), far greater changes ($\Delta f_{\text{BH}} \approx 740$ Hz) and sensitivity $\Delta f_{\text{BH}}/\Delta f_{\text{rock}} \approx 0.93$ than experiment. The experiments of Ref. 14 required f-holes in a rectangular spruce plate to display a prominent BH peak while this model did so without f-holes, an interesting inconsistency that clearly warrants a closer look.

2. The R matrix

In the normal listening situation where listeners can be at differing distances, in different rooms, or listening to a recording, etc., the obvious point that the sound intensity varies yet the general character of sound remains (while not forgetting that the ear's frequency response varies with intensity), supports the general argument that the relative strengths of certain frequency bands, i.e., "acoustic profiles," are more important in determining perceived character. In this context comparing $R(\omega)$ radiativity curves over a large portion of the violin's acoustic range is valuable. On the other hand extracting trends from a visual overlay of 20 such $R(\omega)$ curves created in the bridge trim experiment becomes a daunting interpretive situation. Comparison was somewhat simplified by comparing 250 Hz bands, but there are such a large number of these bands that in practice little has been gained. On the other hand if just one particular frequency band was chosen for the R matrix, it would not provide any comparison with other bands. Our practical compromise was to create for each violin R matrices for only three acoustically significant bands: A0 (~ 280 Hz), BH (~ 2300 Hz, lower portion of the "brilliant, clear" band), and the upper portion of the "brilliant, clear" (B-C) band (2500–4200 kHz). These were all bands where significant changes occurred during the bridge trims. The R matrices for each violin are presented in Fig. 8, where the square sizes for the A0, BH, and B-C bands are proportional to the $\langle R_{\text{part}} \rangle$ values in each band. Moving horizontally in each row shows the effect on $\langle R_{\text{part}} \rangle$ of only varying f_{rock} so that one bridge is trimmed successively. Moving vertically in each column shows the effect on $\langle R_{\text{part}} \rangle$ of only Δm varying and here different bridges were employed (see Sec. II B).

Variations of individual $\langle R_{\text{part}} \rangle$ matrix elements before and after modification in Fig. 8 reflect the actual complexity of bridge trims (and of course any associated comparative quality judgments of violins). For example, even for one violin, changing only Δm (moving in only one f_{rock} column) or f_{rock} (moving in only one Δm row), A0 waxes and wanes; a diagonal move — changing f_{rock} and Δm simultaneously — seems quite an uncertain move for a maker in terms of knowing what will happen to $\langle R_{\text{part}} \rangle$, both in magnitude and relative strength, and to the sound. And this is for just one band. The practical difficulty in evolving a violin from a certain sound character to another, more desirable one via a particular bridge modification is apparent.

The many localized variations — e.g., for the Alf violin $\Delta m = 0.12$ g row, BH magnitudes decrease slightly with increasing f_{rock}, whereas for $\Delta m = 0$ they increase, or in the $f_{\text{rock}} = 2.8$ kHz column $\langle R_{\text{part}} \rangle$ drops off in the 0–0.04 Δm transition, increasing again in the 0.04–0.08 transition — tend to bury overall trends. Yet makers have certain general ideas gained from centuries of experience about what a particular bridge trim will do to the sound. This generalized approach presumably is more fruitful since trends common to both violins do appear, e.g., averaged-across-f_{rock} radiativities increase as wing mass decrements increase, albeit with individual exceptions. The averaging-across-rows/columns of the R matrix will be needed to extract any gen-

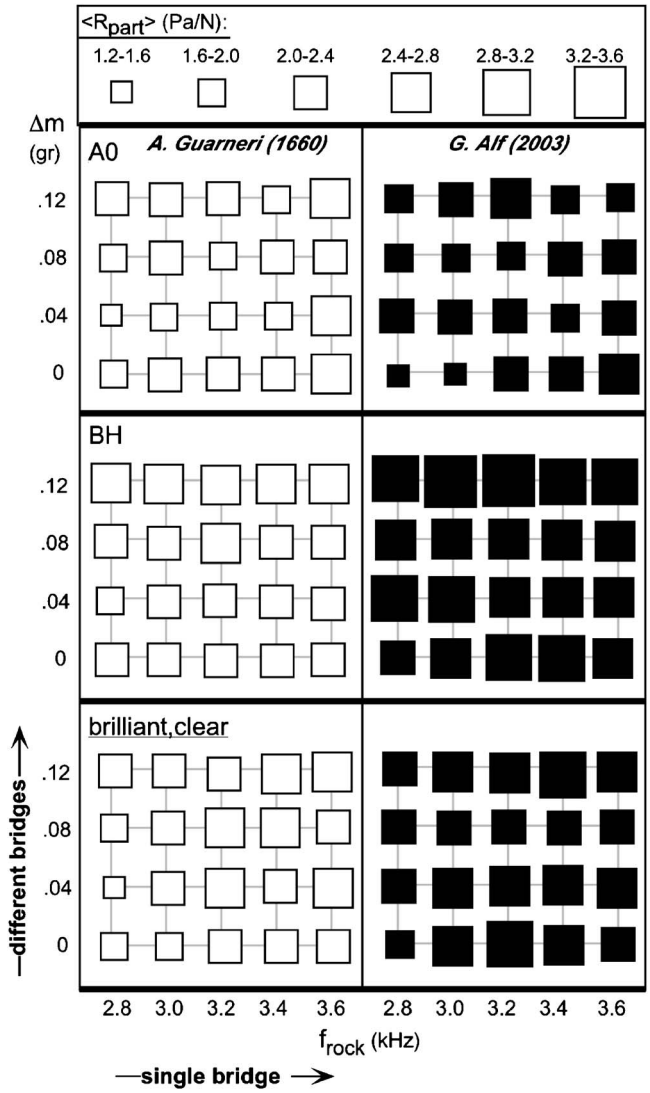


FIG. 8. Δm-f_{rock} R-matrices for A0, BH, and "brilliant, clear" bands for A. Guarneri and G. Alf violins. ($\langle R_{\text{part}} \rangle$ stepped, see key.) Note that f_{rock} changes along a row are for a single bridge, while Δm changes in each column are across four different bridges.

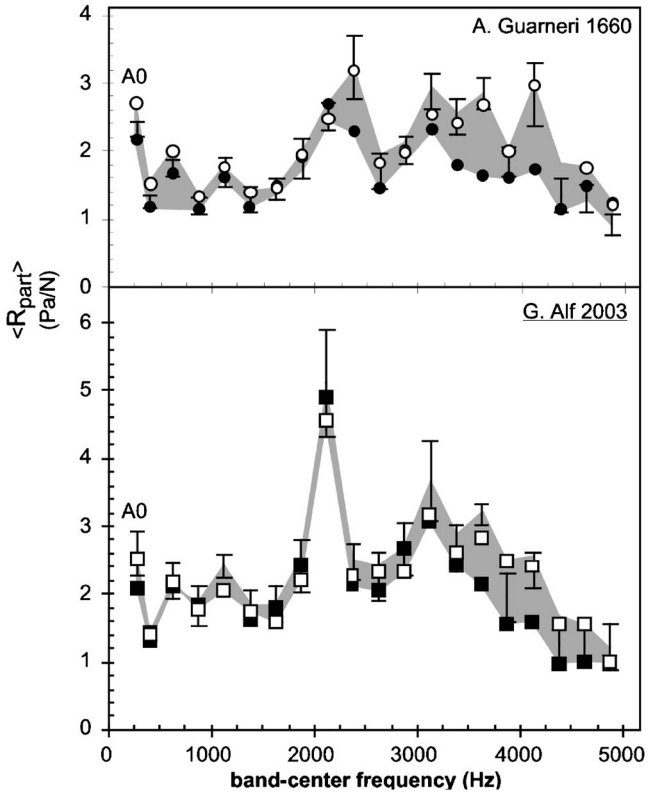


FIG. 9. Effect on partial radiativity of varying f_{rock} from 2.8 (closed symbols) to 3.6 kHz (open symbols), averaged over Δm for four bridges for old Italian (top) and modern violin (bottom). Shaded area min-max curve, all f_{rock}; 1 s.d. errors for 3.2 kHz averages only, no curve). Note — A0 separated from lowest band (see the text).

eral radiativity trends accompanying *just* rocking frequency or wing mass changes.

3. f_{rock} and $\langle R_{\text{part}} \rangle$-$\Delta m$ averaged

The radiativities shown in Fig. 8 showed extensive local fluctuations as f_{rock} or Δm was stepped. To extract generalized trends in $\langle R_{\text{part}} \rangle$ magnitudes and acoustic profiles accompanying changes in f_{rock}, the band-average 2004 data for each f_{rock} value were averaged over all Δm. The averaged radiativity analysis presented in Fig. 9 incorporates: (1) a min-max $\langle R_{\text{part}} \rangle$ shaded curve to highlight regions where waist trims had the largest effect on $\langle R_{\text{part}} \rangle$ and how these evolve with frequency, (2) extreme curves for $f_{\text{rock}}=2.8$ and 3.6 kHz to act as a guide to interpreting generalized trends, and (3) standard deviation error bars for the intermediate 3.2 kHz curve as a nominal statistical measure of variability in the Δm average. Certain general criteria were used to judge the relationship of the extreme curves to the min-max curve: (a) if the extreme curves "bracket" the min-max curve consistently over a region this was taken as *prima facie* evidence for a consistent general trend in this region (e.g., the ~3000 Hz region), (b) if the extreme curves were centered in the min-max curve, further analysis was needed to see if there was a peak in the f_{rock} range, (c) if the 3.2 kHz error bars fell outside the minimum and maximum in a band or region, no significance was attached to the extreme curve variations (e.g., 1875 Hz band), and (d) general trend comments about a wide region could be made irrespective of an individual band's deviation from this trend.

The min-max shaded regions for both violins in Fig. 9 clearly show the most prominent waist trim effects were in the 3000–4200 Hz region where the ear is most sensitive, least prominent near 1500 Hz, and, somewhat surprisingly, rising again at A0. The fact that the extreme f_{rock} curves typically lie at or near the maximum or minimum over important regions leads us to infer certain general trends versus f_{rock}: both A0 and the nominal 2000–4200 Hz region generally weakened as f_{rock} decreased. Waist trims affected the BH position and amplitude more noticeably on the old Italian than on the modern instrument.

Some tendency for increased radiativity above 4200 Hz — the "harsh" region — seen in Fig. 9, is consistent with a remark made in 1979 by Muller who noted that replacing a normal bridge with a solid blank (which would give $f_{\text{rock}} \approx 8$ kHz) brought out the harsh and nasal characteristics of the violin compared with the standard bridge.[23]

Of course averaging over Δm tends to smooth out some

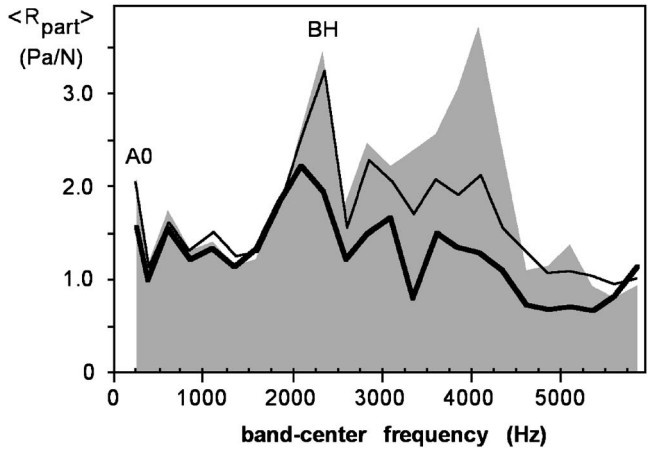


FIG. 10. Effect of varying f_{rock} from 3.4 kHz (shaded) to 3.0 kHz (thin curve) to 2.6 kHz (thick curve) on partial radiativity of the A. Guarneri 1660 violin.

of the $\langle R_{part} \rangle$ variation attending f_{rock} changes. A closer look at the $\langle R_{part} \rangle$ magnitude changes for just one bridge follows.

4. f_{rock} and $\langle R_{part} \rangle$ — quality

The Oberlin 2005 experiment on the A. Guarneri violin narrowed the bridge waist rather more than considered safe for typical playing, stepping f_{rock} from 3.4 to 3.0 to 2.6 kHz, and added qualitative evaluations. This seemingly modest downward extension, which required removing only 2 mm (0.02 g!) from the bridge waist to drop f_{rock} from 3.0 to 2.6 kHz, changed the sound of the Guarneri from that of a good violin to a student (bad) violin! Since corpus normal modes showed no change in frequency, no change would be expected in mode shape or radiation efficiency, which is independent of mode amplitude. The important changes were in the relative mobility amplitudes for the various modes (and hence their radiativity) and consequent band averages. This experiment — where a minimal mechanical alteration not affecting the corpus created a large acoustic effect — provided an unambiguous example of how strong the filter action of the bridge is.

As might be expected from the qualitative evaluations the partial radiativity graph for the 2005 one-bridge experiment including the lowest f_{rock} data is very revealing (Fig. 10). A0 fell off ~25% as f_{rock} decreased from 3.4 to 2.6 kHz, with almost all of the change occurring in the 3.0–2.6 kHz transition. More striking was the overall falloff from 1.6 kHz upward, including the BH peak and 2800–4200 kHz in the "brilliant, clear" region. Relative to BH (averaged from ~1600 to 2600 Hz) both A0 and the 2800–4200 Hz regions fell off when f_{rock} dropped to 2.6 kHz.

Qualitative evaluations tracking the 3.4–3.0–2.6 kHz f_{rock} steps indicated certain consistent acoustic trends such as weaker, more uneven sound that did not carry as well, with the 3.0–2.6 kHz step being more noticeable. The measurements in Fig. 10 show significantly weaker overall partial radiativity at 2.6 kHz, with both 2.6 and 3.0 kHz showing noticeably weaker *relative* values in the 3000–4500 Hz band. Consistent with trends seen in the 2004 experiment (Fig. 8), f_{BH} dropped slightly. The overall acoustic profile also changed dramatically: at 3.4 kHz the radiativity generally increased up to ~4000 Hz and then fell rapidly, whereas at 2.6 kHz the profile showed a general overall decrease. Such large acoustic profile changes clearly show the importance of waist trims to violin sound.

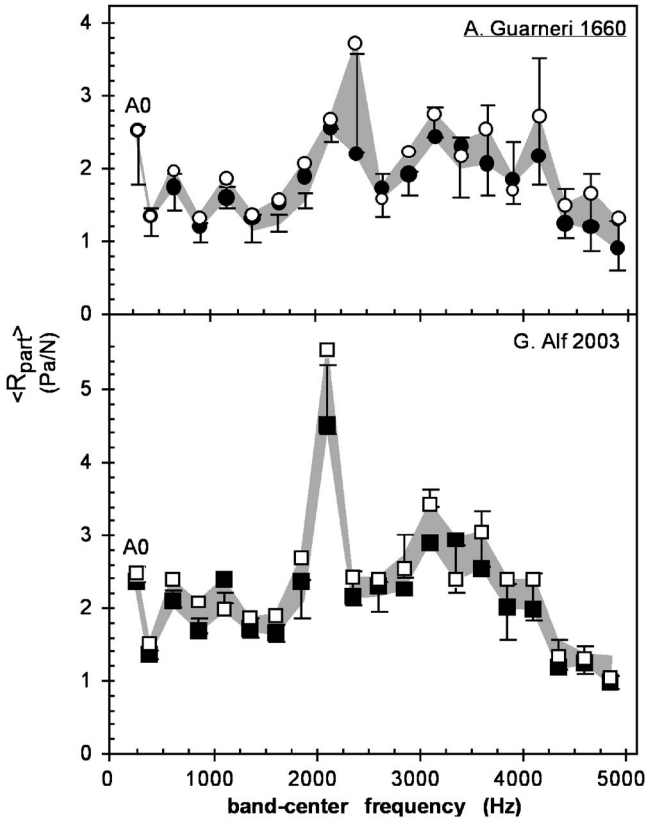


FIG. 11. Effect of decreasing wing mass (by 0.12 g in 0.04 g steps, averaged over f_{rock}) on averaged partial radiativity for four bridges on old Italian (top) and modern violin (bottom). Closed symbols — $\Delta m=0$, open — $\Delta m=0.12$ g; shaded area min-max, all Δm. 1 s.d. errors for $\Delta m=0.04$ mass decrement average only (no curve).

5. Wing mass decrements

By averaging across f_{rock} rows in the data matrix, the generalized effects of wing mass decrements on $\langle R_{part} \rangle$ were extracted. So that a real Δm effect does not get buried by a possible change-of-wood effect, even though these bridges were matched closely, only the trend between $\Delta m(\mathrm{g})=0$ and 0.12 will be discussed. In Fig. 11 the Δm partial radiativity data are presented, again following the format of Fig. 9.

From the min-max curve, where approximately 85% of the bands have their highest radiativity associated with $\Delta m=0.12$ g and approximately 75% their lowest with $\Delta m=0$, we infer that generally the largest mass decrement led to higher overall radiativity. If lower mass bridge tops lead to higher radiativity, then conversely adding mass to the violin bridge top — as in a violin mute — should lead to lower radiativity, which is consistent with experience.

At $f > f_{BH}$ the Δm min-max band was not so wide overall as for f_{rock} in Fig. 9, but appeared slightly wider at $f < f_{BH}$. One significant exception was for A0, which did not change significantly with wing mass decrements. The Guarn-

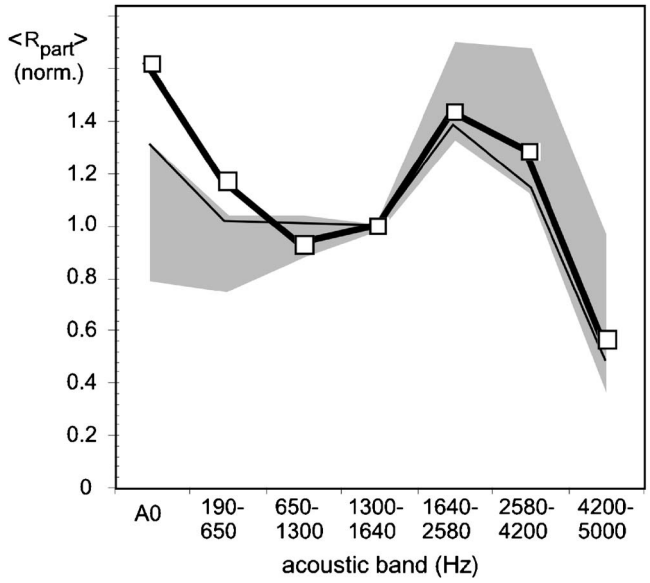


FIG. 12. Acoustic profiles for 20 Alf violin bridge trims, compared to "target" A. Guarneri profile (–□–). (Closest to target - thin line; shaded area denotes min-max region, all profiles normalized to 1300–1640 Hz band).

eri BH region was more sensitive to mass decrements than the Alf violin, similar to behavior seen for f_{rock} variations.

6. Acoustic profiling

Our measurements again confirm that old Italian violins are not necessarily louder (= higher radiativity when no perceptual complications are present) across the board than modern instruments, although in specific cases, for specific violins or specific regions they may well be (e.g., see Table IV, Ref. 24). Of most interest here is the question unintentionally posed by the wide range of f_{rock} and wing mass decrement modifications in the Oberlin experiments. Can an old Italian violin acoustic profile be matched by any one (or more) of the 20 separate modern Alf violin acoustic profiles (bridge modifications only!)?

To test this one A. Guarneri acoustic profile was chosen as the "target" — viz., the f_{rock}=3.0 kHz, Δm=0 acoustic profile. Is this the "best" choice? In our context it does not really matter. The playing tests in the Oberlin 2005 experiment were consistent with it being a good violin, while dropping to f_{rock}=2.6 kHz turned it into a bad violin. More to the point of this heuristic example is that any "exceptional" (= testers' preferred) violin could be measured in the apparatus and its acoustic profile used as the target for bridge modifications on any other violin to reach. (The aforementioned simplified bands 190–650, 650–1300, 1300–4200, 4200–6400 kHz (plus A0 separated out) will be used again, but note that measurements extended only to 5000 Hz, hence this latter band will cover only 4200–5000 Hz.)

Since the region near 1.5 kHz was insensitive to both f_{rock} and Δm variations (see Figs. 9–11) the 1300–1640 Hz band was chosen for normalization purposes for all curves across all measurements for the Alf and the target A. Guarneri curve. The normalized results are shown in Fig. 12. Out of 20 separate f_{rock}-Δm acoustic profiles only one Alf curve came close overall to the Guarneri curve, viz., the f_{rock} =3.6 kHz, Δm=0 curve. Commonly the acoustic profile would be similar below or above 1500 Hz, but not below *and* above. Qualitative tests were not performed in the 2004 measurement series so further remarks are not possible.

IV. CONCLUSIONS

The traditional view of the violin bridge being the string-to-corpus energy conduit was confirmed by modal and acoustical analysis results on 12 quality-rated violins.

The BH structure near 2.3 kHz for all violins was seen in the mobility spectra at every important point in the energy chain for every violin tested, good or bad, ultimately leading to a peak in radiativity; no significant good-bad radiation efficiency change was observed. Waist trims generally had little effect on BH frequency or magnitude. Our experimental scrutiny of the proposed relationship between BH driving point magnitude and violin quality provided no evidence to support this claim.

The complex motions of the bridge above 1.5 kHz and the minimal effect that varying f_{rock} had on f_{BH} confirmed the conclusion of Jansson and collaborators that the BH peak at the driving point arose not from isolated bridge motions at some rocking mode frequency, but rather from local corpus motions. These motions might be enhanced by the bridge. Relatively large f_{rock} changes creating only few-percent f_{BH} changes could be considered experimental evidence for the bridge substructure rocking behavior being subsumed into the overall violin response to excitation.

Bridge trims changed acoustic profiles significantly. The 3000–4200 kHz region was affected most strongly by f_{rock} changes, with decreases in f_{rock} accompanied by decreased radiativity. At the other end of the acoustic spectrum, A0 radiativity weakened substantially when f_{rock} dropped to its lowest values, especially 2.6 kHz — in addition to the above-noted falloff in the 3000–4200 Hz region. Falloff at both ends of the spectrum at the lowest f_{rock} values audibly diminished the sound quality of a violin, unambiguously demonstrating the filter properties of the bridge. Larger wing mass decrements generally increased radiativity, an anticipatable reversal of the effect of a violin mute.

Good-bad radiativity trends observed in the anechoic chamber and in the Oberlin 2005 partial radiativity experiment were similar. Good-bad differences in averaged-over-sphere violin radiativities were not seen in the BH magnitude or frequency, but rather in the relative strengths *below and above BH*: going from good to bad each had weaker A0/400–500 Hz and 3000–4200 Hz responses relative to BH. These partial radiativity changes were quite similar to f_{rock}-induced changes, especially the 3.0–2.6 kHz transition.

Perhaps a stronger argument can be made here. These experiments were *very* different: one compared different violins of different quality with different normal modes, while the other examined quality changes in the same violin arising from changes in the bridge's filter response, with no change in normal mode frequencies, shapes or radiation efficiencies, just relative mode magnitudes. Such similar good-bad acous-

tic trends for two such disparate experiments argues for greater generality than either alone.

There are some important additional implications. Broad-band acoustic profiles appear to offer the potential for directing violin sound quality to some definable goal. And given the comparatively more modest (and much more difficult to achieve) effects possible from modifying the violin vibration→radiation output filter, bridge shape and design modifications seem a more fruitful avenue to achieving good violin sound.

Finally, our results supporting and augmenting prior experiments and simulations signal an important transition in our understanding of the physical origin of the BH peak. It has evolved from the original notion of strong rocking motion of the bridge being transferred to the corpus into one where energy passing through the bridge sets the violin into vibration, and the bridge — no longer an isolated substructure —responds *as part of the violin* to corpus motions at the feet.

ACKNOWLEDGMENTS

I would like to acknowledge the essential contribution to the Oberlin 2004-2005 bridge trim experiment of the violinmaker team of Terry Borman, Alan Copeland, Claire Curtis, Timothy Johnson, Tom King, Don Leister, Guy Rabut, Alkis Rappas, George Yu, and Andreas Zanr, led by Gregg Alf. Portions of this research were supported by the National Science Foundation (DMR- 9802625).

[1]Ed. Heron-Allen, *Violin-Making As It Was And Is* (Ward Lock, London 1885), p. 131.
[2]G. Bissinger, "The role of radiation damping in violin sound," ARLO **5**, 82–87 (2004). http://ojps.aip.org/ARLO
[3]G. Bissinger, "Contemporary generalized normal mode violin acoustics," Acust. Acta Acust. **90**, 590–599 (2004).
[4]*Musical Acoustics, Part I, Violin Family Components*, Benchmark Papers in Acoustics, Vol. **5**, edited by C. M. Hutchins (Dowden, Hutchinson & Ross, Stroudsburg, PA, 1975).
[5]*Research Papers in Violin Acoustics 1975-1993*, edited by C. M. Hutchins (Acoustical Society of America, Woodbury, NY, 1997).
[6]W. Reinicke, "Transfer properties of string-instrument bridges," Catgut Acoust. Soc. Newsletter **19**, 26–34 (1973).
[7]J. A. Moral and E. V. Jansson, "Eigenmodes, input admittance and the function of the violin," Acustica **50**, 329–337 (1982).
[8]E. V. Jansson, N-E. Molin, and H. O. Saldner, "On eigenmodes of the violin — electronic holography and admittance measurements," J. Acoust. Soc. Am. **95**, 1100–1105 (1994).
[9]H. O. Saldner, N.-E. Molin and E. V. Jansson, "Vibration modes of the violin forced via the bridge and action of the soundpost," J. Acoust. Soc. Am. **100**, 1168–1177 (1996).
[10]E. V. Jansson, "Violin frequency response— bridge mobility and bridge feet distance," Appl. Acoust. **65**, 1197–1205 (2004).
[11]E. V. Jansson, B. Niewczyk, and L Frydén, "The BH peak of the violin and its relation to construction and function," *Proceedings of the 17th International Congress on Acoustics 2001*, Vol. **4** (Music #7B.14.03), pp. 10–11.
[12]E. V. Jansson and B. K. Niewczyk, "On the acoustics of the violin: bridge or body hill," Catgut Acoust. Soc. J. **4**, 23–27 (1999).
[13]F. Durup and E. Jansson, "The quest of the violin bridge-hill," Acust. Acta Acust. **91**, 206–213 (2005).
[14]I. P. Beldie, "About the bridge hill mystery," Catgut Acoust. Soc. J. **4**, 9–13 (2003).
[15]G. Bissinger and A. Gregorian, "Relating normal mode properties of violins to overall quality: Signature modes," Catgut Acoust. Soc. J. **4**, 37–45 (2003).
[16]G. Weinreich, "What science knows about violins — and what it does not know," Am. J. Phys. **61**, 1067–1077 (1993).
[17]M. Bailey and G. Bissinger, "Modal analysis study of mode frequency and damping changes due to chemical treatments of the violin bridge," *Proceedings of the 13th International Modal Analysis Conference*- Soc. Exp. Mechanics, Bethel, CT, 1995, pp. 828–833.
[18]J. Woodhouse, "On the bridge-hill of the violin," Acustica Acta Acustica **91**, 155–165 (2005).
[19]W. J. Trott, "The violin and its bridge," J. Acoust. Soc. Am. **81**, 1948–1954 (1987).
[20]H. Dünnwald, "Deduction of objective quality parameters on old and new violins," Catgut Acoust. Soc. J. **1**, 1–5 (1991).
[21]H. Meinel, "Regarding the sound quality of violins and a scientific basis for violin construction," J. Acoust. Soc. Am. **29**, 817–822 (1957).
[22]G. Bissinger and J. C. Keiffer, "Radiation damping, efficiency, and directivity for violin normal modes below 4 kHz, ARLO **4**, 7–12 (2003), online at http://ojps.aip.org/ARLO/top.jsp
[23]H. A. Muller, "The function of the violin bridge," Catgut Acoust. Soc. Newsletter **31**, 19–22 (1979).
[24]F. A. Saunders, "The mechanical action of instruments of the violin family," J. Acoust. Soc. Am. **17**, 169–186 (1946).

High intensity focused ultrasound-induced gene activation in solid tumors

Yunbo Liu
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708

Takashi Kon and Chuanyuan Li
Department of Radiation Oncology, Duke University Medical Center, Durham, North Carolina 27708

Pei Zhong[a)]
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708

(Received 4 January 2006; revised 3 April 2006; accepted 22 April 2006)

In this work, the activation of heat-sensitive trans-gene by high-intensity focused ultrasound (HIFU) in a tumor model was investigated. 4T1 cancer cells (2×10^6) were inoculated subcutaneously in the hind limbs of Balb/C mice. The tumors were subsequently transducted on day 10 by intratumoral injection of a heat-sensitive adenovirus vector (Adeno-hsp70B-Luc at 2×10^8 pfu/tumor). On day 11, the tumors were heated to a peak temperature of 55, 65, 75, or 85 °C within 10–30 s at multiple sites around the center of the tumor by a 1.1- or 3.3-MHz HIFU transducer. Inducible luciferase gene expression was increased from 15-fold to 120-fold of the control group following 1.1-MHz HIFU exposure. Maximum gene activation (120-fold) was produced at a peak temperature of 65–75 °C one day following HIFU exposure and decayed to baseline within 7 days. HIFU-induced gene activation (75 °C-10 s) could be further improved by using a 3.3-MHz transducer and a dense scan strategy to 170-fold. Thermal stress, rather than nonthermal mechanical stress, was identified as the primary physical mechanism for HIFU-induced gene activation *in vivo*. Overall, these observations open up the possibility for combining HIFU thermal ablation with heat-regulated gene therapy for cancer treatment. [DOI: 10.1121/1.2205129]

PACS number(s): 43.80.Gx, 43.80.Cs [FD] Pages: 492–501

I. INTRODUCTION

High-intensity focused ultrasound (HIFU) has emerged as a viable noninvasive therapeutic modality for the treatment of a variety of solid tumors, including liver (Kennedy *et al.*, 2004), prostate (Chapelon *et al.*, 1999; Chaussy and Thuroff, 2003), breast (Hynynen *et al.*, 2001), and soft-tissue sarcoma (Wu *et al.*, 2004). New biomedical applications, such as hemostasis and gene delivery, have also been explored (Vaezy *et al.*, 1999; Miller and Song, 2003). Pioneering studies have demonstrated that HIFU, with spatial-peak temporal-average intensity (I_{SPTA}) between 10^3 and 10^4 W/cm^2, can produce well-defined thermal lesions in deep-sited tissue (ter Haar, 1995). A large volume of tumor tissue can be treated by scanning the HIFU focus in a matrix of positions. The fundamental physical mechanisms of HIFU ablation are coagulative thermal necrosis (>65 °C) and cavitation damage (Bailey *et al.*, 2003).

In addition to thermal ablation, preliminary yet encouraging evidence suggest that HIFU may induce a distinct stress response in sublethally injured tumor cells surrounding the focal lesion. For example, significant up-regulation of heat shock proteins (hsp) has been observed at the border of HIFU-induced necrosis region in patients with benign prostatic hyperplasia (BPH) (Kramer *et al.*, 2004). It has been further postulated that the up-regulation of hsp may play a critical role in eliciting an antitumor immunity (Kramer *et al.*, 2004; Kennedy, 2005). Known as "molecular chaperones," intracellular heat shock (or stress) proteins controlled by a specific family of hsp genes will be dramatically synthesized when cells are exposed to stressful or harmful environments (Morimoto, 1993). In particular, one member of the hsp70 gene family, hsp70B, is strictly stress inducible and absent in unstressed cells (Hildebrandt *et al.*, 2002).

In light of this, ultrasound-induced hyperthermia has been utilized as a noninvasive physical method to achieve spatial and temporal regulation of trans-gene expression under the control of hsp70B promoter both *in vitro* and *in vivo* (Vekris *et al.*, 2000; Guilhon *et al.*, 2003, Zhong *et al.*, 2004). Using ultrasound to regulate transgene expression could be particularly beneficial for site-specific gene therapy when systematic gene dissemination and expression in nontargeted areas are concerned. For example, following exposure to 2 W/cm^2 ultrasound for 20 min, Smith *et al.* (2002) demonstrated a locally induced luciferase or FasL gene expression after systemic delivery of adenoviral constructs under the control of hsp70B promoter. Furthermore, using a MRI-guided focused ultrasound system, trans-gene induction within internal organs was successfully achieved under well-

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

controlled thermal dosage (42 °C, >30 min) in rat livers and in canine prostates (Plathow *et al.*, 2005; Silcox *et al.*, 2005). It is worth noting that heat shock gene expression has also been detected following ultrasound physiotherapy, which is presumably induced by cavitation-associated mechanical stresses (Barnett *et al.*, 1994; Angles *et al.*, 1990).

Most recently, we have explored the physical conditions for HIFU-induced gene activation in sublethally injured cancer cells *in vitro* (Liu *et al.*, 2005). Following exposure to peak temperatures of 50 to 70 °C for 1 to 10 s, the heat shock responses of two cancer cell lines (HeLa and R3230Ac) were investigated. Consistent gene expression in the surviving cell population was detected under various HIFU thermal doses and the maximum gene activation was produced by 60 °C in 5-s heat shock exposure. These exposure conditions are similar to those experienced by the sublethally injured tumor cells surrounding HIFU-induced necrosis lesion. The logical extension of this preliminary *in vitro* study is to confirm *in vivo* that HIFU treatment can indeed induce trans-gene expression under the control of hsp70B promoter in tumor models. If such a hypothesis is confirmed, one could potentially take advantage of this unique biological response to explore a synergistic combination of *in situ* gene therapy with HIFU-produced thermal ablation to improve the overall efficacy and quality of cancer therapy.

The present study is therefore aimed to investigate the feasibility of simultaneous tumor ablation and activation of a transgene under the control of hsp70B promoter using a tumor-bearing mouse model. Specifically, after an intratumaral injection of adenovirus vector (Ad-hsp70B-Luc) for gene transduction, target tumors were scanned by either a 1.1- or 3.3-MHz HIFU transducer under the guidance of B-mode ultrasound imaging. *In vivo* gene expression activities following HIFU exposures using different combinations of peak temperature, treatment duration, and scanning strategy were evaluated. The results were compared with conventional hyperthermia treatment. Furthermore, the roles of two potential physical mechanisms, i.e., thermal stress and cavitation-associated nonthermal mechanical stress, in HIFU-induced gene activation *in vivo* were investigated.

II. MATERIALS AND METHODS

A. Tumor inoculation and gene transduction

All *in vivo* experimental procedures were carried out in accordance with the protocol approved by the Duke University Committee on the Use and Care of Animals. Female Balb/C mice (Charles River Laboratory, Wilmington, MA), 6–8 weeks old, were selected with a compatible mouse mammary carcinoma cancer cell line (4T1) to construct the murine tumor model. 4T1 cells were maintained routinely in DMEM culture medium with 10% heat-inactivated fetal bovine serum and 5% antibiotics in a humidified incubator at 37 °C containing 5% CO_2. For tumor inoculation, 2×10^6 4T1 cells suspended in 50 μL PBS were injected subcutaneously (s.c.) into the shaved right hind limb of the mouse. When the tumor reached a size of ~8 mm in maximum diameter in about 10 days, 30 μL adenoviral luciferase vectors with hsp70B promoters (Ad-hsp70B-Luc) were injected into the center of the tumor using a 30-gauge needle at a dosage of 2×10^8 pfu/tumor, following an established protocol (Wang *et al.*, 2005). After virus injection, the anesthetized animals were sent back to vivarium for virus dissemination and gene transduction inside the tumor for 24 h before HIFU treatment.

B. High-intensity focused ultrasound (HIFU) exposure system

The experiments were carried out utilizing a B-mode ultrasound imaging-guided HIFU exposure system shown in Fig. 1(a). Briefly, a HIFU transducer (H-102, Sonic Concepts, Seattle, WA) with a focal length of 63 mm, operated at either 1.1 MHz (fundamental) or 3.3 MHz (third harmonic), was mounted at the bottom of a heated (37 °C) chamber ($40\times30\times15$ cm^3, L×W×H) filled with degassed water. The transducer was driven by continuous sinusoidal signals produced by a function generator (33120A, Agilent, Palo Alto, CA), connected in series with a 55-dB power amplifier (A150, Electronic Navigation Industries, Rochester, NY). The operation and exposure parameters of the HIFU system were controlled by LabView programs via a GPIB board installed in a PC. The animal was placed in a holder specially designed to facilitate HIFU exposure to the tumor-bearing hind limb while protecting the internal organs of the mouse. Using a 5/7 MHz imaging probe (Terason 2000, Terason Inc., NJ), the target tumor area could be outlined to guide the HIFU treatment [Fig. 1(a)]. The pressure waveform and distribution in the focal plane of the HIFU transducer were measured by using a fiber optical probe hydrophone (FOPH-500, RF Acoustics, Leutenbach, Germany). When the function generator was operated at 0.1 V_{pp}, a peak positive (p^+) /peak negative (p^-) pressure of 3.1/−2.7 MPa and 7.1/−4.5 MPa were measured at the beam focus of the HIFU transducer at 1.1 and 3.3 MHz, respectively [Fig. 1(b)]. The corresponding −6-dB beam sizes at the focus along and transverse to the transducer axis were determined to be 11 mm×1.64 mm at 1.1 MHz and 5 mm×0.6 mm at 3.3 MHz, respectively.

C. Temperature measurement and thermal dose evaluation

A 0.2-mm bare-wire thermocouple (Customer designed IT-23, Physitemp Inc., Clifton, NJ) was used to measure the temperature rise inside the tumor tissue. Temperature output was conditioned by an electronically compensated isothermal terminal block (TC-2190, National Instrument, Austin, TX) and registered at a 6-Hz sampling rate using a Data Acquisition Board (NI4351, National Instrument, Austin, TX) controlled by a LabView program. For alignment, the mouse tumor embedded with the thermocouple was scanned across the acoustical field of the HIFU transducer operated at low intensity using a computer-controlled 3-D step motor (Velmex Inc., Bloomfield, NY) until the maximum temperature rise (<42 °C) was detected at the beam focus. Using this approach, temperature elevations and distribution inside the tumor tissue under the designated HIFU exposures were recorded. In this study, all the HIFU exposures are in con-

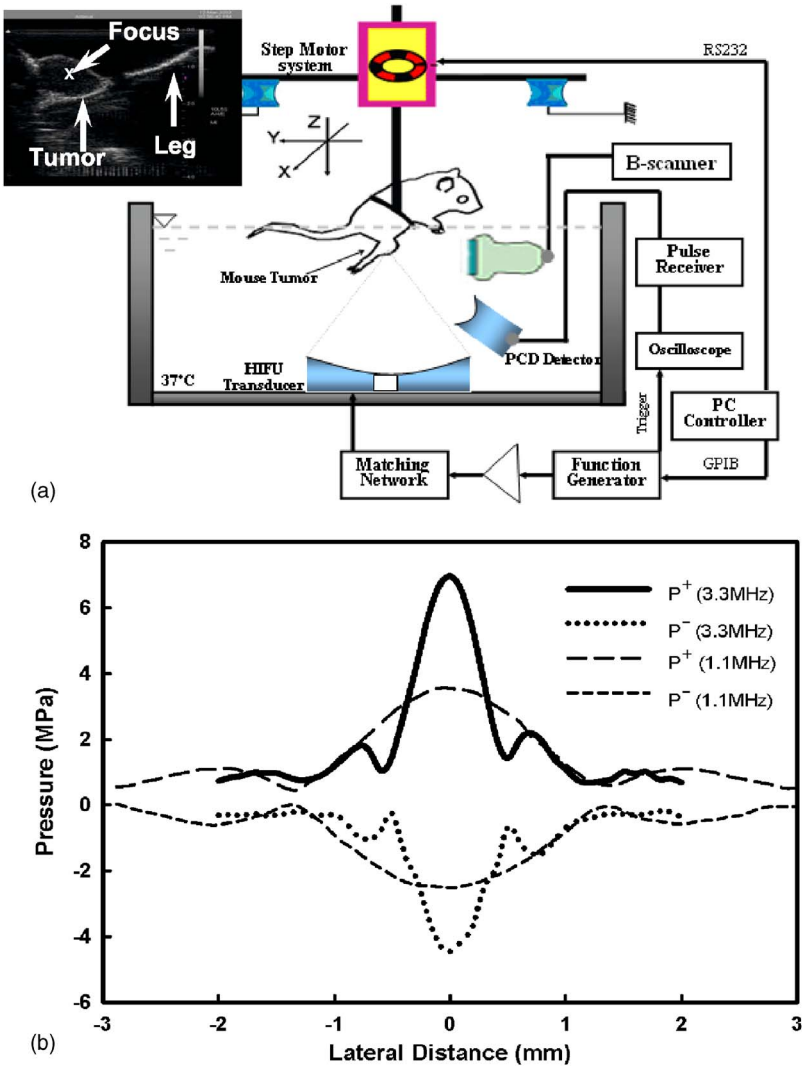


FIG. 1. (a) A schematic diagram of the B-mode ultrasound imaging-guided HIFU system. (b) Pressure distribution across the focal plane of the 1.1/3.3-MHz HIFU transducer in free field. p^+: peak positive pressure and p^-: peak negative pressure.

tinuous wave (CW) mode with an I_{SATA} between 605 and 2657 W/cm^2 calculated based on an established protocol (Harris, 1985). Figure 2 shows the temperature profiles produced at 1.1-MHz HIFU focus inside the tumor. By adjusting the exposure pressure, the temperature at the transducer focus could reach a peak value of 55, 65, 75, and 85 °C within a 10-s exposure duration [p^-=−4.9 to −9.1 MPa, Fig. 2(a)] or the same temperature of 75 °C in 5, 10, 20, and 30 s, respectively [p^-=−9.1 to −5.3 MPa, Fig. 2(b)]. Moreover, the lateral temperature distributions (perpendicular to the HIFU beam axis) under the 75 °C-10 s exposure condition, i.e., under the transducer output conditions for which a temperature rise to 75 °C was reached at the focus in 10 s, were also measured under 1.1- and 3.3-MHz exposure, which revealed a −6-dB lateral temperature beam width of 6 and 1.5 mm, respectively.

Sapareto and Dewey (1984) proposed the concept of equivalent thermal dose in order to elucidate the relationship between thermal dosimetry and biological response. The following equation is defined for single point equivalent thermal dose, also known as equivalent minutes at 43 °C (EM_{43}), as a function of exposure temperature, $T(x,y,z,t)$, and the total treatment duration D. When the temperature history of a point within the tissue is known, the thermal dose at that specific position can be calculated by

$$EM_{43}(x,y,z,T) = \int_0^D R^{43-T(x,y,z,t)}\, dt,$$

$$R = \begin{cases} 0, & T < 37\ ^\circ\mathrm{C}, \\ 0.25, & 37\ ^\circ\mathrm{C} \le T < 43\ ^\circ\mathrm{C}, \\ 0.5, & T \ge 43\ ^\circ\mathrm{C}. \end{cases}$$

Based on the temperature profile measured during HIFU exposure, the spatial distributions of the equivalent thermal dose (represented by EM_{43}) in the focal plane (transverse to the beam axis) and in a longitudinal plane across the beam axis of the transducer were evaluated in order to better understand the role of thermal dose in HIFU-induced gene activation and lesion formation.

D. Experimental design

A total of three series of experiments were conducted to investigate the trans-gene activation in the mouse tumor model during HIFU treatment. The first series were designed to demonstrate the feasibility of HIFU-elicited gene activa-

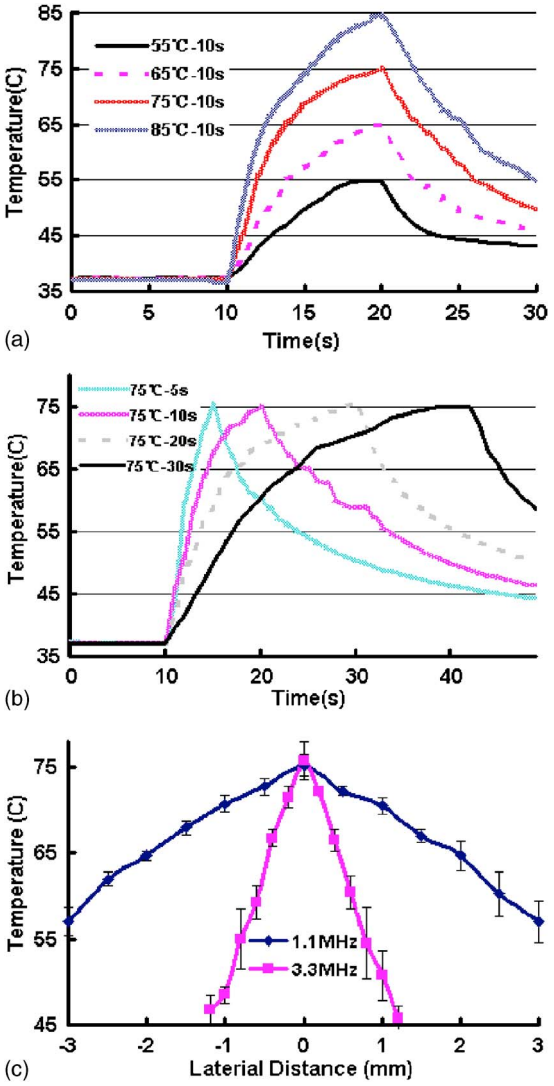


FIG. 2. Temperature profiles produced in mouse tumor tissues under (a) different peak temperatures (55 °C–85 °C) and (b) different exposure durations (5–30 s). (c) Lateral distribution of peak temperature in mouse tumor tissue produced by the 1.1-MHz/3.3-MHz HIFU transducer during 75 °C-10-s exposures.

tion *in vivo*. A total of five spatially distributed 1.1-MHz/10-s HIFU exposures with a peak temperature of 55, 65, 75, and 85 °C were delivered to the center of the tumor at 2-mm spatial interval in order to identify the appropriate HIFU exposimetry for gene activation. Water-bath hyperthermia treatment (42 °C-30 min) was used as positive control throughout this study.

Following the feasibility study, the second series of experiments was conducted to determine the optimal HIFU conditions for gene activation *in vivo*. Specifically, the 1.1-MHz HIFU exposures used were for increasing durations of 5, 10, 20, and 30 s with transducer outputs chosen so that, for each exposure, the peak temperature reached was 65 °C or 75 °C. In addition, the time course of hyperthermia- or HIFU-induced gene activation in a week was monitored and compared. The corresponding tumor growth curve was also monitored for 20 days after HIFU treatment. Once the proper thermal dose was identified, the effect of the scanning strategy on HIFU-induced gene activation (75 °C-10 s) was further investigated, using the 3.3-MHz HIFU transducer and a dense-scan pattern across a 6×6 mm^2 area with 1-mm step size.

In the third series of experiments, the contributions of two potential physical mechanisms, i.e., thermal stress and nonthermal mechanical stress, to HIFU-induced gene activation *in vivo* were compared. The thermal HIFU (75 °C-10 s) was evaluated by scanning the tumor within the focal plane (6×6 mm^2 at 1-mm step size) of the 3.3-MHz HIFU, which enhances thermal absorption while suppressing cavitation in the targeted tissue compared to the 1.1-MHz transducer. Under the same scanning protocol, the mechanical HIFU was evaluated by increasing the acoustic output pressure of the transducer by fourfold while reducing its duty cycle concomitantly from 100% (thermal HIFU) to 6.3% so that the same total acoustic power could be delivered. Using this approach, the accumulated thermal effect could be eliminated because the peak temperature inside the tumor during the 10-s mechanical HIFU exposure was kept below 40 °C when the ambient temperature in the water bath was maintained at 23 °C. Moreover, cavitation activities inside the tumor tissue during both the thermal and mechanical HIFU treatments were monitored by B-mode ultrasound imaging and passive cavitation detection (PCD) technique. For PCD measurements, a 3.5-MHz focused transducer (V380, Panametrics Inc., Waltham, MA) was positioned confocally with the 3.3-MHz HIFU transducer [Fig. 1(a)]. The data acquisition and signal processing protocols used are similar to previously reported procedures (Chen *et al.*, 2004, Rabkin *et al.*, 2005). Briefly, fast Fourier transform (FFT) spectrums of the acoustic emission signals emanated from HIFU-induced cavitation bubbles inside the tumor were averaged and compared at different exposure settings. The rms amplitude of the broadband noise signals for each FFT spectrum between 4.5 and 5.5 MHz was further calculated and presented in time sequence. The inertial cavitation (IC) activity was quantified by computing the cumulated IC dosage (ICD), which is defined by the integration of the rms amplitude (in dBm) of the power spectrum between 4.5 and 5.5 MHz over the entire 10-s exposure period (Chen *et al.*, 2003).

E. Gene expression assay

On the day for gene expression assay, 100 μL of aqueous D-luciferin solution was injected intraperitoneally into the anesthetized animal 20 min before *in vivo* gene expression analysis using a Xenogen *in vivo* bioluminescence imaging system (Xenogen Inc., Alameda, CA). The expressed luciferase protein catalyzes the oxidation of injected D-luciferin (enzyme substrate), resulting in the emission of bioluminescence (Wang *et al.*, 2005). Bioluminescence images were generated by integrating photon emission from the mice tumor during an exposure time of 60 s at a fixed sensitivity. The results were shown in pseudo-colors indicated by the color bar. The final images were represented by superimposing the pseudo-color bioluminescence images on conventional gray scale images taken separately (Wang *et al.*, 2005).

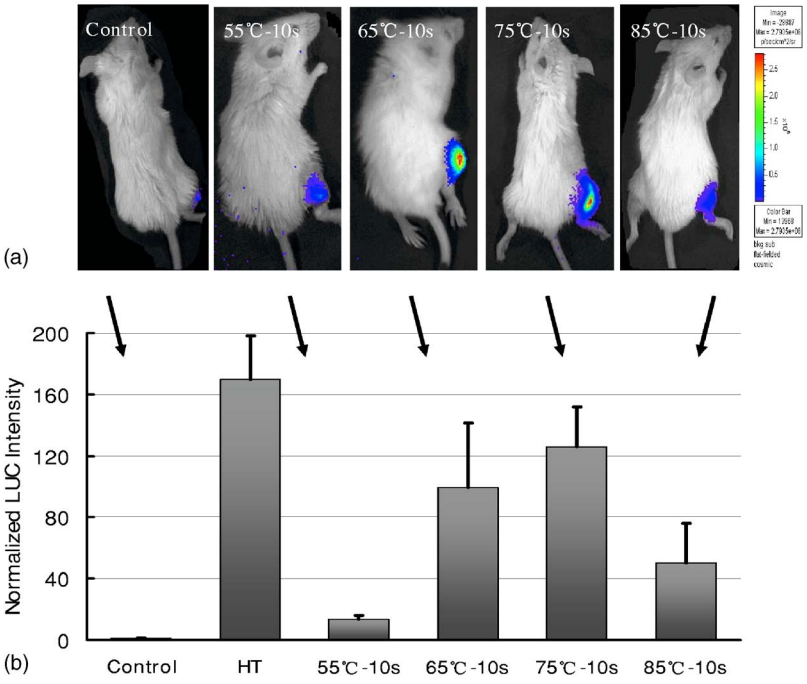


FIG. 3. (a) Representative bioluminescence images of luciferase distribution in the mouse model 24 h after HIFU exposure. (b) Quantitative luciferase intensity after 1.1-MHz/10-s HIFU exposure at peak temperatures of 55 °C–85 °C. HT: hyperthermia (42 °C, 30 min).

F. Tumor growth regression assay

Following HIFU treatment, tumor sizes were measured using an electronic digital caliper every other day and the tumor volume (V) was calculated using the formula for an ellipsoid, $V=(\pi/6)W^2L$, where L is the longest dimension and W is the shortest dimension of the tumor. The tumor growth curves for 20 days were determined and compared with the control group.

G. Statistical analysis

Each experiment data point was averaged from six samples unless otherwise indicated. Student's t test was used to determine the statistical significance and $p<0.05$ was considered to indicate a statistically significant difference between two experimental configurations. All statistical analyses were computed using the Office Excel program (Microsoft Inc., Seattle, WA).

III. RESULTS

A. Feasibility of HIFU-induced gene activation *in vivo*

Enhanced luciferase expression in the tumor was clearly detected in a wide range of temperatures at day 1 following 1.1-MHz/10-s HIFU treatment [Fig. 3(a)]. It is important to note that no transgene expression was observed in other vital organs, such as lung and liver. Compared to the control group, luciferase expression intensity within the tumor volume was found to increase by 15-, 95-, 120-, and 55-fold with a peak HIFU temperature of 55, 65, 75, and 85 °C, respectively [Fig. 3(b)]. Maximum luciferase expression of 120-fold was detected in the 75 °C HIFU treatment group although it is not statistically different from the 65 °C group. Interestingly, higher (85 °C) or lower (55 °C) temperature levels induced less gene expression activity in the solid tumor. These observations of temperature-dependent gene activation during *in vivo* study are consistent with our *in vitro* experiments using HeLa and R3230Ac cell lines (Liu *et al.*, 2005), as well as 4T1 cells (data not shown). Altogether, these results support the notion that an optimal window of peak temperature (65 °C–75 °C) exists for eliciting maximum gene expression following a 10-s HIFU exposure.

B. Effect of treatment duration on HIFU-induced gene activation

As illustrated in Fig. 4(a), gene expression increased initially with the exposure duration and saturated after 10–20-s exposure at a peak temperature of 65 °C or 75 °C. Further increase of exposure duration to 30 s was found to inhibit the overall gene expression intensity at both temperature levels. Compared to the control group, gene expression following 1.1-MHz HIFU at the peak temperature of 75 °C was found to increase by 40-, 120-, 115-, and 60-fold for exposure durations of 5, 10, 20, and 30 s, respectively. Overall, 10–20-s HIFU exposure with a peak temperature between 65 °C and 75 °C led to a maximum gene activation in the target tumor tissue. Furthermore, the time course of gene activation elicited by hyperthermia (42 °C-30 min) and 1.1-MHz HIFU (75 °C-10 s) treatment were compared. As shown in Fig. 4(b), peak luciferase expression level was observed at day 1 following both treatment regimens and decayed gradually to the background level within 7 days. Specifically, luciferase intensity rose to 160- and 58-fold on day 1 and day 3 following water-bath hyperthermia (42 °C-30 min) compared to 120- and 33-fold increase induced by the 1.1-MHz HIFU treatment (75 °C-10 s). The discrepancy might be caused in part by the heterogeneous temperature distribution inside the solid tumor produced by the 1.1-MHz HIFU exposure protocol, in contrast to the relatively uniform temperature field in water-bath hyperthermia.

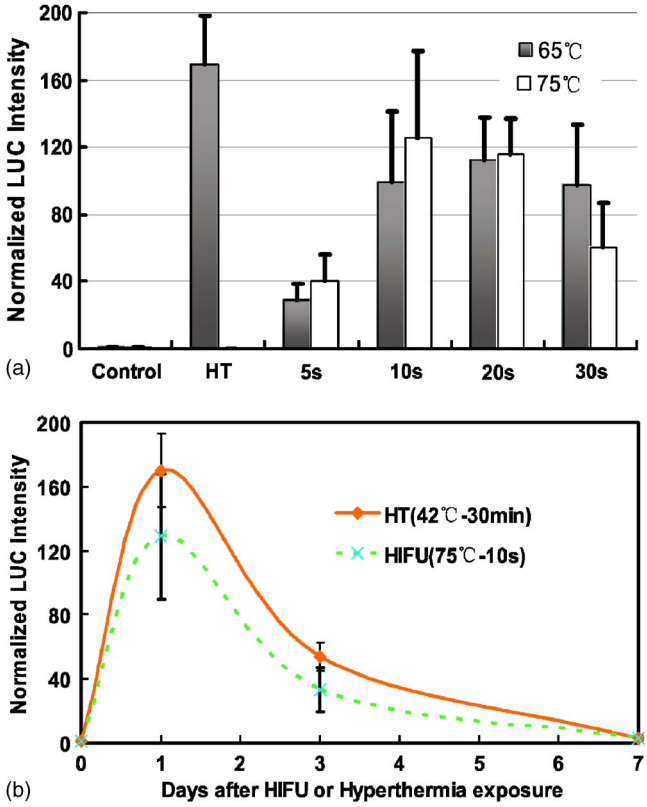


FIG. 4. (a) Effect of HIFU treatment duration on transgene activation (5–30 s) at a peak temperature of 65 °C or 75 °C produced by the 1.1-MHz transducer. (b) Comparison of the time course of luciferase gene expression induced by HIFU and hyperthermia (HT) treatments.

C. Tumor growth regression

Following the HIFU treatment (75 °C-10 s, 1.1 MHz), the tumor volume shrunk to approximately 33% in 12 days while the tumor in the control group increased to 300% of the original size, at which point the animal was euthanized (Fig. 5). The volume of HIFU-treated tumor, however, rebound gradually to 65% of the original size at the end of a 20-day observation period.

D. Effect of scanning strategy on HIFU-induced gene activation

Under the same thermal exposure (75 °C-10 s), the efficiency of HIFU-induced gene activation could be further

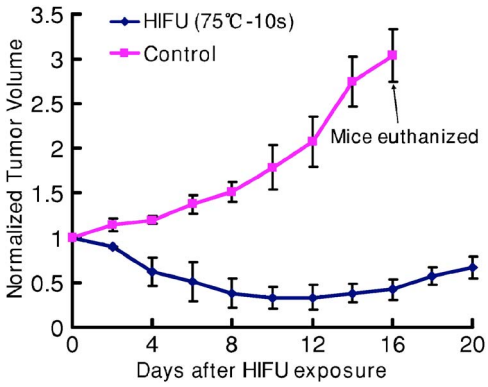


FIG. 5. The time course of tumor growth regression following 1.1-MHz/10 s HIFU treatment (75 °C-10 s).

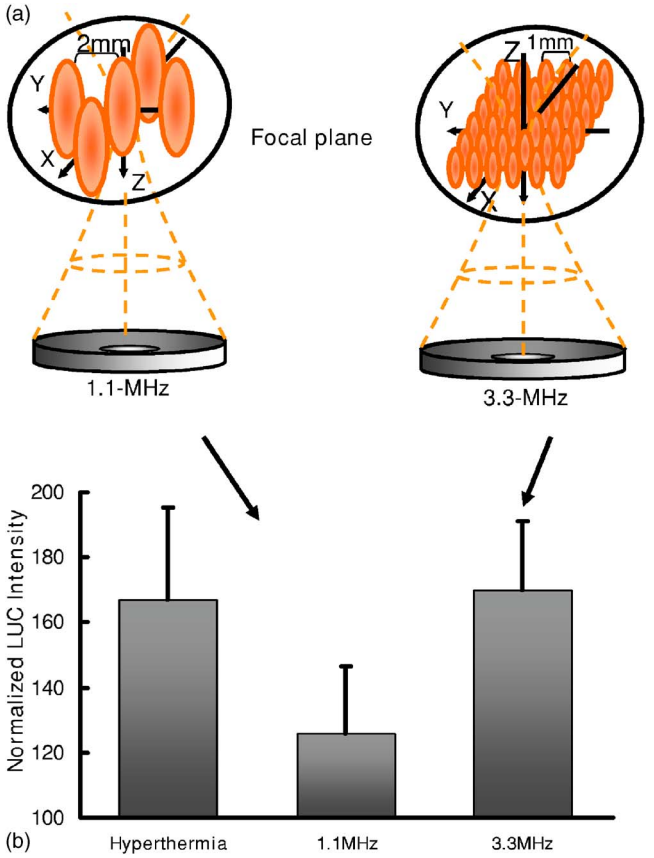


FIG. 6. (a) Schematic illustrations of sparse (1.1 MHz) versus dense (3.3 MHz) HIFU scan strategies and (b) corresponding gene activation efficiency following 75 °C-10 s HIFU exposure.

improved by proper selection of transducer frequency (1.1 vs. 3.3-MHz) and scanning density [5 vs. 36 spots as shown in Fig. 6(a)]. Figure 6(b) shows that HIFU-induced gene activation was elevated from 120-fold at 1.1 MHz with a sparse scan to 170-fold at 3.3 MHz with a dense scan ($p < 0.05$). The latter is statistically comparable to the 168-fold increase induced by hyperthermia ($p > 0.3$). It should be noted, however, because both frequency and scanning density were changed, the current results could not pinpoint which variable is more important in HIFU-induced gene activation.

E. Physical mechanism: Thermal stress versus mechanical stress

Figure 7 shows the equivalent thermal dose (represented by EM_{43}) in tumor tissues, calculated based on temperature measurements in both the focal and beam planes during a single 3.3-MHz HIFU exposure (75 °C-10 s). The calculated iso-exposure contours for 55 °C-10 s, 65 °C-10 s, and 75 °C-10 s were outlined explicitly, together with the empirical EM_{43} (=240 min) contour for thermal necrosis (Damianou *et al.*, 1995). In the focal plane, the 55, 65, and 75 °C iso-exposure contours are concentric circles with diameters of 1.6, 0.8, and 0.3 mm, respectively [Fig. 7(a)]. In comparison, the corresponding contours in the beam plane are close to "cigar" shapes. For instance, the axial and lateral beam diameters for the 65 °C-10 s contour were calculated

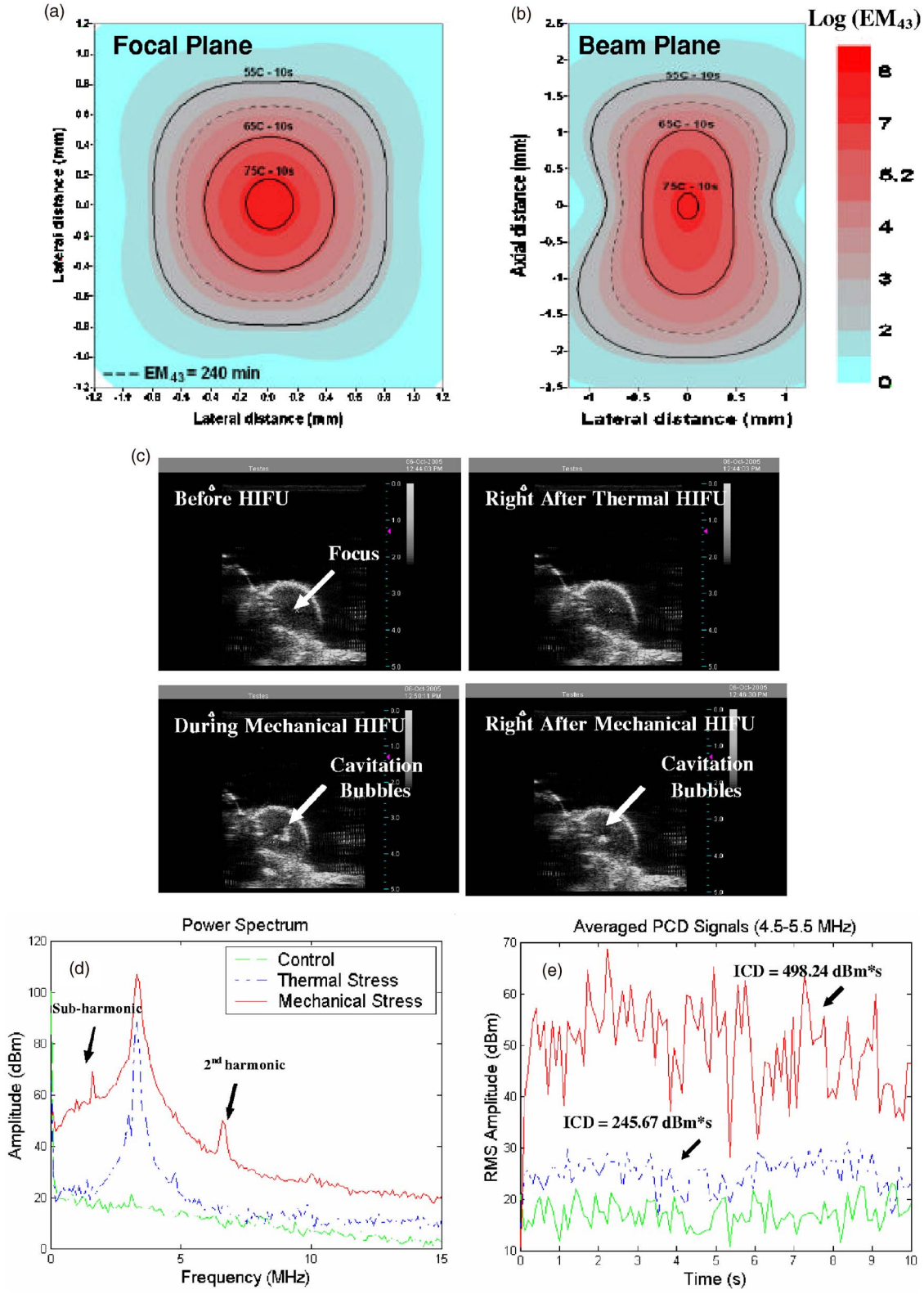


FIG. 7. The distribution of equivalent thermal dose (represented by EM_{43} in logarithmic scale) within (a) focal plane and (b) beam plane of the 3.3-MHz transducer during 75 °C-10 s HIFU exposure. (c) B-mode images of cavitation activity in mouse tumor tissues induced by thermal and mechanical HIFU. (d) Representative frequency spectrum of PCD signals, and (e) time evolution of the averaged broadband noise (4.5–5.5 MHz) in the PCD spectrum during a 10-s HIFU exposure.

to be 2.2 and 1.2 mm [Fig. 7(b)]. The total areas within the 55 °C-10 s contour, which is the primary gene activation zone based on preliminary *in vitro* studies, were calculated to be 2.5 and 7.0 mm^2 in the focal and beam planes, respectively.

Compared to thermal HIFU, mechanical HIFU exposure often generates a bright hyperechoic spot at the beam focus, which is presumably associated with the induction of cavitation bubbles in the target tumor tissue [Fig. 7(c)]. A significant increase in the broadband noise of the acoustic emission

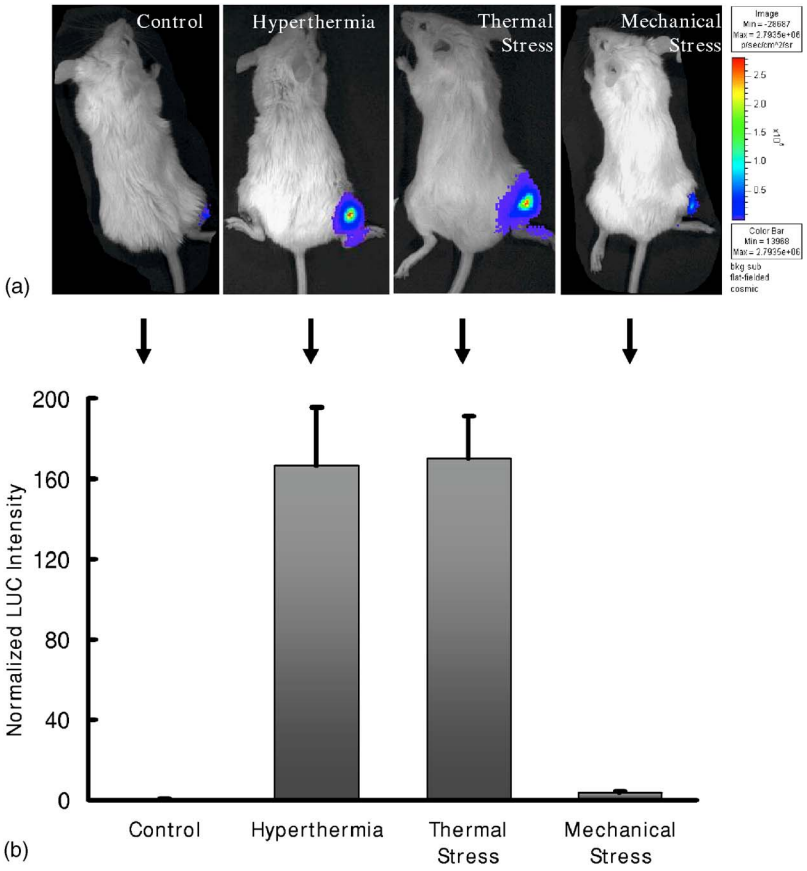


FIG. 8. (a) Representative bioluminescence images in the mouse model and (b) corresponding luciferase intensity induced by 3.3-MHz thermal and mechanical HIFU treatments. Output conditions are I_{SAPA} =430 W/cm^2 at CW mode for 10 s for thermal HIFU (75 °C-10 s) and I_{SAPA}=6849 W/cm^2 with a duty cycle of 6.3% (burst mode) for mechanical HIFU.

signal spectrum was also detected during mechanical HIFU exposure [Fig. 7(d)]. In particular, generation of subharmonic and second harmonic components could be clearly detected, which might be produced by the nonlinear scattering of cavitation bubbles inside the tumor tissue (Miller and Bao, 1998). Figure 7(e) shows the average rms amplitude of the broadband noise in the frequency range of 4.5–5.5 MHz during a 10-s treatment period. The inertial cavitation dosage (ICD) of 498.2±42.5 dBm s produced by the mechanical HIFU exposure is significantly higher than the corresponding value of 245.67±27.3 dBm s generated by the thermal HIFU exposure ($p<0.05$). Altogether these results suggest that significantly higher level of cavitation activities and associated mechanical stresses were generated in the targeted tumor tissue by the mechanical HIFU exposure than its counterpart of thermal HIFU exposure.

As shown in Fig. 8, thermal HIFU was found to stimulate a gene activation of 170-fold in the target tumor tissue compared to the control group, while the mechanical HIFU only produced a fourfold increase. Together with the results shown in Fig. 7, it is concluded that thermal stress is the primary physical mechanism for HIFU-induced gene activation *in vivo*. The mechanical stresses produced during HIFU exposure (primarily associated with cavitation), although strong enough to cause cell lysis and tissue damage, may not be sufficient to elicit a measurable heat shock response (i.e., activation of hsp70B gene promoter). It is worth noting, however, that mechanical and thermal stresses may interact synergistically to enhance the thermal deposition during HIFU treatment (Bailey *et al.*, 2003).

IV. DISCUSSION

In clinical HIFU therapy, the peak temperature at the beam focus could exceed 70 °C within several seconds, leading to coagulative necrosis and lesion formation surrounded by sub-lethally injured tumor tissues (ter Haar, 1995). Up-regulated heat shock response has been detected in the border zone of necrosis lesion during HIFU treatment of BPH, raising the possibility of hsp-mediated immune response following HIFU (Kramer *et al.*, 2004; Kennedy, 2005). The present study demonstrates that HIFU can indeed elicit strong trans-gene activation under the control of hsp70B promoter *in vivo*, presumably in sublethally injured tumor tissues. It is observed that HIFU-induced gene activation depends considerably on exposure conditions, such as peak temperature, duration, frequency, and scan strategy. Furthermore, thermal stress, instead of cavitation-associated mechanical stress, was found to be the primary physical mechanism for HIFU activation of hsp70B promoter *in vivo*, which is consistent with the results of our previous *in vitro* study (Liu *et al.*, 2005).

Under the 1.1-MHz HIFU exposure protocol, maximum gene expression following a 10-s exposure was produced at peak temperature in the range of 65 °C to 75 °C both *in vitro* and *in vivo*. Lower peak temperature (<55 °C) cannot stimulate sufficient heat shock response while higher peak temperature (>85 °C) may activate extensively cellular apoptosis pathways, leading to programmatic cell death and thus hindering the overall gene expression (Barry *et al.*, 1990). This interpretation is further supported by the fact that higher

gene expression was achieved at 10–20-s exposure rather than 5 or 30 s with a fixed peak temperature of 65 °C or 75 °C [Fig. 4(a)]. Heat-sensitive gene expression triggered by HIFU and hyperthermia revealed a similar time course, which was found to peak at day 1 post-treatment and decay gradually within a week [see Fig. 4(b)]. Biologically, the same molecular pathway of heat shock response may be activated by these two treatment regimens through a thermal dose threshold mechanism, despite the dramatically different rate at which the thermal energy is delivered to the target tissue. Our results also suggest that optimization of HIFU transducer frequency and scan strategy is important for achieving maximum gene activation *in vivo* (see Fig. 6).

Because of the nonuniform thermal field produced by a HIFU transducer, it is likely that the gene expression pattern within the targeted tumor volume will be heterogeneous. However, such spatial variation in gene expression cannot be resolved by the Xenogen *in vivo* bioluminescence imaging system used in this study. Future investigations are warranted to determine the spatial distribution of HIFU-induced gene activation in tumor tissues, which may provide critical insight for optimizing HIFU treatment strategies to maximize gene activation. Based on pilot *in vitro* cell studies, we have observed that luciferase positive cells surviving the HIFU treatment were primarily produced between the 55 °C and 75 °C thermal dose contours as outlined in Fig. 7(a) (data not shown). In this range, escalating HIFU thermal exposure levels will increase progressively the resultant cell death while elevating concomitantly the heat shock response (Liu *et al.*, 2005). A combination of cell survival rate and the intensity of heat shock response per cell determines the total accumulated gene expression in the tumor tissue, as shown in Figs. 3 and 8(a).

It is interesting to note that the thermal necrosis criteria (i.e., EM_{43}=240 min) widely quoted in HIFU literature also falls within the thermal exposure range for gene activation [see Fig. 7(a)]. This empirical thermal necrosis criteria were derived based on the histological analysis of muscle tissues in 4–21 days following 44 °C-60 min hyperthermia treatment (Damianou *et al.*, 1995; Jansen and Haveman, 1990). Although majority of the cancer cells or tissue exposed to HIFU at a thermal dose of EM_{43}=240 min will die acutely or gradually as a result of apoptosis or depletion of blood and nutrition supplies, a few percent of HIFU treated cells and/or tissue could survive (Liu *et al.*, 2005; Wu *et al.*, 2004). This observation probably reflects primarily the heterogeneous nature of the biological response of cells and tissues to stress. Therefore, it would be more accurate to interpret the thermal necrosis criteria (EM_{43}=240 min) as a threshold at which coagulative necrosis may result in biological cells and tissues, but not as a criterion to ensure total necrosis. It is also desirable in the future to carry out systematic investigations both *in vitro* and *in vivo* to better determine HIFU-induced stress responses at cellular and molecular levels and their correlation with sublethal heat shock response and tissue necrosis.

HIFU-induced gene activation may also be useful in regulating site-specific gene expression. Systemic dissemination of viral vectors following intratumoral delivery has been found to be a serious problem that may limit its application in cancer gene therapy (Wang *et al.*, 2005). While others have focused on development of novel carriers that can significantly reduce the systemic leakage of viral vectors from the injection site in the solid tumor (Wang *et al.*, 2005), the results of this study suggest that by employing a heat-sensitive hsp-70B promoter HIFU can be used as a noninvasive physical method to control trans-gene expression both spatially and temporally, thus eliminating the adverse effects due to systemic dissemination of the viral vectors.

As shown in Fig. 5, following HIFU treatment the volume of tumor was initially reduced by 70% in 2 weeks and, subsequently, the tumor started to regrow. This recurrence, also frequently observed following clinical HIFU therapy (Wu *et al.*, 2004), may arise from sublethally injured tumor cells that survive the HIFU treatment. Such a limitation of current HIFU cancer therapy may be potentially overcome by a synergistic combination of HIFU thermal ablation with heat-inducible cytotoxic or immunostimulatory gene therapy. Conceptually, HIFU can be used to activate therapeutic genes in sublethally injured tumor cells while thermally debulking the primary tumor mass. Using this strategy, heat-induced cytotoxic gene products or immunostimulatory factors may be produced simultaneously during HIFU therapy to improve the killing of residual or distal metastasis tumor cells via "bystander" effects or an enhanced antitumor immune response (Li and Dewhirst, 2002). Alternatively, a pretreatment of HIFU-induced gene activation at sublethal thermal dose in the target tumor may be applied to boost the host antitumor immune response before a full dose of regular HIFU for thermally ablating the tumor mass is executed. With the advance of transducer technology and real-time thermal dosimetry monitoring by MRI (Hynynen *et al.*, 2001), such new therapeutic strategies are feasible and should be explored in the future to improve the effectiveness of HIFU cancer therapy.

In conclusion, the present work opens up a new paradigm for HIFU-regulated trans-gene activation *in vivo*. Further animal studies are underway to explore the potential for a synergistic combination of HIFU-induced thermal ablation with heat-induced gene therapy to improve the overall quality and effectiveness of cancer therapy.

ACKNOWLEDGMENTS

This work was supported in part by NIH through Grants No. RO1-EB02682, R21-CA91166, and RO1-CA81512.

Angles, J. M., Walsh, D. A., Li, K., Barnett, S. B., and Edwards, M. J. (**1990**). "Effects of pulsed ultrasound and temperature on the development of rat embryos in culture," Teratology **42**, 285–293.

Bailey, M. R., Khokhlova, V. A., Sapozhnikov, O. A., Kargl, S. G., and Crum, L. A. (**2003**). "Physical Mechanisms of the Therapeutic Effect of Ultrasound," Acoust. Phys. **49**, 437–464.

Barnett, S. B., ter Haar, G. R., Ziskin, M. C., Nyborg, W. L., Maeda, K., and Bang, J. (**1994**). "Current status of research on biophysical effects of ultrasound," Ultrasound Med. Biol. **20**, 205–218.

Barry, M. A., Behnke, C. A., and Eastman, A. (**1990**). "Activation of programmed cell death (apoptosis) by cisplatin, other anticancer drugs, toxins and hyperthermia," Biochem. Pharmacol. **40**, 2353–2362.

Chapelon, J. Y., Ribault, M., Vernier, F., Souchon, R., and Gelet, A. (**1999**). "Treatment of localized prostate cancer with transrectal high intensity focused ultrasound," Eur. J. Ultrasound **9**, 31–38.

Chaussy, C., and Thuroff, S. (**2003**). "The status of high-intensity focused ultrasound in the treatment of localized prostate cancer and the impact of a combined resection," Curr. Urol. Rep. **4**, 248–252.
Chen, W. S., Brayman, A. A., Matula, T. J., and Crum, L. A. (**2003**). "Inertial cavitation dose and hemolysis produced in vitro with or without Optison," Ultrasound Med. Biol. **29**, 725–737.
Chen, W. S., Lu, X., Liu, Y., and Zhong, P. (**2004**). "The effect of surface agitation on ultrasound-mediated gene transfer in vitro," J. Acoust. Soc. Am. **116**, 2440–2450.
Damianou, C. A., Hynynen, K., and Fan, X. (**1995**). "Evaluation of accuracy of a theoretical model for predicting the necrosed tissue volume during focused ultrasound surgery," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 183–187.
Guilhon, E., Voisin, P., de Zwart, J. A., Quesson, B., Salomir, R., Maurange, C., Bouchaud, V., Smirnov, P., de Verneuil, H., Vekris, A., Canioni, P., and Moonen, C. T. (**2003**). "Spatial and temporal control of transgene expression in vivo using a heat-sensitive promoter and MRI-guided focused ultrasound," J. Gene Med. **5**, 333–342.
Harris, G. R. (**1985**). "A discussion of procedures for ultrasonic intensity and power calculations from miniature hydrophone measurements," Ultrasound Med. Biol. **11**, 803–817.
Hildebrandt, B., Wust, P., Ahlers, O., Dieing, A., Sreenivasa, G., Kerner, T., Felix, R., and Riess, H. (**2002**). "The cellular and molecular basis of hyperthermia," Crit. Rev. Oncol. Hematol. **43**, 33–56.
Hynynen, K., Pomeroy, O., Smith, D. N., Huber, P. E., McDannold, N. J., Kettenbach, J., Baum, J., Singer, S., and Jolesz, F. A. (**2001**). "MR imaging-guided focused ultrasound surgery of fibroadenomas in the breast: a feasibility study," Radiology **219**, 176–185.
Jansen, W., and Haveman, J. (**1990**). "Histopathological changes in the skin and subcutaneous tissues of mouse legs after treatment with hyperthermia," Pathol. Res. Pract. **186**, 247–253.
Kennedy, J. E. (**2005**). "High-intensity focused ultrasound in the treatment of solid tumors," Nat. Rev. Cancer **5**, 321–327.
Kennedy, J. E., Wu, F., ter Haar, G. R., Gleeson, F. V., Phillips, R. R., Middleton, M. R., and Cranston, D. (**2004**). "High-intensity focused ultrasound for the treatment of liver tumors," Ultrasonics **42**, 931–935.
Kramer, G., Steiner, G., Grobl, M., Hrachowitz, K., Reithmayr, F., Paucz, L., Newman, M., Madersbacher, S., Gruber, D., Susani, M., and Marberger, M. (**2004**). "Response to sublethal heat treatment of prostatic tumor cells and prostatic tumor infiltrating T-cells," Prostate **58**, 109–120.
Li, C. Y., and Dewhirst, M. W. (**2002**). "Hyperthermia-regulated immunogene therapy," Int. J. Hyperthermia **18**, 586–596.
Liu, Y., Kon, T., Li, C., and Zhong, P. (**2005**). "High intensity focused ultrasound-induced gene activation in sublethally injured tumor cells *in vitro*," J. Acoust. Soc. Am. **118**, 3328–3336.
Miller, D. L., and Bao, S. (**1998**). "The relationship of scattered subharmonic, 3.3 MHz fundamental and second harmonic signals to damage of monolayer cells by ultrasonically activated Albunex," J. Acoust. Soc. Am. **103**, 1183–1189.
Miller, D. L., and Song, J. (**2003**). "Tumor growth reduction and DNA transfer by cavitation-enhanced high-intensity focused ultrasound *in vivo*," Ultrasound Med. Biol. **29**, 887–893.
Morimoto, R. I. (**1993**). "Cells in stress: transcriptional activation of heat shock genes," Science **259**, 1409–1410.
Plathow, C., Lohr, F., Divkovic, G., Rademaker, G., Farhan, N., Peschke, P., Zuna, I., Debus, J., Claussen, C. D., Kauczor, H. U., Li, C. Y., Jenne, J., and Huber, P. (**2005**). "Focal gene induction in the liver of rats by a heat-inducible promoter using focused ultrasound hyperthermia: preliminary results," Invest. Radiol. **40**, 729–735.
Rabkin, B. A., Zderic, V., and Vaezy, S. (**2005**). "Hyperecho in ultrasound images of HIFU therapy: involvement of cavitation," Ultrasound Med. Biol. **31**, 947–956.
Sapareto, S. A., and Dewey, W. C. (**1984**). "Thermal dose determination in cancer therapy," Int. J. Radiat. Oncol., Biol., Phys. **10**, 787–800.
Silcox, C. E., Smith, R. C., King, R., McDannold, N., Bromley, P., Walsh, K., and Hynynen, K. (**2005**). "MRI-guided ultrasonic heating allows spatial control of exogenous luciferase in canine prostate," Ultrasound Med. Biol. **31**, 965–970.
Smith, R. C., Machluf, M., Bromley, P., Atala, A., and Walsh, K. (**2002**). "Spatial and temporal control of transgene expression through ultrasound-mediated induction of the heat shock protein 70B promoter in vivo," Hum. Gene Ther. **13**, 697–706.
ter Haar, G. R. (**1995**). "Ultrasound focal beam surgery," Ultrasound Med. Biol. **21**, 1089–1100.
Vaezy, S., Marti, R., Mourad, P., and Crum, L. A. (**1999**). "Hemostasis using high intensity focused ultrasound," Eur. J. Ultrasound **9**, 79–87.
Vekris, A., Maurange, C., Moonen, C., Mazurier, F., De Verneuil, H., Canioni, P., and Voisin, P. (**2000**). "Control of transgene expression using local hyperthermia in combination with a heat-sensitive promoter," J. Gene Med. **2**, 89–96.
Wang, Y., Yang, Z., Liu, S., Kon, T., Krol, A., Li, C. Y., and Yuan, F. (**2005**). "Characterisation of systemic dissemination of nonreplicating adenoviral vectors from tumours in local gene delivery," Br. J. Cancer **92**, 1414–1420.
Wu, F., Wang, Z., Chen, W., Zou, J., Bai, J., Zhu, H., Li, K., Xie, F., Jin, C., Su, H., and Gao, G. (**2004**). "Extracorporeal focused ultrasound surgery for treatment of human solid carcinomas: early Chinese clinical experience," Ultrasound Med. Biol. **30**, 245–260.
Zhong, P., Liu, Y., Li, Z., and Li, C. (**2003**). "Control of gene expression by ultrasound," Proc. of the 3rd Int. Symp. on Therapeutic Ultrasound, Lyon, France, pp. 136–141.

Individual variation in the pup attraction call produced by female Australian fur seals during early lactation

Joy S. Tripovich[a)]
Faculty of Veterinary Science, University of Sydney, Room 225-227, J.D. Stewart Building B01, University of Sydney, New South Wales, 2006, Australia and
Australian Marine Mammal Research Center, Zoological Parks Board of New South Wales, New South Wales / Faculty of Veterinary Science, University of Sydney, New South Wales, 2006, Australia

Tracey L. Rogers
Faculty of Veterinary Science, University of Sydney, New South Wales, 2006, Australia and
Australian Marine Mammal Research Center, Zoological Parks Board of New South Wales, P.O. Box 20, Mosman, New South Wales, 2088, Australia / Faculty of Veterinary Science, University of Sydney, New South Wales, 2006 Australia

Rhondda Canfield
Faculty of Veterinary Science, University of Sydney, Room 309, J.D. Stewart Building B01, University of Sydney, New South Wales, 2006, Australia

John P. Y. Arnould
[1]*School of Life and Environmental Sciences, Deakin University, 221 Burwood Highway, Burwood, VIC, 3125, Australia*

(Received 29 June 2005; revised 16 March 2006; accepted 10 April 2006)

Otariid seals (fur seals and sea lions) are colonial breeders with large numbers of females giving birth on land during a synchronous breeding period. Once pups are born, females alternate between feeding their young ashore and foraging at sea. Upon return, both mother and pup must relocate each other and it is thought to be primarily facilitated by vocal recognition. Vocalizations of thirteen female Australian fur seals (*Arctocephalus pusillus doriferus*) were recorded during the breeding seasons of December 2000 and 2001, when pups are aged from newborns to one month. The pup attraction call was examined to determine whether females produce individually distinct calls which could be used by pups as a basis for vocal recognition. Potential for individual coding, discriminant function analysis (DFA), and classification and regression tree analysis were used to determine which call features were important in separating individuals. Using the results from all three analyses: F_0, MIN F and DUR were considered important in separating individuals. In 76% of cases, the PAC was classified to the correct caller, using DFA, suggesting that there is sufficient stereotypy within individual calls, and sufficient variation between them, to enable vocal recognition by pups of this species. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2202864]

PACS number(s): 43.80.Ka [WWA] Pages: 502–509

I. INTRODUCTION

Recognition between parents and their offspring has been studied extensively in colonial species, e.g., Californian sea lions, *Zalophus californianus* (Schusterman *et al.*, 1992) Mexican free-tailed bats, *Tadarida brasiliensis mexicana* (Balcombe and McCracken 1992); razorbills, *Alca torda* (Insley *et al.*, 2003b), and king penguins, *Aptenodytes patagonicus* (Jouventin *et al.*, 1999). In some colonial species, offspring are usually mobile at a young age and consequently are able to socialize with similar-aged conspecifics and parents must be able to recognize them when they return from foraging trips (Scherrer and Wilkinson, 1993). As parental care promotes the survival of young, thus enhancing the parents' own reproductive success, selection should favor a parent-offspring recognition system (Gubernick, 1981).

In pinnipeds, maternal-offspring recognition appears to be widespread, with most exhibiting some degree of recognition. However, there are some exceptions, for example a lack of maternal recognition exhibited by Hawaiian monk seals (Job *et al.*, 1995). Various sensory modalities such as spatial, vocal, and olfactory cues are also considered important in the reunion process between a mother and her pup (Riedman, 1990). However, in a crowded breeding colony, acoustic signaling is thought to be more effective for long-range communication (Trillmich, 1981).

Breeding and maternal care strategies among the Otariidae (fur seals and sea lions) are generally similar, with mothers giving birth at a natal colony and providing exclusive care to their own young. A few days after birth, females depart on foraging trips offshore and upon their return must relocate their own young within the colony (Riedman, 1990). This process continues until weaning. Once pups are born, there is an initial period of bonding where nuzzling and vocalizing occurs between a mother and her newborn (Ried-

[a)]Electronic mail:joytripovich@hotmail.com

 0001-4966/2006/120(1)/502/8/$22.50

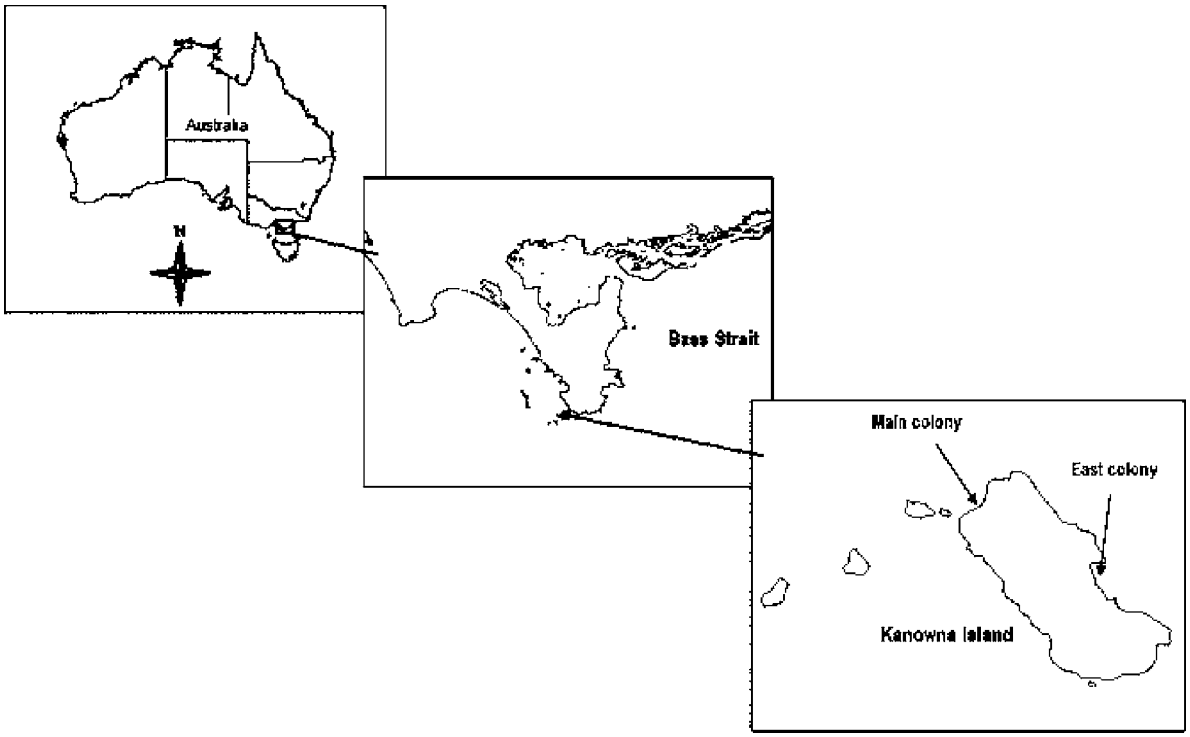


FIG. 1. Map of the study site from which vocalizations of female Australian fur seals were recorded. Main and East Colony indicated on map.

man, 1990). Mothers call to pups using the call termed the pup attraction call (PAC) and pups counter call using the female attraction call. This period of intense vocalizing aids imprinting between a mother and her newborn pup (Riedman, 1990), and occurs within a few days of birth (Charrier *et al.*, 2001).

For vocalizations to be used in individual recognition they must display stereotypy within individuals and significant variation between them (Falls, 1982). Although the presence of individual variation is insufficient evidence that recognition occurs, it is the important initial stage in demonstrating a potential for the recognition process. Individuality of the PAC has been described in several otariid species (Insley *et al.*, 2003a) with the overall structure of the PAC showing general similarities, although slight differences were evident between the species (Page *et al.*, 2002; Stirling and Warneke, 1971).

Preliminary analysis of the PAC produced by female Australian fur seals suggested it was of a lower frequency than that produced by other *Arctocephalines* (Stirling and Warneke, 1971). Detailed information on the structure of the Australian fur seal PAC, however, is lacking such that it is not possible to discern whether it is distinct from, or displays a different degree of individuality than, that of other *Arctocephalines*.

The aims of this study, therefore, were to: (1) establish detailed acoustic parameters that describe the PAC produced by female Australian fur seals; (2) determine the degree of individual variation; and (3) determine the acoustic features that contribute to the individuality of calls.

II. METHODS

A. Study species

Australian fur seals (*Arctocephalus pusillus doriferus*) come ashore between late October and early December giving birth 1 to 2 days later (Warneke and Shaughnessy, 1985; Shaughnessy and Warneke, 1987). Females then alternate between suckling their young ashore and foraging out at sea, with maternal attendance patterns lasting approximately 1.7 days, and foraging trips increasing in duration as lactation progresses (Arnould and Hindell, 2001). Female Australian fur seals suckle pups until they are 10 to 11 months of age (Arnould and Hindell, 2001); with lactation generally varying from 9 to 12 months in the Otariidae (see exceptions Bowen, 1991).

B. Data collection and acoustic analyses

The study was conducted at a breeding colony on Kanowna Island (39°10′S, 146°18′E), Bass Strait, Australia (Fig. 1) This colony has an annual production of ca2300 pups (Kirkwood *et al.*, 2005), and the peak pupping date is 1 December (Warneke and Shaughnessy, 1985). This island has two main colonies: East and Main Colony (Fig. 1). Recordings were made over a one week period during two consecutive breeding seasons (10–16 December 2000 and 6–13 December 2001). Pups during this recording period were aged from newborn to one month of age.

In-air vocalizations of 13 adult female Australian fur seals were recorded using a Sony digital tape recorder (TCD-D8) with a directional K6/ME66 Sennheiser microphone (frequency response 50–20 000 Hz±2.5 dB). Recordings were made at a distance of 5–25 m from the vocalizing

TABLE I. Description of twelve variables measured from the pup attraction call produced by female Australian fur seals.

Pup attraction call	Description
Fundamental frequency (F_0)[a]	As all calls analyzed were harmonically rich, the distance between each harmonic band should be equal. Therefore, the fundamental frequency also equals the distance between two harmonics. To keep measurements uniform, we took all readings of this feature from the center of the call (Hz).
Duration (DUR)	Duration of the first harmonic band (ms)
Initial frequency (IN F)[a]	The start frequency of the first harmonic band (Hz)
End frequency (END F)[a]	Explains the frequency of the last point of the harmonic band (Hz)
Minimum frequency (MIN F)[a]	Minimum frequency of the first harmonic band (Hz)
Maximum frequency (MAX F)[a]	Maximum frequency of the first harmonic (Hz)
Peak frequency (PEAK F1)[b]	This was measured from the center of the call. It describes the location of the energy band or harmonic that has the most energy distributed in it (Hz). When there were multiple peaks of equal energy we only reported the frequency of the first peak.
Peak frequency (PEAK F2)[b]	This was measured from the center of the call. It describes the location of the energy band or harmonic that has the second most energy distributed in it (Hz). When there were multiple peaks of equal energy we only reported the frequency of the first peak.
Peak frequency (PEAK F3)	This was measured from the center of the call. It describes the location of the energy band or harmonic that has the third most energy distributed in it (Hz). When there were multiple peaks of equal energy we only reported the frequency of the first peak.
Peak frequency (PEAK F4)	This was measured from the center of the call. It describes the location of the energy band or harmonic that has the fourth most energy distributed in it (Hz). When there were multiple peaks of equal energy we only reported the frequency of the first peak.
Mean frequency (MEAN F)[a]	This was calculated by dividing the call into 15 intervals (i.e., 16 points). The frequency at each of these points was measured and then averaged (Hz).
Coefficient of frequency modulation (CoFM)	CoFM is as a measure of frequency modulation between consecutive intervals (Harrington, 1989).In this study, 16 data points across H1 were measured and the absolute differences in frequency between these consecutive intervals were summed and then averaged. These averages were then standardized by dividing by the mean fundamental frequency of the PAC and then multiplying by 100 (%).
Peak energy band[c]	The energy band number is given to describe the location of the energy band or harmonic that has the most energy distributed in it (Hz). This was measured from the center of the call. When there were multiple peaks of equal energy we only reported the frequency of the first peak.

[a]Log transformations (log 2) were conducted to normalize variables.
[b]Could not be normalized.
[c]Categorical data.

animal and were conducted during the early morning or afternoon of each day. Individuals were recorded at different locations and sampled during a single recording to avoid re-recording the same focal animal.

Thirteen PACs from thirteen females ($n=156$) with high signal-to-noise ratios were examined, with all calls having rich harmonic structure (Phillips and Stirling, 2001). Vocalizations were analyzed using SIGNAL 3.1 software package (Engineering Design, MA), at a sampling rate of 25 000 Hz, a frequency resolution of 1024-point fast Fourier transforms (FFT), and an analyzing bandwidth of 24.41 Hz (sampling rate/FFT). Monitor settings produced cursor error rates of ±5.36 ms in the time domain and ±25.97 Hz in the frequency domain. Call features analyzed from the PAC are described in Table I and displayed on sonograms [Figs. 2(a)–2(c)] Measurements were made on the first harmonic since the fundamental frequency was not always entirely visible on the spectrogram.

C. Description of the PAC

The description of features and values characterizing the PAC are presented in Tables I and II. As frequency features were measured from the first harmonic band they were divided by two to represent the fundamental frequency. This ensures that results were comparable with other studies as it is more common to represent call features of the fundamental (Caudron *et al.*, 1998; Collins *et al.*, 2005). The features divided by two were: IN F, END F, MIN F, MAX F, and MEAN F.

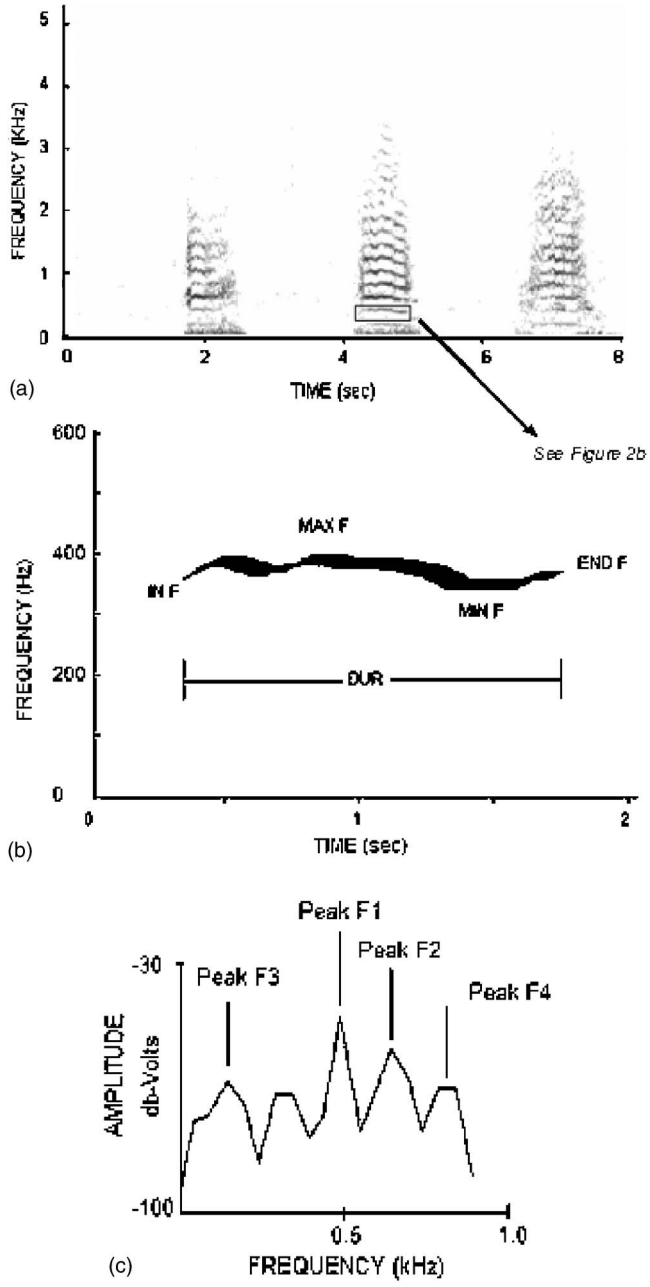


FIG. 2. Sonogram of a pup attraction call (PAC) produced a female Australian fur seal. (b) Harmonic band with associated call features measured from the pup attraction call produced by a female Australian fur seal. Call features measured from sonograms are indicated on diagram. (c) Power spectra of a pup attraction call produced by a female Australian fur seal. Peak frequency indicated on diagram.

D. Statistical analysis of the PAC

1. Potential for individual coding

Potential for individual coding (PIC) (Robisson *et al.*, 1993; Charrier *et al.*, 2002; Charrier *et al.*, 2003a) analysis was used to obtain quantitative information about each variable, allowing the comparison of their potential as individuality markers in the recognition system (i.e., if they are likely or unlikely to be used in the individual recognition process) (Charrier *et al.*, 2003a). This technique determines a ratio of the between-individual variation relative to the within-individual variation. The analysis first calculates the coefficient of variation (CV) for each call feature examined:

$$\mathrm{CV} = \left(\frac{\text{s.d.}}{\text{Mean}}\right) \times 100$$

A corrected CV (CV^*) was calculated following Sokal and Rohlf (1985)

$$\mathrm{CV}^* \left(1 + \frac{1}{4n}\right) \times (\mathrm{CV}),$$

where n=number of individuals.

Both between-individuals (CV_b) and within-individuals (CV_i) CV values were calculated. The (CV_b) was calculated for each characteristic for all individuals. While the $\mathrm{CV}_{i\ \text{grand mean}}$ was calculated for each individual for each characteristic, and a grand mean was generated. A PIC value was used generated using

$$\mathrm{PIC} = \frac{\mathrm{CV}_b}{\mathrm{CV}_{i\ \text{grand mean}}}.$$

The higher the PIC value, the greater its contribution is to the individual coding process (Charrier *et al.*, 2003a).

TABLE II. Characterization, CVs and potential for individual coding (PIC) values of call variables of the pup attraction call produced by thirteen female Australian fur seals (N=156).

Pup attraction call	Mean	s.d.	CV_b	Mean CV_i	PIC
F_0	262.1	34.6	13.5	4.6	2.9
DUR	1030.3	278.5	27.6	19.3	1.4
IN F	226.2	50.6	22.8	15.2	1.5
END F	183.9	48.1	26.7	19.4	1.4
MIN F	178.8	44.4	25.3	18.2	1.4
MAX F	290.9	44.3	15.5	6.6	2.4
PEAK F1	827.7	374.5	46.1	27.1	1.7
PEAK F2	1055.2	546.0	52.7	43.6	1.2
PEAK F3	1258.7	581.0	47.0	44.1	1.1
PEAK F4	1299.5	608.8	47.8	45.3	1.1
MEAN F	251.7	34.9	14.1	5.2	2.7
CoFM	4.4	1.6	36.7	32.4	1.1

2. Discriminant function analysis

In addition to the PIC, discriminate function analysis (DFA) and classification and regression tree (CART) analysis were also used. The DFA compares variation among individuals across several variables at the same time. The analysis is useful as it is likely that combinations of variables are used in the recognition process. The DFA also calculates the percentage of correctly classified calls and therefore determines the ability of the chosen variables to discriminate among individuals (Klecka, 1980).

Following normal transformation, DFA was conducted on the variables identified in the PAC, to investigate inter-individual variation (Table I). Peak F1 and PEAK F2 could not be normalized and were excluded from the DFA. In addition, as part of the computations involved in the DFA, the analyses determined whether any of the call parameters were

redundant. If there are any redundant variables, the analysis will not proceed until one of them is removed. In the current study there were no variables that were redundant and consequently all variables were included in the DFA (Table I).

To examine the stability of the discriminate function a cross-validation procedure was performed on the results. The data were split into two groups; one group (training data) contained half of the replicates for each individual and was used to determine the discriminate function, while the second group (test data) contained the remaining half of the data and was used to evaluate the stability of the classification. This process was repeated swapping the training and test data, ensuring that each call replicate was used in both the test data set and training data set at least once during the cross-validation procedure.

3. Classification and regression tree analysis

There are several assumptions and limitations that are associated with DFA including normality and homogeneity of variance and it is also sensitive to outliers and missing data. On the other hand, CART analysis is a nonparametric technique that does not assume any specific distribution of data (De'ath and Fabricius, 2000) and is therefore more flexible in the variables that can be incorporated in the analysis. Therefore all variables were considered in CART analysis to determine which were important in separating individual seals.

Classification and regression trees explain differences of a single response variable by repeatedly splitting the data into more homogeneous groups, using combinations of variables (De'ath and Fabricius, 2000). Each group is characterized by a value of the response variable, the number of observations, and the values of the variables that describe it (De'ath and Fabricius, 2000).

A subsample of 12 parameters from 7 individuals were used to determine which features were important in splitting individuals. In total, 10 variables were found to be important (all except PEAK F4 and COFM) and these were then used to analyze the complete data set.

4. Peak frequency distribution in the PAC

Preliminary analysis indicated the PACs produced by females were rich in harmonic structure but that the energy was not distributed evenly between the harmonic bands. In most individuals, the majority of energy appeared in only one band, with the occasional individual producing the majority of energy in two or three harmonic bands. Consequently, the peak distribution of energy in harmonic bands was examined in females to determine if this call feature could be used as a basis to separate them.

III. RESULTS

A. Description of PAC

Female Australian fur seals produce loud calls that are frequency modulated and rich in harmonic structure [Figs. 1(a)–1(c)]. Female calls were long, averaging 1.0 s in duration ($N=156$ from 13 females, s.d.=278.49). The majority of call energy is located in the first and second harmonic bands with a fundamental frequency of 262 Hz (Table II).

TABLE III. Number of calls correctly assigned by the discriminant function analysis, proportion and position of peak energy bands in the pup attraction call produced by female Australian fur seal.

Pup attraction call			
Female No.	No. of calls correctly assigned	Harmonic No.	Proportion of calls
1	7	H1	12
2	6	H1	7
3	8	H1	12
4	12	H1	9
5	9	H4	11
6	10	H4	5
7	10	H1	9
8	9	H2, H4	10, 9
9	11	H3, H4	7, 6
10	8	H1	9
11	11	H2	10
12	9	H2	11
13	9	H2	9

B. Inter-individual variation

1. Discriminant function analysis

Ten variables from the PAC were used to discriminate amongst thirteen female Australian fur seals using DFA. There were significant differences in individual PAC amongst females [Wilks′ lambda=0.01, $F(120,1054)=6.35$, $p<0.01$]. Discriminant analysis assigned 76% of the data correctly to individual females, which is greater than would be expected by chance alone ($p<0.0001$). Assigning three or more calls correctly per individual was considered significant at the $p=0.05$ level. All individuals had six or more calls correctly assigned and therefore all produced individually distinct PACs (Table III).

Roots 1–3 account for 91% of the variance of the data, suggesting that these were more important in distinguishing individuals. DUR and MIN F were strongly and positively correlated to Root 1 while F_0, END F, MAX F, and MEAN F were strongly and negatively correlated. DUR and END F were strongly and positively correlated to Root 2 while MEAN F was strongly and negatively correlated. MEAN F was strongly and positively correlated to Root 3 and F_0, IN F, MIN F, CoFM, and PEAK F3 were strongly and negatively correlated (Table IV).

The training cross-validation procedure resulted in 81% of calls being correctly assigned ($p<0.0001$), compared to the test case where 47% of the calls were correctly assigned. The probability of achieving the test percentage by chance is $p<0.01$.

2. Classification and regression tree analysis

Initially a 15-node classification tree was pruned with cross validation. As suggested by Van Opzeeland and Van Parijs (2004) and De'ath and Fabricius (2000) the 1-SE rule was adopted, this is the smallest tree for which the cross-

TABLE IV. Results of the canonical discriminant analysis comparing the pup attraction calls of female Australian fur seals.

Pup attraction call			
Acoustic variable	Root 1	Root 2	Root 3
F_0	−0.44	0.20	−0.48
DUR	0.49	0.82	0.33
IN F	0.15	−0.13	−0.80
END F	−0.44	0.49	0.03
MIN F	0.41	−0.17	−0.88
MAX F	−0.42	0.22	0.16
PEAK F3	0.07	0.32	−0.61
PEAK F4	0.01	0.40	−0.37
MEAN F	−0.49	−0.42	0.99
CoFM	−0.07	0.14	−0.68
Eigenvalue	8.27	0.92	0.59
Cumulative proportion	0.77	0.85	0.91

validated error is within one standard error of the minimum and this produced a 13-node classification tree (Fig. 3). The analysis classified 74% of the calls to individuals in the training set and 51% in the test set. This result is similar to that from the cross-validation procedure in the DFA. From all calls analyzed from this analysis there was a 26% misclassification rate in the train data set and a 49% misclassification rate in the test data set. In this CART the F_0 caused the first major split of the data.

C. Classification of variables

1. Potential for individual coding

From the methods of Charrier *et al.* (2003a), the PIC analysis results were used to rank the variables into 3 groups based on the variables' contribution to the coding of individual distinctiveness. The first group represented a high potential for individual coding (2.5–3.0), the second showed medium potential for individual coding (1.4–1.7) and the third group demonstrated a low potential for individual coding (1.1–1.2). F_0, Max F, and Mean F (PIC=2.94, 2.37 and 2.73, respectively) were classified as having a high potential for individual coding. IN F, PEAK F1, DUR, END F, and MIN F (PIC=1.50, 1.70, 1.43, 1.37, and 1.40, respectively) and PEAK F2, PEAK F3, PEAK F4, and CoFM (PIC = 1.21, 1.07, 1.06, and 1.13, respectively) exhibited a medium and low potential for individual coding, respectively (Table II).

Variables in the first and second groups are likely to be used in the individual recognition process as they were more individualized compared to those of the third group. Variables in the third group were considered unlikely to support any information about the emitter's identity (Charrier *et al.*, 2003a).

2. Discriminant function analysis

In the DFA, Roots 1 and 2 were dominated by the following variables: DUR, MIN F, F_0, END F, MAX F, and MEAN F. These variables accounted for 85% of the data's variance indicating that they were the most important in identifying individuals. The other three variables, INF, CoFM, and PEAK F3, were included in Root 3 and explained a further 6% of the variance. These variables may be important in separating individual females although to a lesser degree. Once again, PEAK F1 and PEAK F2 were not included in the DFA as they could not be normalized.

3. Classification and regression tree analysis

Important variables considered by primary splitters in this CART were F_0, PEAK ENERGY BAND, DUR, PEAK F1, PEAK F3, and MIN F.

There are differences in the variables found to be important by all three analysis techniques, however all agree that the F_0, DUR, and MIN F variables are all important in separating individual seals.

D. Peak frequency distribution in the PAC

Table III displays the most common peak energy band used by each individual and their frequency of occurrence. The results indicate that most individuals analyzed (Females 1, 2, 3, 4, 7, and 10) displayed peak frequency in Harmonic One (H1) in most replicate calls. However, some individuals displayed peak energy in Harmonic Two (H2) (Females 11, 12, 13) and Four (Females 5 and 6). There were also individuals (Females 8 and 9) that equally distributed peak energy in two harmonic bands. As a result this feature alone could only discriminate 2 out of 13 females analyzed in the

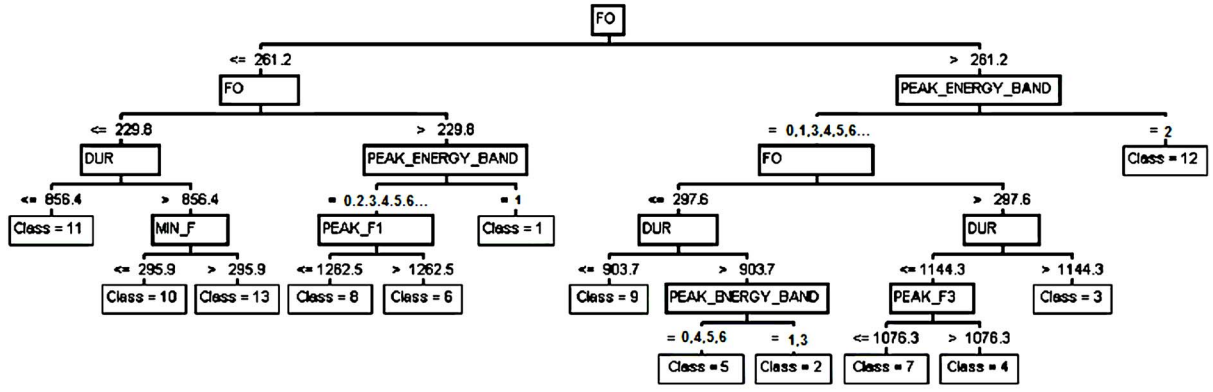


FIG. 3. A 13-node classification tree showing how vocalizations from 13 individual female Australian fur seals split based on 10 vocal parameters.

current study. With the majority of callers using harmonic one (H1), the results suggest that this feature is not a good characteristic to separate individuals.

IV. DISCUSSION

For an individual seal's call to be unique it must vary from that of other seals and be stable within the caller. This study demonstrates that the PAC produced by female Australian fur seals was correctly classified by the DFA in 76% of cases and is consistent with other previous otariid studies (74% in Antarctic fur seals, *Arctocephalus gazella*, Page *et al.*, 2002; 70% in South American fur seals, *Arctocephalus australis*, Phillips and Stirling, 2000; 82% in Northern fur seals, *Callorhinus ursinus*, Insley, 1992). Call features found to be important by this study indicate that pups may rely on a combination of frequency, temporal, and amplitude-related characteristics to differentiate between the calls of their mother and those of other females. Similar results were found with respect to South American (*A. australis*, Phillips and Stirling, 2000) and subantarctic fur seals (*A. tropicalis*, Charrier *et al.*, 2003a).

The overall structure of the PAC in Australian fur seals is generally similar to that of other otariid species. However, differences, can be noted between the species, in particular Australian fur seals have a lower fundamental frequency when compared to the other fur seals (Stirling and Warneke, 1971). This feature and other call structure variations may allow seals to discriminate among species, and result in reduced inter-breeding. Although the sample size in this study was relatively small, call stereotypy in fur seals appears to be fairly consistent with other studies with similar sample sizes (DFA analysis: South American fur seals=70%, Phillips and Stirling, 2000; Antarctic fur seals=74%, Page *et al.*, 2002; Australian fur seals=76%, current study; subantarctic fur seals=84%, Page *et al.*, 2002; New Zealand fur seals=88%, Page *et al.*, 2002). Direct comparisons between studies is difficult for a number of reasons including differences in number of replicate calls per individual, acoustical features measured, and the behavioral context of recordings, all of which can affect the degree of individual distinctiveness (Bee *et al.*, 2001). Nevertheless similarities are noted in the degree of call stereotypy of the Arctocephaline species examined to date, with DFA classification rates ranging between 70% and 88%.

Thirteen acoustic features were analyzed in order to evaluate their importance to call individuality in female Australian fur seals. Six frequency (F_0, MAX F, MEAN F, IN F, END F, and MIN F), one temporal (DUR), and one amplitude related feature (PEAK F1) displayed high to medium PIC values. Five of the preceding features (F_0, MEAN F, END F, MIN F, and DUR) and to a lesser degree accounting for 6% of the variance of the data, IN F, PEAK F3, and COFM, were also important in the DFA. In the CART analysis F_0, PEAK F1, PEAK F3, MIN F, and DUR were important. Given the results, CoFM, PEAK F2, and PEAK F4, were not considered valuable in discriminating individuals. Although a slightly different range of variables were identified as important for call individuality by all three methods (PIC, DFA, and CART), they agree that the F_0, DUR, and MIN F variables were important in separating individual seals. Similar results were reported by Charrier *et al.* (2003a), where in female subantarctic fur seals, the fundamental frequency and the duration of the PAC were identified, by PIC analysis, as important.

In the current study, the initial, end, and minimum frequencies of the first harmonic were found to be important in separating female callers. Previous studies of subantarctic fur seals and king penguins indicate that the start of calls may contain more information encoding an individual's identity than the rest of the call (Charrier *et al.*, 2003b; Jouventin *et al.*, 1999). Although the current study indicates that the features at the start of PACs were important in individual discrimination, it also suggests that the end and minimum frequencies are also important. Playback studies manipulating the PAC would be advantageous to determine whether a pup's ability to recognize its mother is based on these call features.

Previous research on call individuality has reported duration and peak frequency to be important to an animal's identity. In South American (Phillips and Stirling, 2000) and subantarctic fur seals (Charrier *et al.*, 2003a) duration is a key feature in distinguishing female callers, while studies on northern fur seals (*Callorhinus ursinus*) proposed that the duration of a call explained more of an emotive or arousal state of the individual caller rather than providing cues on an individual's identity (Insley, 1992). However, in the current study duration, PEAK F1 and PEAK F3 were found to be important to individuality, while PEAK F2 and PEAK F4 were not considered to be important. This result indicates that PEAK F2 and PEAK F4 are not good individuality markers and may express the emotive state of a caller. Similarly, squirrel monkeys (*Saimiri sciureus*) have been shown to alter the peak frequency of their vocalizations with different behavioral states (Fichtel *et al.*, 2001).

Frequency modulation has been shown to be an important characteristic of individual recognition in king penguins (Jouventin *et al.*, 1999), subantarctic (Charrier *et al.*, 2003a) and South American (Phillips and Stirling, 2000) fur seals. In contrast, the frequency modulation of calls in the present study was not regarded as important in separating female callers when compared to the other variables examined. Further, it is unknown whether frequency modulation is important in individual recognition in other otariid species, as this feature has only been examined in a few studies to date. There may also be cases where the variables measured may not be totally representative of frequency modulation in calls, as was suggested by Charrier *et al.* (2003a).

Functionally vocal recognition has important consequences to pups who are trying to locate their mothers for nourishment. Based on the parent-offspring conflict theory (Trivers, 1974) we expect the burden of reunion to be placed more on pups. Unsuccessful pair reunions may result in a mother's reproductive loss, however, for the pup, pair reunions ultimately means survival or death. For that reason there are distinct selective pressures for reunion between a mother and her young (Insley, 2001). Recent studies have suggested an asymmetry of recognition, however, mutual

recognition has been shown in otariids (Trillmich, 1981; Insley, 2001; Charrier *et al.*, 2002, 2003b). In this study we found that the PAC produced by mothers could be assigned in 76% of cases using DFA, which would suggest that pups have the ability to actively find their mothers. The model presented here in this study may have some shortcomings, as suggested by the low cross validation results in both DFA and CART. This may indicate that pups discriminate female callers by different or additional call variables not included in this study. To further explore this area of vocal recognition, playback studies where a pup's ability to recognize its mother's voice should be tested. Additionally this process would determine those features involved in the vocal recognition process.

ACKNOWLEDGMENTS

The authors would like thank Isabelle Charrier, Kym Collins, Sophie Hall-Aspland, and Sujata Jadhav for the helpful comments on this paper. The authors would also like to thank Parks Victoria, in particular the rangers from the Foster and Tidal River offices, for logistical support in transporting the researchers into the field. We would also like to thank the Veterinary Faculty of the University of Sydney, the Australian Marine Mammal Research Center, Zoological Parks Board of NSW, and the University of Melbourne for the support in establishing and developing this project. Many thanks also to Syntec International, Project AWARE Foundation, The Australian Geographic Society for their generous support with this project.

Arnould, J. P. Y., and Hindell, M. A. (**2001**). "Dive behavior, foraging locations, and maternal attendance patterns of Australian fur seals (*Arctocephalus pusillus doriferus*)," Can. J. Zool. **79**, 35–48.

Balcolme, J. P., and McCracken, G. F. (**1992**). "Vocal recognition in Mexican free-tailed bats: Do pups recognize mothers?," Anim. Behav. **43**, 79–87.

Bee, M. A., Kozich, C. E., Blackwell, K. J., and Gerhardt, H. C. (**2001**). "Individual variation in advertisement calls of territorial male green tree frogs, *Rana clamitans*: Implications for individual discrimination," Ethology **107**, 65–84.

Bowen, W. D. (**1991**). "Behavioral ecology of pinniped neonates," in *The Behavior of Pinnipeds*, edited by D. Renouf (Chapman & Hall, London). pp. 66–127.

Caudron, A. K., Kondakov, A. A., and Siryanov, S. V. (**1998**). "Acoustic structure and individual variation of grey seal (*Halichoerus grypus*) pup calls," J. Mar. Biol. Assoc. U.K. **78**, 651–658.

Charrier, I., Mathevon, N., and Jouventin, P. (**2003a**). "Individuality in the voice of fur seal females: An analysis study of the pup attraction call in *Arctocephalus tropicalis*," Marine Mammal Sci. **19**(1), 161–172.

Charrier, I., Mathevon, N., and Jouventin, P. (**2003b**). "Vocal signature of mothers by fur seal pups," Anim. Behav. **65**, 543–550.

Charrier, I., Mathevon, N., and Jouventin, P. (**2002**). "How does a fur seal mother recognize the voice of her pup? An experimental study of *Arctocephalus tropicalis*," J. Exp. Biol. **205**, 603–612.

Charrier, I., Mathevon, N., and Jouventin, P. (**2001**). "Mother's voice recognition by seal pups. Newborns need to learn their mother's call before she can take off on a fishing trip," Nature (London) **412**, 873.

Collins, K. T., Rogers, T. L., Terhune, J. M., McGreevy, P. D., Wheatley, K. E., and Harcourt, R. G. (**2005**). "Individual variation of in-air female 'pup contact' calls in Weddell seals, *Leptonychotes weddelli*," Behaviour **142**, 167–189.

De'ath, G., and Fabricius, K. (**2000**). "Classification and regression tree: A powerful yet simple technique for ecological data analysis," Ecology **81**, 3178–3192.

Falls, J. B. (**1982**). "Individual recognition by sounds in birds," in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), Vol. **1**, pp. 237–278.

Fichtel, C., Hammerschmidt, K., and Jürgens, U. (**2001**). "On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey," Behaviour **138**, 97–116.

Gubernick, D. J. (**1981**). "Parent and infant attachment in mammals," in *Parental Care in Mammals*, edited by D. J. Gubernick and P. H. Klopfer (Plenum, New York), pp. 243–305.

Harrington, F. H. (**1989**). "Chorus howling by wolves: Acoustic structure, pack size and the Beau Geste effect," Bioacoustics 2, 117–136.

Insley, S. J. (**2001**). "Mother-offspring vocal recognition in northern fur seals is mutual but asymmetrical," Anim. Behav. **61**, 129–137.

Insley, S. J. (**1992**). "Mother-offspring separation and acoustic stereotypy: A comparison of call morphology in two species of pinnipeds," Behaviour **120**(1–2), 103–122.

Insley, S. J., Paredes, R., and Jones, I. L. (**2003b**). "Sex differences in razorbill *Alca torda* parent-offspring vocal recognition," J. Exp. Biol. **206**, 25–31.

Insley, S. J., Phillips, A. V., and Charrier, I. (**2003a**). "A review of social recognition in pinnipeds," Aquat. Mamm. **29**(2), 181–201.

Job *et al.* (1995).

Jouventin, P., Aubin, T., and Lengagne, T. (**1999**). "Finding a parent in a king penguin colony: The acoustic system of individual recognition," Anim. Behav. **57**, 1175–1183.

Kirkwood, R., Gales, R., Terauds, A., Arnould, J. P. Y., Pemberton, D., Shaughnessy, P. D., Mitchell, A. T., and Gibbens, J. (**2005**). "Pup production and population trends of the Australian fur seal (*Arctocephalus pusillus doriferus*)," **21**(2), 260–282.

Klecka, W. R. (**1980**). *Discriminant Analysis* (Sage, Beverly Hills, CA).

Page, B., Goldsworthy, S. D., and Hindell, M. A. (**2002**). "Individual vocal traits of mother and pup fur seals," Bioacoustics 13, 121–143.

Phillips, A. V., and Stirling, I. (**2001**). "Vocal repertoire of South American fur seals, *Arctocephalus Australis*: structure, function, and context," Can. J. Zool. **79**, 420–437.

Phillips, A. V., and Stirling, I. (**2000**). "Vocal individuality in mother and pup South American fur seals, Arctocephalus australis," Marine Mammal Sci. **16**(3), 592–616.

Riedman, M. (**1990**). *The Pinnipeds: Seals, Sea Lions, and Walruses* (University of California Press, Berkley, CA), pp. 64, 192, 200.

Robisson, P., Aubin, T., and Bremond, J. C. (**1993**). "Individuality in the voice of the Emperor Penguin *Aptenodytes forsteri*: Adaptation to a noisy environment," Ethology **94**, 279–290.

Scherrer, J. A., and Wilkinson, G. S. (**1993**). "Evening bat isolation calls provide evidence for heritable signatures," Anim. Behav. **46**, 847–860.

Schusterman, R. J., Hanggi, E. B., and Gisiner, R. (**1992**). "Acoustic signalling in mother-pup reunions, inter-species bonding, and affiliation by kinship in Californian sea lions (Zalophus californianus)," in *Marine Mammal Sensory Systems* edited by J. A. Thomas, R. A. Kastelein (Plenum, New York), pp. 533–551.

Shaughnessy, P. D., and Warneke, R. M. (**1987**). "Australian fur seal, *Arctocephalus pusillus doriferus*," in *Status, Biology, and Ecology of Fur Seals; Proceedings of an International Symposium and Workshop Cambridge, England, 23–27 April 1984*, edited by J. P. Croxal and R. L.Gentry, NOAA Tech. Rep. NMFS 51, Cambridge, pp. 73–77.

Sokal, R. R., and Rohlf, F. J. (**1985**). *Biometry: The Principles and Practice of Statistics in Biological Research* (Freeman, New York), pp. 57–58.

Stirling, I., and Warneke, R. R. (**1971**). "Implications of a comparison of the airborne vocalizations and some aspects of the behavior of the two Australian fur seals. *Artcocephalus* spp, on the evolution and present taxonomy of the genus," Aust. J. Zool. **19**, 227–241.

Trillmich, F. (**1981**). "Mutual mother-pup recognition in Galapagos fur seals and sea lions: Cues used and functional significance," Behaviour **78**, 21–42.

Trivers, R. L., (**1974**). "Parent-offspring conflict," Am. Zool. **14**, 249–264.

Van Opzeeland, I. C., and Van Parijs, S. M. (**2004**). "Individuality in harp seal, Phoca groenlandica, pup vocalizations," Anim. Behav. **68**, 1115–1123.

Warneke, R. M., and Shaughnessy, P. D. (1985). "Arctocephalus pusillus, the South African and Australian Fur Seal: taxonomy, biogeography and life history," in Studies of Sea Mammals in South Latitudes, edited by J. K. Ling and M. M. Bryden (South Australian Museum, Adelaide), pp. 53–57.

Source levels and harmonic content of whistles in white-beaked dolphins (*Lagenorhynchus albirostris*)

M. H. Rasmussen
Institute of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

M. Lammers
Marine Mammal Research Program, Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734

K. Beedholm and L. A. Miller
Institute of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

(Received 18 January 2006; revised 29 March 2006; accepted 10 April 2006)

Recordings of white-beaked dolphin whistles were made in Faxaflói Bay (Iceland) using a three-hydrophone towed linear array. Signals from the hydrophones were routed through an amplifier to a lunch box computer on board the boat and digitized using a sample rate of 125 kHz per channel. Using this method more than 5000 whistles were recorded. All recordings were made in sea states 0–1 (Beaufort scale). Dolphins were located in a 2D horizontal plane by using the difference of arrival time to the three hydrophones, and source levels were estimated from these positions using two different methods (I and II). Forty-three whistles gave a reliable location for the vocalizing dolphin when using method II and of these 12 when using method I. Source level estimates on the center hydrophone were higher using method I [average source level 148 (rms)±12 dB, $n=36$] than for method II [average source level 139 (rms)±12 dB, $n=36$]. Using these rms values the maximum possible communication range for whistling dolphins given the local ambient noise conditions was then estimated. The maximum range was 10.5 km for a dolphin whistle with the highest source level (167 dB) and about 140 m for a whistle with the lowest source level (118 dB). Only two of the 43 whistles contained an unequal number of harmonics recorded at the three hydrophones judging from the spectrograms. Such signals could be used to calculate the directionality of whistles, but more recordings are necessary to describe the directionality of white-beaked dolphin whistles. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2202865]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 510–517

I. INTRODUCTION

Dolphins are known to produce whistles for communication (e.g., Dreher and Evans, 1964; Lilly, 1963). In order to investigate the possible range over which dolphins can communicate using whistles, an on-axis source level must be calculated. This has been done for bottlenose dolphin (*Tursiops truncatus*) whistles (Janik, 2000) and Hawaiian spinner dolphin (*Stenella longirostris*) whistles (Lammers and Au, 2003). To determine source levels of whistles we need to know the position, distance, and orientation to an animal. The former can be estimated from the time of arrival differences (TOADs) of the signal at each receiver in an array using cross correlation and hyperbolic calculations (e.g., Spiesberger and Fristrup, 1990). The sound level at 1 m from the source is called source level (SL) and a minimum of three receivers are needed to calculate the distance to the source in 2D (two dimensions) (e.g., Wahlberg *et al.*, 2001). Calculated source levels of dolphin whistles vary between 109 and 180 dB *re*. 1 μPa (Table 7.2 in Richardson *et al.*, 1995). Source levels of whistles from the Hawaiian spinner dolphin varied from 149.7 (rms)±3.2 dB *re*. 1 μPa to 156.0 (rms)±4 dB *re*. 1 μPa (Lammers and Au, 2003). The mean source level was 158±0.6 dB *re*. 1 μPa for bottlenose dolphin *(Tursiops truncatus)* whistles and the maximum source level was 169 dB (rms) *re*. 1 μPa (Janik, 2000).

Previously, it was reported that dolphin whistles were limited to an upper frequency of 20 kHz (e.g., Popper, 1980) but recently, using broad band equipment, it has been shown that spectrograms of dolphin whistles can extend beyond 20 kHz (e.g., Rasmussen and Miller, 2002; 2004; Lammers *et al.* 2003). The lowest frequency is called the fundamental frequency, and each higher-integer multiple of the fundamental frequency is a harmonic (Yost, 2000). Rasmussen and Miller (2002, 2004) reported white-beaked dolphin whistles with fundamental frequencies up to 35 kHz. Lammers *et al.* (2003) described whistles of Hawaiian spinner dolphins and Atlantic spotted dolphins (*Stenella frontalis*) and reported a mean frequency of the fourth harmonic at 69.0 kHz for spinner dolphins and 54.5 kHz for spotted dolphins. Some of the whistles had the maximum frequency of the fundamental above 20 kHz (Lammers *et al.*, 2003).

The whistles of spinner dolphin's are directional. Lammers and Au (2003) suggest that the high frequency part of the dolphin whistles and the different number of harmonic components can be used by the dolphins as cues for directionality. Miller (2002) also found killer whales' communication calls to be directional. The directionality has been

 0001-4966/2006/120(1)/510/8/$22.50

suggested to be a cue used by the animals to position themselves in the group and useful for keeping group cohesion (Lammers and Au, 2003; Miller, 2002).

White-beaked dolphins are found within a few nautical miles from shore from June to August in Faxaflói Bay, Iceland. These dolphins probably live in a fission-fusion community like that described for bottlenose dolphins (e.g., Smolker *et al.*, 1993; Tyack 1997) and for Hawaiian spinner dolphins (Norris *et al.*, 1994). Brownlee and Norris (1994) suggested the whistles of the Hawaiian spinner dolphins are used as an indicator of activity state, for example the occurrence of many whistles indicates a high activity state. This is most likely the same for white-beaked dolphins. Usually white-beaked dolphins travel in small groups of three to six dolphins (Rasmussen, 1999) that are often widely separated in the bay. However, they can quickly gather into larger groups. How do the dolphins coordinate these gatherings? A good explanation would be that they use their whistles for this purpose.

Directional properties of whistles (and of the receiver) along with *a priori* knowledge of typical source levels are important cues for animals to determine the positions of neighbors. Directional properties of whistles are more difficult to determine than source levels. The aims of this study were to calculate source levels, communication range, and directionality of white-beaked dolphin whistles in coastal Icelandic waters.

II. METHODS

The recordings were made in August 2002 in Faxaflói Bay (Iceland) about 6 miles NNW of Keflavik (64°00.49′N,22°33.37′W). The water depths were between 30 and 50 m in the area of recordings. The ambient noise level was measured from a small boat with the engine turned off and in sea-states 0–1 (Beaufort Scale) using a Reson 4032 hydrophone connected via an amplifier to a lunchbox computer sampling at 800 kHz; except for the hydrophone this setup was the same as that described in Rasmussen *et al.* (2004). Recordings of whistles were made using the same towed linear hydrophone array as that described by Lammers and Au (2003) and consisted of three custom-built hydrophones (A, B, C) spaced 11.5 m apart and separated from the boat by 25 m (see Fig. 1 and Lammers and Au, 2003). The array was towed behind the boat at speeds of 6–7 knots and a custom made "tow-fish" was used to sink the array to a depth of approximately 2 m while a "tattletale" assured the array was stable under water. Hydrophones A and C had a sensitivity of −200 dB *re*. 1 V root mean square (rms) per μPa, and hydrophone B had a sensitivity of −197 dB *re*. 1 V rms per μPa.

The hydrophones had flat frequency responses (±3 dB) up to 150 kHz. The calibration of the hydrophone array was described in Lammers and Au (2003). The hydrophones were connected to a custom built amplifier using a 300-Hz high-pass filter and to a lunchbox computer on board the boat. The sample rate was 125 kHz per channel giving a Nyquist frequency of 62.5 kHz. We did not use antialiasing filters. Had aliasing occurred the result would have appeared as mirror-imaged frequency sweeps of the fundamental in the spectrogram. We saw none of this. For aliased frequency components in echolocation clicks to influence the amplitude of the cross-correlation function (see below), the clicks must be emitted simultaneously with the whistle. We did not test for this.

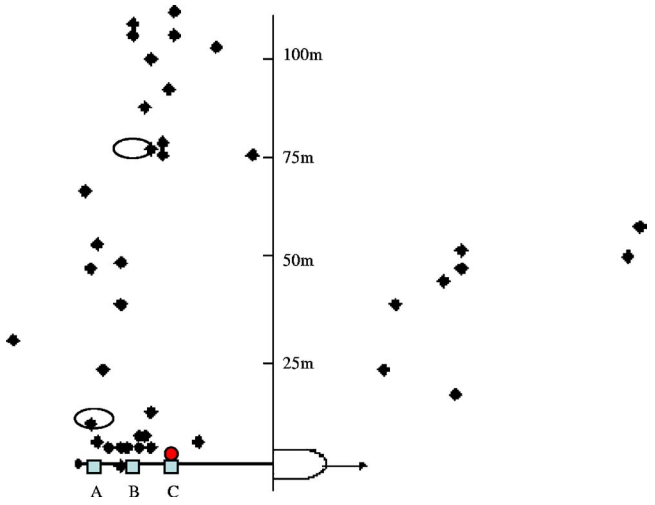


FIG. 1. (Color online) Placement of the 43 whistles (diamonds) relative to the boat and the hydrophone array (squares: A, B, C) when using method II. Note that it is only possible to calculate the position of the dolphins in two dimensions when using three hydrophones. The abscissa has the same dimensions as the ordinate and all positions on the starboard side of the boat are mirrored on the port side. The small circle close to hydrophone C indicates an orange buoy 25 m behind the boat. The circles around two diamonds (two dolphins) corresponds to two whistles with unequal number of harmonics recorded on the three hydrophones [see Fig. 2(a)].

An orange buoy towed 25 m behind the boat marked the location of hydrophone C and was used as a reference point by observers for estimating the distance and bearing of surfacing dolphins relative to the hydrophones. We tried to get the dolphins parallel to the boat during the recordings because this afforded the greatest opportunity for determining both source levels and directionality. Group size estimates were also noted during recordings.

Sound files were opened in Cool Edit Pro (version 2, Syntrillium Software) and band-pass filtered to dampen unwanted tones at 17.33, 37.59, 48.7 kHz on all three channels. Recordings were then scanned to identify times without overlapping whistle contours from different individuals and also for harmonic content. Nonoverlapping whistle contours were preferably used as these represented the best opportunity for localizing an individual and determining the source levels.

We used two different methods for localizing the underwater positions of vocalizing dolphins. Method II was introduced because a low signal-to-noise ratio, particularly on channel A, made cross correlation difficult using method I. Two of the recording channels had a low-power electrical interference signal that was highly correlated. This had the effect that signals were almost always—presumably falsely—localized on a line normal to and equidistant between the two hydrophones. In a correlation analysis of long signals, correlated noise is problematic without making individual judgments from call to call. To avoid this we chose to

TABLE I. Average source level estimates recorded on the three hydrophones (A, B, C), minimum and maximum source level, and minimum and maximum distance of recordings. The estimates are shown both for method I and method II.

Method	Hydrophones	Average SL (rms) ±SD	Min. SL (rms)–Max. SL (rms)	Min. distance–Max distance
	A	148 dB *re*. 1 μPa±12	124–160 dB *re*. 1 μPa	6–182 m
Method I	B	144 dB *re*. 1 μPa±12	124–159 dB *re*. 1 μPa	6–180 m
(N=12)	C	152 dB *re*. 1 μPa±10	136–166 dB *re*. 1 μPa	13–180 m
	A	139 dB *re*. 1 μPa±11	120–153 dB *re*. 1 μPa	7–176 m
Method II	B	139 dB *re*. 1 μPa±10	121–152 dB *re*. 1 μPa	7–175 m
(N=12)	C	140 dB *re*. 1 μPa±10	125–156 dB *re*. 1 μPa	13–174 m
	A	144 dB *re*. 1 μPa±8	126–163 dB *re*. 1 μPa	5–176 m
Method II	B	142 dB *re*. 1 μPa±8	118–163 dB *re*. 1 μPa	7–175 m
(N=43)	C	146 dB *re*. 1 μPa±8	129–167 dB *re*. 1 μPa	13–174 m

analyze only those signals with enough power to be correctly localized. This criterion excludes signals that would have been correctly localized on the equidistant line between the two affected hydrophones.

Method I: This method is described in Lammers and Au (2003) and uses cross-correlation functions to calculate the differences in arrival times of a signal at each of the three hydrophones in the array. Localization was implemented in Matlab (version 6.0, The MathWorks, Inc.) using a custom written script. To determine the sound level, the program steps through the selected whistle in 100-ms steps (12 500 samples). Within each step the voltage values are squared, summed, and divided by the step size (12 500 samples) to get the mean. The square root is then taken of the mean to get the rms value for that step (volts rms). The process is repeated through the selected whistle. Next, the sound pressure level (SPL) is calculated using the following equation: SPL (*re*. 1 μPa)= hydrophone output (dB at 1 V rms)−gain (dB)+20 log (signal (volts rms)/(1 V rms)). Finally, the SL is calculated by taking the transmission and absorption losses (alpha) into account, which are a function of the distance using $\mathrm{SL} = \mathrm{SPL} + 20\log(\mathrm{distance}) + (\mathrm{alpha}(\alpha) \times \mathrm{distance}/1000)$. The absorption coefficient (α) can be expressed as $\alpha = 0.036\, f^{1.5}$ (dB/km) (after Richardson *et al.*, 1995).

Method II: Cross-correlation functions were used to calculate the difference in arrival time of a signal at each of the three hydrophones in the array (as described in method I). The difference of arrival time between channel A and B and the difference between B and C were shown as cross-correlation functions in Matlab. We added another cross-correlation function, the time difference between A and C as a control (Δt_{AC}). This time difference had to equal the sum of the time difference between A and B (Δt_{AB}), and B and C (Δt_{BC}), i.e.,

$$\Delta t_{AC} = \sum \Delta t_{AB} + \Delta t_{BC}.$$

If this was not the case, the whistle was not used since the calculated position would not be reliable. Often we had to choose the second highest peak in the cross-correlation function to get a reliable position of the dolphin and not the highest peak because noise gave the highest peak. (The noise contribution often resulted in a high peak at 0 in the cross-correlation function.)

Method II is similar to the method described by Spiesberger (1998) using the amplitudes of the peaks in the cross-correlation functions to estimate source levels. Spiesberger (1998) solved the equations in general for n transducers, however we used the method with three hydrophones (n =3). The principle is described for three transducers in the Appendix. This is a method to minimize the contribution of the noise when calculating source levels.

Finally the locations calculated using methods I and II were compared with the locations of dolphin groups noted during visual observations in the field. If the locations were comparable, whistles were used, if not they were discarded.

III. RESULTS

Whistles from white-beaked dolphins were recorded during the evenings of 19 and 21 August 2002. More than 5000 whistles were recorded in about 6 h (from 17:12 to 20:41 on 19 August and from 18:13 to 20:39 on 21 August). In both cases whistles were not detected during the first half hour of recording, then occasional whistles were heard, and finally continuous whistling was recorded. In each case when recordings were linked to visual observations, group size varied over time (min. 3 and max. 15 animals), with a tendency for group size to increase during the recording and decrease at the end of recordings.

A. Source levels of whistles

The selection criteria outlined in Sec. II allowed for 43 whistles to be reliably found using method II and of these 12 whistles were found using method I (Fig. 1). These calculated positions matched visual observations. This gave a total of 129 source levels estimations from the recordings when using method II and 36 source level estimates when using method I. Source levels ranged from 118 to 167 dB (rms) *re*. 1 μPa (Table I). Source levels were higher when using method I than when using method II and the difference was significant [N=12, one way ANOVA, F=12.3, $p<0.05$, (Zar, 1996) see Table I]. This can be explained by the fact

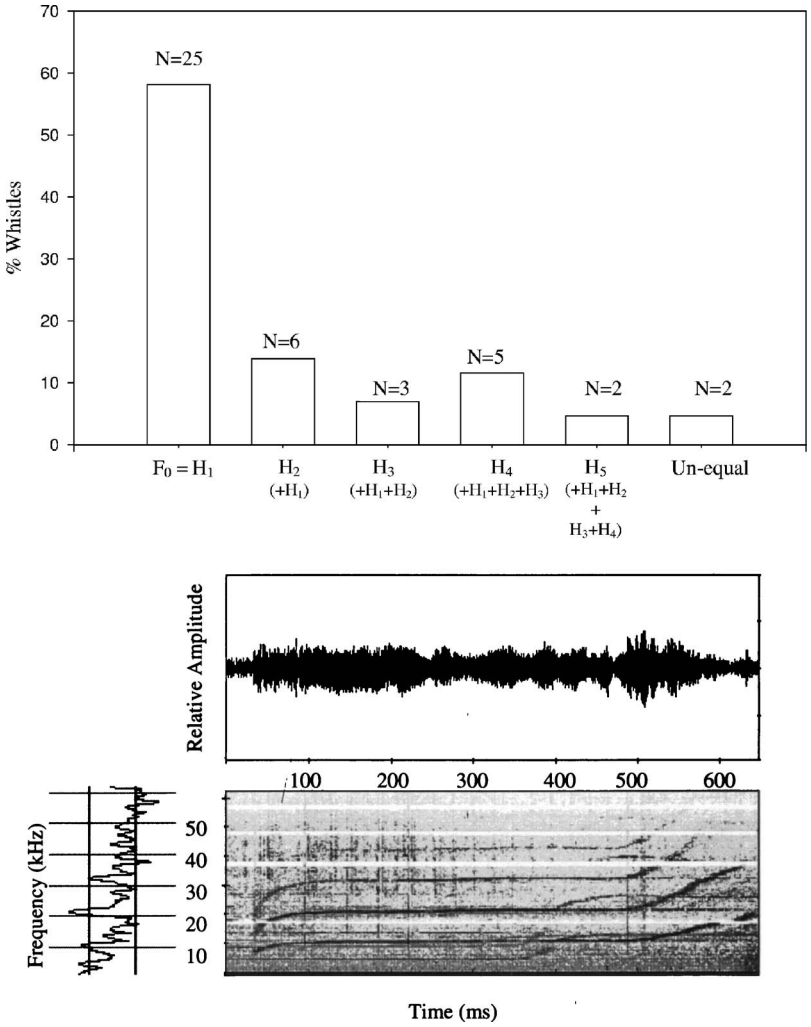


FIG. 2. (a) The percentages of whistles containing the fundamental and higher harmonic components judged from visual inspection in spectrograms recorded on the three hydrophones ($F_0=H_1$ whistles containing only the fundamental or the first harmonic, H_1+H_2=whistles containing the first and second harmonic, $H_1+H_2+H_3$ =whistles containing the first, second, and third harmonic, $H_1+H_2+H_3+H_4$=whistles containing the first, second, third, and fourth harmonic, $H_1+H_2+H_3+H_4+H_5$=whistles containing the first, second, third, fourth, and fifth harmonic). Note that only two whistles (5%) contained unequal harmonic components recorded on the three hydrophones. One of the two whistles had three harmonics on hydrophones A and B, but only two harmonics on hydrophone C. The other had five harmonics on hydrophone B and four harmonics on hydrophones A and C (see Figs. 1 and 4). (b) Example of a whistle containing four harmonics ($H_1+H_2+H_3+H_4$ =whistle containing the first, second, third, and fourth harmonics). The whistle is shown as the time signal (above), the spectrogram (below), and the power spectrum (to the left). The spectrum is an average power spectrum using the whole time signal. Note the frequency content extends to frequencies above 50 kHz.

that source levels estimated using method I include ambient noise, whereas when using method II we were using the amplitude of the peak in the cross correlation and the noise is reduced in the source level estimates (see the Appendix). A significant difference was found both between the methods [mixed model, $F=101.62$, $p<0.05$, (Hand, 1997)] and the hydrophones [mixed model, $F=10.2$, $p<0.05$ (Hand, 1997)] and also an interaction between the methods and hydrophones [mixed model, $F=4.5$, $p<0.05$ (Hand, 1997)]. This indicates that the differences in source levels recorded on the three hydrophones are not the same when using the two different methods. The noise levels are reduced when using method II resulting in less variation of source levels at the three hydrophones when compared to method I.

B. Harmonic content

The percentages and numbers of the 43 whistles with and without harmonics, judged from visual inspection of spectrograms in Cool Edit, are shown in Fig. 2(a). Fifty-eight percent of the whistles contained only the first harmonic or the fundamental frequency and the rest (42%) had higher harmonic components. Only 5% (or two whistles) had unequal numbers of harmonics recorded on the three hydrophones. Figure 2(b) shows an example of a whistle with four harmonics containing frequencies to at least 50 kHz.

The number of harmonics plotted against average source level (±SD) and the maximum source levels on the three channels (A, B, C) along with distances are shown in Figs. 3(a)–3(c) of whistles when using method II. No relationship was found between source level and the number of harmonics, suggesting that energy in the fundamental dominates. The number of harmonics recorded decreased with increasing distance (linear regression, $r^2=0.84$, one way ANOVA, $p<0.005$; for hydrophone A; $r^2=0.94$, one way ANOVA, $p<0.005$, for hydrophone B; $r^2=0.95$, one way ANOVA, $p<0.05$ for hydrophone C). Whistles with five harmonics were only recorded when the dolphins were close to the array. The maximum distance to an animal was 31 m when five harmonics were recorded and 176 m when the fundamental (first harmonic) was recorded on all three hydrophones (Fig. 3).

The fundamental frequency of all the whistles with harmonic components varied between 7 and 13 kHz (average =10.7 kHz, SD=1.5 kHz). Most whistles were upsweeps and had a fundamental frequency of about 10 kHz (61%), giving the second harmonics at 20 kHz, the third at 30 kHz, the fourth at 40 kHz, and the fifth at 50 kHz. We therefore used a simulated third octave filter (made in Cool Edit) using a center frequency close to these frequencies and bandwidth as described in the Brüel and Kjær handbook (1985) in hopes

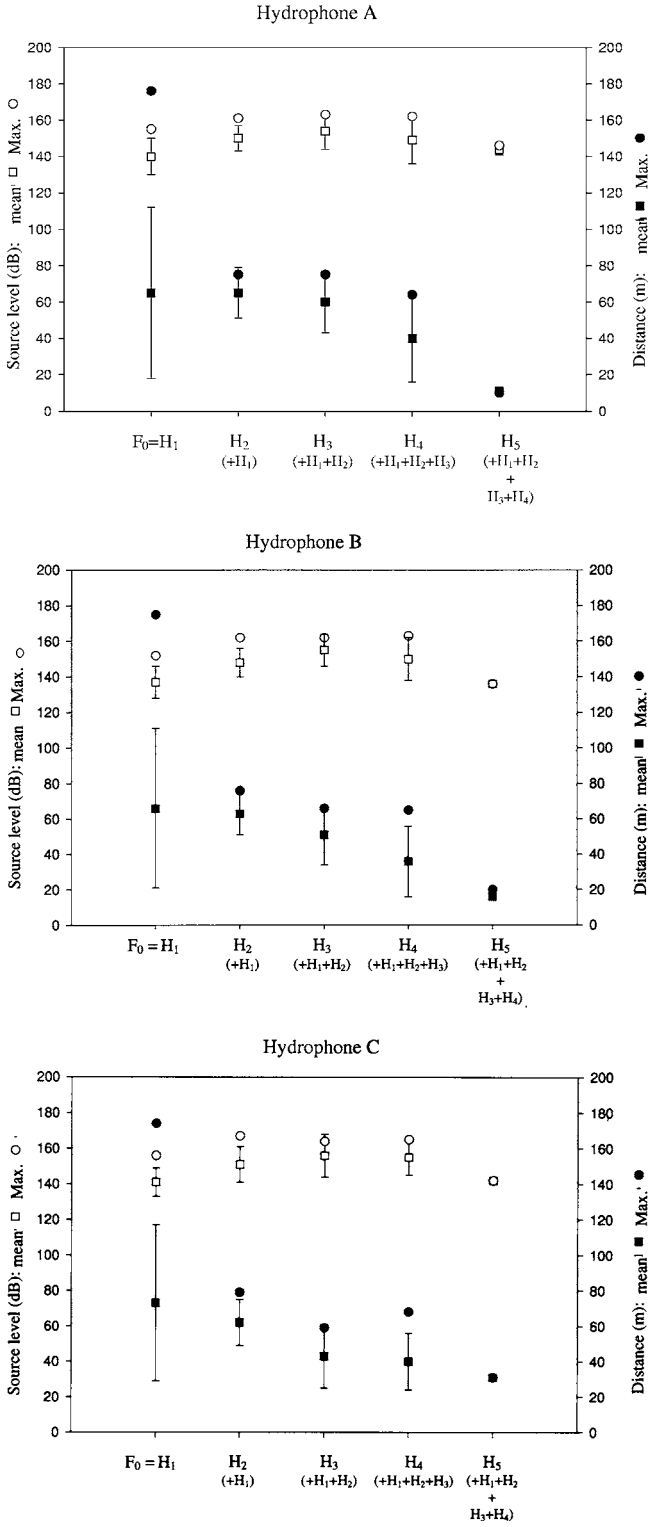


FIG. 3. Distribution of the whistles containing a different number of harmonics plotted with the average source level (±SD) (white squares) and the maximum source level (white circles) using method II. In addition the average distance (±SD) (black squares) and the maximum distance (black circles) are included in (a) hydrophone A, (b) hydrophone B, and (c) hydrophone C. See Fig. 2 for the number of signals in each category. Note the lack of a trend between source level and distance, and that the number of harmonics is inversely proportional with distance.

of detecting directionality effects, especially at higher frequencies. No trend was found. The signal-to-noise ratio (S/N) varied within a whistle when looking at the different third octave bands (Table II), but a large variation was also found between the different channels. However, in general the signal-to-noise ratio decreased with the increasing number of harmonics, with good S/N for whistles containing only the fundamental frequency and poor S/N for the fourth harmonic in whistles with multiple harmonics.

IV. DISCUSSION

Signal intensity and directionality are important acoustic communication parameters. Despite recording about 5000 whistles, only a few (43) satisfied our criteria for determining source levels. The source levels of white-beaked dolphin whistles [118–167 dB *re*. 1 μPa (rms)] are similar to those found for other dolphin species (e.g., Janik, 2000; Lammers and Au, 2003), independent of the methods we used for the calculations. The most difficult aspect in estimating the source levels of dolphin whistles is obtaining reliable positions of individuals. Freitag and Tyack (1993) discuss this problem with respect to multiple reflections both from the surface and from the bottom as well as the problem of poor signal-to-noise ratio. Dolphins must deal with the same problems, but it may not be as important for them to know the exact position of another whistling dolphin or of a vocally active distant group, reducing the need for high signal-to-noise ratios. However, whistles may be used to identify individuals, so-called "signature whistles" (Caldwell *et al.*, 1990). "Signature whistles" have been suggested to be cohesion calls as to keep contact between dispersed group members (Janik and Slater, 1998) and they may be important for maintaining contact between a mother and a calf (Smolker *et al.*, 1993). In this case it must be important for the mother and/or the calf to locate the other precisely from the whistles, especially in murky waters and at night. The dolphins use multi-harmonic whistles and since directionality of the emitted signal and of the auditory receiver increases with increasing frequency, higher harmonics could be used by a dolphin for positioning a caller (Au, 1993; Lammers and Au, 2003). Determining directionality from harmonics of the 43 signals was not possible with our recording methods. But it should be trivial for a dolphin to determine the direction and distance to a whistling conspecific. They are continuously moving and whistling when acoustically active and they can presumably identify the signals of other group members and perhaps those of individuals in other groups. With this and their directional hearing, as well as directionality of the signal, an animal should have sufficient information for localization. Our monitoring methods do not offer these advantages.

The extensive recordings of whistles from white-beaked dolphins using broad band recording equipment presented here reveal whistles rich in harmonics that were not reported earlier (Rasmussen and Miller, 2002, 2004). As mentioned above, harmonics could be used to determine direction to a source, at least at closer rangers. Figure 4 shows the position of a dolphin estimated from two whistles. These two whistles were recorded with an interval of about 10s and could be produced by the same animal. One whistle had five harmonics on all three hydrophones and the other had five harmonics on hydrophone B, the middle hydrophone, and four har-

TABLE II. Third octave analyses of three whistles including the fundamental, second, third, and fourth harmonics. The distance from each hydrophone to the dolphin is included in the table. The whistles were filtered using a third octave filter constructed in Cool Edit centered at 10, 20, 31.5, and 40 kHz. The rms value in dB of the filtered portion of the signal at each hydrophone relative to the unfiltered whistle and the noise in each third octave band relative to the full band width noise were measured in Cool Edit. In addition the signal-to-noise ratio (S/N) is noted in each third octave band. Note poor S/N at higher frequencies; this cannot be explained by loss due to absorption at higher frequencies, but rather by less energy in higher frequencies, more noise, and maybe some directionality. Note the poorer S/N for the higher harmonics, a factor that makes it difficult to determine whistle directionality.

Whistle no. (and duration)	Hydrophones	Distance (m)	10 kHz Signal (rms) (dB)	10 kHz Noise (rms) (dB)	10 kHz S/N (dB)	20 kHz Signal (rms) (dB)	20 kHz Noise (rms) (dB)	20 kHz S/N (dB)	31.5 kHz Signal (rms) (db)	31.5 kHz Noise (rms) (dB)	31.5 kHz S/N (dB)	40 kHz Signal (rms) (dB)	40 kHz Noise (rms) (dB)	40 kHz S/N (dB)
1 (600 ms)	A	22	−35	−50	15	−45	−50	5	−44	−51	7	−44	−51	7
1	B	32	−37	−56	19	−52	−60	8	−50	−60	10	−50	−60	10
1	C	43	−34	−56	22	−43	−49	6	−39	−49	10	−41	−51	10
2 (1 s)	A	11	−28	−49	21	−27	−42	15	−32	−35	3	−34	−35	1
2	B	20	−33	−54	21	−32	−52	20	−38	−48	10	−42	−45	3
2	C	31	−36	−48	12	−39	−50	11	−44	−51	7	−46	−49	3
3 (500 ms)	A	54	−26	−52	26	−32	−53	21	−46	−55	9	−51	−52	1
3	B	46	−28	−56	28	−34	−59	25	−49	−63	14	−58	−61	3
3	C	40	−28	−53	25	−31	−57	26	−47	−59	12	−53	−57	4

monics on hydrophones A and C. Unfortunately it was not possible to estimate the beam pattern of the whistle owing to few usable signals recorded with this type of array. Increasing the number of hydrophones in the array would decrease range error (e.g., Spiesberger and Fristrup, 1990; Wahlberg *et al.*, 2001) and provide more data for determining the beam pattern of white-beaked dolphin whistles.

Sound production in odontocetes has been evaluated with a piston model (e.g., Au, 1993). Lammers and Au (2003) described a theoretical beam pattern of bottlenose dolphin whistles using a 4-cm piston radius. The beam pattern of white-beaked dolphin clicks is narrower than that of bottlenose dolphin clicks (Rasmussen *et al.*, 2004). A piston radius of 6 cm gave the best fit for white-beaked dolphin clicks so we used 6 cm to model the beam pattern for whistles. At 30 kHz the 10-dB beam is 40°, at 40 kHz it is 30°, and at 50 kHz we get 20°. These frequencies correspond closely to those in the third, fourth, and fifth harmonics of white-beaked dolphin whistles. Figure 4 illustrates the beam pattern of two consecutive whistles separated by a 10-s interval. Assuming the whistles came from the same animal, it is possible for a dolphin at a distance of 11 m from hydrophone A to ensonify all hydrophones with five harmonics (dashed lines, 30°), but with another whistle, and by changing the beam width, to only ensonify hydrophone B with five harmonics (solid lines, 20°) and hydrophones A and C with four harmonics of its beam. The bottlenose dolphin shows considerable control over its echolocation beam width (Dankiewicz *et al.*, 2005).

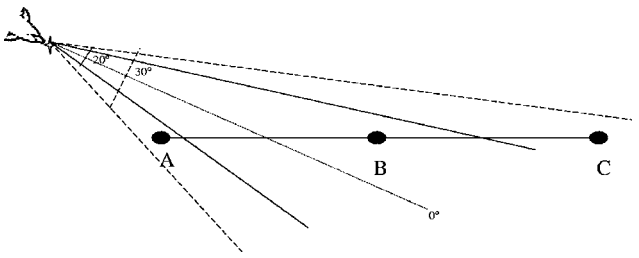


FIG. 4. Two whistles recorded at 10-s intervals giving the same position; one whistle contained unequal harmonics and the other equal harmonics. The former had five harmonics on hydrophone B and four on hydrophones A and C. The latter had five harmonics on all three hydrophones. The calculated distances to the dolphin were 11, 20, and 31 m from hydrophones A, B, and C, respectively (see Fig. 1). The two whistles could have come from the same dolphin or two different dolphins. We used a radius of 6 cm in the piston model to calculate the directionality of white-beaked dolphin whistles. The calculated beam width for the fifth harmonic is 20° and for the fourth harmonic is 30°. See text for further explanation.

Contact ranges are also important factors in communication. These can be calculated by using signals with highest (167 dB) and lowest (118 dB) source levels and some assumptions. If we assume white-beaked dolphins have similar hearing sensitivity as bottlenose dolphins (Au, 1993), then at 10 kHz the detection threshold (DT) should be 65 dB *re*. 1 μPa. The ambient noise level (NL) in Faxaflói Bay (sea state 0–1) measured using a $\frac{1}{3}$-oct filter centered at 10 kHz (Spectra Plus, Sound Technology Inc.) was about 75 dB *re*.1 μPa (Fig. 5). In this case the hearing of the dolphins is limited by the ambient noise level. When using a source level of 167 dB (SL), the maximum transmission loss before the sound becomes inaudible will be 167 dB(SL) −75 dB (NL)=92 dB. At 10 kHz the loss due to absorption is 1.1 dB/km and the transmission loss of 92 dB corresponds approximately to a range of 10.5 km. When using a minimum source level of 118 dB, the transmission loss equals 43 dB, giving a communication range of up to about 140 m.

Assuming that our distance determinations using acoustic and visual methods are reasonable, how can we account for the large variation in source level? We found the source level of white-beaked dolphin whistles to vary between 118 and 167 dB (n=129) using method II and between 124 and 166 dB (n=36) using method I (see Table I), resulting in

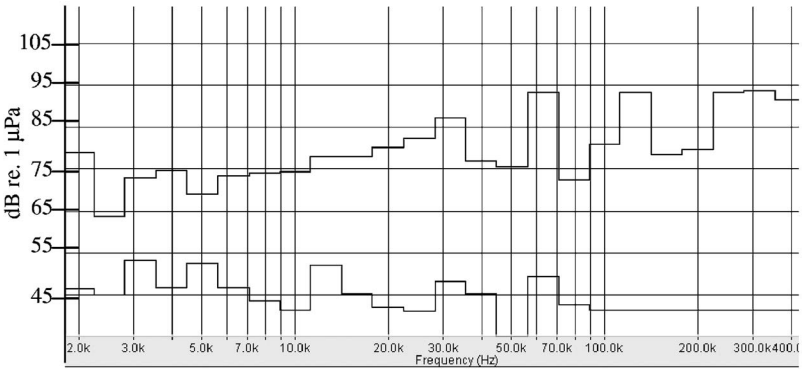


FIG. 5. System noise level (light gray line) with frequency plotted using $\frac{1}{3}$ octave filter and ambient sea noise level (black line) in Faxafloí Bay, Iceland, recorded at Beaufort 0-1.

source level variations of 42 to 49 dB. This variation cannot be explained by the orientation of the animals. The first harmonic of the white-beaked dolphin whistles contains the main part of the energy (Table II), which is about 10 kHz. Dolphin whistles are almost omni-directional at 10 kHz with a 3-dB beam width of 180° (Lammers and Au, 2003).

The variation in source levels may be explained by dolphins communicating at different ranges. On the first recording day we initially recorded quiet whistles with a single group of dolphins around our boat. The whistles became louder shortly before new dolphins joined the group and finally became more quiet at the end of the recording session with only a single group of dolphins in view. This scenario suggests that louder whistles were attracting distant individuals. Consequently, individuals communicating within the group may use quieter whistles while those communicating between distant groups may use louder whistles.

In conclusion this study shows that white-beaked dolphin whistles could be used both for short-range (up to about 140 m) and for long-range communication (up to about 10.5 km). In addition, our study indicates that white-beaked dolphin whistles are directional, an important factor for acoustic communication.

ACKNOWLEDGMENTS

This study was supported by the Oticon Foundation and the Danish National Research Foundation. The authors would like to acknowledge Dr. Whitlow Au for his assistance in the development of the hydrophone array system used in this study. Thanks also to Gisli Vikingsson at the Marine Research Institute and Jörundur Svarvarsson at the Institute of Biology, Iceland University in Reykjavik, for their cooperation, to Troels Jacobsen for assisting in the field, and to the captain and boat owner, David Thor Olafsson. We also thank Ulrik Nørum, Institute of Biology and Pia Veldt Larsen, Department of Statistics, University of Southern Denmark, for helping with the statistical analyses. Thanks to René Swift and Kimie Salo for reviewing the manuscript. Thanks to Dr. Paul Nactigall and Dr. Vincent Janik along with two referees for useful comments to improve the manuscript.

APPENDIX: SOURCE LEVEL ESTIMATES FOR THREE TRANSDUCERS

First, we observe that the peaks of the cross-correlation functions (CCFs) between the signals at the three channels (a,b,c) are readily obtainable. We use the delay estimates derived from these CCFs to align the signals.

The three aligned signals,

$$x_j(t) = s_j(t) + n_j(t), \quad j \in \{a,b,c\}, \tag{A1}$$

then represent the observed signals, delayed so that the correlation between them is maximal. The noise terms, n_j, are assumed to be independent in the three channels. The "signal" terms, s_j, represent attenuated copies of the unknown signal ultimately to be estimated. The alignments mean that, for instance, for signals x_a and x_b,

$$\int x_a(t)x_b(t)\,dt = \max[\mathrm{CCF}(x_a,x_b)] \equiv P_{ab} \tag{A2}$$

with similar definitions for P_{ac} and P_{bc}. The sought after energies, E_j, of the signals, s_j, are given by

$$E_j = \int S_j^2(t)\,dt. \tag{A3}$$

The signals s_j can be factorized so that

$$s_j(t) = \sqrt{E_j}s_j(t), \quad \text{with} \int s_j^2(t)\,dt = 1. \tag{A4}$$

Since the variation between the signals s_j is dominated by a scaling factor, we can write (A1) as

$$x_j(t) = \sqrt{E_j}s(t) + n_j(t), \tag{A5}$$

that is, as three amplitude factors times the common unit energy signal plus an uncorrelated noise term for each signal. Inserting (A5) in (A2) gives

$$\begin{aligned} P_{ab} &= \int \left(\sqrt{E_a}s(t) + n_a(t)\right)\left(\sqrt{E_b}s(t) + n_b(t)\right)\,dt \\ &= \int \sqrt{E_a}s(t)\sqrt{E_b}s(t) + n_a(t)n_b(t) + \sqrt{E_a}s(t)n_b(t) \\ &\quad + \sqrt{E_b}s(t)n_a(t)\,dt. \end{aligned} \tag{A6}$$

In (A6), the noise terms at the three receivers are uncorre-

lated with each other and the signals are uncorrelated with noise, so—using (A4)—only

$$\approx \int \sqrt{E_a}s(t)\sqrt{E_b}s(t)\,dt = \sqrt{E_a}\sqrt{E_b} \tag{A7}$$

remains. The amplitude of each peak of the CCFs is the product of the square roots of the signal energies. With the three cross-correlation peaks we can therefore find the energies of s_j as

$$E_a = \frac{P_{ab}P_{ac}}{P_{bc}}, \quad E_b = \frac{P_{ab}P_{bc}}{P_{ac}}, \quad E_c = \frac{P_{ac}P_{bc}}{P_{ab}}. \tag{A8}$$

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Brownlee, S. M., and Norris, K. S. (**1994**). "The Acoustic Domain," in *The Hawaiian Spinner Dolphin*, edited by K. S. Norris, B. Würsig, R. S. Wells, and M. Würsig (Univ. of California, Berkeley, CA).

Brüel and Kjær (**1985**). "Noise and vibration," *Pocket Handbook*, Lyngby, Denmark.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). "A review of the signature whistle hypothesis for the Atlantic bottlenose dolphin, *Tursiops truncatus*" in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. Reeves (Academic, London).

Dankiewicz, L. A., Houser, D. S., and Moore, P. W. B. (2005). "Sonar beam control, beam steering, and off-axis target detection by a bottlenose dolphin," in *16th Biennial Conference on the Biology of Marine Mammals*, p. 68.

Dreher, J., and Evans, W. E. (**1964**). "Cetacean communication," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), Vol. 1.

Freitag, L. E. and Tyack, P. L. (**1993**). "Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks," J. Acoust. Soc. Am. **93**, 2197–2205.

Hand, D. J. (**1997**). *Multivariate Analysis of Variance and Repeated Measures: A Practical Approach for Behavioural Scientists* (Chapman and Hall, London).

Janik, V. M. (**2000**). "Source levels and the estimated active space of bottlenose dolphin (*Tursiops truncatus*) whistles in the Moray Firth, Scotland," J. Comp. Physiol., A **186**, 673–680.

Janik, V. M., and Slater, P. J. B. (**1998**). "Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.

Lammers, M. O., and Au, W. W. L. (**2003**). "Directionality in the whistles of Hawaiian Spinner dolphins (*Stenella Longirostris*): A signal feature to cue direction of movement," Marine Mammal Sci. **2**, 249–264.

Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signalling behaviour of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1630–1639.

Lilly, J. C. (**1963**). "Distress call of the bottlenose dolphin: Stimuli and evoked behavioral responses," Science **139**, 116–118.

Miller, P. J. O. (**2002**). "Mixed-directionality of killer whale stereotyped calls: A direction of movement cue?", Behav. Ecol. Sociobiol. **52**, 262–270.

Norris, K. S., Würsig, B., Wells, R. S., and Würsig, M. (**1994**). *The Hawaiian Spinner Dolphin* (Univ. of California, Berkeley, CA).

Popper, A. N. (**1980**). "Sound emission and detection by Delphinids," in *Cetacean Behaviour: Mechanisms and Function*, edited by L. M. Herman (Wiley-Interscience, New York), pp. 1–52.

Rasmussen, M. H. (1999). "Hvidnæsens lydproduktion, adfærd samt udbredelse," Masters thesis. Institute of Biology, University of Southern Denmark, Odense (in Danish).

Rasmussen, M. H., and Miller, L. A. (**2002**). "Whistles and clicks from white-beaked dolphins, *Lagenorhynchus albirostris* recorded in Faxaflói Bay," Aquat. Mamm. **28**, 78–89.

Rasmussen, M. H., and Miller, L. A. (**2004**). "Echolocation and social signals from white-beaked dolphins, *Lagenorhynchus albirostris*, recorded in Icelandic water," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (Univ. of Chicago, Chicago), pp. 50–53.

Rasmussen, M. H., Wahlberg, M., and Miller, L. A. (**2004**). "Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris*)," J. Acoust. Soc. Am. **116**, 1826–1831.

Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Smolker, R., Mann, J., and Smuts, B. (**1993**). "Use of signature whistles during separations and reunions by wild bottle nosed dolphin mothers and infants," Behav. Ecol. Sociobiol. **33**, 393–402.

Spiesberger, J. L. (**1998**). "Linking auto and cross-correlation functions with correlation equations: Application to estimating the relative travel times and amplitudes of multipath," J. Acoust. Soc. Am. **104**, 300–312.

Spiesberger, J. L., and Fristrup, K. M. (**1990**). "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," Am. Nat. **135**, 107–153.

Tyack, P. (**1997**). "Development and social functions of signature whistles in bottlenose dolphins, *Tursiops truncatus*," Bioacoustics 8, 21–46.

Wahlberg, M., Møhl, B., and Madsen, P. T. (**2001**). "Estimating source position of a large-aperature hydrophone array for bioacoustics," J. Acoust. Soc. Am. **109**, 397–406.

Yost, W. A. (**2000**). *Fundamentals of Hearing* (Academic, New York).

Zar, J. H. (**1996**). *Biostatistical Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

Source-to-sensation level ratio of transmitted biosonar pulses in an echolocating false killer whale

Alexander Ya. Supin[a)]
Institute of Ecology and Evolution of the Russian Academy of Sciences, 33 Leninsky Prospect, 119071 Moscow, Russia

Paul E. Nachtigall[b)] and Marlee Breese
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

(Received 6 December 2005; revised 7 April 2006; accepted 10 April 2006)

Transmitted biosonar pulses, and the brain auditory evoked potentials (AEPs) associated with those pulses, were synchronously recorded in a false killer whale *Pseudorca crassidens* trained to accept suction-cup EEG electrodes and to detect targets by echolocation. AEP amplitude was investigated as a function of the transmitted biosonar pulse source level. For that, a few thousand of the individual AEP records were sorted according to the spontaneously varied amplitude of synchronously recorded biosonar pulses. In each of the sorting bins (in 5-dB steps) AEP records were averaged to extract AEP from noise; AEP amplitude was plotted as a function of the biosonar pulse source level. For comparison, AEPs were recorded to external (in free field) sound pulses of a waveform and spectrum similar to those of the biosonar pulses; amplitude of these AEPs was plotted as a function of sound pressure level. A comparison of these two functions has shown that, depending on the presence or absence of a target, the sensitivity of the whale's hearing to its own transmitted biosonar pulses was 30 to 45 dB lower than might be expected in a free acoustic field.
 [DOI: 10.1121/1.2202862]

PACS number(s): 43.80.Lb [WWA] Pages: 518–526

I. INTRODUCTION

The biosonar of odontocetes (toothed whales, dolphins, and porpoises) has been a subject of interest over the past few decades (Nachtigall and Moore, 1988; Au, 1993; Thomas *et al.*, 2004), but many of the basic mechanisms underlying echolocation functioning remain uninvestigated. In particular, the problem of the avoidance of masking of faint perceived echoes by the much more intense transmitted biosonar pulses remains unsolved. When an animal echolocates, it hears not only the echo but also its own transmitted acoustic pulse. When a target is small and distant, the echo is many times weaker than the outgoing pulse. Due to high sound velocity in water, the delay between the transmitted pulse and its echo may be very short, down to a few milliseconds. Normally, in these conditions, one would expect strong forward masking of the echo but successful performance of the biosonar of odontocetes indicates that this sort of masking is negligible. For understanding the mechanisms of releasing from self-masking by transmitted pulses, among other features of the biosonar, it needs to be known how loudly the transmitted pulses are heard by the animal. In the present study, we tried to clarify this issue.

Some mechanisms that serve to avoid or diminish the forward-masking effect in the odontocete's auditory system are already known. In particular, their auditory system has a very high temporal resolution as demonstrated by a variety of psychophysical (Au, 1993; Nachtigall *et al.*, 2000; Helweg *et al.*, 2003) and physiological (evoked response) data (Supin and Popov, 1995a, b; Dolphin *et al.*, 1995; Popov and Supin, 1997; Popov *et al.*, 2001; Supin *et al.*, 2001; Supin and Popov, 2004; Mooney *et al.*, 2006).

Another possible way to avoid the self-masking in the biosonar may be to concentrate acoustic energy into the transmitted beam directed to a target and to dampen the acoustic energy reaching the animal's ears. Indeed, the role of the skull and melon as concentrators of transmitted sounds in front of the head is well known (Cranford, 2000; Ketten, 2000) along with the presence of sound-muffling structures inside of the head, in particular, air-filled peribullar and pterygoid sinuses. One more way to avoid the self-masking may be a short-term suppression of sensitivity during and immediately after the emission of each biosonar pulse with a subsequent quick releasing of this suppression. Mechanisms like this, based on the stapeadial reflex of the middle ear, have been demonstrated to occur in echolocating bats (Suga and Jen, 1975), but it is not known yet whether there is any contribution of any of these mechanisms in avoiding the self-masking effect in odontocetes because the toothed whale's ability to hear its own outgoing clicks, as compared to external sounds of similar amplitudes, has not yet been measured.

One way to examine this issue is to record the brain auditory evoked potentials (AEPs) during natural echolocation in dolphins. AEPs indicate the magnitude of the brain response to sounds. Recording AEPs during echolocation can show how the brain responds to both the emitted click and echo. Studies in a false killer whale (Supin *et al.*, 2003,

[a)]Electronic mail: alex_supin@sevin.ru
[b)]Electronic mail: nachtiga@hawaii.edu

 0001-4966/2006/120(1)/518/9/$22.50

2004, 2005) have shown that this approach is feasible. During active echolocation experiments, a set of AEPs was recorded containing responses to both the transmitted sounds and to the echoes. In the present study we used this method to evaluate the sensation level of the transmitted biosonar pulses in a toothed whale. For that, the AEPs provoked by the transmitted biosonar pulses of varying intensity were compared with AEPs provoked by external stimuli of known intensities. Assuming that (keeping other conditions equal) stimuli of equal sensation levels produce equal AEP amplitude, we attempted to assess how much the hearing of the whales own transmitted biosonar pulses was reduced when compared to externally generated pulses.

II. MATERIALS AND METHODS

A. Subject and experimental conditions

1. General

The experiments were carried out in facilities of the Hawaii Institute of Marine Biology, Marine Mammal Research Program. The subject was a false killer whale *Pseudorca crassidens*, an approximately 30-year-old female kept in a wire-net enclosure in Kaneohe Bay, Hawaii. The animal was trained to accept soft latex suction cups containing EEG electrodes to pick up the evoked potentials, to ensonify and recognize targets by echolocation, and to report the target presence or absence using a go/no-go reporting paradigm. Two experimental procedures were used: (1) recording AEP to biosonar clicks and (2) recording AEP to external stimuli.

2. AEP to biosonar clicks

The experimental facilities were laid out as follows [Fig. 1(a)]. The experimental enclosure was constructed of a floating pen frame (1), 8×10 m^2 in size, supported by floats and bearing an enclosing wire net. This enclosure (the animal section) linked to a target section—another floating frame (2), 6×8 m^2 in size that served to mount targets and did not bear net. In the net divider separating these two sections, there was an opening bounded by a hoop (3), 55 cm in diameter, that served as a hoop station for the animal. In front of the hoop, a hydrophone (4) was positioned 1 m from the level of the animal's blowhole to record the echolocation pulses. A target (5) was hung from a thin monofilament line at a distance of 3 m from the animal's head and could be pulled up out of water and lowered down into water. The targets were hollow aluminum cylinders with an outer diameter of 38 mm (1.5 in.) and 25.5 mm (1 in.) inner diameter, axis vertical. Two targets were used: 180 and 32 mm long, their target strengths were −22 and −37 dB, respectively. The hoop station (3), the hydrophone (4), and the lowered target (5) were in a horizontal straight line, all at a depth of 80 cm. In front of the animal, there was a movable baffle (6). When pulled up, this baffle screened the target section from the animal positioned in the hoop station; when it was lowered down, it opened the space in front of the animal. Behind the baffle, there was a screen (7) made thin black polyvinylchloride film that was sound-transparent but not light-transparent. This screen served to prevent visual detection of the target. Near the hoop station, a response ball (8) was mounted above the water surface serving as a target-present response indicator. The trainer kept a position (9) to give instructions to the animal and to reward it with fish for correct responses. The animal's position in the stationing hoop was monitored through an underwater video camera (10). The electronic equipment and the operator were housed in a shack (11).

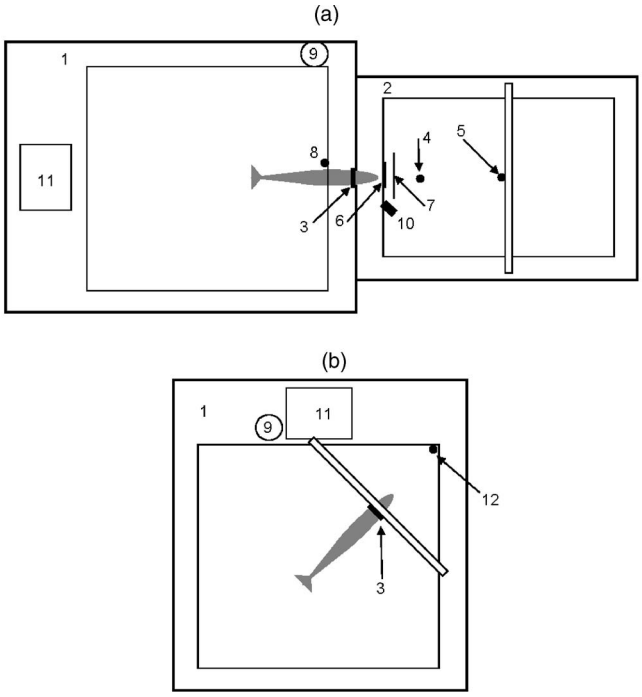


FIG. 1. Experimental conditions (a) for recording AEP to biosonar clicks and (b) for recording AEP to external stimuli. 1—experimental enclosure, 2—target section, 3—hoop station; 4—hydrophone, 5—target, 6—baffle, 7—screen, 8—response ball, 9—trainer's position, 10—video camera, 11—operator's shack, and 12—transducer.

3. AEP to external stimuli

In these experiments [Fig. 1(b)], the experimental enclosure (1) was supplied with a hoop station (3). At a distance of 2 m in front of the hoop station, there was a sound-transmitting transducer (12). The trainer kept a position (9) to give instructions to the animal and to reward it with fish for correct performance. The electronic equipment and the operator were housed in a shack (11).

B. Experimental procedure

1. AEP to biosonar clicks

Each session included an equal number of target-present and target-absent trials. The experimental procedure was as follows.

(*i*) Each session began with the trainer attaching suction-cup electrodes for AEP recording (see below for detail). (*ii*) The animal was given a signal to go to the hoop station. During the animal positioning, the baffle (6 in Fig. 1) screened the target from the animal. The target was either lowered down into water (a target-present trial) or pulled up out of water (a target-absent trial) in advance. (*iii*) As soon as the animal took the position in the hoop station, the baffle was lowered down, thus opening the space in front of the

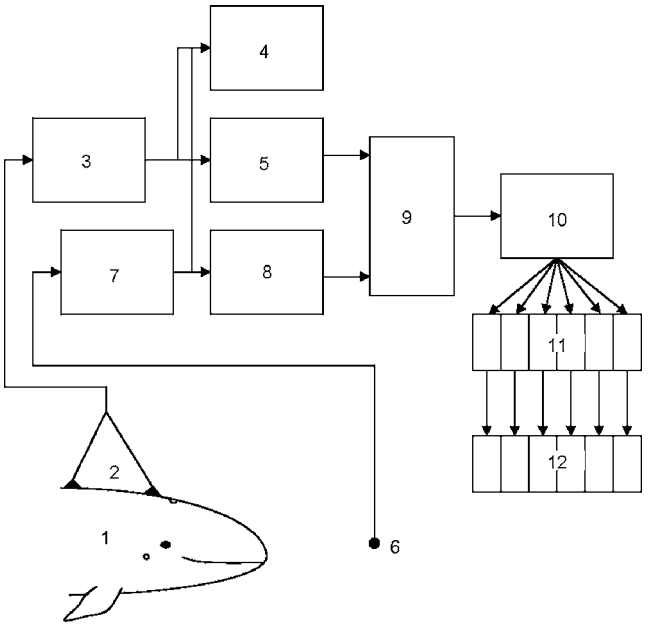


FIG. 2. Instrumentation for recording AEP to biosonar clicks. 1—subject, 2—electrodes, 3—EEG amplifier, 4—oscilloscope, 5 and 8—A/D converters, 6—hydrophone, 7—sound amplifier, 9—PC memory, 10—off-line sorting, 11—memory bins for sorted records, and 12—memory bins for averaged records.

animal. Immediately after that, the animal emitted a train of echolocation clicks; as a rule there were 20 to 50 clicks in a train. (*iv*) If the target was present, the animal was required to signal its detection by leaving the hoop and touching the signal ball, then coming to the trainer for the fish reward. During the no-target trails, the animal was required to wait until it was signaled to leave the hoop and come for the fish reward.

Each session consisted of 40 trials, 20 target-present and 20 target-absent, randomly alternated. During each trial, the sound and AEP-acquisition system (see below) was turned on as soon as the baffle was lowered, and it was kept on until the end of the echolocation pulse train. Thus, both transmitted biosonar pulses and AEPs provoked by these pulses were collected.

2. AEP to external stimuli

Each session began with the trainer attaching suction-cup electrodes for AEP recording. In each trial the animal was sent to a hoop station. As soon as the animal held the proper position, external stimuli (sound pulses) began to be played through the transducer, and AEP to these stimuli were collected. The animal was required to stay in the hoop for approximately 1 min while 1000 stimuli at a rate of 20/s were presented and the AEP were collected. After that, the animal was called back to the trainer for reward.

C. Instrumentation and data collection

1. AEP to biosonar clicks

The recording equipment was designed as shown in Fig. 2. Brain potentials were picked up from the subject (1) by EEG electrodes (2) which were gold-plated disks 10 mm in diameter mounted within rubber suction cups 60 mm in diameter. The active electrode was attached with conductive gel at the dorsal head surface, at the midline, 5–7 cm behind the blowhole. The reference electrode was also attached along with conductive gel on the animal's back near the dorsal fin. Brain potentials were led by shielded cables to a balanced EEG amplifier (3) and amplified by 2.5×10^4 within a frequency range from 200 to 5000 Hz. The amplified signal was monitored by an oscilloscope Tektronix TDS1002 (4) and entered into a 12-bit analog-to-digital converter (5) of a data acquisition card DAQ-6062E (National Instruments) installed in a standard laptop computer. Signals from the sound-recording B&K 8103 hydrophone (6) were amplified by a custom-made 40-dB amplifier (7), monitored by the same oscilloscope (4), and led to another analog-to-digital converter (8) of the same data acquisition card. Sampling rates were 25 kHz for the EEG-recording channel and 250 kHz for the sound-recording channel.

Data acquisition process was controlled by a custom-made program designed on the base of LabVIEW software (National Instruments). The program continuously monitored the sound-recording input, and each time when the signal exceeded a predetermined triggering level, a 10-ms window of the EEG-recording channel and 0.2-ms window of the sound-recording channel were stored in computer memory (9); the sound-recording window included 0.02-ms pretrigger time. The triggering level to detect biosonar pulses was compression of 150 dB *re* 1 μPa; for majority of pulses, the *one-peak* wave corresponded to a *peak-to-peak* level of around 155 dB. Lower triggering level was not used since it resulted in false triggering by clicks of snapping shrimps inhabiting the Kaneohe bay. Consequently, the used triggering level captured only biosonar pulses from the animal.

To extract low-amplitude AEPs from background brain-wave noise, an off-line averaging procedure was used. For that, all EEG records were sorted according to the peak-to-peak level of accompanying biosonar clicks (10); the sorted records were stored in separate memory bins (11). The sorting was done in 5-dB steps of the clicks, namely, 160 ± 2.5 dB, 165 ± 2.5 dB peak-to-peak, etc., up to the highest available click amplitude. In each of the sorting bins, brain-wave records were averaged. This resulted in click-related AEPs extracted from the background noise and stored in separate memory cells (12), with each resulting AEP waveform corresponding to a certain biosonar-click level with a tolerance of ±2.5 dB.

2. AEP to external stimuli

The equipment for AEP collection included the same electrodes, amplifier, and data acquisition card as for biosonar-click related AEP. The brain-wave amplifier gain, passband, and recording window were also the same: 2.5×10^4, 200–5000 Hz, and 10 ms, respectively. Unlike the biosonar-click-related AEP collection, the card was programmed for on-line averaging to extract AEP from background noise. The averaging was triggered by the external stimuli presented at a rate of 20/s. AEP were collected by averaging 1000 individual records.

Sound stimuli were digitally generated by the same card and played through a 12-bit digital-to-analog converter,

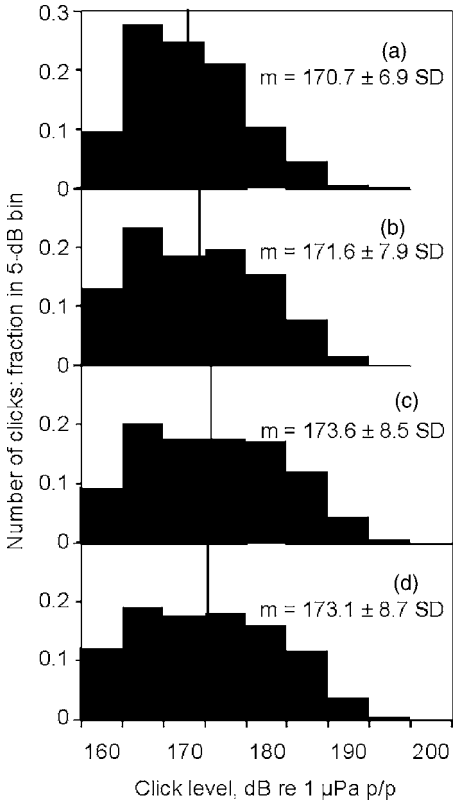


FIG. 3. Distributions of biosonar click peak-to-peak amplitudes. (a) Large target present (−22 dB strength). (b) Small target present (−37 dB strength). (c) Target absent in large-target sessions. (d) Target absent in small-target sessions. Means (m) and standard deviations (SD) are indicated near the histograms.

custom-made power amplifier-attenuator with a passband up to 1 MHz, and ITC-1032 spherical transducer (International Transducer Corporation). The stimuli were clicks produced by activation of the transducer by 7-μs rectangular pulses; pilot measurements showed that with such activation this particular transducer produces acoustic pulses similar to biosonar pulses of the experimental subject.

The stimulus intensity was calibrated with a B&K 8103 hydrophone placed in the center of the hoop station in the absence of the animal. This intensity was adopted as SPL next to the animal's head. Below it is specified in dB *re* 1 μPa peak-to-peak sound pressure span.

III. RESULTS

A. Biosonar click levels, waveform, and spectrum

The collected biosonar clicks varied in amplitude from 155 to 205 dB *re* 1 μPa peak-to-peak sound pressure span. Since the receiving hydrophone was at a distance of 1 m from the click source, the recorded values can be adopted as source levels of the clicks. The statistical distribution of click levels was analyzed separately for four conditions: the large target present and absent and the small target present and absent (Fig. 3). These four distributions featured some differences. With the small target present, the proportion of higher-level pulses slightly exceeded those with large target present [Figs. 3(a) and 3(b)], e.g., the proportion of pulses of 180 dB and higher was 16% with the large target and 25% with the small target. In target-absent conditions, the proportion of high-level pulses exceeded those in either of the target-present conditions [Figs. 3(c) and 3(d)]; (32% and 35%, respectively, of pulses of 180 dB and higher). However, because of rather small proportions of the highest-level pulses, these differences little influenced the means of the distributions: 170.7 dB at large target present, 171.6 dB at small target present, and 173.1 to 173.6 in target-absent conditions.

To characterize the biosonar click waveform and spectrum, all the recorded clicks were sorted according to their intensity in 5-dB bins, from 160±2.5 dB to 190±2.5 dB peak-to-peak source level, and in each bin all the click waveforms (a few hundreds to a few thousands, as available) were averaged. All the averaged clicks were bipolar (compression-rarefaction) of an overall duration about 60 μs [Fig. 4(a)]. This waveform was little dependent on the click level, except for a somewhat less prominent rarefaction phase at lower levels as compared to higher levels. Respectively, their frequency spectra were monomodal with the peak at 25–30 kHz, but the low-frequency "tail" was better pronounced at lower, rather than at higher, intensities [Fig. 4(b)].

B. Biosonar click-related AEP

The sorting and averaging procedure described above resulted in the ability to extract the AEP related to transmitted biosonar clicks. Preliminary sorting and AEP extraction was done separately for the four trial types: target present and target absent, each in sessions with large (−22 dB) and small (−37 dB) targets. However, a preliminary evaluation showed negligible difference between results obtained in sessions with large and small targets, so these results were combined and averaged. Thus, the final results were obtained by sorting and averaging separately for two trial types: target (either large or small) present and target absent. Although the maximum click intensity exceeded 200 dB *re* 1 μPa, AEP extraction was done only for clicks up to 185±2.5 dB, since the number of more intensive clicks was not large enough to be able to get a good AEP-to-noise ratio. The final AEP waveforms obtained in target-present trials are presented in Fig. 5, and those obtained in target-absent trials are presented in Fig. 6.

The mean background noise level in our records was estimated on average as 1.1 μV rms. The averaging procedure reduced the noise level by a factor from 31.6 (at 996 averaged records) to 63.3 (at 4011 averaged records) i.e., down to 35 to 19 nV rms. As shown in Figs. 5 and 6, the majority of the obtained final AEP waveforms well exceeded the noise level. These AEPs were mainly positive-negative-positive waveforms with the onset latency of 2.7 to 3 ms after triggering by the biosonar click and duration of each wave about 1 ms. The first positive wave had two peaks.

In both trial-present and target-absent trials the extracted AEPs displayed an obvious dependence on click intensity: the AEP amplitude was maximal at clicks of 180–185 dB and gradually decreased at lower click intensities. However, this dependence was different for the two trial types: in

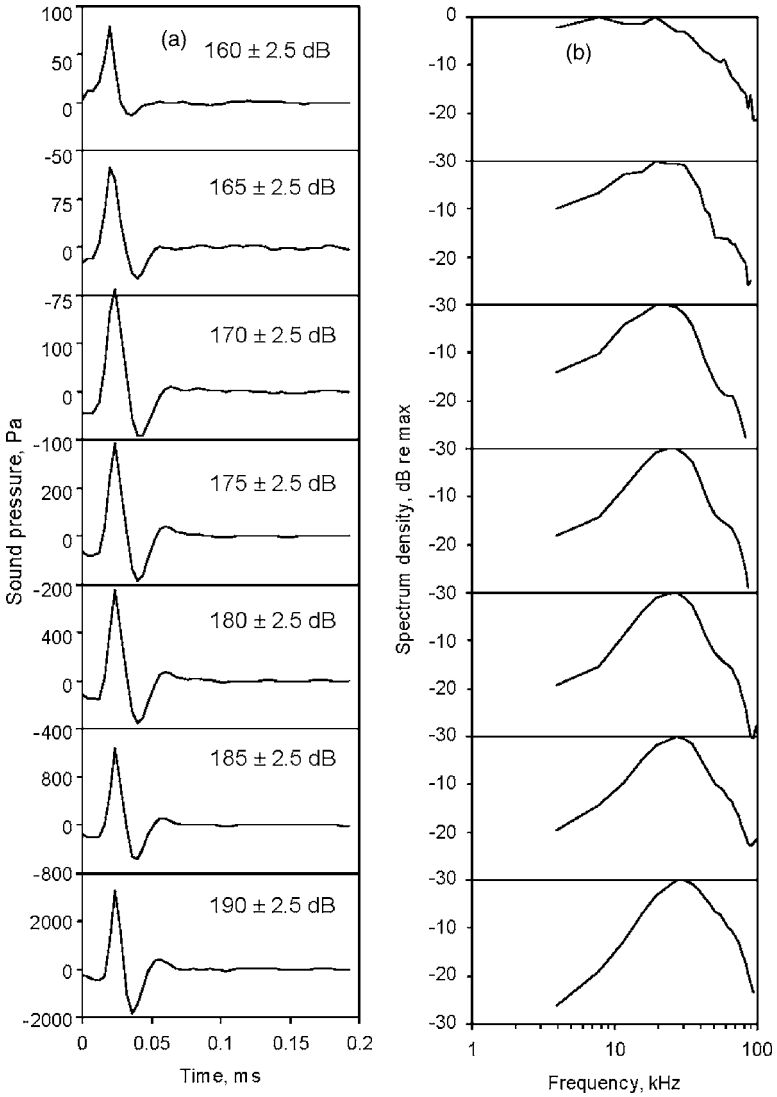


FIG. 4. Biosonar click waveforms and spectra. (a) Waveforms of clicks sorted by peak-to-peak amplitude in 5-dB bins. Peak-to-peak source levels are indicated near the waveforms. (b) Frequency spectra of the corresponding waveforms.

target-present trials, noticeable AEP could be detected at click intensities of 165 dB and higher (Fig. 5); whereas in target-absent trials, a rather large AEP appeared at the lowest of detectable click intensity of 160 dB (Fig. 6).

C. AEP produced by external clicks

The waveform and spectrum of the clicks used for external free-field stimulation are presented in Fig. 7. Comparison with Fig. 4 shows that they did not exactly reproduce the waveform and spectrum of transmitted biosonar clicks. The difference in the waveforms was due to the larger number of alternating compression-rarefaction waves in the external clicks. In the frequency-spectrum representation, this waveform difference manifested itself in a lower level of lower frequencies in the external-click spectrum as compared to the biosonar-click spectra. Nevertheless, there was significant similarity between the biosonar and external clicks: little differing peak frequency (near 30 kHz) and similar cut-off frequency (90–95 kHz at a −30-dB level).

Pilot monitoring had shown that the animal echolocated while it moved to the hoop station and positioned itself in the hoop but stopped echolocation while it stood motionless in the hoop during the data collection, so no interference between the external stimuli and animal's own transmitted click could be expected.

The AEPs provoked by the external clicks of various intensities are presented in Fig. 8. Their latencies were 4 to 4.5 ms (depending on stimulus intensity) after generating the sound pulse, but excluding the acoustic delay of 1.4 ms for the transducer distance of 2 m, the true physiological latencies were 2.6 to 3.1 ms, respectively. Similar to the biosonar-click-related AEPs, AEPs to external stimuli were composed of alternating positive and negative waves, each lasting about 1 ms; however, their waveforms were more structured with the positive wave split to two components. These waveforms correspond well to those described in many other odontocete species (Supin *et al.*, 2001). The AEP amplitude was intensity dependent within the investigated range of intensities. The dependence was the most steep (more than 0.02 μV/dB) within a range from 125 dB (threshold) to 145 dB *re* 1 μPa peak-to-peak.

D. Quantitative AEP dependence on click intensity

For quantitative presentation, biosonar-related AEP peak-to-peak amplitude was plotted as a function of click

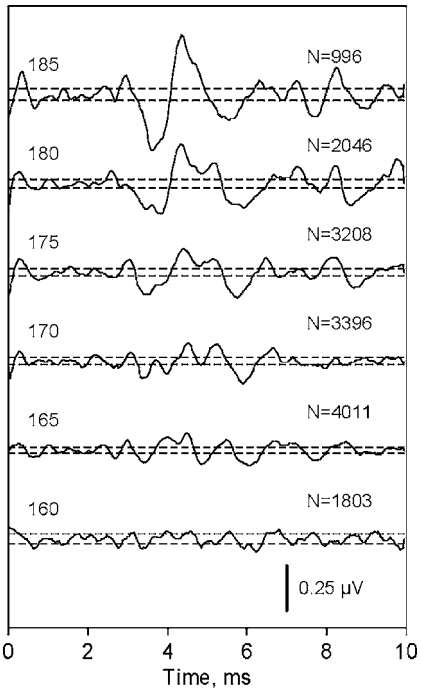


FIG. 5. AEP to biosonar clicks of various amplitudes in target-present trials. Mean click peak-to-peak amplitude (±2.5 dB) is indicated near the records in dB *re* 1 μPa. Straight dashed lines—±rms values of background noise. N—number of averaged records.

intensity specified as peak-to-peak *source level* for target-present [Fig. 9(a)] and target-absent [Fig. 9(b)] trial types. The two plots could be satisfactorily approximated by straight regression lines ($r^2 > 0.9$) with similar slopes (0.022 μV/dB) but were shifted relative one another by roughly 15 dB, target-absent trials featuring better AEP sensitivity.

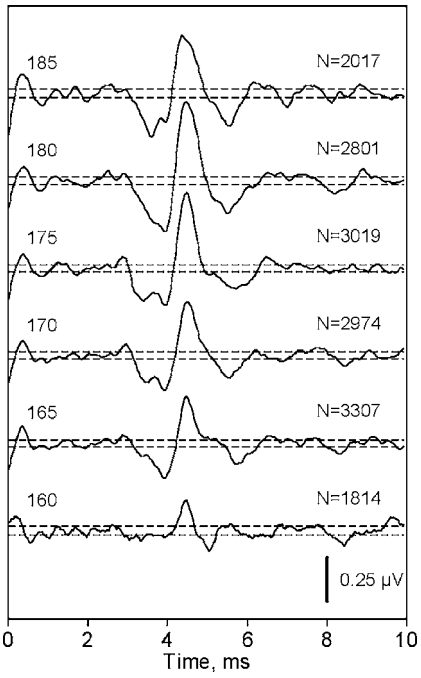


FIG. 6. AEP to biosonar clicks of various amplitudes in target-absent trials. Designations are the same as in Fig. 5.

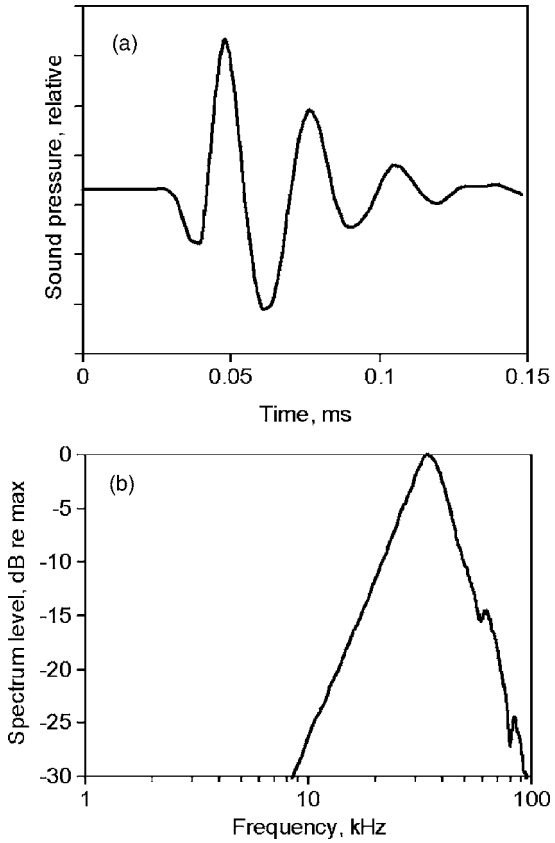


FIG. 7. Waveform (a) and spectrum (b) of external clicks.

Similarly, AEP dependence on external-click intensity was plotted as a function of external-click intensity, however specified as peak-to-peak *sound pressure level* [Fig. 9(c)]. This function also could be satisfactorily approximated by a regression line ($r^2 = 0.99$) and had almost the same slope

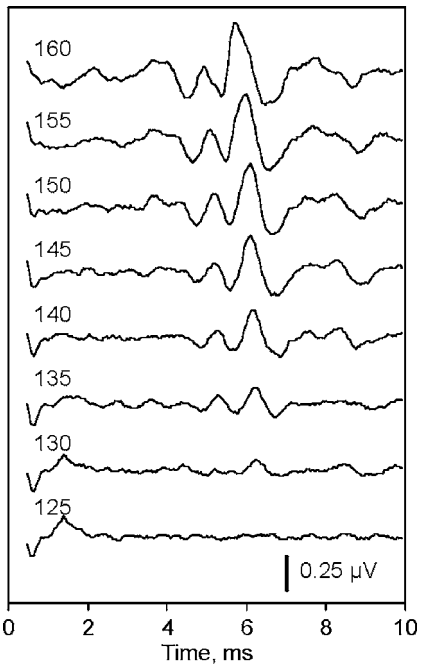


FIG. 8. AEP to external clicks. Click peak-to-peak amplitude is indicated near the records in dB *re* 1 μPa.

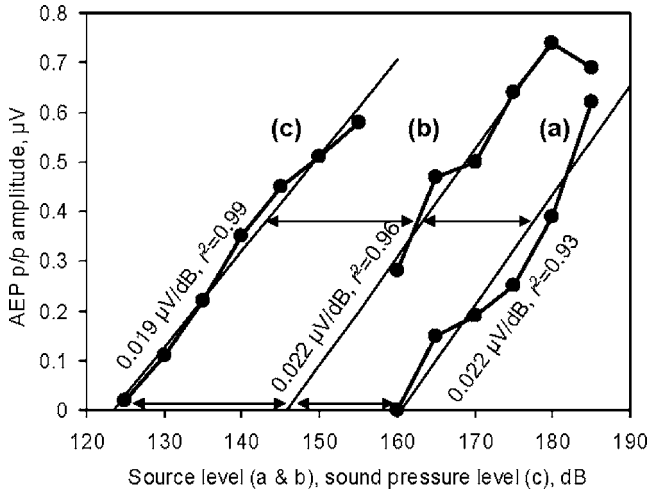


FIG. 9. AEP peak-to-peak amplitude dependence on sound click intensity. (a) Responses to biosonar click in target-present trials. (b) Responses to biosonar click in target-absent trials. (c) Responses to external stimuli. Solid lines with dot symbols—experimental data; thin straight lines—approximations of oblique parts of the plots by regression lines. Click intensity is specified as peak-to-peak source level for (a) and (b) and as peak-to-peak sound pressure level for (c) in dB *re* 1 μPa. Double-headed arrows show shift of plots relative to one another at a 0- and 0.35-μV amplitude levels.

(0.19 μV/dB) as AEP dependence on biosonar click intensity. Its position on the *sound-pressure level* scale was markedly shifted downward relative to the functions for biosonar click-related AEP [Figs. 4(a)] on the *source level* scale.

To characterize this shift quantitatively, we compared positions of regression lines at a zero level (threshold estimates) and at a level of 0.35 μV arbitrarily taken as a middle point of the dynamic range of the plots. At the zero level, the threshold estimates were 160.6, 145.9, and 123.5 dB for the plots (a)–(c), respectively. Thus, the shift of the external-click threshold [plot (c)] relative to the target-present biosonar-click threshold (a) was 37.0 dB, and its shift relative to the target-absent biosonar-click threshold was 22.4 dB. At the level of 0.35 μV, positions of the regression lines of the plots (a)–(c) were 176.3, 162.0, and 141.6 dB, respectively. Thus, the shift of the external-click plot (c) relative to the target-present plot (a) was as 34.7 dB, and its shift relative to the target-absent plot (b) was 20.4 dB.

IV. DISCUSSION

A. Comparison of AEPs to biosonar and external clicks

AEPs provoked by external sound clicks are typical auditory brainstem responses (ABRs) described in a number of odontocete species under similar stimulation and recording conditions (rev. Supin *et al.*, 2001, Nachtigall *et al.*, 2004). AEPs to biosonar clicks are obviously of the same nature, i.e., represent ABR provoked by transmitted clicks. They have similar physiological latency (around 3 ms), duration, and amplitude when recorded from one and the same point of the head surface.

A noticeable difference between the AEP produced by external and biosonar clicks is a more structured waveform of the responses produced by external stimuli. It may be explained by a lesser degree of synchronization of responses to biosonar clicks as compared to those to external clicks. Indeed, external clicks were entirely stereotyped during the AEP collection, whereas biosonar clicks were not. Being detected at a rather low level of sound pressure (which was necessary to detect low-amplitude clicks), biosonar clicks could reach their maximum at different delays (within a range of 10–30 μs) after the triggering instant. This asynchronicity might smooth out the AEP structure. Another source of desynchronization may be a complicated sound field reaching the ear through head tissues.

It is noteworthy that both external click-related and biosonar click-related AEP featured almost the same slope of the amplitude-versus-intensity functions, notwithstanding the incomplete similarity of external and biosonar click waveform and spectrum. It is not surprising taking into consideration that the dissimilarity of biosonar and external clicks mostly concerned the representation of lower frequencies. As was shown (Popov and Supin, 2000), lower frequencies contribute little to ABR in dolphins. Equality of slopes of the amplitude-versus-intensity functions means that both biosonar and external click of equal sensation levels (i.e., equally exceeding the threshold level) evoke equal excitation of the AEP-generating structures. This provides a basis for comparison of physiological intensities of external and biosonar clicks of a certain level.

B. Sound channeling into the biosonar beam and protection of ears

For correct interpretation of the obtained results, first of all, it should be stressed again that intensities of the external and biosonar clicks were necessarily specified herein in different measures. The external click intensity was specified as *sound-pressure level* (SPL), which is the sound pressure next to the receiving surface of the subject's head. The biosonar click intensity was specified as *source level*, which is the sound pressure at a standard (1 m) distance from the sound-generation structures. However, a comparison of these two measures would allow us to estimate the effectiveness of a system (or systems) to concentrate the sound energy into the biosonar beam and protect the animal's ears.

One possible way for such an estimation is to compare the experimental results with a situation expected in the absence of any sound-channeling or ear-protection system, i.e., in a free uniform acoustic field. In free 3-D field, sound intensity decreases with distance at a rate of 6 dB per distance doubling. In the false killer whale, the distance from the biosonar sound source (which is located in odontocetes between the blowhole and melon, as shown by Cranford, 2000) and the ears was 25–30 cm. Therefore, in free field, sound intensity at a distance of 1 m from the source (where it was monitored in our experiments) should be 10–12 dB *lower* than at a distance of 25–30 cm (where ears are located). Actually, in target-present trials, the click intensity in the biosonar beam was 35 to 37 dB (depending on the response amplitude taken for comparison) *higher* than an equally effective (in terms of evoked-potential amplitude) sound intensity near the ears; in target-absent trials it was

20–22 dB higher. Thus, in terms of comparison with a free field, the efficiency of the sound-channeling and ear-protection system is *not less than* 45–49 dB in the target-present and 30–34 dB in the target-absent conditions. Obviously this is a rather high effectiveness of sound concentration in the biosonar beam and isolation from the animal's ears even in the target-absent condition and a very high effectiveness in the target-present condition.

One way to concentrate sound energy is focusing into a rather narrow beam. Directivity index of the transmitted beam is more than 20–25 dB in bottlenose dolphins (Au *et al.*, 1986) and a false killer whale (Au *et al.*, 1995) and more than 30 dB in a beluga (Au *et al.*, 1987). Anatomical structures concentrating sound energy into a focused biosonar beam are well known and described. First, the role of the skull and melon as concentrators of transmitted sounds in front of the head should be mentioned (Cranford, 2000; Ketten, 2000). On the other hand, it is quite probable that the whale's ears are additionally screened from the transmitted biosonar pulses. In particular, air-filled peribullar and pterygoid sinuses also may play a role as sound-muffling structures (Houser *et al.*, 2004). In this context, it may be noted that investigations of binaural hearing in dolphins (Popov and Supin, 1992; Supin and Popov, 1993) have shown interaural intensity differences as large as 20 dB, which means that some head tissues are capable of shadowing sound spread across the head as much as 20 dB. Maybe the same, or similar, structures shadow the dolphin's ears from its sound-emitting devices.

Apart from sound channeling based on the head anatomy, the functional regulation of hearing sensitivity cannot be excluded as a mechanism to decrease hearing sensitivity to emitted biosonar pulses. It is commonly known that hearing sensitivity may be regulated at both conductive (the stapedial reflex) and sensorineural levels (adaptation). These mechanisms are known as being provoked by acoustical stimuli themselves, reducing the hearing sensitivity to high-level sounds. Is it possible that in whales and dolphins similar regulations of sensitivity are triggered in another way: not by sounds themselves but by the echolocation activity? We cannot exclude this possibility.

A hypothesis of regulated hearing sensitivity deserves attention, especially when one considers the fact that sensitivity to the transmitted biosonar pulses was different in target-present and target-absent trials. This difference could be achieved in two ways: either by regulation of sound-damping mechanisms within the head or by regulation of hearing sensitivity. The physiological significance of the second of these two options is easier to explain: in the absence of a stronger echo, the increase of hearing sensitivity may be a way to search for a weaker echo.

Anyway, we can posit that some mechanisms provide a much lower sensitivity of the whale's auditory system to transmitted biosonar pulses rather than to sound arriving from outside. Among the last, there are returning echoes during echolocation. Lower sensitivity to a transmitted signal should reduce masking of echo by the preceding biosonar pulse.

ACKNOWLEDGMENTS

This study was supported by the Office of Naval Research Grants Nos. N00014-98-1-0687 and N00014-05-1-0738, for which the authors thank Robert Gisiner. Support was also provided by the Russian Ministry of Science and Education Grants NSh-2152.2003.4 and NSh-7117.2006.4. Work was conducted under U.S. Marine Mammal Permit 978-1567 issued to Paul E. Nachtigall. This is contribution number 1231 of the Hawaii Institute of Marine Biology.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., Moore, P. W. B., and Pawloski, D. (**1986**). "Echolocation transmitting beam of the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **80**, 688–691.

Au, W. W. L., Pawloski, J. L., and Nachtigall, P. E. (**1995**). "Echolocation signals and transmission beam pattern of a false killer whale (Pseudorca crassidens)," J. Acoust. Soc. Am. **98**, 51–59.

Au, W. W. L., Penner, R. H., and Turl, C. W. (**1987**). "Propagation of beluga echolocation signals," J. Acoust. Soc. Am. **82**, 807–813.

Cranford, T. W. (**2000**). "In search of impulse sound sources in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. F. Fay (Springer, New York), pp. 109–155.

Dolphin, W. F., Au, W. W. L., and Nachtigall, P. (**1995**). "Modulation transfer function to low-frequency carriers in three species of cetaceans," J. Comp. Physiol., A **177**, 235–245.

Helweg, D. A., Moore, P. W. B., Dankewicz, L. A., Zafran, J. M., and Brill, R. L. (**2003**). "Discrimination of complex synthetic echoes by an echolocating bottlenose dolphin," J. Acoust. Soc. Am. **113**, 1138–1144.

Houser, D. S., Finneran, J., Carder, D., Van Bonn, W., Smith, C., Hof, C., Mattrey, R., and Ridgway, S. (**2004**). "Structural and functional imaging of bottlenose dolphin (Tursiops truncates) cranial anatomy," J. Exp. Biol. **207**, 3657–3665.

Ketten, D. R. (**2000**). "Cetacean Ears," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. F. Fay (Springer, New York), pp. 43–108.

Mooney, T. A., Nachtigall, P. E., and Yuen, M. E. (**2006**). "Temporal resolution of the Risso's dolphin, Grampus griseus, auditory system," J. Comp. Physiol., A **192**, 373–380.

Nachtigall, P. E., and Moore, P. W. B. (**1988**). *Animal Sonar: Processes and Performance* (Plenum, New York).

Nachtigall, P. E., Lemonds, D. W., and Roitblat, H. L. (**2000**). "Psychoacoustic Studies of Whale and Dolphin Hearing," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. J. Fay (Springer, New York), pp. 330–364.

Nachtigall, P. E., Supin, A. Ya., Pawloski, J. L., and Au, W. W. L. (**2004**). "Temporary threshold shifts after noise exposure in a bottlenosed dolphin (Tursiops truncatus) measured using evoked auditory potentials," Marine Mammal Sci. **20**, 673–687.

Popov, V. V., and Supin, A. Ya. (**1992**). "Electrophysiological study of the interaural intensity difference and interaural time-delay in dolphins," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 257–267.

Popov, V. V., and Supin, A. Ya. (**1997**). "Detection of temporal gaps in noise in dolphins: Evoked-potential study," J. Acoust. Soc. Am. **102**, 1169–1176.

Popov, V. V., and Supin, A. Ya. (**2000**). "Contribution of various frequency bands to ABR in dolphins," Hear. Res. **151**, 250–260.

Popov, V. V., Supin, A. Ya., and Klishin, V. O. (**2001**). "Auditory brainstem recovery in the dolphin as revealed by double sound pulses of different frequencies," J. Acoust. Soc. Am. **110**, 2227–2233.

Suga, N., and Jen, P. H.-S. (**1975**). "Peripheral control of acoustic signals in the auditory system of echolocating bats," J. Exp. Biol. **62**, 277–311.

Supin, A. Ya., and Popov, V. V. (**1993**). "Direction-dependent spectral sensitivity and interaural spectral difference in a dolphin: Evoked potential study," J. Acoust. Soc. Am. **96**, 3490–3495.

Supin, A. Ya., and Popov, V. V. (**1995a**). "Envelope-following response and modulation transfer function in the dolphin's auditory system," Hear. Res. **92**, 38–46.

Supin, A. Ya., and Popov, V. V. (**1995b**). "Temporal resolution in the dolphin's auditory system revealed by double-click evoked potential study," J. Acoust. Soc. Am. **97**, 2586–2593.

Supin, A. Ya., and Popov, V. V. (**2004**). "Temporal processing of rapidly following sounds in dolphins: evoked-potential study," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. Vater (Univ. Chicago, Chicago), pp. 153–161.

Supin, A. Ya., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer, Boston).

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2004**). "The interaction of outgoing echolocation pulses and echoes in the false killer whale's auditory system: evoked-potential study," J. Acoust. Soc. Am. **115**, 3218–3225.

Supin, A. Ya., Nachtigall, P. E., Au, W. W. L., and Breese, M. (**2005**). "Invariance of echo-responses to target strength and distance in an echolocating false killer whale: evoked potential study," J. Acoust. Soc. Am. **117**, 3928–3935.

Supin, A. Ya., Nachtigall, P. E., Pawloski, J., and Au, W. W. L. (**2003**). "Evoked potential recording during echolocation in a false killer whale Pseudorca crassidens," J. Acoust. Soc. Am. **113**, 2408–2411.

Thomas, J. A., Moss, C. F., and Vater, M. (**2004**). *Echolocation in Bats and Dolphins* (Univ. Chicago, Chicago).

Generalized perceptual linear prediction features for animal vocalization analysis

Patrick J. Clemins[a] and Michael T. Johnson
Speech and Signal Processing Laboratory, Marquette University, P.O. Box 1881, Milwaukee, Wisconsin 53233-1881

(Received 30 June 2005; revised 31 March 2006; accepted 18 April 2006)

A new feature extraction model, generalized perceptual linear prediction (gPLP), is developed to calculate a set of perceptually relevant features for digital signal analysis of animal vocalizations. The gPLP model is a generalized adaptation of the perceptual linear prediction model, popular in human speech processing, which incorporates perceptual information such as frequency warping and equal loudness normalization into the feature extraction process. Since such perceptual information is available for a number of animal species, this new approach integrates that information into a generalized model to extract perceptually relevant features for a particular species. To illustrate, qualitative and quantitative comparisons are made between the species-specific model, generalized perceptual linear prediction (gPLP), and the original PLP model using a set of vocalizations collected from captive African elephants (*Loxodonta africana*) and wild beluga whales (*Delphinapterus leucas*). The models that incorporate perceptional information outperform the original human-based models in both visualization and classification tasks. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2203596]

PACS number(s): 43.80.Lb, 43.66.Gf [WWA] Pages: 527–534

I. INTRODUCTION

One of the primary tasks when analyzing animal vocalizations is determining and measuring acoustically relevant features. Currently, many features used in bioacoustic analysis are based on the entire vocalization, often extracted by hand from spectrogram plots (Fristrup and Watkins, 1992; Leong *et al.*, 2002; Owren *et al.*, 1997; Riede and Zuberbühler, 2003; Sjare and Smith, 1986). Some of the features commonly used for analysis include duration, fundamental frequency measures, amplitude information, and spectral information such as Fourier transform coefficients. These traditional features are unable to capture temporally fine details of vocalizations because each feature has only one value for the entire vocalization. In addition, these features are often susceptible to researcher bias because the features are determined interactively. An alternative to this feature extraction paradigm is to divide signals into frames and extract features automatically on a frame basis. This generates a feature matrix for each vocalization that captures information about how the vocalization changes over time. Another limitation of traditional features, either global or frame based, is that they typically do not use information about the perceptual abilities of the species under study explicitly in the feature extraction process.

The generalized perceptual linear prediction (gPLP) model introduced here is a frame-based feature extraction model that uses perceptual information about the species under study to calculate features that are relevant to that species. The gPLP model is applicable to different species by incorporating experimental data from available perceptual tests. Furthermore, the gPLP model can significantly decrease the time spent analyzing vocalizations and generates features with finer temporal resolution that are largely uncorrelated and not subject to researcher bias.

The gPLP feature extraction model generates features based on the source filter model of speech production. Although this model was originally developed for human speech processing, it has been shown to be applicable to the vocalizations of terrestrial mammals for the purposes of describing vocal production mechanisms (Fitch, 2003). The source excitation, modeled as a pulse train for voiced sound or white noise for unvoiced sound, is produced by physiology such as the glottis in land mammals, the tympaniform membrane in birds, or air sacs in marine animals. This excitation then propagates through a filter consisting of the vocal tract and nasal cavity in terrestrial animals or the body cavity and melon in marine animals.

The gPLP model presented here is designed to suppress excitation information and quantify the vocal tract filter characteristics of the vocalizations. Excitation information includes the fundamental frequency contour, while vocal tract characteristics are represented by formant information. Vocal tract features carry the majority of the information in human speech, but there are a number of languages in which the fundamental frequency contour discriminates between units of speech with similar vocal tract characteristics. There is reason to believe that excitation information is also important to the discrimination of animal vocalizations. In fact, many studies have used fundamental frequency measures in order to classify vocalizations (Buck and Tyack, 1993; Darden *et al.*, 2003). Excitation information such as fundamental frequency measures can be added to the gPLP feature vector to include excitation information.

[a]Electronic-mail: patrick.clemins@marquette.edu

The gPLP feature extraction model generates features in the discrete cepstral domain. The discrete cepstral domain is defined as

$$c[n] = F^{-1}\{\log[F(s[n])]\}, \quad (1)$$

where F is the discrete Fourier transform and $s[n]$ is the original sampled time domain signal. This domain is preferred for speech processing systems because the general shape of the spectrum is accurately described by the first few cepstral coefficients, yielding an efficient signal representation. The cepstral domain is particularly appropriate for source filter model analysis because the logarithm operation effectively separates the excitation from the vocal tract filter (Deller *et al.*, 1993, p. 355). Finally, because cepstral values tend to be relatively uncorrelated with each other because of their orthonormal set of basis functions (Deller *et al.*, 1993, p. 377), the coefficients are good for statistical analysis methods.

The following section of this paper will describe the gPLP model in detail. Examples of the use of the gPLP model in vocalization analysis follow. Visualization, vocalization classification, and statistical testing tasks will be presented.

II. METHODS

A. Generalized perceptual linear prediction (gPLP)

The gPLP model is based on the perceptual linear prediction (PLP) model developed by Hermansky (1990). The goal of the original PLP model is to describe the psychophysics of human hearing more accurately in the feature extraction process. The gPLP model incorporates frequency warping to account for nonlinear frequency perception along the basilar membrane, critical bandwidth analysis to model frequency masking, equal-loudness normalization using audiogram information, and intensity-loudness power normalization. A block diagram of the gPLP method is shown in Fig. 1. The gPLP model includes the same components as the PLP, but incorporates experimentally acquired perceptual information as shown in Fig. 1 to tailor the feature extraction process to the species under study. The components designated by dotted boxes indicate where species-specific perceptual information is incorporated into the model. The various components of the model are discussed in detail in the following sections.

1. Preprocessing

The vocalization is first filtered using a preemphasis filter of the form

$$s'[n] = s[n] - ks[n-1], \quad (2)$$

where k is typically chosen to be between 0.95 and 0.99. This preemphasis filter gives greater weight to higher frequencies to emphasize the higher frequency formants and reduce spectral tilt (Deller *et al.*, 1993, p. 330). It also reduces the dynamic range of the spectrum so that the spectrum is more easily approximated by the autoregressive modeling component.

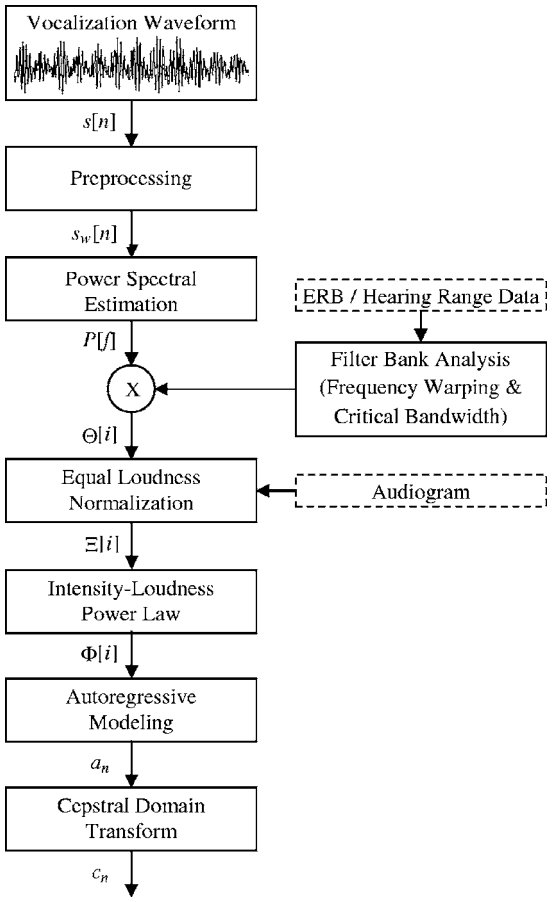


FIG. 1. PLP feature extraction block diagram. This original wave form is filtered and windowed in the preprocessing component. The power spectrum is then estimated for each frame of the vocalization. The power spectrum is convolved with a number of filters to generate filterbank energies which effectively smoothes and down samples the power spectrum. The filterbank energies are multiplied by the equal-loudness curve and cube-root compressed to account for the physiology of the ear. The down-sampled, normalized power spectrum is modeled by a set of autoregressive coefficients which are then converted to cepstral coefficients to take advantage of the cepstral domain.

The vocalization is then broken into frames and windowed using the Hamming window function (Oppenheim *et al.* 1999, p. 465). The frame size is usually chosen to include several fundamental frequency peaks, which is typically about 30 ms for human speech but may vary for other species' vocalizations. The vocalization is broken into frames so that the spectral estimation can be performed on quasistationary segments of the signal to ensure the precision of the spectral estimation. More information about the effects of windowing can be found in Oppenheim *et al.* (1999, p. 465).

2. Power spectral estimation

Once the signal is divided into windowed frames, the power spectrum is estimated. The discrete fast Fourier transform is used to estimate the power spectrum in this work, but other spectral estimation methods could also be used (Stoica and Moses, 1997). The discrete-time power spectrum $P[f]$ is estimated using

$$P[f] \approx abs\{F(s_w[n])\}^2, \tag{3}$$

where F is the discrete Fourier transform and $s_w[n]$ is the wth windowed frame of the signal.

3. Filter bank analysis

The next few components of the gPLP model transform the power spectrum to take into account various psychoacoustic phenomena. The filter bank analysis component accounts for two such phenomena, frequency masking and the nonlinear mapping between cochlear position and frequency sensitivity. Greenwood (1961) found that the cochlear-frequency map could be described logarithmically in many animal species with the equation

$$f = A(10^{ax} - k), \tag{4}$$

where f is frequency in Hz, x is the position on the basilar membrane that perceives that frequency, and A, a, and k are species-specific constants. Functions to convert between real frequency f and perceived frequency f_p can be created by replacing the basilar membrane position variable with perceived linear frequency as follows:

$$F_p(f) = (1/a)\log_{10}(f/A + k), \text{ and} \tag{5}$$

$$F_p^{-1}(f_p) = A(10^{af_p} - k), \tag{6}$$

where $F_p(f)$ converts from real frequency to perceived frequency and $F_p^{-1}(f_p)$ converts from perceived frequency to real frequency. The Mel-frequency scale, commonly used in speech processing, is a specific implementation of this warping function, using constant values of $A=700, a=1/2595 = 3.85\times10^{-4}$, and $k=1$. Values of A, a, and k can be determined for various species by fitting Eq. (6) to frequency-position data (Greenwood, 1990).

In cases where frequency-position data is not available, there are two other ways to acquire values for the constants. The first and most accurate method is to use equal-rectangular bandwidth (ERB) data (Zwicker and Terhardt, 1980). If the ERB data is fit by an equation of the form

$$\text{ERB} = \alpha(\beta f + \delta), \tag{7}$$

then the appropriate values of A, a, and k can be determined using the equations

$$A = \frac{1}{\beta},$$

$$a = \alpha\beta \log(e), \text{ and}$$

$$k = \delta. \tag{8}$$

where e is Euler's constant, the natural logarithm base. These equations are derived by taking the integral of the reciprocal of Eq. (7). The derivation of these equations is in the Appendix.

An alternative method for determining appropriate values for the constants requires an estimate of the hearing range of the species ($f_{\min}$ and $f_{\max}$). LePage (2003) noted that most mammals have a value of k near 0.88 and showed that this value is an optimal value when the tradeoffs between high frequency resolution, loss of low frequency resolution, minimization of map nonuniformity, and map smoothness are considered (LePage did not include non-mammalian species in the analysis, therefore using 0.88 as the value for k for those species may not be appropriate). Using the assumption that $k=0.88$, values for A and a can be determined using the equations

$$A = \frac{f_{\min}}{1-k}$$

$$a = \log_{10}\left(\frac{f_{\max}}{A} + k\right). \tag{9}$$

If this method is used, the lower bound of the filter bank must be greater than $f_{\min}$, otherwise negative values of f_p result.

The second psychoacoustic phenomenon the filter bank takes into account is frequency masking. The original PLP model (Hermansky, 1990) constructed the filter bank using filters, Ψ_i, shaped like the critical band masking filters described by Fletcher (1940). These exponential-shaped masking filters are based on human sound perception and are computationally complex. Because of this complexity, the gPLP model implemented in this work uses triangular-shaped filters to approximate the critical band masking curve. Triangular-shaped filters can be described by the equation

$$\Psi_i[f] = 1 - \left|\left(\frac{2}{f_H - f_L}\right)f - \left(\frac{f_H + f_L}{f_H - f_L}\right)\right|, \tag{10}$$

where f_L and f_H are the low and high cutoff frequencies of each filter. This approximation is common in human speech processing feature extraction models (Davis and Mermelstein, 1980). Another reason for using a simple filter shape is that there is little data on the auditory filter shapes of animals other than humans, so more complex filter shapes are not necessarily more accurate.

The number of filters contained in the filter bank should be determined so that the bandwidth of each filter approximates the critical bandwidth of each species. However, because of the limitations on the resolution of the Fourier spectral estimate, this is not always possible. The lower frequency filters in the filter bank can become very narrow due to the Greenwood frequency warping. If too many filters are specified for the filter bank, the low frequency filters become narrow enough that they do not contain any points, or frequency bins, of the spectral estimate. The maximum number of filters the filter bank can contain before some filters contain no spectral points is a function of window size and the range of the filter bank (Clemins *et al.*, 2005).

As an example of the incorporation of perceptual information into filter bank design, the filter bank for the Indian elephant, is shown in Fig. 2. Perceptual data from Heffner and Heffner (1982) is used to determine the Greenwood equation constants. The equal loudness curve, discussed below, is applied to the filter bank in the figure which results in the variable height of the individual filters. Using the filter bank, filter energies $\Theta[i]$ are calculated with

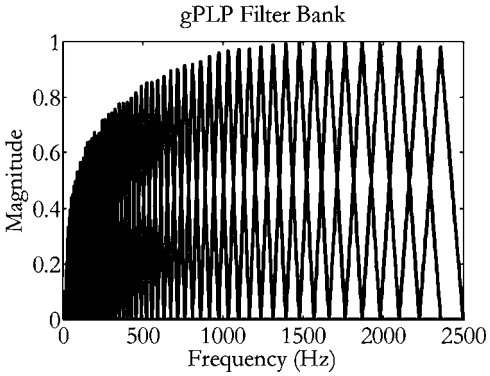


FIG. 2. Perceptual filterbank for an Indian elephant. The filters are logarithmically spaced according to the Greenwood cochlear map function. The constants A, a, and k are computed assuming the optimal $k=0.88$ for mammals as calculated by LePage (2003) and the approximate range of hearing for an Indian elephant (10–10 000 Hz). The equal loudness curve has been applied to the filter bank magnitudes to show its effect. Perceptual data is from Heffner and Heffner (1982).

$$\Theta[i] = \sum_{f=f_L}^{f_H} P[f]\Psi_i[f], \tag{11}$$

where $P[f]$ is the power spectrum, and f_L and f_H are the low and high cutoff frequencies of each filter $\Psi_i[f]$.

4. Equal loudness normalization

Once the filter bank energies are calculated, an equal-loudness curve is used to normalize the filter bank energies. Hermansky (1990) originally used a filter transfer function based on human sensitivity at about the 40-dB level adopted from Makhoul and Cosell (1976). For other species, an equal-loudness curve $E[f]$ can be approximated from the audiogram $A[f]$ of a species, which is much more widely available, using

$$E[f] = T - A[f], \tag{12}$$

where T is 60 dB for species which acquire sound through the air, and 120 dB for species that acquire sound in water. The different values are a result of the different propagation properties of sound waves and different reference dB pressures in the different mediums (Ketten, 1998). A polynomial curve is then fitted to $E[\log(f)]$ in order to interpolate for the frequency values sampled in $\Theta[i]$. A fourth-order curve has been found to adequately model most equal-loudness curves when $\log(f)$ is used. The constraint that $E(f)$ is always positive is maintained by setting all negative values of $E(f)$ to zero. The equal loudness curve is applied by multiplying the filter bank energies by the fitted curve using the equation

$$\Xi[i] = \Theta[i]E(f_i), \tag{13}$$

where $\Xi[i]$ are the equal loudness normalized filter bank energies and f_i is the center frequency of the ith filter. The multiplication of the filter bank energies, in linear units, by the equal loudness curve, in decibel units, results in filter bank energies in arbitrary units. The resulting energy scale is relative to the perceptual abilities of the species at that frequency.

5. Intensity-loudness power law

The last psychoacoustic related operation is the application of the intensity-loudness power law

$$\Phi[i] = \Xi[i]^{1/3}, \tag{14}$$

where $\Phi[i]$ are the power law and equal loudness normalized filter bank energies. Stevens (1957) found this cube root relationship between the intensity of sound and its perceived loudness in humans. Although this exact relationship may not hold for other species, it is likely that the structural similarities between species yield a comparable correspondence between power and loudness. This relationship may also be different for marine species because of the differences in the propagation of sound through air and water. Regardless of the appropriate power coefficient, this operation is beneficial from a mathematical modeling sense because it reduces the spectrum's dynamic range to make the normalized filter bank energies $\Phi[i]$ more easily modeled by a low-order autoregressive all-pole model.

6. Autoregressive modeling

The last two components of the gPLP model transform the filter bank energies into more mathematically robust features. First, $\Phi[i]$ is approximated by an all-pole model using the autocorrelation method and the Yule-Walker equations as specified in Makhoul (1975). A fifth-order model has been shown to be adequate to model the first two formants of human speech and suppress interspeaker details of the auditory spectrum (Hermansky, 1990). The appropriate order of the LP analysis for other species is dependent on the number of harmonics present in the vocalization, the relative complexity of the power spectrum, and the task being performed.

7. Cepstral domain transform

The autoregressive coefficients a_n from the LP analysis can be transformed directly into equivalent cepstral coefficients c_n using a recursive formula (Deller *et al.*, 1993, p. 376). The primary reason to transform autoregressive coefficients into the cepstral domain is that Euclidean distance is perceptually meaningful in the cepstral domain (Deller *et al.*, 1993, p. 377), whereas a more complex distortion measure such as Itakura distance must be used for autoregressive coefficients to maintain consistency (Itakura, 1975). Cepstral coefficients are generally less correlated with each other than autoregressive coefficients because they are based on an orthonormal set of functions (Deller *et al.*, 1993, p. 377).

B. Greenwood frequency cepstral coefficients (GFCCs)

As an alternative to gPLP it is possible to apply similar techniques to the Mel frequency cepstral coefficients (MFCC) feature extraction model. The MFCC feature extraction model was made popular by Davis and Mermelstein (1980) and has been the most commonly used feature extraction method in human speech processing for many years. While the MFCC model is still widely used because of its computational efficiency, PLP is sometimes preferred because of its robustness and more accurate modeling of the

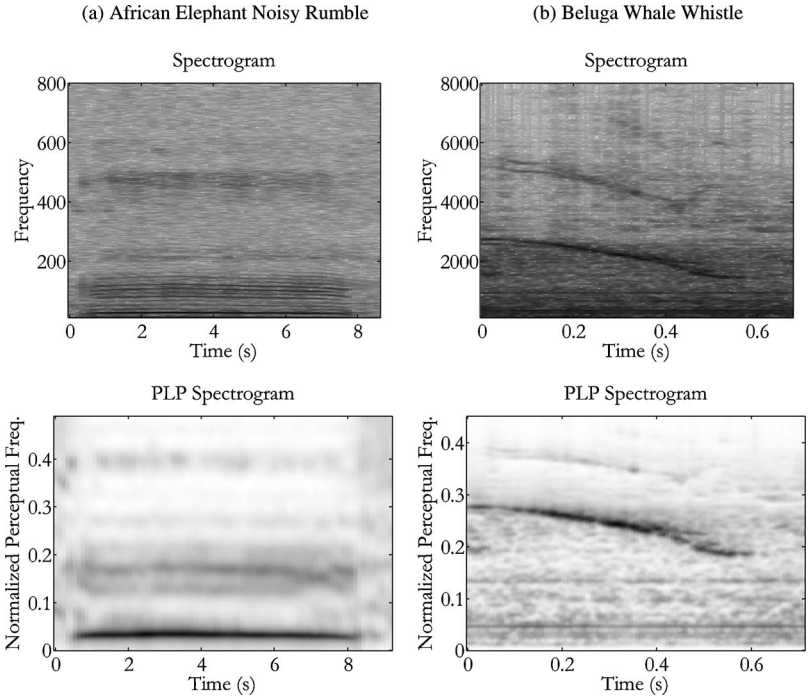


FIG. 3. Perceptual spectrograms. The top plots are traditional FFT-based spectrograms, while the bottom plots are perceptual spectrograms created using gPLP features. The left plots are of an African elephant's noisy rumple, and the right plots are of a beluga whale's whistle. Notice how the perceptual spectrogram enhances the peaks and valleys of the spectrum and warps the frequency axis according to the Greenwood cochlear map function.

human auditory system (Milner, 2002). The application of the Greenwood warping function to the MFCC model results in Greenwood Frequency Cepstral Coefficients (GFCCs) (Clemins *et al.*, 2006, p. 162). Although the GFCC model does not contain all of the psychophysical components of the gPLP model, the same filter bank design details presented here for gPLP are used to design the filter bank in the GFCC model. The main difference between the gPLP model and the GFCC model is the method used for calculating the cepstral coefficients. gPLP uses linear predictive coding (LPC)-derived cepstral coefficients, while GFCC calculates the cepstral coefficients directly from the filter bank energies using a discrete cosine transform. For more details on the GFCC feature extraction model, see Clemins *et al.* (2006).

III. EXAMPLES

The features generated by the gPLP model outlined above can be used to perform many types of analyses on animal vocalizations. Three typical types of analysis are described below utilizing gPLP features.

A. Visualization with perceptual spectrograms

Spectrograms have become an important analysis and visualization tool in the field of bioacoustics. They are useful for many different species and help researchers determine the differences between vocalizations. However, spectrograms do not incorporate any information about the perceptual abilities of the animal and are sometimes dominated by fundamental frequency content rather than spectral shaping. gPLP features can be used to generate perceptual spectrograms, incorporating information about the animal's perceptual abilities into the spectrogram. Because of the incorporation of perceptual data, perceptual spectrograms more closely represent the sound as the animal would hear it.

Figure 3 shows perceptual spectrograms of an African elephant's noisy rumble and a beluga whale's down whistle along with traditional FFT-based spectrograms. These two species were chosen to show that the gPLP model can be used to analyze vocalizations which include formants (elephant rumble) as well as vocalizations with harmonics (beluga down whistle). The perceptual spectrograms are plots of the linear prediction spectrum of each frame of the signal generated directly from LPC coefficients instead of transforming the coefficients into the cepstral domain (Deller *et al.*, 1993, p. 336).

For the perceptual spectrograms, a frame size of 300 ms with 100 ms step size was used to calculate 18 autoregressive coefficients from a filter bank of 50 filters. The perceptual data for the African elephant was taken from Heffner and Heffner (1982), while the data for the beluga whale was acquired from Ketten (1998) and Scheifele (2003). All of the plots, spectrograms, and perceptual spectrograms, are normalized so that pure white represents the absence of spectral energy and pure black represents the peak spectral energy of the vocalization.

In the two examples, notice the frequency warping that occurs in each perceptual spectrogram. The logarithmic warping as dictated by the Greenwood cochlear map function causes the lower frequencies to make up a larger portion of the perceptual spectrogram's horizontal axis. This warping makes small changes in the low frequency components of the vocalization more visible in the perceptual spectrogram. This effect can be seen in the whistle vocalization by examining the dynamics of the first (lowest) harmonic.

In a spectrogram, the excitation signal, which consists of the fundamental frequency and its harmonics, typically masks the response of the vocal tract filter in the spectrum. In contrast to this, the gPLP method enhances the spectral envelope's peaks and valleys and smoothes out the harmonics of the fundamental frequency. This can be seen best in the rumble vocalization's perceptual spectrogram in Fig. 3(a)

where the fundamental frequency harmonics are no longer present. These harmonics, the dark horizontal lines spaced about 20 Hz apart in the fast Fourier transform (FFT)-based spectrogram, distract from the formants (at 40, 100, and 220 Hz) which are much easier to see in the perceptual spectrogram along with their relative magnitudes.

The whistle example in Fig. 3(b) shows how the gPLP extraction model can track signals with quickly changing spectral characteristics. Although the whistle's harmonics move throughout the vocalization, the perceptual spectrogram tracks these changes as well as the FFT-based spectrogram. As with FFT-based spectrogram analysis, smaller window sizes can be used to better track faster moving spectral dynamics.

The gPLP spectrograms improve the contrast between the vocalization energy and the background noise in the spectrogram, making the vocalization easier to visualize. Both vocalizations have a much lighter background in the gPLP spectrogram when compared to the FFT-based spectrogram. However, the darkness of the vocalization energy stays the same, enhancing the contrast between the background noise and the vocalization energy. The gPLP spectrograms also enhance the visualization of narrow-band noise as seen in the beluga whale whistle in two places near 0.05 perceptual Hz and 0.14 perceptual Hz. On the other hand, the noise sources are not as dark as the vocalization energy because of the equal loudness curve applied to the filter bank energies.

B. Classification

Features obtained from gPLP analysis can also be effectively used in various machine learning classification systems. Since gPLP coefficients are perceptually relevant and largely uncorrelated they are a good choice for these tasks. To demonstrate the effectiveness of features generated from the gPLP model in a classification system, an example speaker identification task is performed on a set of vocalizations. This task is appropriate since the spectral characteristics of the rumble continuously change during the vocalization. The data set consists of 143 rumbles from five different African elephants, one male and four females. For more information on the data collection procedure, see Leong *et al.* (2002).

The classification model used for this experiment is a hidden Markov model (HMM). A HMM is a statistical classification model that can represent both the temporal and spectral characteristics of a signal. For more information about the HMM, refer to Clemins *et al.* (2005) and Rabiner (1989). Three state HMMs were used to model the rumble of each elephant and an additional three state HMM was used to model the silence before and after each rumble.

Table I shows the speaker identification accuracies for both MFCC and PLP features as various psychophysical signal processing methods are applied to the feature extraction process. Eighteen coefficients were extracted from 50 filter bank energies using an 18th-order autoregressive model for all trials. The total energy in each frame was also included in the feature vector. The vocalizations were framed using a window size of 300 ms and a window step size of 100 ms. These parameters choices were chosen empirically based on the perceptual information about the species.

TABLE I. Speaker identification accuracies. This table shows the effect of the various psychophysical signal processing components of the gPLP model on the classification accuracy of a speaker identification task.

	Filter bank range	
	10–3000 Hz (%)	10–500 Hz (%)
MFCC	46.9	72.7
PLP with Mel warp, human EQL	49.0	76.2
PLP with Mel warp, elephant EQL	49.0	75.5
PLP with Greenwood warp, human EQL	63.6	74.1
PLP with Greenwood warp, elephant EQL	68.5	81.8

Two different filter bank ranges are used: 10–3000 Hz and 10–500 Hz, to show the effect of limiting the filter bank range for each set of parameters. It is expected that since the range of most of the vocal energy of an elephant rumble is contained in the 10–500 Hz range, the use of that range for the filter bank should result in higher accuracies because it filters out noise in the other frequencies. This hypothesis is verified by the experimental results.

The rows of Table I represent various trials of the experiment with different feature extraction parameters. The first two rows show the change in accuracy when the cepstral coefficients are derived using autoregressive coefficients (gPLP) as opposed to a direct discrete cosine transform (GFCC) of the filter bank energies. Although the use of autoregressive coefficients results in slightly higher accuracies, the significance of the improvement is marginal.

The third row shows the accuracies when the human equal loudness curve is replaced by the derived African elephant equal loudness curve using audiogram data from Heffner and Heffner (1982). In this trial, the Mel-frequency scale was used to place the filters in the filter bank. The incorporation of the elephant equal loudness curve, when used with the Mel-frequency scale, does little to improve accuracy and in one case, decreases classification accuracy.

The fourth row shows the effect of using the Greenwood warping function instead of the Mel-frequency scale. The Greenwood constants were calculated using the optimal $k=0.88$ as suggested by LePage (2003) and the approximate hearing range for the African elephant, 10–10 000 Hz. The human equal loudness curve is used in this trial. While the use of the Greenwood warping function greatly improves accuracy for the larger filter bank range, it does little to improve accuracies for the smaller filter bank range. This suggests that the Greenwood warp helps to focus the analysis on the perceptually important parts of the vocalization when too large of a filter bank range is chosen.

The bottom row combines both the African elephant equal loudness curve and the Greenwood warping function as derived for the African elephant hearing range. When all available species-specific data is incorporated in to the feature extraction process, the classification accuracies improve significantly over the trials in which parameters based on human perception are used. While the Greenwood warping

TABLE II. Results of MANOVA analysis. MANOVA results Wilk's Λ statistic. Each row represents a different experimental setup.

	MANOVA results
All frames	$F_{95,13426}=142.8, P<0.001$
Middle frame	$F_{95,143}=5.81, P<0.001$
Average of all frames	$F_{95,143}=7.09, P<0.001$

function had the most effect when the larger filter bank range was used, the biggest increase in accuracy for the smaller filter bank range occurred when the African elephant equal loudness curve was incorporated into the feature extraction process. It is interesting to note that when used with the Mel-frequency warping scale, the species-specific equal loudness curve decreased the accuracy. However, when the appropriate warping function for that species is used, the species-specific equal loudness curve improved the classification accuracy.

C. Statistical tests

Features generated by the gPLP model can also be used as dependent variables in various statistical tests such as multivariate analysis of variance (MANOVA) or the multivariate *t* test. Since the cepstral coefficients generated by the gPLP model are orthogonal and relatively uncorrelated with each other, techniques such as principle components analysis (PCA) and linear discriminant analysis (LDA) are not necessary as preprocessing steps. To demonstrate the effectiveness of gPLP features in a statistical analysis scenario, a speaker identification experiment using MANOVA is presented.

The main issue with using frame-based features, such as those derived using the gPLP feature extraction model, with statistical tests is that frame-based features generate a feature matrix for each data example instead of a feature vector. Although repeated measures statistical tests might at first seem like an appropriate solution for handling the multiple feature vectors for each vocalization, the vocalizations are the result of a time-varying vocal production system. Therefore, the assumption that the system is unchanging, required for repeated measures tests, is invalidated. Three different methods for overcoming this issue are presented. Each method's advantages and disadvantages are also discussed. Other approaches are discussed in Clemins (2005).

The MANOVA analysis is performed on the same African elephant speaker identification data set used for the classification example. As in the classification example, 18 gPLP coefficients were extracted from each frame of the vocalizations using 50 filters spaced between 10 and 500 Hz along with the energy in each frame.

The first MANOVA analysis uses all of the frames of data from each vocalization in the analysis. The second analysis uses only the feature vector from the middle frame for each vocalization. Finally, the third analysis uses the average feature values across the entire vocalization.

Table II shows the results for the three different trials of the MANOVA analysis. The trial using all of the frames of data had the highest *F* value. The two trials that use one frame of data for each vocalization had substantially lower *F* values. It is interesting to note that using the average value of each feature of all frames in each vocalization resulted in a slightly higher *F* value as compared to using the features from the middle frame of each vocalization. This suggests that there is additional information in other parts of the vocalization besides the middle that could help separate the vocalizations by speaker.

Each of these methods for determining the variables to use has its own advantages and disadvantages. In the first method, the number of observations is much larger than the actual number of vocalizations because each vocalization generates a number of data points, one for each frame of the vocalization. However, because the spectral characteristics of the vocalizations vary over time, this first method more completely quantifies each vocalization. The last two methods have the advantage that they give more reasonable (i.e., lower) *F* values in the analysis because there is only one observation for each vocalization. On the other hand, it is difficult to determine which frame of the vocalization should be used to quantify the vocalization because the spectral characteristics of the vocalization can change dramatically. Therefore, for highly dynamic vocalizations, it might be better to use all of the observed frames instead of picking one frame for analysis as long as the higher *F* values are noted.

IV. CONCLUSIONS

The gPLP model generates perceptually meaningful features for animal vocalizations by incorporating psychophysical information about each species' sound perception. Physically, gPLP coefficients represent the shape of the vocal tract filter during vocalization production. gPLP coefficients are relatively uncorrelated and perceptually meaningful in a Euclidean space. They are also efficient in that a small number of coefficients can model a vocalization frame accurately. These features can be utilized for various types of animal vocalization analyses including visualization, classification, and statistical tests.

gPLP spectrograms are shown to enhance the spectral peaks and suppress broadband background noise. For the speaker identification task, the perceptual information included in the gPLP feature extraction model improves classification accuracy. Finally, the MANOVA analysis shows that the elephants produce significantly different vocalizations, which is consistent with the speaker identification task.

The features generated by the gPLP model can augment or replace traditional frequency-based features. gPLP coefficients can be added to a feature vector of traditional features before a statistical analysis and because they are relatively uncorrelated with each other, they can be added before or after principal component analysis (PCA) or a related technique. Finally, gPLP coefficients have no interpretive bias and decrease analysis time because they can be automatically extracted from the vocalization. Because of its efficiency and adaptability to various species' perceptual abilities, the gPLP model for feature extraction is an innovative and valuable addition to current tools available for bioacoustic signal analysis.

ACKNOWLEDGMENTS

The authors would like to thank the staff of the Wildlife Tracking Center and the Elephant Team at Disney's Animal Kingdom™ for the collection and organization of the acoustic data used in this research.

APPENDIX

The constant values A, a, and k for the frequency warping function, Eq. (5), can be derived from an ERB function of the form in Eq. (7) by taking the integral of the inverse as follows (Zwicker and Terhardt, 1980).

$$f_p = \int \frac{1}{\alpha(\beta f + 1)}, \tag{A1}$$

$$f_p = \frac{1}{\alpha} \int \frac{1}{\beta f + 1}, \tag{A2}$$

$$f_p = \frac{1}{\alpha\beta} \ln(\beta f + 1) + C. \tag{A3}$$

The integration constant C is then set to 0 in order to meet the constraint that $f_p = 0$ when $f = 0$. The base of the logarithm is then changed to 10 in order to match the base in Eq. (9).

$$f_p = \frac{1}{\alpha\beta \log(e)} \log(\beta f + 1). \tag{A4}$$

The equation is in the same form as Eq. (5) and the constant values can be read directly as

$$A = \frac{1}{\beta},$$

$$a = \alpha\beta \log(e),$$

$$k = 1. \tag{A5}$$

Buck, J. R., and Tyack, P. L. (**1993**). "A quantitative measure of similarity for *tursiops truncatus* signature whistles," J. Acoust. Soc. Am. **94**(5), 2497–2506.

Clemins, P. J. (2005). Automatic Classification of Animal Vocalizations. Ph.D. dissertation, Marquette University, Milwaukee, WI.

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**(2), 956–963.

Clemins, P. J., Trawicki, M., Adi, K., Tao, J., and Johnson, M. T. (**2006**). "Generalized perceptual feature for vocalization analysis across multiple species," *Proceedings of ICASSP*, Toulouse, France, May 14–19, 2006, in press.

Darden, S., Dabelsteen, T., and Pedersen, S. B. (**2003**). "A potential tool for swift fox (Vulpes velox) conservation: Individuality of long-range barking sequences," J. Mammal. **84**(4), 1417–1427.

Davis, S. B., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **28**(4), 357–366.

Deller, J. R., Proakis, J. G., and Hansen, J. H. L. (**1993**). *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, New York.

Fitch, W. T. (**2003**). "Mammalian vocal production: Themes and variation," *Proceedings of First International Conference on Acoustic Communication by Animals*, University of Maryland, College Park, MD, July 27–30, 2003, pp. 81–82.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Fristrup, K. M., and Watkins, W. A. (1992). Characterizing Acoustic Features of Marine Animal Sounds, Technical Report WHOI-92-04 Woods Hole Oceanographic (Woods Hole, MA, Institution.)

Greenwood, D. D. (**1961**). "Critical bandwidth and the Frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**(10), 1344–1356.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am. **87**(6), 2592–2605.

Heffner, R. S., and Heffner, H. E. (**1982**). "Hearing in the elephant (*Elephas maximus*): Absolute sensitivity, frequency discrimination, and sound localization," J. Comp. Physiol. Psychol. **96**(6), 926–944.

Hermansky, H. (**1990**). "Perceptual linear predictive (PLP) analysis for speech recognition," J. Acoust. Soc. Am. **87**(4), 1738–1752.

Itakura, F. (**1975**). "Minimum prediction residual principle applied to speech recognition," IEEE Trans. Acoust., Speech, Signal Process. **23**(1), 67–72.

Ketten, D. R. (1998). "A summary of audiometric and anatomical data and its implications for underwater acoustic impacts," *NOAA Technical Memorandum*.

Leong, K. M., Ortolani, A., Burks, K. D., Mellen, J. D., and Savage, A. (**2002**). "Quantifying acoustic and temporal characteristics of vocalizations of a group of captive African elephants (*Loxodonta africana*)," Bioacoustics 13(3), 213–231.

LePage, E. L. (**2003**). "The mammalian cochlear map is optimally warped," J. Acoust. Soc. Am. **114**(2), 896–906.

Makhoul, J. (**1975**). "Spectral linear prediction: properties and application," IEEE Trans. Acoust., Speech, Signal Process. **23**, 283–296.

Makhoul, J., and Cosell, L. (**1976**). "LPCW: An LPC vocoder with linear predictive spectral warping," *Proceedings of 1976 International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, pp. 466–469.

Milner, B. (**2002**). "A comparison of front-end configurations for robust speech recognition," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 13–17. Vol. **1**, pp. 797–800.

Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (**1999**). *Discrete-Time Signal Processing 2nd ed.* Prentice-Hall, Upper Saddle River, NJ.

Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): implications for production processes and functions," J. Acoust. Soc. Am. **101**(5), 2951–2963.

Rabiner, L. R. (**1989**). "Tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE **77**, 257–286.

Riede, T., and Zuberbühler, K. (**2003**). "The relationship between acoustic structure and semantic information in Diana monkey alarm vocalizations," J. Acoust. Soc. Am. **114**(2), 1132–1142.

Scheifele, P. M. (2003). Investigation into the response of the auditory and acoustic communication systems in the beluga whale (Delphinapterus leucas) of the St. Lawrence River estuary to noise, using vocal classification, Ph.D. dissertation, University of Connecticut, Hartford, CT.

Sjare, B. L., and Smith, T. G. (**1986**). "The vocal repertoire of white whales, *Delphinapterus leucas*, summering the Cunningham Inlet, Northwest Territories," Can. J. Zool. **64**, 407–415.

Stevens, S. S. (**1957**). "On the psychophysical law," Psychol. Rev. **64**, 153–181.

Stoica, P., and Moses, R. L. (**1997**). *Introduction to Spectral Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**(5), 1523–1525.

Sonoelastographic imaging of interference patterns for estimation of shear velocity distribution in biomaterials

Zhe Wu[a] and Kenneth Hoyt
ECE Department, University of Rochester, Hopeman Building 204, Rochester, New York 14627-0126

Deborah J. Rubens
Department of Radiology, University of Rochester, Rochester, New York 14627

Kevin J. Parker
ECE Department, University of Rochester, Hopeman Building 204, Rochester, New York 14627-0126

(Received 29 September 2005; revised 17 April 2006; accepted 18 April 2006)

The authors have recently demonstrated the shear wave interference patterns created by two coherent vibration sources imaged with the vibration sonoelastography technique. If the two sources vibrate at slightly different frequencies ω and $\omega+\Delta\omega$, respectively, the interference patterns move at an apparent velocity of $(\Delta\omega/2\omega)*v_{\text{shear}}$, where v_{shear} is the shear wave speed. We name the moving interference patterns "crawling waves." In this paper, we extend the techniques to inspect biomaterials with nonuniform stiffness distributions. A relationship between the local crawling wave speed and the local shear wave velocity is derived. In addition, a modified technique is proposed whereby only one shear wave source propagates shear waves into the medium at the frequency ω. The ultrasound probe is externally vibrated at the frequency $\omega-\Delta\omega$. The resulting field estimated by the ultrasound (US) scanner is proven to be an exact representation of the propagating shear wave field. The authors name the apparent wave motion "holography waves." Real-time video sequences of both types of waves are acquired on various inhomogeneous elastic media. The distribution of the crawling/holographic wave speeds are estimated. The estimated wave speeds correlate with the stiffness distributions. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2203594]

PACS number(s): 43.80.Jz [CCC] Pages: 535–545

I. INTRODUCTION

It is well known that changes in tissue mechanical properties are possible disease markers. In modern medicine, digital palpation is a routine screening method in physical examinations. In recognizing the significance, many scientists and researchers are developing various imaging modalities to visualize one or more parameters of the tissue mechanical properties qualitatively. Among many material properties parameters, shear wave propagation speed (and/or wavelength) has many researchers' interests, because the shear wave speed is closely related to shear modulus of elastic media (Love 1944). Parker and Lerner (1992) related shear wave speed to the production of eigenmodes in homogeneous biological materials. Yamakoshi *et al.* (1990) estimated both the vibration amplitude and phase due to external excitation with ultrasonic probing beams. Levinson *et al.* (1995) measured shear wave propagation speed in skeletal muscle *in vivo* with phase based ultrasonic techniques. Muthupillai *et al.* (1995) visualized the physical response of a material to harmonic mechanical excitation with phase encoded magnetic resonance imaging (MRI). Sandrin *et al.* (1999) developed an ultrafast imaging system (up to 10 000 frames/s), which is able to image the propagation of the low-frequency transient shear waves. Vibration displacements are measured using cross correlation of the ultrasonic signals (Sandrin *et al.* 1999). Dutt *et al.* (2000) measured small cyclic displacements (submicrometer level) caused by propagating shear waves in tissuelike media with a phase-based ultrasound method. Jenkyn and Ehman (2003) estimated the shear wave wavelength in skeletal muscles with magnetic resonance elastography (MRE). Shear wave wavelengths were found to increase with increasing tissue stiffness and increasing tissue tension (Jenkyn and Ehman 2003).

II. THEORY

A. Sonoelastography

The methods we propose are based on an ultrasonic imaging modality called sonoelastography (Lerner *et al.* 1988). Sonoelastography estimates the peak displacements of particle motion under audio frequency excitations by analyzing the power spectrum variance of the U.S. echoes, which is proportional to the local vibration amplitude (Huang *et al.* 1990, Taylor *et al.* 2000). Vibration fields are then mapped to a commercial ultrasound scanner's screen. Since this technique utilizes the existing Doppler hardware on most modern U.S. scanners, the frame rate of sonoelastography is as high as other Doppler modalities. Regions where the vibration amplitude is low are displayed as dark green, while regions with high vibration are displayed as bright green. Unless additional phase estimators are employed (Huang *et al.* 1992), sonoelastography ignores the phase information of

[a]Electronic address: wuzhe@ece.rochester.edu

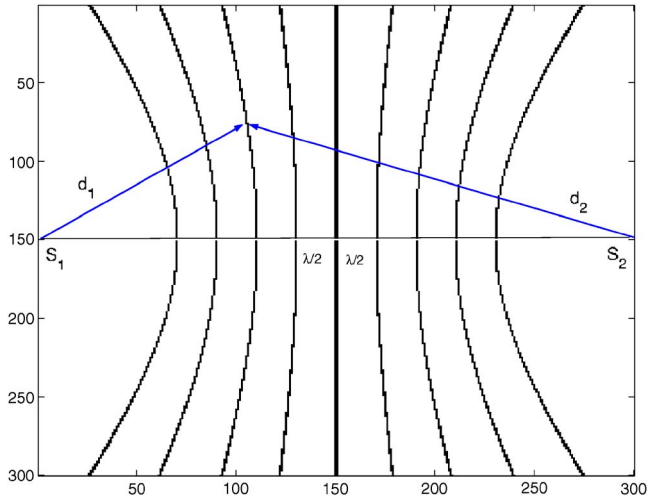


FIG. 1. S_1 and S_2 are the shear wave sources. If they vibrate in phase at equal amplitude, the antinodes lines of the interference pattern can be represented by a family of hyperbolas where $|d_1-d_2|=n\lambda$, n being integer numbers. Along the line S_1S_2, the spacing between the antinode lines is equal to $\lambda/2$.

shear wave propagations. Sonoelastography has been applied to visualize shear wave transducers' beam patterns or interference patterns (Wu *et al.* 2002).

B. Static shear wave interference patterns (a review)

As we reported in the previous paper (Wu *et al.* 2004), coherent shear wave sources create shear wave interference patterns in the media and the patterns can be visualized by sonoelastography in real time. Assuming the medium is homogeneous and isotropic, the phase of an arbitrary point in the vibration field is proportional to the distance between this point and the wave source

$$\phi = kd, \tag{1}$$

where k is the shear wave number and d is the distance from the field point to the wave source. In such a medium of infinite size, if there exists two coherent shear wave sources, then interference patterns appear. If the two sources are in phase, the antinode lines, which correspond to high vibration amplitudes, reside in such locations that the distance d_1 and d_2 have constant differences which equal integer multiple of the shear wavelength. The definition of d_1 and d_2 is depicted in Fig. 1,

$$|\phi_1 - \phi_2| = 2n\pi$$

$$|kd_1 - kd_2| = 2n\pi$$

$$|d_1 - d_2| = n\lambda \tag{2}$$

where n is an integer.

In other words, the interference patterns can be represented as a family of hyperbolas (Fig. 1). Assuming the distance between the two wave sources is D, the function of the family of hyperbolas can be expressed as

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

where $a = \frac{n\lambda}{2}$

and $b^2 = \left(\frac{D}{2}\right)^2 - a^2$ (3)

If we set $y=0$, Eq. (3) yields

$$x = \pm \frac{n\lambda}{2}. \tag{4}$$

Equation (4) states that along the central axis, the distance between the x intercepts of this family of hyperbolas equals half of the shear wave wavelength. If the two shear wave sources are far away compared to the size of the field of view, the family of hyperbolas can be approximated as parallel lines. This is equivalent to assuming the waves from each source are plane waves.

C. Moving shear wave interference patterns (a review)

Again in homogeneous and isotropic media, if one of the two sources vibrates at the frequency ω and the other source vibrates at $\omega+\Delta\omega$, where $\Delta\omega \ll \omega$, the interference patterns no longer remain static. They move toward the source with the lower frequency. Throughout the rest of this paper, we assume the two sources vibrate in such a fashion unless otherwise stated. This is referred to as the frequency difference condition. Following the derivation in the previous section, the location of the patterns should follow the equation:

$$|kd_1 - (kd_2 + \Delta\omega t)| = 2n\pi. \tag{5}$$

Following the same derivation in Eq. (4), the family of hyperbolas intercept the x axis at

$$x = \pm \frac{n\lambda}{2} + \frac{\Delta\omega \cdot \omega}{2\omega \cdot k} t$$

$$x = \pm \frac{n\lambda}{2} + \frac{\Delta\omega}{2\omega} t v_{\text{shear}}. \tag{6}$$

Therefore, the x intercepts of the hyperbolas move at the speed of $(\Delta\omega)/(2\omega)v_{\text{shear}}$. If we further assume the waves from each source are plane waves, then the whole interference pattern moves at this speed. We name the interference pattern motion crawling waves. Accordingly, the interference fringes are named crawling wave fronts; the spacing between the interference fringes is the crawling wavelength.

D. Crawling wave in inhomogeneous media

One important element to generalize the above results into elastic property estimation of inhomogeneous media is to understand the crawling wave's behavior *locally*. Despite the fact that the interference fringes' locations depend on the path integral of phase from each source, the following discussion proves that the crawling wave velocity depends solely on the local property [to a certain approximation in the two-dimensional (2D) case].

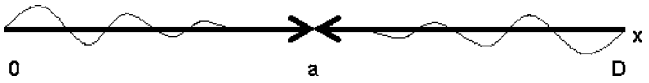


FIG. 2. Two shear wave sources are at 0 and D, respectively.

1. One-dimensional problem

We start with the one-dimensional problem. As depicted by Fig. 2, there are two wave sources at 0 and D, respectively. Without the loss of generality, we assume the two sources are in phase at the initial time ($t=0$). One source has frequency ω and the other source has frequency $\omega+\Delta\omega$. As we stated in the previous section, the locations of the antinodes are dependent on time. For instance, if one antinode resides at a, at $t=T, T>0$, we can write an equation on the phase relation

$$\int_0^a k(x)dx = \int_D^a -k(x)dx + \Delta\omega t + 2n\pi, \tag{7}$$

where $k(x)$ is the local wave number of the shear wave and $-k(x)$ indicates the wave propagation at the negative direction.

Now we take the derivative relative to a, on both sides of Eq. (7)

$$k(x) = -k(x) + \frac{d}{da}\Delta\omega t = -k(x) + \Delta\omega\frac{dt}{da}. \tag{8}$$

Note that da/dt is the apparent velocity of the antinode v_{pattern}

$$v_{\text{pattern}}(x) = \frac{\Delta\omega}{2k(x)} = \frac{\Delta\omega}{2\omega}v_{\text{shear}}(x). \tag{9}$$

This proves that the speed of the interference motion is directly proportional to the local velocity of the shear waves. The ratio of the apparent velocity to the local shear velocity is $\Delta\omega/2\omega$.

According to the equation above, there are clearly some advantages to investigating the interference patterns motions. First, this technique virtually slows down the shear wave propagation by a controllable factor $\Delta\omega/2\omega$. This enables available U.S. systems to visualize the wave propagation. Second, the local shear wave velocity can be recovered once the pattern motion is analyzed.

2. Two-dimensional problem

One of the challenges to generalize the previous derivation into two-dimensional (2D) domain is that in generalized media, according to Fermat's principle, waves do not necessarily travel along straight lines. Therefore, the phase at any particular field point is a complicated line integral over a curve. However, we can consider an infinitesimal region where the material property is approximately homogeneous and all waves can be approximated as plane waves (Greenleaf *et al.* 2003). A geometrical analysis in such infinitesimal regions is given in Appendix A. We find in the 2D case, in addition to the local shear speed, the crawling wave speed is also related to the angle between the wave fronts from each source

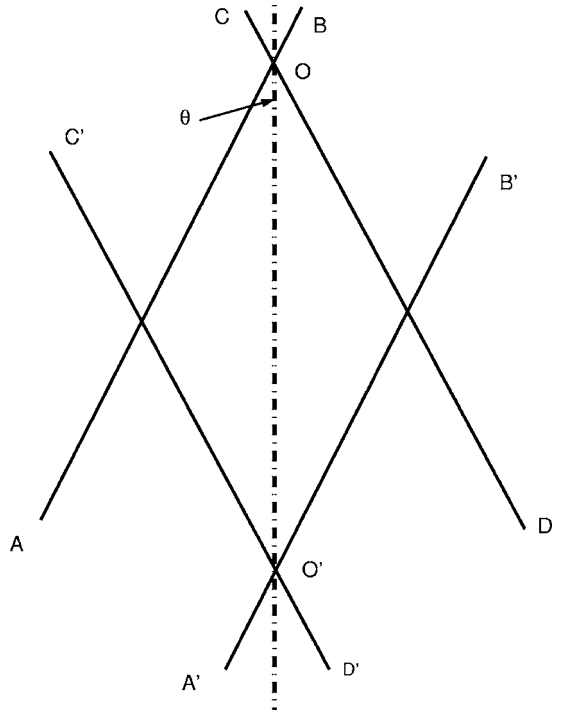


FIG. 3. The geometry of shear wave wave fronts in an infinitesimal region. The shear wave sources (not shown) are vibrating at the same frequency. At $t=0$, AB is the wave front from source 1 and CD is the wave front from source 2; at $t=T$, $A'B'$ is the wave front from source 1 and $C'D'$ is the wave front from source 2.

$$v_{\text{pattern}} = \frac{\Delta\omega}{2\omega}\cdot\frac{v_{\text{shear}}}{\cos\left(\frac{\theta}{2}\right)}, \tag{10}$$

where θ is defined in Fig. 3.

If the region of interest (ROI) is in the far field of both the two shear wave sources, $\theta/2$ is a small quantity, and $\cos(\frac{\theta}{2})$ is close to 1. The case where the lesion size is larger than the shear wave wavelength is out of the scope of this paper but discussed in detail by Ji and Mclaughlin (2004).

When $\cos(\frac{\theta}{2})$ is close to 1

$$v_{\text{pattern}} \approx \frac{\Delta\omega}{2\omega}\cdot v_{\text{shear}}. \tag{11}$$

E. Holographic wave

The technique of crawling waves requires two coherent shear wave sources from the opposing two sides of the region of interest. Sometimes, this particular configuration is not easy to achieve in practice. Meanwhile, as discussed previously, the interference fringes closely approximate the shear wave wave fronts only when the perturbation of the elasticity in the medium is small. The application of these techniques is thus limited. To overcome this drawback and to visualize the *exact* wave fronts of shear waves, another technique is proposed, which only requires one shear wave source touching the testing samples.

In this technique, the ultrasound probe, which is the observer and the frame of reference, is vibrated while the shear wave source transmits the waves. As Fig. 4 depicts, the shear

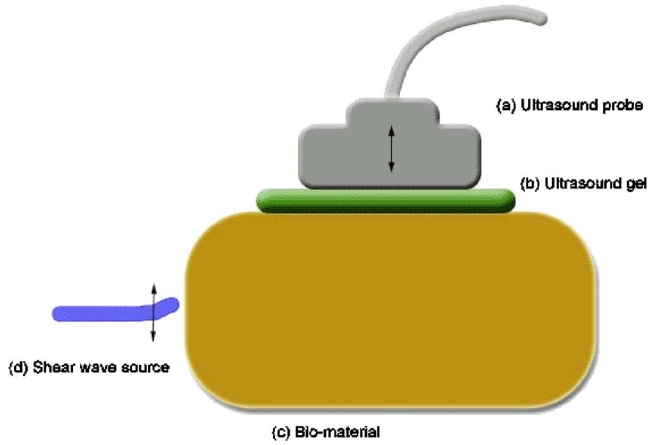


FIG. 4. Shear wave holography setup: (a) the ultrasound probe is externally vibrated at the frequency $\omega-\Delta\omega$, where $\Delta\omega \ll \omega$; (b) the thick layer of liquid ultrasound gel isolates the probe motion so it does not penetrate into the elastic material; (c) the elastic material; (d) the shear wave source vibrating at the frequency ω. The black arrows indicate the motion direction.

wave source vibrates at the frequency ω. Induced by the shear wave source, each particle in the medium oscillates at the same frequency with a spatially dependent phase term. The derivative of the phase term is the velocity of the shear wave. Assuming the shear wave source is stationary and harmonic, the vibration field can be written as $f(x)$ where $f(x)$ is a function of the 2D spatial variable x. In addition, we allow the spatial dependent part of the wave equation to be any arbitrary function of x subject to the wave equation. In other words, we do not require the wave to take any particular form (such as plain wave, spherical wave, etc.). Let the spatial dependent term be $s(x)$ [in the plain wave case, $s(x) = k \cdot x$], the vibration field can be written as

$$U(x,t) = f(x) * \exp\{i[\omega t - s(x)]\}. \tag{12}$$

The wave fronts of $U(x,t)$ are determined by assigning $[\omega t - s(x)]$ to a constant

$$\omega t - s(x) = \phi,$$

$$\omega t = s(x) - \phi,$$

where ϕ is a constant phase.

Taking the time derivatives on both sides

$$\omega = \text{grad}[s(x)] * \frac{dx}{dt}. \tag{13}$$

Multiplying both sides by $1/|\text{grad}\, s|$

$$\frac{\omega}{|\text{grad}(s)|} = \vec{n} * \frac{dx}{dt}, \tag{14}$$

where $\vec{n}$ is the unit vector normal to the wave front and it is along the direction of shear wave propagation. Therefore

$$v_s = \vec{n} * \frac{dx}{dt} = \frac{\omega}{|\text{grad}(s)|}, \tag{15}$$

where v_s is the shear wave phase velocity and $|\text{grad}(s)|$ is the gradient of $s(x)$.

While transmitting and receiving ultrasound signals, the ultrasound probe is also externally vibrated. The ultrasound probe is carefully positioned above the biomaterial without touching it. A thick layer of the ultrasound gel is applied to acoustically connect the probe and the biomaterial. In this setup, the ultrasound probe does *not* propagate shear waves into the biomaterial. Since the liquid ultrasound gel does not support shear waves, it isolates the ultrasound probe motion from penetrating into the medium.

Because the ultrasound probe is the observer and thus the frame of reference, the particle motion relative to the ultrasound probe is estimated by the sonoelastography algorithm. Therefore the *estimated* shear wave field is modulated by the probe motion. Instead of the shear wave source frequency ω the ultrasound probe is tuned to vibrate at a slight different frequency $\omega-\Delta\omega$, where $\Delta\omega \ll \omega$. Therefore, the motion of the ultrasound probe is

$$R(t) = A * \exp[i(\omega - \Delta\omega)t], \tag{16}$$

where A is a constant.

Hence the vibration field relative to the ultrasound probe is

$$P(x,t) = U(x,t) - R(t) = f(x) * \exp\{i[\omega t - s(x)]\} - A * \exp[i(\omega - \Delta\omega)t]. \tag{17}$$

The square of $P(x,t)$'s envelope is

$$\begin{aligned} |P(x,t)|^2 &= P * P^* = [U(x,t) - R(t)] * [U(x,t)^* - R(t)^*] \\ &= (f(x) * \exp\{i[\omega t - s(x)]\} - A * \exp[i(\omega \\ &\quad - \Delta\omega)t]) * (f(x) * \exp\{-i[\omega t - s(x)]\} \\ &\quad - A * \exp[-i(\omega - \Delta\omega)t]) = f(x)^2 + A^2 \\ &\quad - A \cdot f(x)\exp[i\Delta\omega t - is(x)] - A \cdot f(x) \\ &\quad \times \exp[-i\Delta\omega t + is(x)] = f(x)^2 + A^2 \\ &\quad - 2A \cdot f(x)\cos[\Delta\omega t - s(x)], \end{aligned} \tag{18}$$

where P^* is the complex conjugate of P.

We name $|P^2(x,t)|$ the holographic wave. Similar to Eq. (15), taking $\Delta\omega$ as the equivalent frequency and the velocity of the holographic wave is

$$v_m = \frac{\Delta\omega}{|\text{grad}(s)|}. \tag{19}$$

Comparing Eq. (19) with Eq. (15), it is clear that

$$v_m = \frac{\Delta\omega}{\omega} \cdot v_s. \tag{20}$$

Also notice that because the US probe motion is only a function of time, the mechanical modulation does not interfere with the spatial component of Eq. (12), i.e., $s(x)$. Therefore the exact shear wave appearance is preserved. Apart from a dc shift in the amplitude and a change in the velocity, the shear wave propagation is exactly represented by the holographic wave. With the proposed technique, the shear wave can be slowed down by an arbitrary yet controllable factor $\Delta\omega/\omega$. Therefore, with the mechanical modulation, the phase

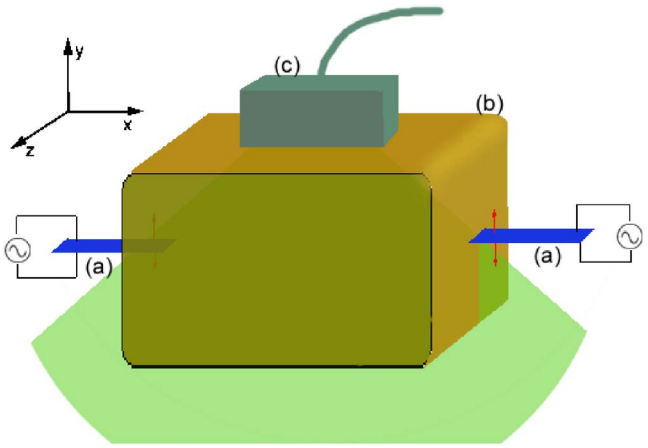


FIG. 5. Schematic drawing of the experiment setup. Two bimorphs (a) are in close contact with the phantom (b). The arrows indicate the motion vectors of the tips of the bimorphs. The sector shape depicts the imaging plane of the ultrasound probe (c).

of the shear waves is recovered and the speed is reduced to be studied by the ordinary US scanners with sonoelastography modifications.

F. Pattern motion speed reconstruction

The interference pattern motion velocity may be reconstructed with many existing wave motion inversion techniques such as discussed in Dutt *et al.* (1997), Catheline *et al.* (1999), Oliphant *et al.* (2000), Bishop *et al.* (2000), and Braun *et al.* (2001). In this article, the local spatial frequency estimator (LFE) is selected to demonstrate the feasibility to apply the crawling wave and holographic wave techniques as an elasticity imaging technique. The local spatial frequency is obtained by passing the images through a bank of 2D lognormal filters and averaging the outputs, as originally proposed by Knutsson *et al.* (1994) and modified by Manduca *et al.* (1996) to be employed in MRE estimations as a Helmholtz inversion technique. Note the local shear wave speed is inversely proportional to the local spatial frequency k_l: $v_{\text{crawl}}=\omega/k_l$. Recent advances in velocity estimations have been developed by Ji *et al.* (2003).

III. EXPERIMENTS AND RESULTS

In the validating experiments, two bending piezoelements known as bimorphs (Piezo Systems, Cambridge, MA) are applied as the vibration sources. A double channel signal generator (Tektronix AFG320) produces two monochrome low frequency signals at slightly different frequencies. A GE Logiq 700, which has been specially modified to implement the sonoelastography functions, is applied to visualize the "crawling wave" propagation. A schematic drawing of the experiment setup is depicted in Fig. 5.

A. Crawling wave experiments

Two experiments were conducted to validate the theory of crawling waves. First, a double-layer gelatin phantom was constructed. The hard layer of the phantom was made of 1000 g H_2O, 70 g gelatin (Knox™), 100 g glycerin, and 10 g graphite. The soft layer of the phantom was made of 1000 g H_2O, 49 g gelatin, 100 g glycerin, and 10 g graphite. The phantom was placed in such a position that the boundary

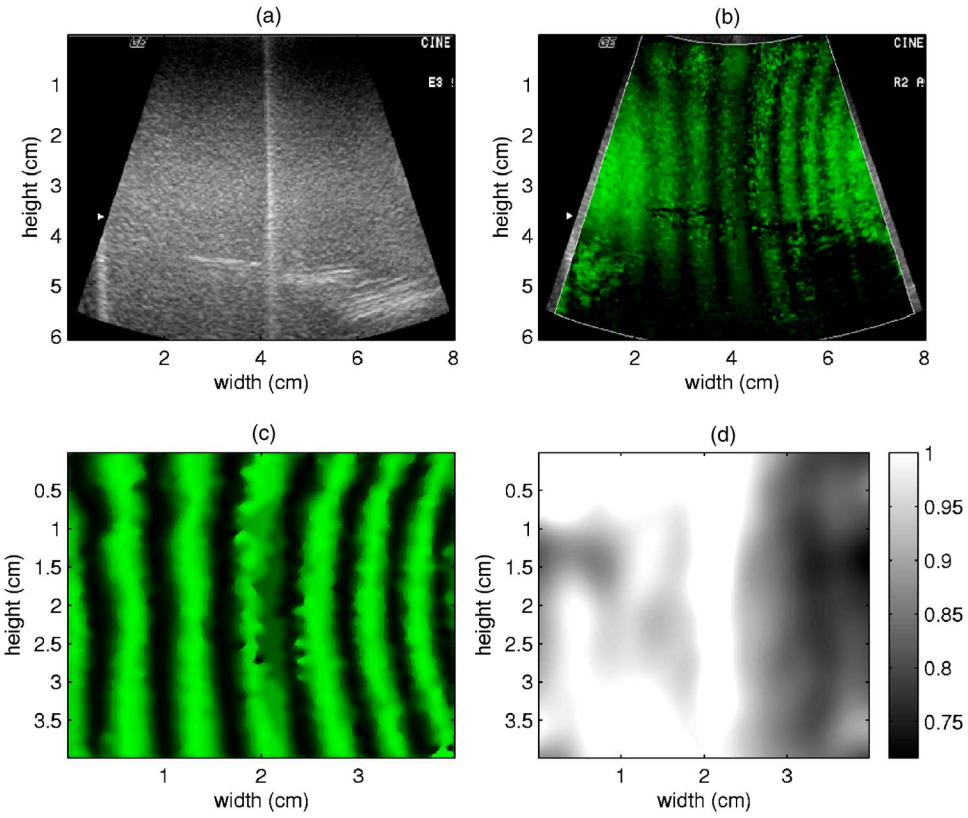


FIG. 6. Two-layer gelatin phantom experiment. (a) The B-mode ultrasound image of the phantom. The hard layer is on the left and the soft layer is on the right. (b) One frame of the crawling wave video taken with sonoelastography. The brightness of each pixel corresponds to the amplitude of the local vibration. Please note that the dc background is subtracted from the image for clarity. It can be seen that the crawling wave wavelength is larger in the left half of the phantom than that in the right half. (c) The cropped and enhanced sonoelastography image of the crawling waves with the amplitude information removed. (d) The normalized crawling wave speed estimation based on LFE algorithm.

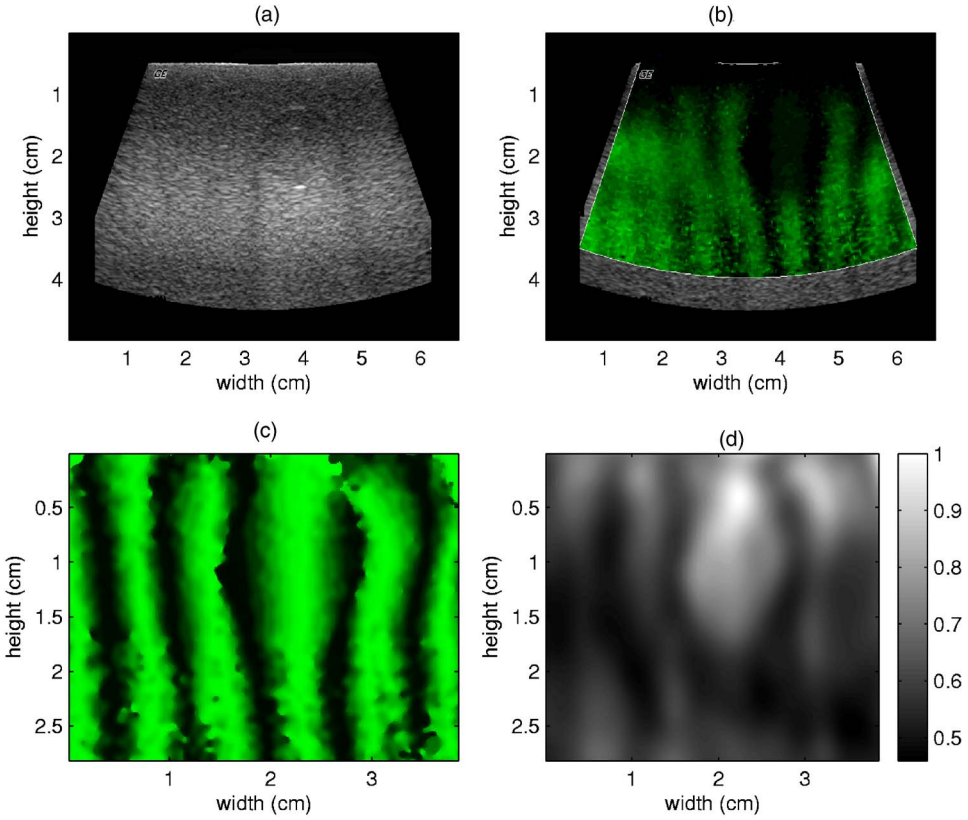


FIG. 7. Commercial (Zerdine) phantom experiment. (a) The B-mode ultrasound image of the phantom with a stiff inclusion. (b) One frame of the crawling wave video taken with sonoelastography. Please note that the dc background is subtracted from the image for clarity. It can be seen that the crawling wave wavelength is larger in the inclusion than outside. (c) The enhanced sonoelastography image with the amplitude information removed. (d) The normalized crawling wave speed estimation based on LFE algorithm.

of the two layers was vertical with the hard layer on the left and the soft layer on the right. The B-mode ultrasound image of the phantom is shown in Fig. 6(a).

The shear wave velocities of the hard and the soft layer of the phantom are estimated, respectively, with the method described in Wu *et al.* (2004). The estimated shear wave speed in the hard layer is 3.3 m/s. The estimated shear wave speed in the soft layer is 1.9 m/s.

The bimorphs are placed in close contact with the left side and the right side of the phantom. The two bimorphs vibrate at 300 and 300.2 Hz, respectively. The frequency difference $\Delta\omega$, namely, 0.2 Hz, was selected to achieve the best video quality with the Doppler frame rate (roughly seven frames per second). The shear wave interference pattern (one frame of the video) is shown in Fig. 6(b). It can be seen that the crawling wave wavelengths in the two layers are different. The wavelength in the hard half is longer and the wavelength in the soft half is shorter. The video sequence is cropped around the layer interface and enhanced by fitting the time trace at each pixel to a cosine curve [Fig. 6(c)]. The bank of LFE filters were applied on each of the 60 frames of the crawling wave video and the average of the outcome is shown in Fig. 6(d). The brightness of this image corresponds to the local crawling wave speed, which is inversely proportional to the local spatial frequency. The speed estimation was high (bright) in the hard layer and was low (dark) in the soft layer.

Moreover, a Zerdine tissue phantom (CIRS Norfolk, VA) is applied in the experiment. The phantom is bowl shaped and approximately $15\times15\times10$ cm in size. With the exception of a small spherical inclusion out of the imaging plane, the phantom is isotropic and homogeneous with uniform shear modulus. The tissue-mimicking material has a sound speed near 1540 m/sec. The background material has a Young's modulus of 20 KPa and a shear modulus of 6.67 KPa. The stiff spherical inclusion is 1.3 cm in diameter and approximately seven times as stiff as the background. A B-mode ultrasound image of the phantom is shown in Fig. 7(a). The inclusion is vaguely visible in the B-mode image. Two bimorphs are placed against the sides of the phantom, vibrating at 210 and 210.1 Hz, respectively. One frame of the crawling wave propagation video over the stiff lesion is shown in Fig. 7(b). The distortion of the crawling wave wave fronts around the stiff inclusion is visible. Because of the symmetry of the two sources, the distortions appear both before and after the lesion. The video is cropped around the lesion and enhanced by fitting the time trace at each pixel to a cosine curve Fig. 7(c). By applying the LFE upon each frame of the propagation video and taking the average of the outputs, we have an estimation of the crawling wave velocity distribution in the phantom, shown in Fig. 7(d). The brightness of Fig. 7(d) corresponds to the local shear wave speed, which is inversely proportional to the local spatial frequency. The region with high crawling wave speed agrees with the location of the stiff inclusion.

B. Holographic wave experiments

Similar experiments are performed with the holographic wave technique. One channel of the signals (199.9 Hz) drives a bending piezo elements known as Thunder (Face International Corporation, Norfolk, VA) which is applied to vibrate the US transducer. The other channel of the signal (200 Hz) drives a shear wave actuator (Piezo system, MA),

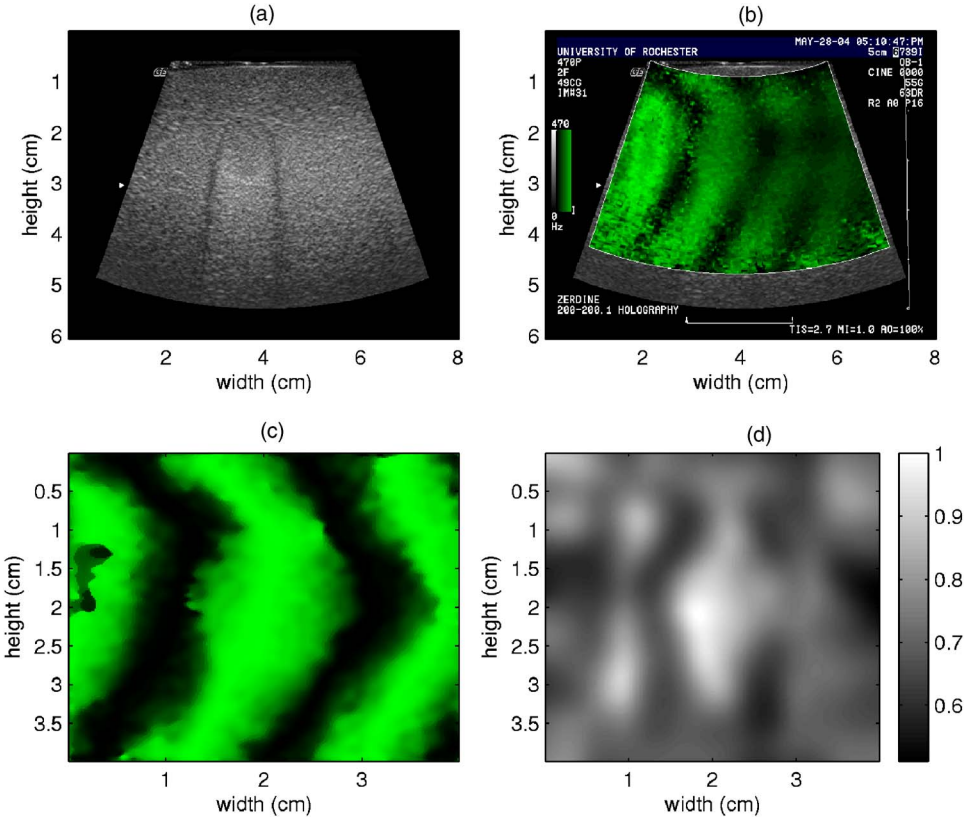


FIG. 8. (a) The B-mode ultrasound image of the phantom with a stiff inclusion. (b) One frame of the holography wave video taken with sonoelastography. It can be seen that the holography wave wavelength is larger in the inclusion than outside. (c) The enhanced sonoelastography image with the amplitude information removed. (d) The normalized holography wave speed estimation based on LFE algorithm.

which propagates shear waves into a Zerdine phantom. With the real-time visualization, the shear waves are virtually "slowed down" so that the local and subtle behaviors of the waves can be examined closely. The different wave speeds within and outside of the lesion can be perceived by human eye. One frame of the "modulation wave" propagation is shown in Fig. 8(b). The shear wave wave fronts are visibly distorted by the hard inclusion and thus the size and the location of the lesion may be estimated. One estimation of the stiffness distribution with LFE is shown in Fig. 8(d).

IV. DISCUSSION

In order for Eq. (10) ($v_{\text{pattern}}=\Delta\omega/2\omega\cdot v_{\text{shear}}/\cos(\theta/2)$) to be valid, we have to assume the medium is locally homogeneous, in other words, if there is an abrupt changes in the shear modulus, Eq. (A2) is not valid. Near the boundary of such abrupt changes, the assumption that $\cos(\theta)\approx 1$ in Eq. (A2) may not be valid either (The holographic wave does not require this condition). Therefore, the shear wave speed estimation close to the media boundary is not exact. A reasonable approximation of the size of the transition zone near the boundary is on the order of the wavelength of the crawling waves. Therefore, increasing the frequency of the crawling wave, thus reducing the wavelength, may increase the reconstruction resolution of the shear wave velocity. In particular, if the size of the lesion is less than the crawling wave wavelength, the shear wave speed in the whole lesion may be misestimated, thus the reconstructed shear wave speed may not exactly reflect the physical stiffness contrast. Also, the LFE filters' region of support plays a role in setting the lower limit of resolution. However, even in this case, the proposed technique is able to qualitatively indicate the location of the stiff region with reasonable estimation of its size and shape.

The theory of holographic wave requires fewer assumptions. The only major assumption is that the medium is linear so that no higher harmonics other than the input signal exist. However, due to the reduced signal strength beyond the obstacles, the holographic wave method provides less accurate estimation than the crawling wave method in the far field.

The accurate shear wave estimates partly depend on the signal strength. Estimation errors and imaging artifacts may occur in deep tissues or where ultrasound shadows occur. In such locations, sonoelastographic signals are too weak to present correct spatial frequency and may thus be estimated as high crawling/holographic speed regions. Similarly, regions close to the shear wave sources may saturate the sonoelastographic estimator and thus also present low spatial frequency.

Like many other imaging modalities, there exists a tradeoff between the estimation accuracy and the image resolution of either the crawling wave or the holographic wave reconstructions. The relation is formulated in Appendix B with a realistic example provided. This tradeoff exists because of the noisy nature of the signals and the finite size of US foci. Please note that this appendix intends to find the lower (finer) bound of the resolution without considering any particular estimator. This lower bound may not be achievable.

V. CONCLUSION

We developed two experimental procedures to induce and visualize the shear wave interference patterns in tissue

mimicking soft materials in real time. The interference patterns caused by two shear wave sources or one wave source and vibrating U.S. transducer move at a certain speed, which is proved to be related to the local shear wave velocity. The local interference pattern speeds are estimated in both a two-layer gelatin phantom and a commercial phantom with a stiff inclusion. The shear wave velocity and thus the shear modulus of each phantom are therefore reconstructed off line. This technique provides a real-time visualization of crawling waves with quantitative assessment of local elastic properties.

ACKNOWLEDGMENTS

The authors are thankful to Professor Nicholas George for informative discussion. This work was supported in part by the NSF/NYS Center for Electronic Imaging Systems, NIH Grant No. 2 R01 AG16317-01A1, the University of Rochester Departments of Radiology and Electrical and Computer Engineering, and the General Electric Company (GE).

APPENDIX A: INTERFERENCE FRINGES IN 2-D SPACE

As drawn in Fig. 3, the wave front AB from the left source and the wave front CD from the right source intersects at O. For the sake of convenience, the interference pattern is redefined as *the set of intersection points of AB and CD as they evolve in time.*

Theorem 1. OP is the interior angle bisector of angle $\angle$AOD.

Proof. Suppose wave front AB and wave front CD intersect at O at time 0. After a short period of time, AB moves to $A'B'$ and CD moves to $C'D'$ and they intersect at O'. OO' is the local segment of the interference pattern, as drawn in Fig. 3. Since AB and CD move at the same speed $EO=OF=FO'=O'E$. Therefore, $\triangle OEO'$ and $\triangle OFO'$ are congruent. Hence, OO' divides $\angle AOD$ in equal halves. Since the source frequency difference does not change the orientation of the wave fronts nor the interference pattern at a particular location, this relation is valid for both the static interference pattern case, and the crawling wave case.

In Fig. 10, suppose the wave fronts are at AB and CD, respectively, at $t=0$ and the interference fringe is at OP. Then, after a small period of time, at $t=\Delta T$, AB advances to $A'B'$ with $k_l \cdot d=\omega \cdot \Delta T$. Should CD have the same frequency, it would advance to the dashed line $C''D''$. However, since CD has a slight different frequency, it advances to $C'D'$ instead with

$$d' \cdot k_l = \omega \cdot \Delta T + \Delta\omega \cdot \Delta T, \tag{A1}$$

where k_l is the local wave number of shear wave, and ΔT is an infinitesimal time interval.

It is easy to prove that without the frequency difference, $A'B'$ and $C''D''$ produce the same interference fringe OP. The frequency difference $\Delta\omega$, however, causes a fringe displacement from OP to $O'P'$.

The distance between CD and $C''D''$ is $d'-d=\Delta\omega \cdot \Delta t/k_l$. Let $\angle O'OD''$ be θ, then $|OO'|=(d'-d)/\sin(\theta)$.

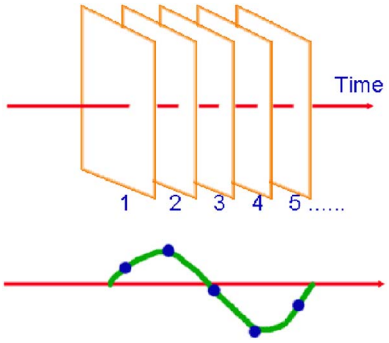


FIG. 9. Phase of the time trace of the pixel value of the hologram video is estimated.

The distance between OP and $O'P'$ is $d''=|OO'|*\sin(\theta/2)$, according to Theorem 1. So

$$d'' = \left|OO'\right| * \sin(\theta/2) = \frac{d'-d}{\sin(\theta)} * \sin(\theta/2) = \frac{d'-d}{2\cos\left(\frac{\theta}{2}\right)}$$

$$= \frac{\Delta\omega \cdot \Delta T}{2\cos\left(\frac{\theta}{2}\right)k} = \frac{\Delta\omega \cdot \Delta T \cdot v_{\text{shear}}}{2\omega\cos\left(\frac{\theta}{2}\right)}$$

$$\frac{d''}{\Delta T} = \frac{\Delta\omega \cdot v_{\text{shear}}}{2\omega\cos\left(\frac{\theta}{2}\right)}$$

$$v_{\text{pattern}} = \frac{\Delta\omega}{2\omega} \cdot \frac{v_{\text{shear}}}{\cos\left(\frac{\theta}{2}\right)}. \tag{A2}$$

Comparing Eq. (A2) with Eq. (9), we see that in the 2D domain, an extra factor of $1/\cos(\frac{\theta}{2})$ is introduced. Other than this factor, the velocity of the interference pattern is solely dependent on the local shear wave velocity and the frequencies relation. We further notice that if the two shear wave sources are far separated comparing to the size of the ROI, $\theta/2$ is a small quantity and $\cos(\frac{\theta}{2})$ is close to 1, thus

$$v_{\text{pattern}} \approx \frac{\Delta\omega}{2\omega} \cdot v_{\text{shear}}. \tag{A3}$$

APPENDIX B: ESTIMATOR ACCURACY CONSIDERATIONS

In the proceeding shear wave velocity estimation procedures, it is obvious that the final estimation results rely extensively upon the phase estimation of the local vibration. The local vibration phase is estimated by tracking the brightness variation at each pixel as shown in Fig. 9. In this appendix, the lower bound of the crawling/holographic wave velocity error is formulated and an example with realistic values is given.

According to Eq. (18), the modulation wave equation is

$$|P(x,t)|^2 = f(x)^2 + A^2 - 2A \cdot f(x)\cos[\Delta\omega t - s(x)].$$

At a given location x_o, the pixel value is

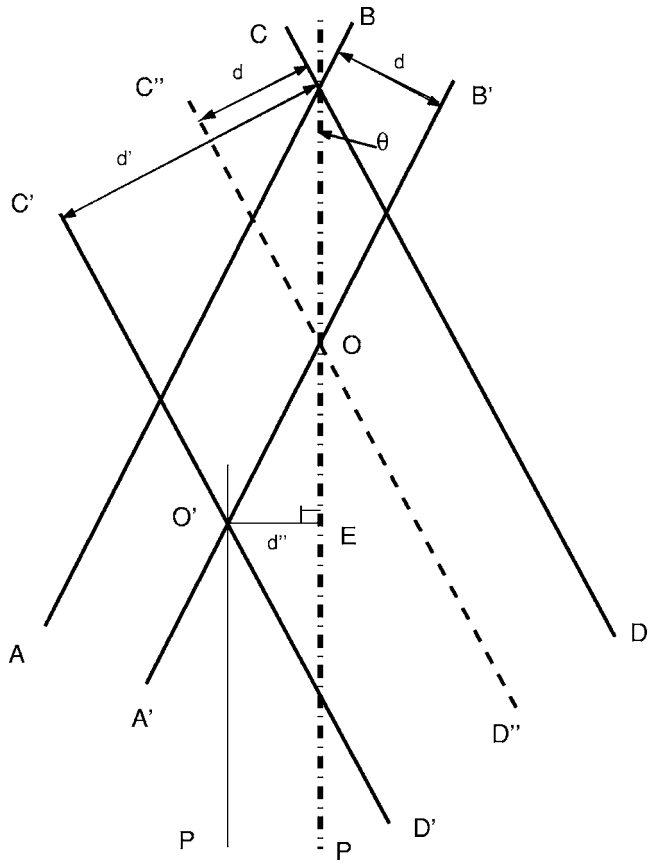


FIG. 10. The geometry of shear wave wavefronts in an infinitesimal region. The shear wave sources (not shown) are vibrating at slightly different frequencies.

$$B(t) = f(x_o)^2 + A^2 - 2A \cdot f(x_o)\cos[\Delta\omega t - s(x_o)]. \tag{B1}$$

Assume the signal is in white Gaussian noise, the discrete pixel value over multiple observations (multiple frames of the wave video) can be written as

$$x[n] = C + D \cdot \cos(\Delta\omega n + \phi) + w[n], \tag{B2}$$

where $C=f(x_o)^2+A^2$, $D=-2A\cdot f(x)$, $\phi=-s(x_o)$, and $w(n)=\mathcal{N}(0,\sigma^2)$, a zero mean Gaussian distribution with standard deviation σ.

Therefore, the likelihood function is

$$p(x;\phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \cdot \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1} \{x[n] - C - D \cdot \cos(\Delta\omega n + \phi)\} \right]. \tag{B3}$$

Taking the first and second derivatives of the natural logarithm of the likelihood function yields

$$\frac{\partial p(x;\phi)}{\partial \phi} = -\frac{1}{\sigma^2} \cdot \sum_{n=0}^{N-1} [x[n] - C - D \cdot \cos(\Delta\omega n + \phi)] \cdot D \sin(\Delta\omega n + \phi) \tag{B4}$$

and

$$\frac{\partial^2 p(x;\phi)}{\partial \phi^2} = -\frac{D}{\sigma^2} \cdot \sum_{n=0}^{N-1} [(x[n] - C) \cdot \cos(\Delta\omega n + \phi) - D\cos(2\Delta\omega n + 2\phi)]. \tag{B5}$$

Taking the negative expected value, we have

$$\begin{aligned} -E\left[\frac{\partial^2 p(x;\phi)}{\partial \phi^2}\right] &= \frac{D}{\sigma^2} \cdot \sum_{n=0}^{N-1} [(x[n] - C) \cdot \cos(\Delta\omega n + \phi) - D\cos(2\Delta\omega n + 2\phi)] \\ &= \frac{D}{\sigma^2} \cdot \sum_{n=0}^{N-1} [D\cos^2(\Delta\omega n + \phi) - D\cos(2\Delta\omega n + 2\phi)] = \frac{D^2}{\sigma^2} \cdot \sum_{n=0}^{N-1} \left[\frac{1}{2} + \frac{1}{2}\cos(2\Delta\omega n + 2\phi) - \cos(2\Delta\omega n + 2\phi) \right]. \end{aligned}$$

If we acquire integer or half integer number of cycles in experiments by choosing $\Delta\omega N = m\pi$, m being an integer, the expected value of the cos term is zero

$$E[\cos(2\Delta\omega n + 2\phi)] = 0. \tag{B6}$$

Thus,

$$-E\left[\frac{\partial^2 p(x;\phi)}{\partial \phi^2}\right] = \frac{ND^2}{2\sigma^2}. \tag{B7}$$

We notice that the inverse of Eq. (B7) gives the Cramer-Rao lower bound of the phase estimation

$$\mathrm{var}(\hat{\phi}) \geq \frac{1}{-E\left[\frac{\partial^2 p(x;\phi)}{\partial \phi^2}\right]} = \frac{2\sigma^2}{ND^2} \tag{B8}$$

The local shear wave velocity estimation is equivalent to estimating the local slope of the phase function. At this stage, the tradeoff of image resolution and estimation accuracy has to be considered. If we set the image resolution to be the size of M pixels, the accuracy of the slope estimation is bounded by a function of M. If we model the problem as a line fitting problem, and assume the phase function is in the form of

TABLE I. Summary of the Major parameters to produce the point spread function simulation.

Parameter	Value
Sampling frequency	150e6 Hz
Speed of sound	1540 m/s
Central frequency	7.5e6 Hz
Relative Bandwidth	30%
Number of element	128
F number	3

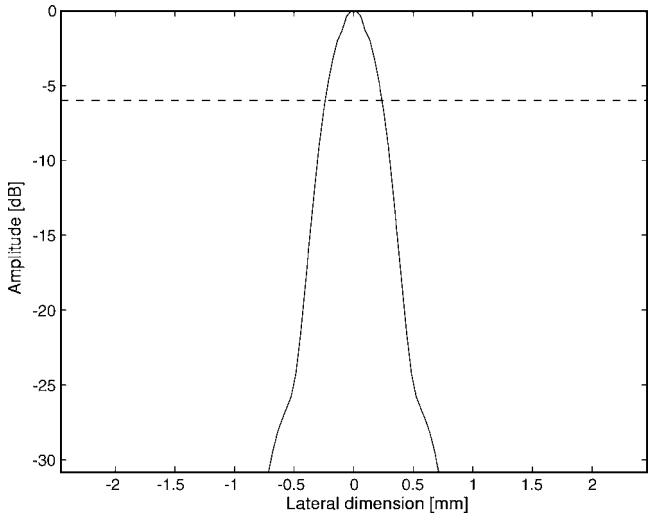


FIG. 11. 6*dB* width of the ultrasound scanner's point spread function simulated with Field II tool box.

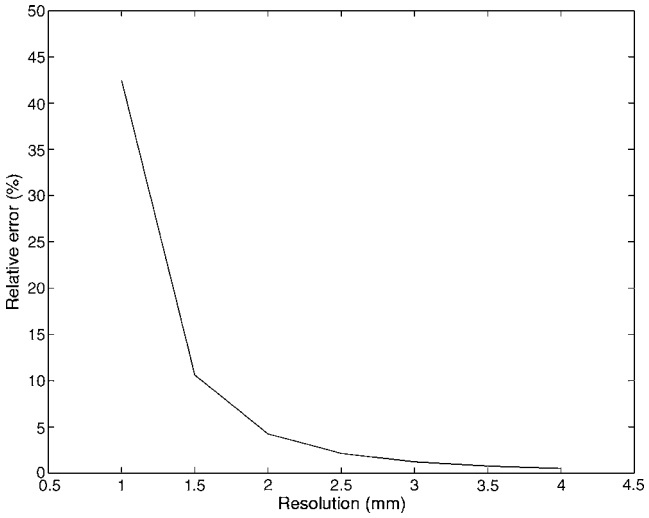


FIG. 12. The tradeoff between the image resolution and the relative error.

$$\phi[m] = F + G \cdot m + w[m], \tag{B9}$$

where $w[m]$ is a zero mean Gaussian distribution with variance determined by Eq. (B8). With the independent observations at these M pixels, we may obtain the slope estimation G with variance

$$\mathrm{var}(\hat{G}) \geq \frac{12 \cdot \mathrm{var}(\hat{\phi})}{M(M^2-1)}. \tag{B10}$$

Because the stiff regions are generally of more importance than the normal background, we pay more attention to the estimator accuracy in the stiff regions. The vibration amplitude is low in these regions due to the sonoelastography effect, the signal-to-noise ratio (SNR) is thus also low. An empirical estimate of the SNR in the stiff regions is 1. In our experiments, a typical number of frames of the shear wave propagation video is 60. Thus in Eq. (B8), the variance of the phase estimation is approximately 1/30.

M in Eq. (B10) refers to the number of independent measurements. The ultrasound scanner determines that only one independent measurement can be made within the width of the point spread function. A point spread function is simulated with the Field II tool box Jensen and Svendsen 1992. The imaging system parameters are selected from a typical experiment setting and are summarized in Table I.

The simulation shows that the 6 dB width of the point spread function in the lateral direction is approximately 0.5 mm (Fig. 11). Assuming a realistic shear wave speed of 4 m/s and a driving frequency at 200 Hz, we proceed to discuss the relation between the elasticity estimation resolution and the estimation relative error.

Assume that we choose the resolution to be 2 mm, then there are four independent measurements within this length. According to Eq. (B10)

$$\mathrm{var}(\hat{G}) \geq \frac{12 \cdot \mathrm{var}(\hat{\phi})}{M(M^2-1)} = \frac{12 \cdot 1/30}{4(4^2-1)} = 0.0067. \tag{B11}$$

Since the phase increase is 2π over one shear wave wavelength, the phase slope may be estimated by

$$\mathrm{slope}_{\phi} = 2\pi/\lambda = 2\pi/(20 * 2) = 0.1571. \tag{B12}$$

Thus the relative error is 4%.

The tradeoff between the elasticity image resolution and the estimation relative error is plotted in Fig. 12. Please note that Eq. (B10) provides a lower bound of the estimation accuracy. In practice, this lower bound may not be achievable.

Braun, J., Buntkowsky, G., Bernarding, J., Tolxdorff, T., and Sack, I. (**2001**). "Simulation and analysis of magnetic resonance elastography wave images using coupled harmonic oscillators and gaussian local frequency estimation," Magn. Reson. Imaging **19**, 703–713.

Bishop, J., Samani, A., Sciarretta, J., and Plewes, D. B. (**2000**). "Two-dimensional mr elastography with linear inversion reconstruction: methodology and noise analysis," Phys. Med. Biol. **45**, 2081–2091.

Catheline, S., Wu, F., and Fink, M. (**1999**). "A solution to diffraction biases in sonoelasticity: the acoustic impulse technique," J. Acoust. Soc. Am. **105**, 2941–2950.

Dutt, V., Kinnick, R. R., Muthupillai, R., Oliphant, T. E., Ehman, R., and Greenleaf, J. F. (**2000**). "Acoustic shear-wave imaging using echo ultrasound compared to magnetic resonance elastography," Ultrasound Med. Biol. **26**, 397–403.

Dutt, V., Manduca, A., Muthupillai, R., Ehman, R., and Greenleaf, J. (**1997**). "Inverse approach to elasticity reconstruction in shear wave imaging," *Proceeding of IEEE Ultrasonics Symposium* Vol. **2**, pp. 1415–1418.

Greenleaf, J. F., Fatemi, M., and Insana, M. (**2003**). "Selected methods for imaging elastic properties of biological tissues," Annu. Rev. Biomed. Eng. **5**, 57–78.

Huang, S. R., Lerner, R. M., and Parker, K. J. (**1990**). "On estimating the amplitude of harmonic vibration from the doppler spectrum of reflected signals," J. Acoust. Soc. Am. **88**, 2702–2712.

Huang, S. R., Lerner, R. M., and Parker, K. J. (**1992**). "Time domain Doppler estimators of the amplitude of vibrating targets," J. Acoust. Soc. Am. **91**, 965–974.

Jenkyn, T. R., and Ehman, R. L., "An KN 2003 Noninvasive muscle tension measurement using the novel technique of magnetic resonance elastography (MRE)," J. Biomech. **36**, 1917–1921.

Jensen, J. A., and Svendsen, N. B. (**1992**). "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267.

Ji, L., and McLaughlin, J. (**2004**). "Recovery of the lame parameter mu in biological tissues," Inverse Probl. **20**, 1–24.

Ji, L., McLaughlin, J. R., Renzi, D., and Yoon, J. R. (**2003**). "Interior elastodynamics inverse problems: Shear wave speed reconstruction in transient elastography," Inverse Probl. **19**, 1–29.

Knutsson, H., Westin, C. J., Granlund, G. (**1994**). "Local multiscale frequency and bandwidth estimation," *Proc. IEEE Intl Conf on Image Processing*, Vol. **1**, pp. 36–40.

Lerner, R. M., Parker, K. J., Holen, J., Gramiak, R., and Waag, R. C. (**1988**).

"Sono-elasticity: Medical elasticity images derived from ultrasound signals in mechanically vibrated targets," Acoustical Imaging. Vol. **16**. *Proceedings of The 16th International Symposium*, pp. 317–327.

Levinson, S. F., Shinagawa, M., and Sata, T. (**1995**). "Sonoelastic determination of human skeletal-muscle elasticity," J. Biomech. **28**, 1145–1154.

Love, A. E. H. (**1944**). *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York), Chap. 13.

Manduca, A., Dutt, V., Borup, D. T., Muthupillai, R., Ehman, R. L., and Greenleaf, J. F. (**1998**). "Reconstruction of elasticity and attenuation maps in shear wave imaging: An inverse approach," *Medical Image Computing and Computer-assisted Intervention-Miccai'98*, pp. 606–613.

Manduca, A., Muthupillai, R., Rossman, P. J., Greenleaf, J. F., Ehman, R. L. (**1996**). "Image processing for magnetic resonance elastography," Proc. SPIE **2710**, 616–623.

Manduca, A., Oliphant, T. E., Dresner, M. A., Mahowald, J. L., Kruse, S. A., Amromin, E., Felmlee, J. P., Greenleaf, J. F., and Ehma, R. L., (**2001**). "Medical image analysis: Non-invasive mapping of tissue elasticity," Antivir. Chem. Chemother. **5**, 237–254.

Muthupillai, R., Lomas, D. J., Rossman, P. J., Greenleaf, J. F., Manduca, A., and Ehman, R. L. (**1995**). "Magnetic-resonance elastography by direct visualization of propagating acoustic strain waves," Science **269**, 1854–1857.

Oliphant, T. E., Kinnick, R. R., Manduca, A., Ehman, R. L., Greenleaf, J. F. (**2000**). "An error analysis of helmholtz inversion for incompressible shear, vibration elastography with application to filter-design for tissue characterization," *Ultrasonics Symposium, 2000 IEEE*, Vol. **2**, pp. 1795–1798.

Parker, K. J., Huang, S. R., Musulin, R. A., and Lerner, R. M. (**1990**). "Tissue-response to mechanical vibrations for Sonoelasticity imaging," Ultrasound Med. Biol. **16** (3) 241–246.

Parker, K. J., and Lerner, R. (**1992**). "Shear waves ring a bell," J. Ultrasound Med. **11**, 387–392.

Sandrin, L., Catheline, S., Tanter, M., Hennequin, X., and Fink, M. (**1999**). "Time-resolved pulsed elastography with ultrafast ultrasonic imaging," Ultrason. Imaging **21**, 259–272.

Taylor, L. S., Porter, B. C., Rubens, D. J., and Parker, K. J. (**2000**). "Three-dimensional sonoelastography: Principles and practices," Phys. Med. Biol. **45**, 1477–1494.

Wu, Z., Lawrance, T. S., Rubens, D. J., and Parker, K. J. (**2004**). "Sonoelastographic imaging of interference patterns for estimation of the shear velocity of homogeneous biomaterials," Phys. Med. Biol. **49**, 911–922.

Wu, Z., Taylor, L. S., Rubens, D. J., Parker, K. J. (**2002**). "Shear wave focusing for three-dimensional sonoelastography," J. Acoust. Soc. Am. **111**, 439–446.

Yamakoshi, Y., and Sato, J., and Sato, T. (**1990**). "Ultrasonic imaging of internal vibrations of soft tissue under forced vibration," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 45–53.

Impact of local attenuation approximations when estimating correlation length from backscattered ultrasound echoes

Timothy A. Bigelow[a)]
Department of Electrical Engineering, University of North Dakota, Box 7165, Grand Forks, North Dakota 58202

William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana—Champaign, 405 N. Mathews, Urbana, Illinois 61801

(Received 30 November 2005; revised 21 April 2006; accepted 7 May 2006)

Estimating the characteristic correlation length of tissue microstructure from the backscattered power spectrum could improve the diagnostic capability of medical ultrasound. Previously, size estimates were obtained after compensating for source focusing, the frequency-dependent attenuation along the propagation path (total attenuation), and the frequency-dependent attenuation in the scattering region (local attenuation). In this study, the impact of approximations of the local attenuation on the scatterer size estimate was determined using computer simulations and theoretical analysis. The simulations used Gaussian impedance distributions with an effective radius of 25 μm randomly positioned in a homogeneous half-space sonified by a spherically focused source ($f/1$ to $f/4$). The approximations of the local attenuation that were assessed neglected local attenuation (i.e., assume 0 dB/cm-MHz) neglected frequency dependence of the local attenuation, and assumed a finite frequency dependence (i.e., 0.5 dB/cm-MHz) independent of the true attenuation of the medium. Errors in the scatterer size estimate due to the local attenuation approximations increased with increasing window length, increasing true local attenuation and increasing f number. The most robust estimates were obtained when the local attenuation was approximated by a tissue-independent attenuation value that was greater than 70% of the largest attenuation expected in the tissue region of interest. [DOI: 10.1121/1.2208456]

PACS number(s): 43.80.Vj, 43.80.Ev, 43.80.Qf [FD] Pages: 546–553

LIST OF SYMBOLS

a_{eff} = effective radius, or correlation length, of scatterer
A_{comp} = generalized attenuation-compensation function including focusing effects along the beam axis
$a_{\text{eff}\,j}$ = estimated effective radius of scatterer found from one set (i.e., 25 averaged rf echoes) of simulated backscatter waveforms
$\bar{a}_{\text{eff}}$ = mean value of estimated effective radius from all sets of backscattered waveforms (i.e., $\bar{a}_{\text{eff}}=\Sigma_{\forall j} a_{\text{eff}\,j}/\Sigma_{\forall j} j$)
ASD = average squared difference term minimized when solving for a_{eff}
$E[\,]$ = expected value with respect to scattering random process
f = frequency
$F_\gamma(\omega, a_{\text{eff}})$ = form factor related to the scatterer geometry and effective radius
g_{win} = windowing function used to gate the time domain waveforms
k = wave number in tissue
L = equivalent length (in mm) of the windowing function used to gate the time domain waveforms
V_{plane} = backscattered voltage spectrum from rigid plane placed at the focal plane
w_z = equivalent Gaussian depth of focus of velocity potential field in focal region
V_{scat} = backscattered voltage spectrum from tissue containing scatterers
$X, \bar{X}$ = terms used in minimization scheme when solving for a_{eff}
z_T = distance from aperture plane to focal plane of ultrasound source
α_{high} = largest value of tissue attenuation expected
α_{loc} = frequency-dependent local attenuation in the scattering region of interest
α_{low} = smallest value of tissue attenuation expected
α_{ref} = frequency-dependent attenuation value selected for the local attenuation that is independent of the true value of α_{loc}
α_{tot} = frequency-dependent total attenuation along the propagation path for all tissue between focal plane and aperture plane
$\delta\alpha_{\text{loc}}$ = error between true α_{loc} value of the tissue and assumed α_{loc} value used in A_{comp} due to approximations in the value of α_{loc}

[a)]Electronic mail: timothybigelow@mail.und.nodak.edu

$\sigma_{a_{\text{lower}}}$ = percent deviation in values of scatterer effective radius for sizes smaller than the mean size (i.e., $a_{\text{eff}\,j} < \bar{a}_{\text{eff}}$)

$\sigma_{a_{\text{upper}}}$ = percent deviation in values of scatterer effective radius for sizes larger than the mean size (i.e., $a_{\text{eff}\,j} > \bar{a}_{\text{eff}}$)

ω = radian frequency

I. INTRODUCTION

Quantifying the underlying tissue structure using the information contained in the frequency spectrum of backscattered ultrasound echoes has been extensively studied (Chivers and Hill, 1975; Hall *et al.*, 1996; Insana *et al.*, 1990; Lizzi *et al.*, 1983; Nassiri and Hill, 1986; Oelze *et al.*, 2004). Typically, the spectra of the backscattered echoes are compared to a reference spectrum in order to estimate the characteristic correlation length (scatterer size) of the scatterers in the tissue while assuming a form factor that describes the correlation function. Once the correlation length has been determined, other parameters such as acoustic concentration (i.e., scatterer number density times their relative impedance change squared) can be estimated. Before the scatterer size can be determined, the backscattered spectra need to be compensated for frequency-dependent attenuation along the propagation path (total attenuation), frequency-dependent attenuation in the scattering region of interest (local attenuation), and focusing (Bigelow and O'Brien, 2004b).

Earlier publications (Bigelow and O'Brien, 2004a, b) demonstrated that the impact of attenuation and focusing can be compensated if the values of attenuation and effective Gaussian depth of focus are known. Although it may be possible to measure the effective Gaussian depth of focus for a source *a priori*, the attenuation varies drastically between patients even for the same tissue type. For example, many different attenuation coefficients for fat taken at different frequencies have been reported (Goss *et al.*, 1978). In particular, one study (Dussik and Fritch, 1956) gives attenuation values of 0.6 ± 0.2 dB/cm at 1 MHz and 2.3 ± 0.7 dB/cm at 5 MHz. Hence, for a change in frequency of 4 MHz, the change in attenuation varied from 0.8 to 2.6 dB/cm [i.e., $(2.3-0.7)$ dB/cm minus $(0.6+0.2)$ and $(2.3+0.7)$ dB/cm minus $(0.6-0.2)$ dB/cm], yielding attenuation slopes between 0.2 and 0.65 dB/cm-MHz. For this reason, an algorithm termed the Spectral Fit algorithm that estimated scatterer size and total attenuation simultaneously from the backscattered spectrum was developed (Bigelow *et al.*, 2005). The algorithm initially assumed weakly focused sources and utilized small window lengths to gate the time-domain echoes in order to reduce the importance of local attenuation and focusing. As a result, neither local attenuation nor focusing were included in the algorithm.

Before expanding the Spectral Fit algorithm to include focusing and local attenuation, the effect of different methods for correcting for these parameters when estimating the scatterer size needs to be determined. The impact of different corrections for focusing when the attenuation is known has already been quantified (Bigelow and O'Brien, 2004a, b). In this paper, the impact of different approximations of local attenuation on the scatterer size estimate is quantified using computer simulations and theoretical derivations. In the simulations, the total attenuation and effective Gaussian depth of focus are known, and three different approximations are used for the local attenuation value. First, local attenuation was completely neglected by assuming the local attenuation was zero when solving for the scatterer size. Second, the frequency dependence of the attenuation was neglected by assuming the local attenuation was the mean value of the attenuation over the frequency range of interest. Third, the local attenuation was assumed to have some finite frequency dependence (i.e., 0.5 dB/cm-MHz) independent of the true attenuation of the medium. The impact of each approximation of the local attenuation was then assessed by comparing the resulting scatterer size estimates for each against the size estimates that were obtained using the correct value of local attenuation for the same attenuation and degree of focusing.

Of the three estimates of attenuation considered in the analysis, the third case is of special interest because it is equivalent to the reference phantom technique that has also been implemented to correct for focusing (Gerig *et al.*, 2003). The reference phantom technique obtains time-gated signals about the focus from a phantom for use as a reference spectrum. All changes in the ultrasound field relevant to the focal region, including local attenuation, are included in the reference spectrum. Therefore, the technique is equivalent to assuming a specific frequency-dependent local attenuation equal to the attenuation of the phantom for all biological tissues independent of the true attenuation for the tissue. However, when using the reference phantom technique it is still critical to correct for total attenuation on a tissue-specific basis.

II. SIMULATION SETUP

A. Review of scatterer size estimation for focused sources

The scatterer radius, a_{eff}, is related to the backscattered power spectrum by (Bigelow and O'Brien, 2004b)

$$E[|V_{\text{scat}}|^2] \propto \frac{|k|^4 |V_{\text{plane}}(\omega)|^2}{A_{\text{comp}}(\omega)} F_\gamma(\omega, a_{\text{eff}}). \tag{1}$$

$V_{\text{plane}}(\omega)$ is the voltage spectrum that would be returned from a rigid plane placed at the focal plane in a water bath and is obtained independently to calibrate the echoes from the tissue. $F_\gamma(\omega, a_{\text{eff}})$ is the form factor describing the correlation function for the tissue, and the functional form of the form factor must be assumed before the scatterer size can be obtained. $A_{\text{comp}}(\omega)$ is a generalized attenuation-compensation function that corrects for focusing, local attenuation, and total attenuation and is given by (Bigelow and O'Brien, 2004b)

$$A_{\text{comp}} = \frac{e^{4\alpha_{\text{tot}} z_T}}{\int_{-L/2}^{L/2} ds_z (g_{\text{win}}(s_z) e^{4s_z^2/w_z^2} e^{\alpha_{\text{loc}} s_z})}. \quad (2)$$

Assuming that α_{tot}, α_{loc}, and w_z are known, the scatterer size can be determined by finding the value of a_{eff} that minimizes (Bigelow and O'Brien, 2004b),

$$\text{ASD} = \underset{\omega}{\text{mean}}[(X(\omega, a_{\text{eff}}) - \bar{X}(a_{\text{eff}}))^2], \quad (3)$$

where

$$X = \ln(E[|V_{\text{scat}}|^2]) - \ln(k^4 |V_{\text{plane}}|^2 F_\gamma(\omega, a_{\text{eff}})/A_{\text{comp}}),$$

$$\bar{X} = \underset{\omega}{\text{mean}}[X(f, a_{\text{eff}})]. \quad (4)$$

An estimate for $E[|V_{\text{scat}}|^2]$ is obtained by averaging the amplitude of the power spectra from adjacent echoes windowed in the time domain. The windowing restricts the depth resolution along the beam axis to the current tissue region of interest. Subtracting by $\bar{X}$ removes the effects of any multiplicative constants allowing for the estimation of scatterer size independent of the acoustic concentration.

B. Simulation parameters

The impact of different approximations for α_{loc} was tested by simulating the echoes generated by scatterers with Gaussian correlation functions [i.e., form factor of $F_\gamma(f, a_{\text{eff}}) = \exp(-0.827(ka_{\text{eff}})^2)$] randomly positioned in a homogeneous attenuating half-space. The attenuation of the half-space was varied from 0.05 to 1 dB/cm-MHz to test the impact of the approximations for different attenuation values. The scatterers were placed at a density of 35/mm^3 throughout the three-dimensional focal region with an a_{eff} of 25 μm. The sources used in the simulations had a focal length of 5 cm and f numbers of 1, 2, or 4, yielding 0.3, 1.2, and 4.8 scatterers per resolution cell, respectively. However, earlier simulations demonstrated that the scatterer size estimate was not strongly dependent on the number of scatterers per resolution cell (Bigelow and O'Brien, 2004b). The f number was varied to assess the impact of focusing in conjunction with the impact of α_{loc}. For all of the sources, the spectrum returned from a plane placed at the focal plane was

$$|V_{\text{plane}}(f)| \propto |f|^2 \exp\left(-2\left(\frac{f - 8\ \text{MHz}}{6\ \text{MHz}}\right)^2\right), \quad (5)$$

yielding ka_{eff} values corresponding to the maximum of the backscattered power spectrum ranging from ~1.3 to ~0.35 depending on the attenuation of the half-space. The optimal range for ka_{eff} values is between 1.2 and 0.5 (Insana and Hall, 1990).

For each f number and half-space attenuation value, 1000 different random distributions of scatterers were generated. The resultant echoes from each distribution were then grouped into sets of 25 waveforms for a total of 40 sets. Each waveform was then windowed in the time domain (providing resolution along the beam axis) and Fourier transformed to obtain the frequency spectrum. The spectra for all 25 waveforms in a set were then averaged to obtain $E[|V_{\text{scat}}|^2]$ for all 40 sets. Hence, 40 independent estimates of scatterer size were obtained for each simulated case. After averaging, $E[|V_{\text{scat}}|^2]$ needed to be compensated for spectral distortions (i.e., convolution with the windowing function in the frequency domain) introduced by the windowing. The convolution-related spectral distortions are different from the windowing effects compensated by A_{comp} in Eq. (2). The windowing compensation involves approximating $E[|V_{\text{scat}}|^2]$ and the Fourier transform of the windowing function as Gaussian functions. Then, the effect of windowing can be removed by multiplying the measured backscattered power spectrum by an appropriate Gaussian transformation (Bigelow and O'Brien, 2005b).

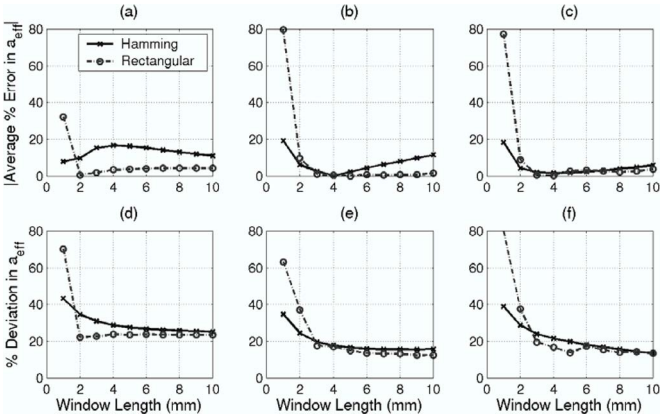


FIG. 1. Comparison of rectangular and Hamming windowing functions in terms of (a) accuracy and (d) precision of a_{eff} estimates for an $f/1$ transducer, (b) accuracy and (e) precision of a_{eff} estimates for an $f/2$ transducer, and (c) accuracy and (f) precision of a_{eff} estimates for an $f/4$ transducer for a half-space attenuation of 1 dB/cm-MHz without any approximations in the attenuation values.

C. Impact of different windowing functions

Prior to assessing the impact of the approximations for α_{loc}, the performance of two different windowing functions, the Hamming window and the rectangular window, were compared. A Hamming windowing function had been used when investigating the Spectral Fit algorithm (Bigelow and O'Brien, 2005b; Bigelow *et al.*, 2005) while a rectangular windowing function was used when developing the generalized attenuation-compensation function for use with focused sources (Bigelow and O'Brien, 2004a, b). The results for both types of windowing function are shown in Fig. 1 versus resolution along the beam axis. The half-space attenuation for these simulations was 1 dB/cm-MHz while α_{tot}, α_{loc}, and w_z were known exactly when applying A_{comp}. The average error was found by comparing the average value of all 40 estimates to the true value of a_{eff}. Likewise, the deviation was found by adding the standard deviations for estimates above and below the average value (i.e., $\sigma_{a_{\text{lower}}} + \sigma_{a_{\text{upper}}}$) as given by

$$\sigma_{a_{\text{upper}}} = \frac{100}{a_{\text{eff}}|_{\text{Theory}}} \sqrt{\frac{\sum_{\forall a_{\text{eff}j} > \bar{a}_{\text{eff}}} (a_{\text{eff}j} - \bar{a}_{\text{eff}})^2}{\sum_{\forall a_{\text{eff}j} > \bar{a}_{\text{eff}}} j}},$$

$$\sigma_{a_{\text{lower}}} = \frac{100}{a_{\text{eff}}|_{\text{Theory}}} \sqrt{\frac{\sum_{\forall a_{\text{eff}j} < \bar{a}_{\text{eff}}} (a_{\text{eff}j} - \bar{a}_{\text{eff}})^2}{\sum_{\forall a_{\text{eff}j} < \bar{a}_{\text{eff}}} j}}. \quad (6)$$

The calculation for the deviations was done using Eq. (6) because the deviations above the mean are typically different from the deviations below the mean (Bigelow and O'Brien, 2005a).

Except for small window lengths ($\leqslant$2 mm), the rectangular windowing function has better accuracy (smaller average % error) and better precision (smaller % deviation) than the Hamming windowing function, especially for the highly focused sources. The improvement with the rectangular windowing function may be related to the assumptions involved with the derivation of A_{comp} (Bigelow and O'Brien, 2004b) where it was assumed that the windowing function weighted the influence of the scatterers away from the focus along the beam axis. However, a complete analysis of the best windowing function is beyond the scope of this paper. Based on these results, a rectangular windowing function of length from 2 to 10 mm was selected for the remainder of the simulations.

III. RESULTS

A. Different approximations for α_{loc}

The impact of the different approximations for α_{loc} was investigated by solving for the scatterer size while using the approximate value for α_{loc} in the expression for A_{comp}. The size estimate from the approximate value for α_{loc} could then be compared to the size obtained for the correct value of α_{loc} for the same window length, half-space attenuation, and f number. Three different approximations for α_{loc} were considered. First, α_{loc} was neglected by setting it to 0 dB/cm-MHz in the calculation of A_{comp} regardless of the true value of α_{loc}. Second, the frequency dependence of α_{loc} was neglected by setting $\alpha_{\text{loc}} = \text{mean}_{\omega}(\alpha_{\text{loc}}(\omega))$ (i.e., constant α_{loc} without any frequency dependence). Third, α_{loc} was set to 0.5 dB/cm-MHz regardless of the true attenuation value.

The simulation results for all three variations in α_{loc} are shown with the results using the correct value of α_{loc} for window lengths of 3 and 6 mm in Figs. 2 and 3, respectively. Once again, the average error was found by comparing the average value of all 40 estimates to the true value of a_{eff}. Likewise, the deviation was found by adding the standard deviations for estimates above and below the average value (i.e., $\sigma_{a_{\text{lower}}} + \sigma_{a_{\text{upper}}}$) as given by Eq. (6). For the smaller window length of 3 mm (Fig. 2), the average errors in the size estimates [Figs. 2(a) and 2(c)] are all less than 5% except when the frequency dependence of α_{loc} was neglected [i.e., assumed $\alpha_{\text{loc}} = \text{mean}_{\omega}(\alpha_{\text{loc}}(\omega))$] for the $f/1$ transducer shown in Fig. 2(a). Similarly, the precision of the estimates for the different α_{loc} approximations [Figs. 2(d)–2(f)] is the same as the precision of the estimates when α_{loc} is known exactly

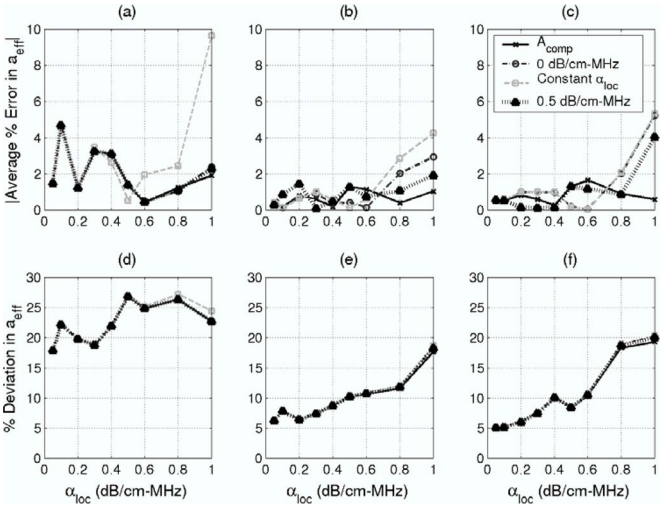


FIG. 2. Comparison between scatterer size estimates obtained when the local attenuation is known exactly (denoted A_{comp}), the local attenuation is neglected (denoted 0 dB/cm-MHz), the frequency dependence of the local attenuation is neglected (denoted constant α_{loc}), and the local attenuation is approximated by a tissue-independent value of 0.5 dB/cm-MHz in terms of (a) accuracy and (d) precision of a_{eff} estimates for an $f/1$ transducer, (b) accuracy and (e) precision of a_{eff} estimates for an $f/2$ transducer, and (c) accuracy and (f) precision of a_{eff} estimates for an $f/4$ transducer for a window length of 3 mm.

(denoted as A_{comp} in the figures). Hence, for small window lengths, the value used for α_{loc} does not affect the estimate of scatterer size.

For the larger window length of 6 mm (Fig. 3), there is a clear increase in the average errors in the size estimates as the true value of α_{loc} is increased for the $f/2$ and $f/4$ source. Hence, errors when approximating α_{loc} have a greater impact as the value of local attenuation increases. Except for when the frequency dependence of α_{loc} is neglected [i.e., assumed $\alpha_{\text{loc}} = \text{mean}_{\omega}(\alpha_{\text{loc}}(\omega))$], the average errors resulting from the

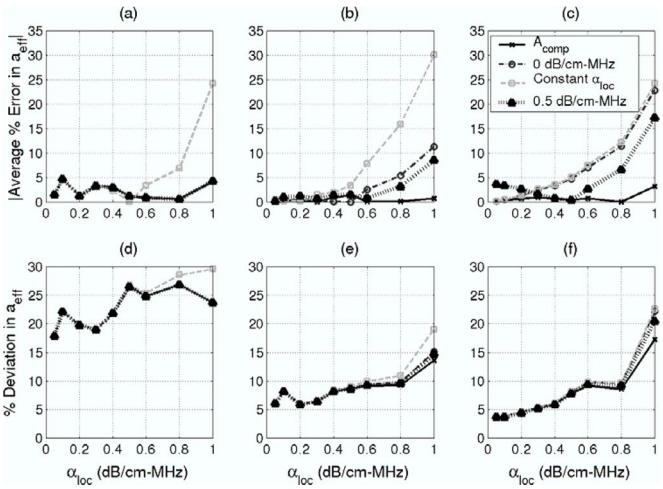


FIG. 3. Comparison between scatterer size estimates obtained when the local attenuation is known exactly (denoted A_{comp}), the local attenuation is neglected (denoted 0 dB/cm-MHz), the frequency dependence of the local attenuation is neglected (denoted constant α_{loc}), and the local attenuation is approximated by a tissue-independent value of 0.5 dB/cm-MHz in terms of (a) accuracy and (d) precision of a_{eff} estimates for an $f/1$ transducer, (b) accuracy and (e) precision of a_{eff} estimates for an $f/2$ transducer, and (c) accuracy and (f) precision of a_{eff} estimates for an $f/4$ transducer for a window length of 6 mm.

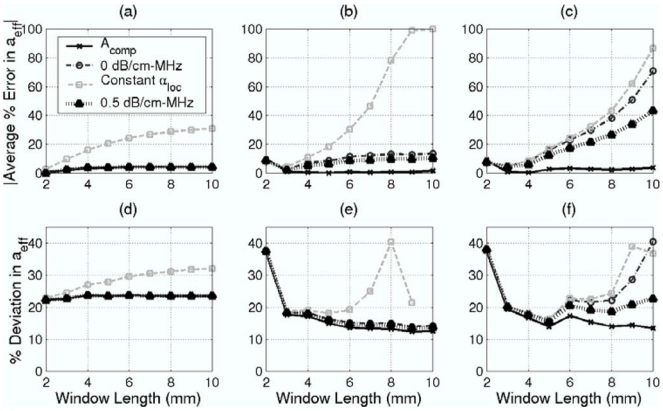


FIG. 4. Comparison between scatterer size estimates obtained when the local attenuation is known exactly (denoted A_{comp}), the local attenuation is neglected (denoted 0 dB/cm-MHz), the frequency dependence of the local attenuation is neglected (denoted constant α_{loc}), and the local attenuation is approximated by a tissue-independent value of 0.5 dB/cm-MHz in terms of (a) accuracy and (d) precision of a_{eff} estimates for an $f/1$ transducer, (b) accuracy and (e) precision of a_{eff} estimates for an $f/2$ transducer, and (c) accuracy and (f) precision of a_{eff} estimates for an $f/4$ transducer for a half-space attenuation of 1.0 dB/cm-MHz.

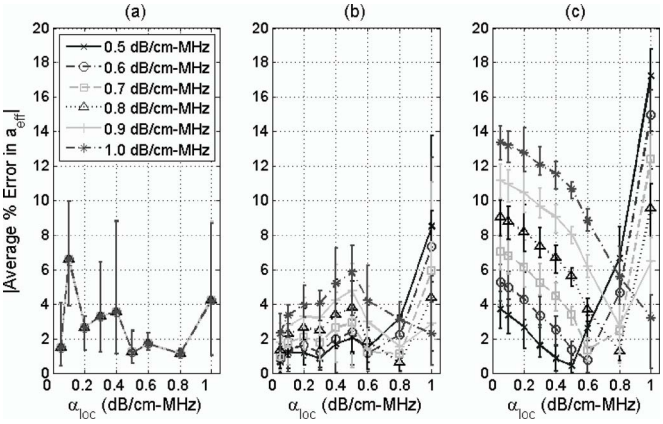


FIG. 5. Comparison of the accuracy of scatterer size estimates obtained when the local attenuation is approximated by a tissue-independent attenuation value of 0.5 to 1 dB/cm-MHz (denoted in legend) for an (a) $f/1$ transducer, (b) $f/2$ transducer, and (c) $f/4$ transducer for a window length of 6 mm.

approximations also decrease as the f number decreases (i.e., ~5% for $f/1$, ~10% for $f/2$, ~20% for $f/4$ for an α_{loc} of 1 dB/cm-MHz). Hence, the size estimates for the $f/1$ and $f/2$ sources exhibit a smaller dependence on the α_{loc} approximations. For the different approximations of α_{loc}, the errors are largest when the frequency dependence of α_{loc} is neglected (i.e., assumed $\alpha_{\text{loc}}=\text{mean}_{\omega}(\alpha_{\text{loc}}(\omega))$), and the next largest errors occur when α_{loc} is completely ignored (i.e., $\alpha_{\text{loc}}=0$ dB/cm-MHz). However, for the strongly focused sources, the errors when the frequency dependence of α_{loc} is neglected are much larger than the other errors while for the weakly focused sources all of the errors are comparable, indicating something different is happening for the weaker focused source. A tissue-independent attenuation value of 0.5 dB/cm-MHz gives the smallest errors for all of the true α_{loc} values and transducer f numbers.

The improved accuracy of the tissue-independent attenuation value and smaller f numbers can also be illustrated by the simulation results shown in Fig. 4 for a true α_{loc} value of 1 dB/cm-MHz for the different window lengths. Once again, the tissue-independent attenuation value of 0.5 dB/cm-MHz gives the smallest errors when compared to the estimates obtained with the other approximations [i.e., $\alpha_{\text{loc}}=\text{mean}_{\omega}(\alpha_{\text{loc}}(\omega))$ or $\alpha_{\text{loc}}=0$ dB/cm-MHz]. Also, the estimates obtained with the strongly focused sources are considerably better than the estimates obtained with the weaker focused source when α_{loc} is approximated except when the frequency dependence of α_{loc} is neglected, once again indicating that the source of error for weakly focused sources may be different than the source of error for the strongly focused sources. However, regardless of the desired resolution and true α_{loc} value, when approximating α_{loc}, using a strongly focused source and setting α_{loc} to 0.5 dB/cm-MHz in the attenuation-compensation calculation results in the most accurate estimates.

B. Variations in tissue-independent attenuation value

Although setting α_{loc} to 0.5 dB/cm-MHz was shown to give the most accurate estimates in the initial simulations when approximating the α_{loc} value, other values of tissue-independent attenuation might further improve the accuracy of the estimates. In this set of simulations, six different values for the tissue-independent attenuation were compared by finding the accuracy of the size estimates for true α_{loc} values from 0.05 to 1 dB/cm-MHz, window lengths from 2 to 10 mm, and f numbers of 1, 2, and 4. The values selected for the tissue-independent attenuation were 0.5, 0.6, 0.7, 0.8, 0.9, and 1 dB/cm-MHz. For these simulations, the variation in the accuracy was of interest so the 40 estimates were averaged in sets of ten estimates, each yielding four values for the average error of the size estimates.

The results for a window length of 6 mm for all of the different true α_{loc} values are shown in Fig. 5. The points correspond to the average percentage error of the four accuracy values while the error bars show the largest and smallest percentage errors of the four values. Once again, the most accurate estimates when α_{loc} is approximated are obtained for the smaller f-number transducers. However, the best value for the tissue-independent attenuation depends on the true α_{loc} value. Larger tissue-independent attenuation values give more accurate estimates when the true value of α_{loc} is large, while the smaller values for the tissue-independent attenuation give better estimates when the true value of α_{loc} is small. The smallest error in the estimates is obtained when the true value of α_{loc} is the same as the selected tissue-independent attenuation. Unfortunately, it is impossible to select an attenuation value that is always identical to the attenuation value of tissue due to biological variability.

To determine the best choice for the tissue-independent attenuation value when approximating α_{loc}, the maximum of the average errors over all values of α_{loc} (i.e., 0.05 to 1 dB/cm-MHz) is plotted versus window length in Fig. 6. The 40 estimates were still averaged in sets of ten to yield four values for the accuracy for each window length, tissue-independent attenuation value, and value of α_{loc}.

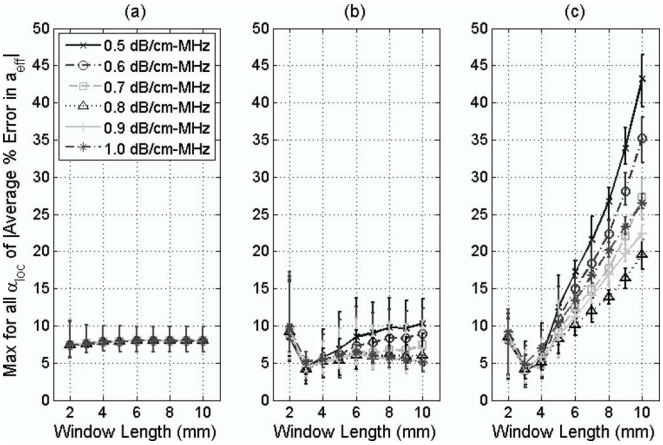


FIG. 6. Maximum average % error for the true α_{loc} values of 0.05 to 1 dB/cm-MHz when the local attenuation is approximated by a tissue-independent attenuation value of 0.5 to 1 dB/cm-MHz (denoted in legend) for an (a) $f/1$, (b) $f/2$, and (c) $f/4$ transducer.

Therefore, four values of the maximum of the average errors over all values of α_{loc} were obtained for each window length and tissue-independent attenuation value. The points in Fig. 6 correspond to the average of these four values while the error bars give the maximum and minimum.

For the $f/1$ and $f/2$ transducers [Figs. 6(a) and 6(b), respectively], all of the tissue-independent attenuation values give comparable error values with the possibility of subtle improvement for tissue-independent attenuation values of 0.7 to 1 dB/cm-MHz at larger window lengths. For the $f/4$ transducer [Fig. 6(c)], a tissue-independent attenuation value of 0.8 dB/cm-MHz gives consistently lower error values. Hence, 0.8 dB/cm-MHz would be the best choice for the tissue-independent attenuation when the biological variability is between 0.05 and 1 dB/cm-MHz. Notice that the best tissue-independent attenuation value is not in the middle of the range of the true α_{loc} values (i.e., not $\sim$0.5 dB/cm-MHz). This is probably due to the increased importance of errors in local attenuation at larger local attenuation values as was observed in Fig. 3.

C. Theoretical analysis of local attenuation approximations

The dependence of the error in the scatterer size estimate on approximations of the local attenuation value can also be analyzed by a theoretical analysis of the generalized attenuation-compensation function. For rectangular windowing functions, the influence of local attenuation on the estimate of the scatterer size is compensated by $\int_{-L/2}^{L/2} ds_z (e^{-4s_z^2/w_z^2} e^{4\alpha_{\mathrm{loc}} s_z})$. Because the performance of the strongly focused $f/1$ and $f/2$ sources in the simulations differed significantly from the performance for the $f/4$ source, an approximation for this integral was found for both highly focused and weakly focused sources.

First consider the limiting case when the length of the windowing function, L, is much greater than w_z as would be approximately true for the strongly focused $f/1$ and $f/2$ sources. The integral would then be given by

$$\int_{-L/2}^{L/2} ds_z \left(e^{\frac{4s_z^2}{w_z^2}} e^{4\alpha_{\mathrm{loc}} s_z} \right) \cong \frac{w_z\sqrt{\pi}}{2} e^{\alpha_{\mathrm{loc}}^2 w_z^2} = \frac{w_z\sqrt{\pi}}{2} e^{(\alpha_{\mathrm{loc}}^2|_{\mathrm{true}} + 2\delta\alpha\alpha_{\mathrm{loc}}|_{\mathrm{true}} + \delta\alpha^2) w_z^2}, \tag{7}$$

where $\delta\alpha$ is the error introduced by the approximation for α_{loc}. However, in order to impact scatterer size, $\delta\alpha$ must change the frequency dependence of the backscattered spectrum. Therefore, $(2\alpha_{\mathrm{loc}}|_{\mathrm{true}}\delta\alpha + \delta\alpha^2) w_z^2$ would need to have a dependence on frequency. Because w_z in the simulations and real tissue is approximately proportional to wavelength (Bigelow and O'Brien, 2004a, b), as long as $\delta\alpha$ and $\alpha_{\mathrm{loc}}|_{\mathrm{true}}$ are approximately proportional to frequency, the approximation would not impact the scatterer size estimate regardless of the difference between the assumed and true local attenuation value. This explains why the $f/1$ and $f/2$ sources yielded small errors in the scatterer size regardless of the approximation except when the approximation neglected the frequency dependence of the attenuation.

Now consider the limiting case when the windowing function, L, is much smaller than w_z as would be a rough approximation for the $f/4$ source. Under this limit, the integral would be given by

$$\int_{-L/2}^{L/2} ds_z (e^{-4s_z^2/w_z^2} e^{4\alpha_{\mathrm{loc}} s_z}) \cong \int_{-L/2}^{L/2} ds_z (e^{4\alpha_{\mathrm{loc}} s_z}) = \frac{e^{4\alpha_{\mathrm{loc}} L/2} - e^{-4\alpha_{\mathrm{loc}} L/2}}{4\alpha_{\mathrm{loc}}} = \frac{1}{2\alpha_{\mathrm{loc}}} \sinh(2\alpha_{\mathrm{loc}} L). \tag{8}$$

Because $2\alpha_{\mathrm{loc}} L \ll 1$, Eq. (8) can be further simplified to yield

$$\frac{1}{2\alpha_{\mathrm{loc}}} \sinh(2\alpha_{\mathrm{loc}} L) \cong L\left(1 + \frac{(2\alpha_{\mathrm{loc}} L)^2}{6}\right). \tag{9}$$

Therefore, the error in the scatterer size estimate due to the approximation for α_{loc} would be on the order of $L^2 2/3 |2\alpha_{\mathrm{loc}}|_{\mathrm{true}}\delta\alpha + \delta\alpha^2|$, which equals $L^2 2/3 |\alpha_{\mathrm{ref}} - \alpha_{\mathrm{loc}}|_{\mathrm{true}}| (\alpha_{\mathrm{ref}} + \alpha_{\mathrm{loc}}|_{\mathrm{true}})$ when written in terms of the tissue-independent attenuation value α_{ref}. Notice that this term increases as the true value of local attenuation and window length increase for the same $\delta\alpha$ as was also observed for the error in the size estimates in the computer simulations. The error term $C_{\mathrm{mse}} |\alpha_{\mathrm{ref}} - \alpha_{\mathrm{loc}}|_{\mathrm{true}}| (\alpha_{\mathrm{ref}} + \alpha_{\mathrm{loc}}|_{\mathrm{true}})$ for a α_{ref} of 0.7 dB/cm-MHz is also plotted with the corresponding curve from Fig. 5(c) in Fig. 7. C_{mse} is a scaling constant used to plot the curves on the same scale and was found by minimizing the mean squared error between the two curves. There is reasonable agreement between the theoretical error term and the simulated error curve except at higher attenuation values where the simulation error is larger than expected from the theory.

Assuming that the error due to the approximations for α_{loc} is approximately described by $L^2 2/3 |\alpha_{\mathrm{ref}} - \alpha_{\mathrm{loc}}|_{\mathrm{true}}|$

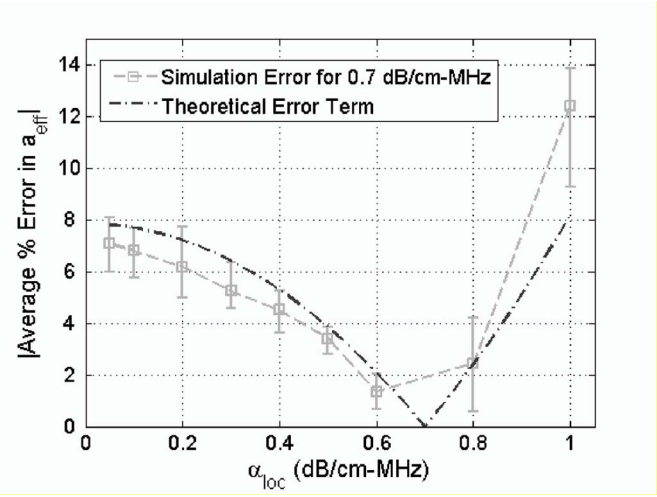


FIG. 7. Comparison between theoretical error term, $C_{mse}(|\alpha_{ref}-\alpha_{loc}|_{true}|(\alpha_{ref}+\alpha_{loc}|_{true}))$, and the accuracy of scatterer size estimates obtained when the local attenuation is approximated by a tissue-independent attenuation value of 0.7 dB/cm-MHz for an $f/4$ transducer for a window length of 6 mm.

$(\alpha_{ref}+\alpha_{loc}|_{true})$, it is possible to calculate the best value for α_{ref} given a range of expected $\alpha_{loc}|_{true}$ values. A plot of the $\max_{\forall \alpha_{loc}|_{true}}(|\alpha_{ref}-\alpha_{loc}|_{true}|(\alpha_{ref}+\alpha_{loc}|_{true}))$ vs. α_{ref} is shown in Fig. 8 for $\alpha_{loc}|_{true}$ values between 0.05 and 1 dB/cm -MHz. The plot consists of the intersection of two curves with the minimum value of $\max_{\forall \alpha_{loc}|_{true}}(|\alpha_{ref}-\alpha_{loc}|_{true}|(\alpha_{ref}+\alpha_{loc}|_{true}))$ occurring at the intersection. The curve with the negative slope corresponds to when the maximum error is due to large $\alpha_{loc}|_{true}$ values [i.e., $\max_{\forall \alpha_{loc}|_{true}}(|\alpha_{ref}-\alpha_{loc}|_{true}|(\alpha_{ref}+\alpha_{loc}|_{true}))=(\alpha_{high}^2-\alpha_{ref}^2)$ where $\alpha_{high}>\alpha_{ref}$], while the curve with the positive slope corresponds to when the maximum error is due to a small $\alpha_{loc}|_{true}$ values [i.e., $\max_{\forall \alpha_{loc}|_{true}}(|\alpha_{ref}-\alpha_{loc}|_{true}|(\alpha_{ref}+\alpha_{loc}|_{true}))=(\alpha_{ref}^2-\alpha_{low}^2)$ where $\alpha_{low}<\alpha_{ref}$]. α_{high} and α_{low} correspond to the largest and smallest values of local attenuation possible for a tissue region of interest, respectively. Because the minimum theoretical error value corresponds to the intersection, the minimum error should occur when

$$(\alpha_{high}^2-\alpha_{ref}^2)=(\alpha_{ref}^2-\alpha_{low}^2)\Rightarrow\alpha_{ref}=\sqrt{\frac{\alpha_{high}^2+\alpha_{low}^2}{2}}. \tag{10}$$

For the simulations presented in this paper, Eq. (10) would yield an optimal tissue-independent attenuation value of 0.71 dB/cm-MHz, which is slightly smaller than the optimal value of 0.8 dB/cm-MHz that was observed in Fig. 6. The small difference can be attributed to the theoretical error term being smaller than the real error value for the higher attenuation values as was observed in Fig. 7. Because the error due to the higher attenuation values will decrease as the tissue-independent attenuation value increases, it is reasonable to expect that the optimal tissue-independent attenuation value should be slightly larger than the value given by Eq. (10).

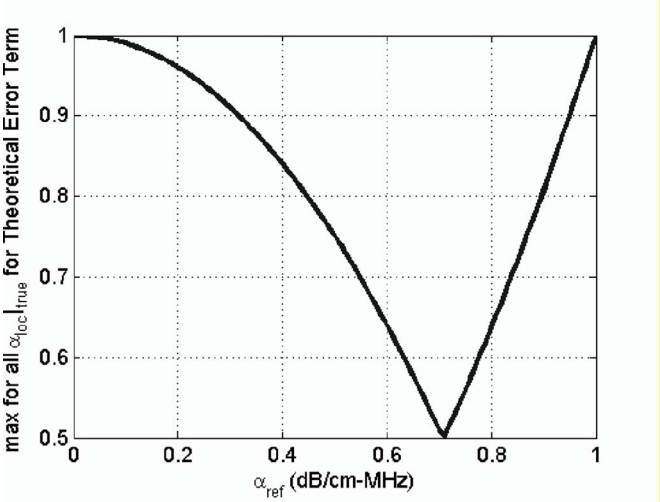


FIG. 8. Value of $\max_{\forall \alpha_{loc}|_{true}}(|\alpha_{ref}-\alpha_{loc}|_{true}|(\alpha_{ref}+\alpha_{loc}|_{true}))$ for α_{loc} values of 0.05 to 1 dB/cm-MHz when the local attenuation is approximated by a tissue-independent attenuation value, α_{ref}.

IV. CONCLUSIONS

Although it is well known that correcting for the frequency-dependent attenuation both along the propagation path (total attenuation) and in the scattering region (local attenuation) is critical when estimating the correlation length of tissue microstructure, the impact of approximations of the local attenuation on the estimate has not been addressed. In this investigation, three different types of approximations for the local attenuation were assessed using computer simulations and theoretical analysis. First, the local attenuation was completely neglected (i.e., $\alpha_{loc}=0$ dB/cm-MHz regardless of the true attenuation value). Second, the frequency dependence of the local attenuation was neglected [i.e., $\alpha_{loc}=\text{mean}_{\omega}(\alpha_{loc}(\omega))$]. Third, the local attenuation was approximated by a tissue-independent attenuation value (i.e., $\alpha_{loc}=0.5$ dB/cm-MHz regardless of the true attenuation value). Errors in the scatterer size estimate due to the approximations were shown to increase with increasing window length, increasing true local attenuation, and increasing f number provided that the frequency dependence of the attenuation was not ignored. The most robust estimates were obtained when the local attenuation was approximated by a tissue-independent attenuation value.

After demonstrating that the tissue-independent attenuation yielded the best results when the local attenuation was being approximated, the optimal choice for the tissue-independent attenuation was determined using computer simulations and theoretical calculations. In the computer simulations, six different reference attenuation values (0.5, 0.6, 0.7, 0.8, 0.9, and 1 dB/cm-MHz) were compared for true attenuation values ranging from 0.5 to 1 dB/cm-MHz. The largest error over all of the true local attenuation values for each value of the tissue-independent attenuation was then used to quantify the performance of each tissue-independent attenuation value. The calculations and simulations showed that the optimal value for the tissue-independent attenuation was slightly larger than $\sqrt{(\alpha_{high}^2+\alpha_{low}^2)/2}$ where α_{high} and α_{low} correspond to the largest and smallest values of local

attenuation expected for a tissue region of interest, respectively. The optimal tissue-independent attenuation value being slightly larger than $\sqrt{(\alpha_{\text{high}}^2+\alpha_{\text{low}}^2)/2}$ results from the current theory not adequately capturing the error performance at higher attenuation values. However, $\sqrt{(\alpha_{\text{high}}^2+\alpha_{\text{low}}^2)/2}$ still serves as a good choice for the tissue-independent attenuation value, especially when operating at smaller window lengths and/or using strongly focused sources.

ACKNOWLEDGMENTS

This work was supported by the University of Illinois Research Board and the University of North Dakota School of Engineering and Mines.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2004a**). "Scatterer size estimation in pulse-echo ultrasound using focused sources: Calibration measurements and phantom experiments," J. Acoust. Soc. Am. **116**, 594–602.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2004b**). "Scatterer size estimation in pulse-echo ultrasound using focused sources: Theoretical approximations and simulation analysis," J. Acoust. Soc. Am. **116**, 578–593.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2005a**). "Evaluation of the spectral fit algorithm as functions of frequency range and Δka_{eff}," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 2003–2010.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2005b**). "Signal processing strategies that improve performance and understanding of the quantitative ultrasound SPECTRAL FIT algorithm," J. Acoust. Soc. Am. **118**, 1808–1819.

Bigelow, T. A., Oelze, M. L., and O'Brien, W. D., Jr. (**2005**). "Estimation of total attenuation and scatterer size from backscattered ultrasound waveforms," J. Acoust. Soc. Am. **117**, 1431–1439.

Chivers, R. C., and Hill, C. R. (**1975**). "A spectral approach to ultrasonic scattering from human tissue: methods, objectives and backscattering measurements," Phys. Med. Biol. **20**, 799.

Dussik, K. T., and Fritch, D. J. (1956). "Determination of sound attenuation and sound velocity in the structure constituting the joints, and of the ultrasonic field distribution within the joints on living tissues and anatomical preparations, both in normal and pathological conditions," Public Health Service, Natl. Inst. Health Project A454, Prog. Rep. (15 September, 1956).

Gerig, A., Zagzebski, J., and Varghese, T. (**2003**). "Statistics of ultrasonic scatterer size estimation with a reference phantom," J. Acoust. Soc. Am. **113**, 3430–3437.

Goss, S. A., Johnston, R. L., and Dunn, F. (**1978**). "Comprehensive compilation of empirical ultrasonic properties of mammalian tissues," J. Acoust. Soc. Am. **64**, 423–457.

Hall, T. J., Insana, M. F., Harrison, L. A., and Cox, G. G. (**1996**). "Ultrasonic measurement of glomerular diameters in normal adult humans," Ultrasound Med. Biol. **22**, 987–997.

Insana, M. F., and Hall, T. J. (**1990**). "Parametric ultrasound imaging from backscatter coefficient measurements: Image formation and interpretation," Ultrason. Imaging **12**, 245–267.

Insana, M. F., Wagner, R. F., Brown, D. G., and Hall, T. J. (**1990**). "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192.

Lizzi, F. L., Greenebaum, M., Feleppa, E. J., Elbaum, M., and Coleman, D. J. (**1983**). "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1373.

Nassiri, D. K., and Hill, C. R. (**1986**). "The use of angular acoustic scattering measurements to estimate structural parameters of human and animal tissues," J. Acoust. Soc. Am. **79**, 2048–2054.

Oelze, M. L., O'Brien, W. D., Jr., Blue, J. P., and Zachary, J. F. (**2004**). "Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quantitative ultrasound imaging," IEEE Trans. Med. Imaging **23**, 764–771.