CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 123, No. 3 March 2008

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

ERRATA

JASA EXPRESS LETTERS

CUMULATIVE AUTHOR INDEX

Improved prewhitening method for linear frequency modulation reverberation using dechirping transformation

Byung Woong Choi, Eun Hyon Bae, Jeong Soo Kim, and Kyun Kyung Lee
School of Electrical Engineering and Computer Science, Kyungpook National University, Daegu 701-702, Republic of Korea
bwng@ee.knu.ac.kr, jamespool@ee.knu.ac.kr, kimnobi@ee.knu.ac.kr, kklee@ee.knu.ac.kr

Abstract: This letter presents an autoregressive (AR) prewhitener for linear frequency modulation (LFM) reverberation to enhance the target signal. The proposed method uses a dechirping transformation to inversely compensate the frequency chirp rate of the LFM and give the LFM reverberation a stationary frequency property in each data block. The left or right beam signal adjacent to the current beam is then used as the reference signal, and the frequency response of each data block modeled using the AR coefficients. Finally, these coefficients are used to implement the inverse filter and efficiently prewhiten the LFM reverberation of the current beam.

PACS numbers: 43.60.Gk, 43.30.Vh, 43.60.Hj [JC]
Date Received: October 21, 2007 **Date Accepted:** December 4, 2007

1. Introduction

Improving the performance of target detection in a reverberation environment is one of the most important issues when using active sonar for target detection. However, detecting a target in reverberation is difficult, as reverberation is highly correlated with the transmitted signal. Several attempts have already been made to reduce reverberation. For example, Kay *et al.* developed a suboptimal detector for cw reverberation using a prewhitener based on autoregressive (AR) modeling of reverberation.[1] With this method, the reverberation in the current block is whitened based on modeling the reverberation in the previous block. However, to whiten reverberation using the above-noted methods, local stationarity is assumed, where the frequency variation is a decisive factor.[1] Thus, in the case of a linear frequency modulation (LFM) signal with a nonstationary frequency that varies linearly with time, it is difficult to whiten LFM reverberation efficiently. A LFM reverberation whitening method based on the AR model was already suggested by Camillet *et al.*,[2] yet this method has a potential modeling error for LFM reverberation prewhitening, as the frequency variation of the LFM signal is not considered. Accordingly, this letter proposes a method for making the LFM signal in each block stationary by applying a dechirping transformation, thereby allowing the dechirped LFM reverberation signal to be regarded as a "tonal-like" signal and conventional AR modeling methods to be adopted for the LFM whitening problem.

For validation, the proposed method is applied to real oceanic reverberation data acquired using an active towed array sonar system (TASS).

2. Conventional prewhitening

Spectral estimation of reverberation is an important procedure in prewhitening processes. Yet, estimating the spectrum precisely is difficult, as reverberation is highly nonstationary. Thus, reverberation data are cut into blocks, where the length of each block corresponds to that of the emitted signal, and a spectral estimation method based on these short data blocks is required to overcome this difficulty, as used by the AR model.[1]

To whiten reverberation using the AR model, a reference signal is necessary, which can be chosen from a data block adjacent to the current block temporally[1,2] or spatially. For example, suppose the reference signal is chosen from a spatially adjacent beam. If the rever-

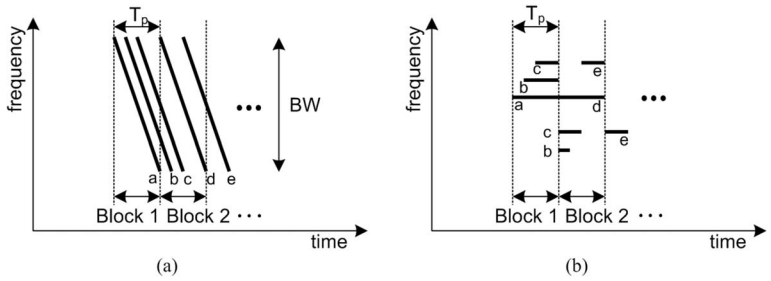


Fig. 1. Concept of the dechirping LFM reverberation. (a) Original time-frequency characteristic of LFM reverberation. (b) Dechirped time-frequency characteristic of LFM reverberation.

beration to be whitened is positioned at the mth time block of the lth beam, given by $x_{(m,l)}(t)$, the AR coefficients $\{1, a_1, \ldots, a_p\}$ are extracted from modeling the reverberation of the adjacent beam $x_{(m,l-1)}(t)$ or $x_{(m,l+1)}(t)$ using a pth order AR model. From these AR coefficients, the inverse filter $\hat{A}(z)$ is defined as Eq. (1) and applied to $x_{(m,l)}(t)$, then the whitened signal is obtained using Eq. (2),

$$\hat{A}(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \tag{1}$$

$$y_{(m,l)}(t) = x_{(m,l)}(t) + \sum_{k=1}^{p} a_k x_{(m,l)}(t-k). \tag{2}$$

3. Proposed prewhitening with dechirping transformation

Prewhitening using the AR model relies on the reverberation being locally stationary. First, it is assumed that the signal is stationary in two consecutive blocks, that is, the spectrum of the reverberation does not vary significantly from one block to the next. Second, it is also assumed that the signal is stationary within each block, allowing the spectrum to be estimated correctly.[1]

For the first assumption, the stationarity is validated by evaluating the Itakura distance between the spectral densities of two signals in blocks adjacent to each other.[2,3] The second assumption is no problem in the cw case, as there is only a minimal frequency variation in each block. However, an LFM signal has a linearly varying frequency, which undermines the second assumption, thereby affecting the accuracy of the spectral estimation of the reverberation and degrading the whitening performance when a conventional prewhitening method is applied without modification.

Base on this analysis, this letter proposes a dechirping transformation to convert the frequency of LFM signal into tonal-like frequency in each block. When a LFM signal with a center frequency f_0 and frequency chirp rate α is described by Eq. (3), the dechirped signal $x_D(t)$ is defined by Eq. (4) in each block and the frequency variation compensated inversely,

$$x(t) = \exp\left(j2\pi\left(f_0 t + \tfrac{1}{2}\alpha t^2\right)\right), \tag{3}$$

$$x_D(t) = x(t)\exp\left(-j2\pi\tfrac{1}{2}\alpha t^2\right). \tag{4}$$

Figure 1 depicts the concept of a dechirping transformation of the dominant LFM reflection signals from multipaths, where the pulse length is T_p and the bandwidth is BW. As seen, each LFM reflection signal is mapped into frequency lines based on the dechirping transformation. However, a single LFM reflection signal can give different frequency components

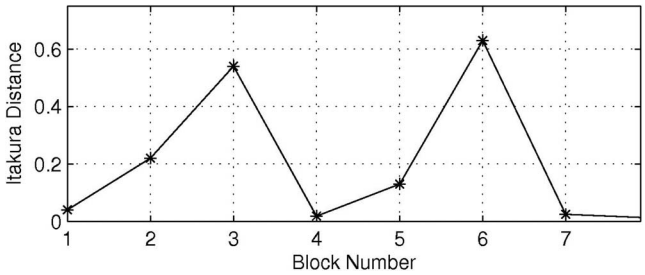


Fig. 2. Averaged Itakura distances dechirped reverberation in the current beam and the adjacent beam.

for adjacent time blocks, as the duration of a LFM reflection signal and the dechirping transformation time block cannot be synchronized. Nonetheless, it can still be assumed that the dechirped signal $x_{D(m,l)}(t)$ for the current beam and $x_{D(m,l-1)}(t)$ or $x_{D(m,l+1)}(t)$ in the spatially neighboring blocks have a similar spectrum, as these signals are received through similar ray paths. Therefore, the spatially neighboring block in the left or right adjacent beam $x_{D(m,l-1)}(t)$ or $x_{D(m,l+1)}(t)$ is adopted as the reference signal to whiten $x_{D(m,l)}(t)$ in the current beam.

The reference dechirped signal in the adjacent beam is then modeled by AR coefficients, and the inverse filer designed and applied to current signal as described earlier, to produce the whitened output. The proposed prewhitening is summarized as the following procedure:

(1) Cut the current beam and adjacent beam signals into blocks, where blocks of current beam are synchronized in time with those of adjacent beams.
(2) Apply the dechirping transformation for each block of current and adjacent beams.
(3) To whiten the reverberation positioned at the mth time block of the lth beam, given by $x_{D(m,l)}(t)$, extract the AR coefficients $\{1, a_1, \ldots, a_p\}_D$ from modeling the reference $x_{D(m,l-1)}(t)$ or $x_{D(m,l+1)}(t)$.
(4) From these AR coefficients, define the inverse filter $\hat{A}_D(z)$
(5) Apply the inverse filter to $x_{D(m,l)}(t)$, then obtain the whitened signal $y_{D(m,l)}(t)$
(6) Because $y_{D(m,l)}(t)$ is dechirped, rechirp it to restore its original state and whitened output $y_{(m,l)}(t)$ is obtained.
(7) Apply matched filter to detect a target

4. Experiment and results

For comparison simulated targets 1 and 2, located at 1.2 and 3.5 km with 0 and 3 kn of relative Doppler, respectively, were injected into real oceanic reverberation data sets acquired using an active TASS in the littoral water of South Korea. Target 1 represents the detection problem under high reverberation environment, while target 2 is in a background noise limited environment. Both targets were located in the broadside beam direction of the TASS. The transmitted signal was LFM with 800 Hz bandwidth 1200 Hz center frequency, 1 s pulse duration, and received by a uniform linear array of 48 sensors with a 4096 Hz sampling frequency. For stochastic analysis, we acquired 50 reverberation data sets which lasted about 12 min. The sensor signals were shifted to the baseband and downsampled by a factor of 4. A conventional delay-sum beamformer was used with 49 beams.[4]

Figure 2 shows the Itakura distance which demonstrates the similarity of the dechirped AR spectrums between the current beam and the adjacent beam. The distance value was evaluated on temporally synchronized blocks of beams. The evaluation was repeated and averaged for the target injected 50 reverberation data sets. Blocks 3 and 6 show large Itakura distance

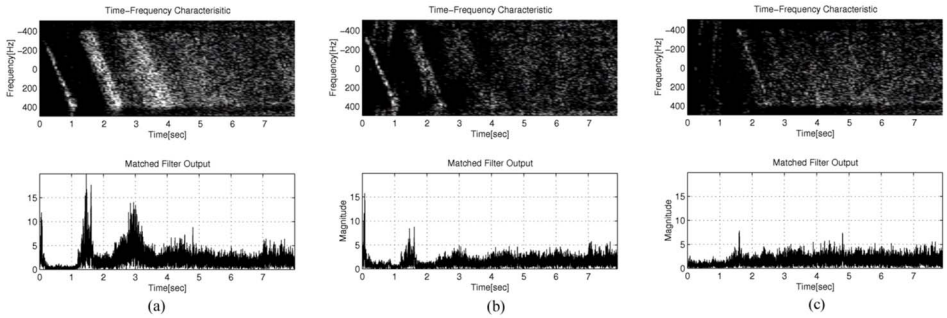


Fig. 3. Typical target detection result. (a) Without prewhitening. (b) Conventional prewhitening. (c) Proposed prewhitening.

because of the injected targets 1 and 2 in the current beam, respectively. Otherwise, the Itakura distance was small demonstrating the similarity of the dechirped AR spectrums.

Figure 3 shows typical target detection result using each prewhitening method selected from 50 detection results. Figure 3(a) shows the time-frequency characteristic and matched filter output for a current beam signal without prewhitening. Although the targets were indicated by slanted lines at intervals of 1.5–2.5 s and 4.8–5.8 s, refer to the time-frequency characteristic of Fig. 3(a), they were not recognizable, and it was difficult to detect the targets in the matched filter output refer to the lower plot of Fig. 3(a). Figure 3(b) shows the results when using conventional spatial prewhitening based on AR modeling. In this experiment, the AR order was selected as $p=10$ while each beam signal cut into eight blocks and the data block spatially adjacent and temporally synchronized with the current block was used as the reference. The length of each data block was 1024 samples, corresponding to the pulse length. Although the targets were visible, there were also many false alarms, due to the remaining presence of some high-intensity reverberations. Figure 3(c) shows the results when using the proposed prewhitening method with a dechirping transformation. The experiment parameters were the same as those used with the conventional method. As seen in time-frequency characteristic of Fig. 3(c) each target was clearly indicated by slanted lines at the appropriate positions. The matched filter output, shown in the lower plot of Fig. 3(c), also revealed the target peaks very prominently.

Finally we present the detection performance of the prewhitening methods. By performing 50 Monte Carlo runs on 50 reverberation data sets, the experimental receiver operating characteristic (ROC) curves are plotted in Fig. 4. These curves represent detection probability versus false-alarm probability for each fixed signal-to-noise ratio. The detection criterion was as follows. When targets were absent, we evaluated the probability of false alarms on the 1024 points around the expected position of the targets as a function of detection threshold. When targets were injected, the maximum peak positions of matched filter output above the detection threshold within acceptable error interval around the true positions, 20 ms in this experiment were counted as a detection. This procedure was repeated and the results were averaged for the 50 Monte Carlo runs. Figure 4(a) represents the ROC curves for target the 1 under high reverberation. By comparing the left and right plot of Fig. 4(a), we could conclude that the proposed prewhitening method improved the detection performance significantly. Figure 4(b) represents the ROC curves for target 2 under noise limited environment. Although the performance enhancement was not as remarkable as in the reverberation limited environment, we could still observe some improvement over the conventional method.

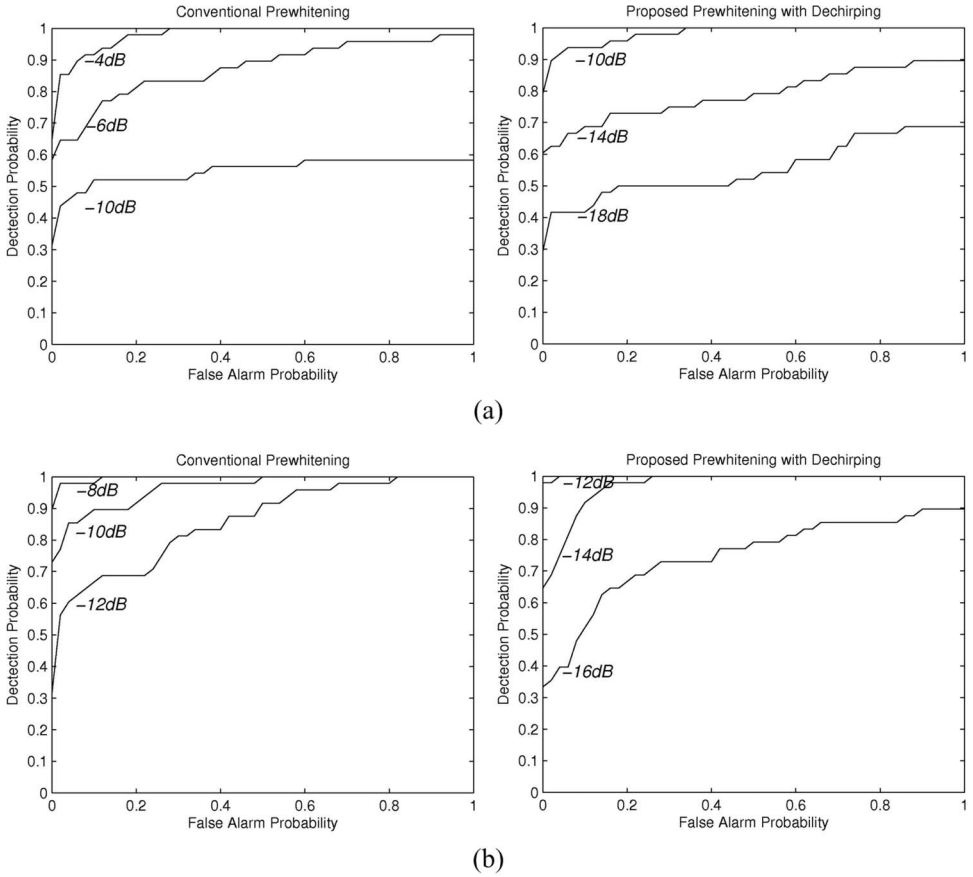


Fig. 4. Experimental ROC curves. (a) Conventional prewhitening (left) and proposed prewhitening method with dechirping (right) for target 1. (b) Conventional prewhitening (left) and proposed prewhitening method with dechirping (right) for target 2.

5. Conclusion

A prewhitening method using a linear dechirping transformation was proposed for LFM reverberation. Experiments with real oceanic data verified that the proposed method provided an improved whitening performance. Consequently, it is expected that the proposed method can be applied for the detection of real underwater objects.

Acknowledgment

This work was supported by the BK21 Project.

References and links

[1]S. Kay and J. Salisbury, "Improved active sonar detection using autoregressive prewhiteners," J. Acoust. Soc. Am. **87**(4), 1603–1611 (1990).
[2]V. Camillet, P. Amblard, and G. Jourdain, "Detection of phase- or frequency-modulated signals in reverberation noise," J. Acoust. Soc. Am. **105**(6), 3375–3389 (1999).
[3]F. Itakura, "Minimum prediction residual principle applied to speech recognition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-23**(1), 67–72 (1975).
[4]R. Nielson, *Sonar Signal Processing* (Artech House, Boston, 1991).

Small-scale seismic inversion using surface waves extracted from noise cross correlation

Pierre Gouédard, Philippe Roux, and Michel Campillo
Laboratoire de Géophysique Interne et de Tectonophysique, UMR CNRS 5559, Université Joseph Fourier, Grenoble, France
Philippe.Roux@obs.ujf-grenoble.fr, Pierre.Gouedard@obs.ujf-grenoble.fr, Michel.Campillo@obs.ujf-grenoble.fr

Abstract: Green's functions can be retrieved between receivers from the correlation of ambient seismic noise or with an appropriate set of randomly distributed sources. This principle is demonstrated in small-scale geophysics using noise sources generated by human steps during a 10-min walk in the alignment of a 14-m-long accelerometer line array. The time-domain correlation of the records yields two surface wave modes extracted from the Green's function between each pair of accelerometers. A frequency–wavenumber Fourier analysis yields each mode contribution and their dispersion curve. These dispersion curves are then inverted to provide the one-dimensional shear velocity of the near surface.

PACS numbers: 43.40.Ph, 43.60.Fg [JM]
Date Received: September 22, 2007 **Date Accepted:** December 10, 2007

1. Introduction

It has been demonstrated both theoretically and experimentally that records of seismic noise, equidistributed in angle, yield the Green's functions between receiver pairs after a time-averaged cross-correlation process. Since 2002, several papers have been published on this noise cross-correlation process in geophysics applied from a hundred of meter scale to a hundred of kilometers scale.[1–4]

One necessary ingredient to extract Green's function from noise correlation is the angle isotropy of seismic noise or, at least, the knowledge of the noise directivity on the seismic network. This strong hypothesis is rarely verified in practical situations when dealing with high frequency seismic noise (above 5 Hz) and small scales (less than 1 km) where noise sources have human origin. A way to get around the noise source distribution ambiguity is to perform the correlation process from a set of user-defined noise sources. When the seismic network is defined as a line of geophones, one way to deal with user-defined noise sources is to spread them in range on 10–50 wavelengths along the receiver alignment.

This technique presents several advantages. First we do not need an active device as a source. The strength of the source is often a problem when working in environments (such as an urban area, for example) where the use of potentially destructive devices such as explosives is prohibited. Here it is the duration of the record that governs the available energy of the signal. Furthermore, no synchronization of source and acquisition is required. However, using a source to perform velocity measurement between two stations is well known and currently used in seismology but it generally leads to apparent velocities which must be corrected according to the source location. Here by using a spatial distribution of sources, whose contributions are eventually averaged, we extract a response between the stations that is independent of the paths source stations, as far as the illumination is sufficient for the Green's function emergence. This condition requires in practice that the source, i.e., the steps, cover a region corresponding to the "end fire lobes" of the receiver line.[5] In this case, no angular correction relying on a speculative velocity model is needed. In conclusion, extracting the Green's function from user-defined noise sources makes the correlation technique a trade-off between an active and a passive method—active since we do not rely on the natural ambient noise in the medium, but passive since the correlation process naturally achieves synchronization and averaging.

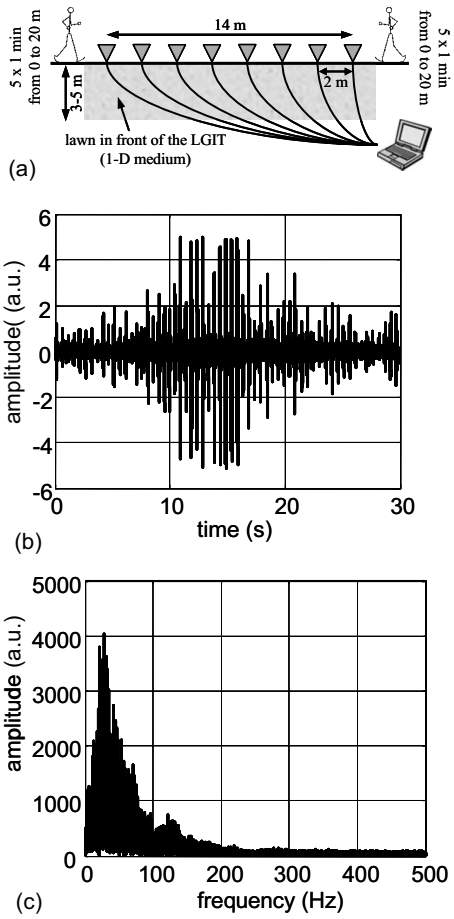


Fig. 1. (a) Experimental setup. The acquisition system is performed from a National Instrument board that gives access to the recordings in real time. The array deployment involved eight geophones but can be adapted to various configurations using up to 16 accelerometers or geophones. (b) Typical 30-s-long time-domain recording and (c) its frequency spectrum. The foot step amplitude depends on the range to the geophone. Frequency spectrum ranges from 4 to about 150 Hz.

The goal of this work is to perform an experimental demonstration of the correlation process of user-defined noise sources at the meter scale using a linear array of geophones. This letter is divided into two parts. In Sec. 2, we present the experimental setup, the recording of user-defined noise sources, and the averaging correlation process that leads to the final seismic section. In Sec. 3, a classical geophysical inversion is performed from the phase velocity dispersion curves of two Rayleigh modes extracted from the seismic section. Finally, the shear velocity profile is inverted over the first 5 m.

2. Retrieving a seismic section from user-defined noise cross correlation

A 14-m-long line of eight evenly spaced vertical geophones has been used to record human steps [Fig. 1(a)]. Two persons walked in the alignment of the accelerometers line, five times 1 min on each side, from 0 to about 30 m away from the geophone array. For a matter of convenience, the studied medium was the lawn in front of the laboratory.

The experimental setup has been designed to be versatile. The array configuration may include up to 16 one-component seismic stations. These seismic sensors could be accelerometers or geophones depending on the expected frequency bandwidth. Similarly, the array length is adjustable to the surface wave wavelength.

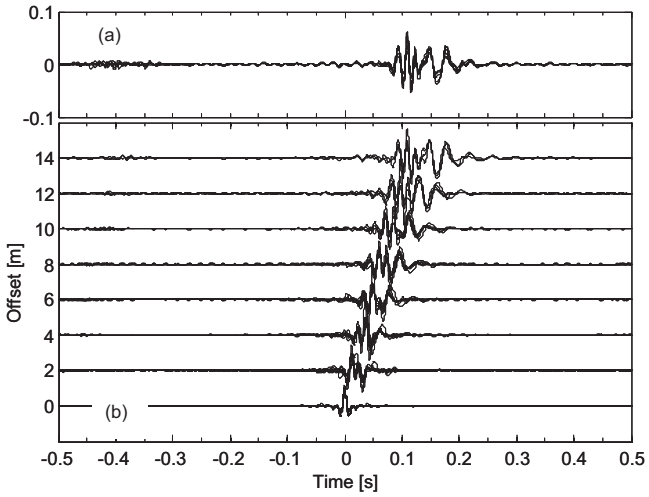


Fig. 2. (a) Superposition of five 1-min-long correlations between a geophone pair separated by 14 m. (b) Superposition of seismic sections obtained from correlation between accelerometer pairs separated by the same range. The signal is normalized at each range.

The main advantage of this system is to be easy and fast to setup. Our ambition was to perform a complete deployment, acquisition, and inversion in approximately 30 min. The system design makes it very convenient for local and near surface measurements. Figure 1(b) shows a typical time-domain recorded signal with its frequency spectrum. Each event in the signal is a step whose amplitude depends on the range to the receiver. The energy spectrum of the 30-s-long record is spread up to 150 Hz. Given the frequency response of the accelerometers and the spatial extension of the array, a frequency interval ranging from 10 to 100 Hz was selected for the analysis.

Since the frequency spectrum of the steps [Fig. 1(c)] is not flat in the frequency interval of interest, and as correlating is mathematically equivalent to a spectrum product, only the most energizing frequencies will emerge in the correlation signal. To enlarge the effective frequency bandwidth, the spectrum of the records will be equalized in the selected frequency interval $[10-100\ \text{Hz}]$ before the correlation process.

To check the robustness of the correlation process, 5 1-min long records were correlated separately for each receiver pair. Figure 2(a) shows, for a given pair, that the five time-domain correlations superimpose in phase. Thus correlation is robust and does not depend on the person who walked. As those correlations signals superimpose, they are stacked to increase the signal-to-noise Ratio (SNR). The residual signal observed at large lag times in the correlation (for time ~ -0.5 s) corresponds to the correlation between two consecutive steps. This signal is incoherent over the five realizations and vanishes in the 5-min average.

The advantage of the correlation process is to perform an average on the different sources by stacking them with the appropriate time lag without the need for any synchronization between the source and the receivers. The superposition of the correlations of 1-min long signals is thus just a verification of the repeatability of the steps. Stacking the five correlations of 1-min-long records is equivalent to directly correlating a 5-min-long signal.

To obtain a seismic section from the correlation process, each signal recorded at one receiver is correlated with the signal recorded by geophone 1 or 8 located at each extremity of the line array. The seismic section clearly shows wave propagation along the geophone array [Fig. 2(b)]. Taking geophone 1 or 8 as the reference signal does not modify the seismic section. This shows that seismic propagation from left to right is identical to propagation from right to left on the 14-m-long seismic array. The medium can then be assumed as one dimensional (1D) in the frequency bandwidth of the recordings.

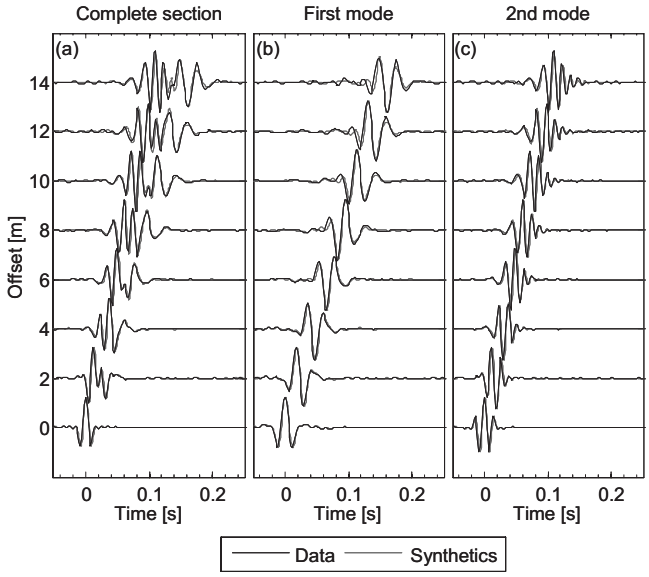


Fig. 3. (a) Seismic section obtained from the correlation process after all averaging operations. The signal-to-noise ratio is above 30 dB for each trace. Phase dispersion is clearly observed for two surface waves with group velocity of about 90 and 120 m/s. (b), (c) The two propagating modes have been extracted from the $F-K$ plot in Fig. 5. Dispersion of each mode is clearly observed. Synthetics obtained from the velocity model in Fig. 6(a) show a good agreement with experimental data.

The 1D argument can be pushed even further. Each receiver pair spaced by the same range should give the same correlation function as the propagation does not depend on the receiver locations but only on the separation between receivers. For each range, there are several equivalent receiver pairs. Figure 2(b) shows the superposition of the time-domain correlations of the equivalent pairs. For shorter ranges, there are more equivalent pairs (7 for 2 m, 6 for 4 m, etc.), which results in an efficient stacking and a better SNR. Figure 4(a) shows the seismic section obtained after all averaging operations. After a 10-min total recording, SNR is above 30 dB. Both phase dispersion and geometrical spreading during propagation are retrieved.

The final seismic section clearly reveals the presence of two surface waves, with mean group velocities of about 90 and 120 m/s. Those low group velocities are good indications of two surface waves identified as Rayleigh modes.

We insist on the fact that this section was obtained from 10 min of unsynchronized human steps only, which makes it nearly a passive method. To get the same result with classical active seismic techniques, much more time would have been needed to synchronize numerous sledgehammer blows. The "passive" method presented here is thus (1) easy to implement, as there is a large flexibility in the array configuration, (2) fast, as it takes only about 30 min to complete the array deployment and the recording, and (3) simple, as there is no synchronization task and processing is performed in real time.

3. Geophysics inversion

Given the high SNR of the seismic section obtained from the correlation process, we completed the analysis by a classical surface wave seismic inversion.[6,7]

A frequency—wave-number ($F-K$) transform was applied to the seismic section [Fig. 3(a)] to measure phase velocity dispersion curves [Fig. 4]. As the geophones are evenly spaced by a 2-m distance, the largest wave number satisfying the aliasing criterion is $2\pi/\mathrm{d}=\pi$. In Fig. 4, this corresponds to the black vertical line. Higher k's are wrapped, and appear as low wave numbers. In this simple case, the wave-number spectrum can be extended by unwrapping the k axis.

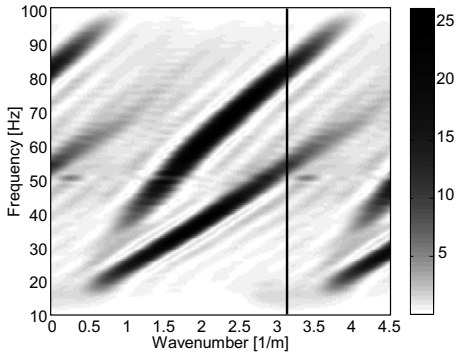


Fig. 4. Frequency–wave-number ($F-K$) transform of the seismic section obtained in Fig. 4(a). The largest measurable wave number according to the aliasing criterion is $2\pi/d$ with $d=2$ m (black vertical line). The aliasing in the $F-K$ diagram is resolved by unwrapping the k axis. The shape of the two modes on the $F-K$ diagram reveals dispersive modes.

In the $F-K$ diagram, modes are separated and extracted by masking a region of the $F-K$ plot and returning to the range/time domain [Fig. 3(b) and 3(c)]. This operation cannot be performed directly in the range/time domain since the two modes are mixed at short ranges.

From Fig. 4, phase velocity curves are measured for each mode separately using the relation $k=2\pi f/c$, where k is the wavelength, f is the frequency, and c is the phase velocity [Fig. 5(b)]. A Monte Carlo inversion is performed simultaneously on the two modes to obtain the most-likely shear velocity profile in the medium[7] [Fig. 5(a)]. The minimum frequency of the surface waves being larger than 20 Hz for a maximal wavelength of 12 m, the model is not constrained for depth deeper than 6 m. Similarly, the P-velocity profile and the density were poorly constrained by the surface wave dispersion curves and arbitrarily set to 900 m/s and 1.5 g/cm^3, respectively.

To improve confidence in the model, synthetic sections are computed separately for the two Rayleigh modes [Fig. 3(b) and 3(c)].[8] The computed traces are very close to the extracted modes. To check the relative weight of each mode, their sum is compared to the original seismic section before mode separation [Fig. 3(a)]. As the complete seismic section is well retrieved, the relative weights of the computed and experimental modes are identical, which

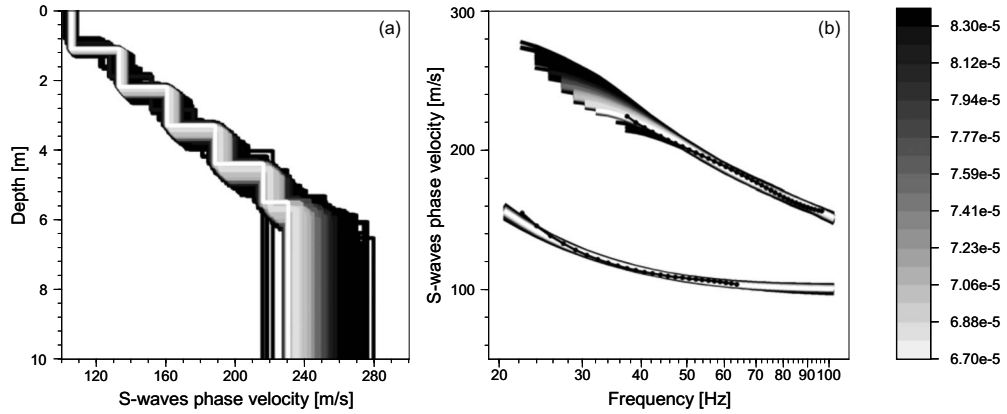


Fig. 5. (a) Distribution of shear velocity profiles obtained from a Monte Carlo inversion using the dispersion curves of the two surface-wave modes. (b) Distribution of phase dispersion curves for the two surface modes in the Monte Carlo inversion. Black spots correspond to the experimental phase-velocity curves obtained from Fig. 5. Colorbar scales correspond to a misfit increase of 24%.

confirms the quality of the inversion model. To increase the precision of the model, a full wave inversion could be performed that would benefit from the high SNR of the seismic section [Fig. 3(a)]. In this case, the inversion kernel should be modified to include both amplitude and time dispersion information.

4. Conclusion

A new method to retrieve surface waves at small scales has been developed. This method, which combines passive processing and active sources is simpler, faster, and easier to implement than classical active seismic techniques. It is also versatile and can be easily adapted to various bottom configurations and expected wavelengths.

A classical seismic inversion was performed using the observed seismic section. The high SNR allows us to get an accurate model for shear velocity in the $(0-7\ \mathrm{m})$ depth.

Future works will investigate the case of two-dimensional (2D) media. In this case the stacking process performed in a 1D case won't be possible, but differences between intercorrelations at the same range but at different places could provide a 2D (range + depth) inversion result.

Acknowledgment

The authors are grateful to Marc Wathelet for his advice in the use of the Monte Carlo inversion software GEOPSY.

References and links

[1]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[2]K. Wapenaar, "Synthesis of an inhomogeneous medium from its acoustic transmission response," Geophysics **68**, 1756–1759, (2003).
[3]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[4]P. Roux, K. G. Sabra, P. Gerstoft, W. A. Kuperman, and M. C. Fehler, "P-waves from cross-correlation of seismic noise," J. Geophys. Res. **32**, 19303 (2005).
[5]P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wavefronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am., **116**, 1995–2003, (2004).
[6]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).
[7]M. Wathelet, D. Jongmans, and M. Ohrnberger, "Surface wave inversion using a direct search algorithm and its application to ambient vibration measurements," Near Surface Geophysics **2**, 211–221 (2004).
[8]R. B. Herrmann, "Computer programs in seismology," Department of Earth and Atmospheric Sciences, Saint Louis University, http://www.eas.slu.edu/People/RBHerrmann/ComputerPrograms.html.

Evoking biphone neighborhoods with verbal transformations: Illusory changes demonstrate both lexical competition and inhibition

James A. Bashford, Jr.,[a] **Richard M. Warren, and Peter W. Lenz**
Department of Psychology, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201-0413
bashford@uwm.edu; rmwarren@uwm.edu; plenz@uwm.edu

Abstract: When a recorded verbal stimulus repeats over and over, perceptual changes occur and listeners hear competing forms. These verbal transformations (VTs) were obtained for a phonemically related set of 24 consonant-vowel syllables that varied widely in frequency-weighted neighborhood density (FWND). Listener's initial transformations involving substitution of consonants versus vowels were strongly correlated with the lexical substitution neighborhood [$r = +0.82$, $p < 0.0001$]. Interestingly, as stimulus FWND increased, average time spent hearing illusory forms substantially decreased [$r = -0.75$, $p < 0.0001$]. These results suggest that VTs not only reveal underlying competitors, but also provide a highly sensitive measure of lexical inhibition.

PACS numbers: 43.71.An, 43.71.Es [EL]
Date Received: September 25, 2007 **Date Accepted:** December 21, 2007

1. Introduction

Modern theories of spoken word recognition typically assume that acoustic-phonetic input activates sets of phonologically similar verbal representations in memory, which compete with one another in the process of word recognition. These activation-competition models of speech perception are supported by evidence from a variety of paradigms (e.g., shadowing, perceptual identification, and lexical decision), which have shown that processing speed and accuracy for a verbal stimulus are influenced by both its neighborhood density (i.e., the number of phonologically similar, lexical neighbors) and by its neighborhood frequency (the sum of the word frequencies of its lexical neighbors). Words having large numbers of high-frequency lexical neighbors are generally processed more slowly and less accurately than words having only a few, low-frequency lexical neighbors (e.g., Goldinger *et al.*, 1989; Luce and Pisoni, 1998; Vitevitch and Luce, 1998, 1999).

Studies dealing with the effects of neighborhood density and frequency typically employ a computational procedure that considers the effective neighbors to be those words in the lexicon that differ from the stimulus by the addition, deletion, or substitution of a single phoneme (Landauer and Streeter, 1973; Luce and Pisoni, 1998). The present study uses an auditory illusion known as the Verbal Transformation Effect (Warren, 1961) as a direct means of identifying salient neighbors, as well as examining their competitive interaction. When listeners are presented with a recorded verbal stimulus, such as a syllable or word, which repeats over and over without change, they typically hear abrupt and compelling illusory changes to other words or syllables (Warren, 1961). These verbal transformations (VTs) appear to result from the operation of two simultaneous processes (Warren, 1996): (1) a repetition-induced adaptation ef-

[a]Author to whom correspondence should be addressed.

fect, which lowers the activation level of the dominant neural representation best fitting the stimulus; and (2) a repetition-induced summation effect that progressively increases the activation level of neural representations that are structurally similar to the stimulus.[1] It is considered that VTs occur when the diminished activation level of the stimulus representation is exceeded by that of the most highly activated competing representation. Hence, it is suggested that reports of apparent changes may provide a direct means of accessing the sets of competing representations that have been inferred to influence the speed and accuracy of performance in a variety of psycholinguistic tasks, as well as in everyday spoken word recognition.

A recent experiment (Bashford, Warren, and Lenz, 2006) explored the use of VTs in the study of neighborhood competition. Listeners were presented with repeating lexical and nonlexical consonant-vowel-consonant (CVC) stimuli that were either very low or very high in frequency-weighted neighborhood density (FWND). About 75% of the initial VTs evoked by high-FWND words and nonwords were lexical neighbors of the stimuli. In contrast, the low-FWND stimuli evoked relatively few lexical neighbors (14%), but did evoke a large percentage (47%) of lexical non-neighbors, which typically differed from the stimuli by 2 phonemes and had very high word frequencies (averaging about 880 wpm). These findings indicate that VTs are strongly influenced by two factors considered to be of central importance in modern theories of word recognition: the structural similarity and word frequency of stimulus competitors. The present study uses VTs to measure the effects of a third factor, lexical inhibition, which is incorporated as an important component of the competitive process in several word-recognition models [e.g., TRACE (McClelland and Elman, 1986), SHORTLIST (Norris, 1994), and PARSYN (Luce *et al.*, 2000)]. The capacity of word representations to inhibit each other, in proportion to their goodness of fit with speech input, could aid word recognition by amplifying activation-level differences among competing representations. However, this mutual inhibition would also lower the activation level of all competitors, including the stimulus representation, thus accounting at least in part for the elevations in response latency and error rate observed for high-density stimuli in a variety of psycholinguistic tasks. Paradoxically, lexical inhibition should improve perceptual accuracy in the VT paradigm. Specifically, it was predicted that the amount of time repeating stimuli are heard accurately should increase with increasing stimulus FWND, producing a decrease in the strength of the VT illusion, since greater numbers of high-frequency neighbors would be expected to exert greater amounts of inhibition, which in turn should lower the activation level of the stimulus representation, and thereby reduce the adapting effects of repeated stimulation. In order to test this prediction, it was decided to use a series of naturally produced stimuli having minimal phonetic complexity: a set of 24 consonant-vowel (CV) syllables, 11 words and 13 nonwords, which represent a factorial pairing of the six English stop consonants [three voiced: /b/, /d/, /g/; and three unvoiced: /p/, /t/, /k/] with each of four steady-state English vowels: /ɑ/, /i/, /u/, and /ɛ/. An additional set of 6 vowel-consonant (VC) syllables paired the series of six English stops with the preceding vowel /u/; this series provided control comparisons to be discussed in the results section. These biphone stimuli cover a large range of FWND (sums of the log frequencies ranging from 1.301 to 90.42), which permits a strong test of the lexical inhibition hypothesis.

The 24 CV stimuli employed in the main experiment not only varied in FWND, but also differed critically in the phonemic composition of their lexical substitution-neighborhoods (i.e., those subsets of neighbors differing from the stimuli via phoneme substitution rather than phoneme addition or deletion). For 12 of the CV syllables, more substitution neighbors differed from the stimulus at the consonant position, and for the remaining 12 syllables, more substitution neighbors differed at the vowel position. Based on the hypothesis that VTs are determined chiefly by lexical neighborhood competition, rather than some form of sublexical processing, it was predicted that initial substitution transformations would generally involve the stimulus phoneme conflicting with the greatest number of lexical substitution neighbors.

2. Method

The 300 listeners in this study (10 groups of 30) were undergraduate students at the University of Wisconsin-Milwaukee who were paid for their participation in sessions lasting about 30 min.

All listeners were native monolingual English speakers who reported having no hearing problems and had normal bilateral hearing, as measured by pure tone thresholds of 20 dB hearing level or better at octave frequencies from 250 to 8000 Hz.

The 24 CV syllables and 6 VC syllables described in the introduction were digitally recorded (44.1 kHz sampling, 16 bit quantization) by a highly practiced announcer having an average voicing frequency of approximately 100 Hz and speech patterns typical of southeast Wisconsin. The monosyllables were produced to fit within a 340 ms capture window, and an additional 250 ms segment of digital silence was then added to each capture to emphasize syllable boundaries and minimize any tendency for perceptual resegmentation [e.g., /ti/ heard as /it/] (see Bashford *et al.*, 2006). The 340 ms stimulus capture, along with the added 250 ms silent gap, was digitally iterated to produce a 5 min test stimulus that provided 508 repetitions.

The 240 listeners employed in the main experiment were randomly divided into 8 groups of 30, and each group received a different set of three CV stimuli, having either the three voiced or three unvoiced stop consonants paired with one of the four vowels. The remaining two groups of 30 listeners, who were presented with the VC stimuli, received the vowel /u/ paired with either the three voiced or three unvoiced stop consonants. The order of stimulus presentation within groups was pseudorandom, with the restriction that each stimulus was presented an equal number of times in each serial position across listeners in a group. Testing was performed in a sound-attenuating chamber, with the VT stimuli delivered diotically through Sennheiser HD 250 Linear II Headphones at a slow-rms peak level of 70 dBA sound pressure level. For each stimulus presented, listeners were instructed to call out what the voice was saying at stimulus onset, and then to call out what the voice was saying anytime a change was heard. Listeners' responses during the 5 min stimulation periods were transcribed by the experimenter.

3. Results and discussion

As in our earlier study (Bashford *et al.*, 2006), listeners' initial reported transformations were used for the primary analysis of forms. Nearly all (99.8%) of these initial forms reported for the 24 CV syllables were either English words or phonotactically legal nonwords. They were categorized as one of four types: (1) lexical neighbors (words differing from the stimulus by a single phoneme); (2) nonlexical neighbors; (3) lexical non-neighbors (words differing by more than one phoneme); and (4) nonlexical non-neighbors. Reports of stimulus neighbors comprised 77% of all initial VT responses, with 48% being lexical neighbors and 29% nonlexical neighbors. Reports of non-neighbors as VTs comprised 23% of initial responses, with 13% being lexical and 10% nonlexical. [2] The frequency of reports for non-neighbors was not correlated with stimulus FWND, either for lexical responses [$r=0.18$, $F(22)=0.74$, $p>0.5$] or for nonlexical responses [$r=0.08$, $F(22)=0.1$, $p>0.70$]. In contrast, report frequencies for lexical and non-lexical neighbors of the stimuli did vary reciprocally with FWND: There was a strong direct correlation between FWND and the report frequency for lexical neighbors [$r=+0.735$, $F(22)=25.9$, $p<.0001$] and a strong inverse correlation with the report frequency for nonlexical neighbors [$r=-0.834$, $F(22)=50.4$, $p<0.0001$].

The increase in reports of lexical neighbors with increasing stimulus FWND is consistent with earlier results from this lab (Bashford *et al.*, 2006), and is also consistent with the hypothesis that VTs reveal the most salient neighbors of repeating stimuli. However, reports of lexical neighbors also would be expected to increase with FWND to some extent by chance alone if VTs were due to random errors occurring at a phonetic level of processing. If transformations are indeed determined by interactions between lexical representations, then the phonetic transformations observed for VT stimuli should be consistent with the composition of their lexical neighborhoods. Specifically, it was predicted that substitution transformations of stimulus consonants versus vowels should be correlated with the proportion of lexical neighbors that differ from the stimuli at the consonant versus vowel positions. Strong confirmation of this predicted pattern was obtained through a regression analysis that compared the proportion of consonant versus vowel substitution transformations for the 24 CV syllables with the proportion of lexical neighbors differing from the stimuli by consonant versus vowel substitution

[(consonant substitution neighbors – vowel substitution neighbors) / total substitution neighbors]. The proportion of initial transformations involving consonant changes increased in a very strong linear fashion ($r=+0.82$, $F(22)=45.2$, $p<0.0001$) as the proportion of consonant-substitution neighbors increased relative to vowel-substitution neighbors. This correlation accounts for about 67% of the variance in consonant versus vowel substitution transformations, despite the inclusion of data for non-neighbor transformations that involved changes in both the consonant and vowel components. Hence, a substantial portion of VTs are clearly not the result of random errors in sublexical processing, but rather are strongly influenced by interactions between lexical representations.

Although substitution transformations were the most frequent initial responses (comprising 63.5% of reported changes), transformations involving the addition of a vowel or consonant (comprising about 30% of responses) occurred with sufficient frequency to permit an examination of the predictability of consonant-addition versus vowel-addition transformations based on the proportion of consonant- versus vowel-addition neighbors. Interestingly, and in contrast with the results obtained for substitution transformations, the latter correlation did not approach significance ($r=0.163$, $F(22)=0.6$, $p>0.4$). This disparity in the effects of substitution- versus addition-neighbors was also found in the analysis of overall illusion strength discussed below.

The final prediction tested in this study was that the strength of the VT illusion, measured in terms of the amount of time stimuli were heard nonveridically during the 300 s repetition period, would decrease with increasing stimulus FWND. This prediction is based on the hypothesis that increasing FWND should result in increased lexical inhibition, which should reduce the activation levels of stimulus representations and thereby attenuate the adapting effects of their repeated activation in the VT paradigm. The results obtained for the 24 CV stimuli clearly conform to this prediction: As FWND increased, there was a highly reliable linear decrease in the average amount of time listeners reported hearing the stimuli nonverdically. This strong negative correlation ($r=-0.746$, $F(22)=27.5$, $p<0.0001$) accounted for approximately 56% of the variance in illusion strength. Additional regression analyses separately examining the data for lexical and nonlexical stimuli confirmed that the negative correlation between FWND and illusion strength was reliable for both the 11 lexical CVs ($r=-0.67$, $F(9)=6.3$, $p<0.03$) and the 13 nonlexical CVs ($r=-0.57$, $F(11)=5.3$, $p<0.05$).

Although the decline in VT strength observed with increasing FWND is consistent with the predicted inhibitory effect of lexical competition, it is necessary to consider a possible sublexical effect of phonotactic probability. Previous studies have shown that speech processing is influenced not only by the density and frequency of lexical neighors (FWND), but also by the position-specific frequency with which stimulus phonemes and phoneme sequences occur within words and syllables of the listener's language (e.g., Vitevitch and Luce, 1998, 1999). Accordingly, these phonotactic probabilities were obtained for the 24 CV stimuli (see Vitevitch and Luce, 2004) and submitted to regression analyses, which indicated that the strength of the VT effect (i.e., nonverdical percept time) was not correlated with the probability of either the consonants ($r=-0.144$, $F(22)=0.46$, $p>0.50$) or the CV biphones ($r=+0.11$, $F(22)=0.264$, $p>0.6$). Correspondingly, there was no correlation between FWND and the probability of the consonants ($r=+0.03$, $F(22)=0.02$, $p>0.90$) or biphones ($r=+0.02$, $F(22)=0.01$, $p>0.90$). However, there was a strong negative correlation between FWND and vowel probability for this set of stimuli ($r=-0.91$, $F(22)=93.3$, $p<0.0001$) and, correspondingly, there was a strong correlation between vowel probability and VT strength ($r=+0.79$, $F(22)=35.7$, $p<0.0001$). Given the absence of correlation between VT strength and the probabilities of either the consonants or biphones, it appeared likely that the significant correlation with vowel probability was the spurious consequence of its very high correlation with FWND. This hypothesis was tested by combining the data for the 6 CV syllables ending with the vowel /u/ with those obtained for the 6 VC syllables that were comprised of the same phonemes in reverse order. For this set of 12 syllables, the correlation between vowel probability and FWND was also very high, but was positive rather than negative in slope ($r=+0.917$, $F(10)=52.9$, p

<0.0001). As expected, the correlation between vowel probability and VT strength was also reversed in direction ($r=-0.69$, $F(10)=8.9$, $p<0.02$), while, in contrast, the correlation between FWND and VT strength was unchanged in direction. Stimulus FWND, the apparent causal factor, accounted for approximately 68% of the variance in VT strength for the matched set of CV and VC syllables ($r=-0.827$, $F(10)=21.6$, $p<0.001$).

The dominance of FWND over phonotactic probability in this study is not surprising given the extended stimulation period employed in the VT paradigm. Effects of phonotactic probability may override those of FWND when stimuli, especially nonwords, are presented in speeded tasks that do not require lexical analysis (e.g., phoneme identification); but FWND has been found to dominate processing for both words and nonwords when lexical analysis (e.g., lexical decision) is required (e.g., Vitevitch and Luce, 1998, 1999). The 5 min stimulus-repetition periods in the present study certainly provided ample time for the activation of lexical representations and the consequent buildup of inhibition.

It should be noted that the effect of FWND on VT strength was not only highly reliable but also substantial: The five CV stimuli having the lowest FWND values (sums of the log frequencies of lexical neighbors averaging 20.8) were heard nonverdically for an average of 248 s (standard error=15.3 s) during the 300 s repetition interval, whereas the five CVs highest in FWND (averaging 86.2) were heard nonveridically for an average of only 155 s (standard error=11.1 s). This mean shift of nearly 40% suggests that the VT strength (i.e., illusion time) measure may provide an especially sensitive index of lexical inhibition.

A final analysis of the VT-strength data for the complete set of 30 biphones examined the relative salience of competition effects exerted by substitution neighbors versus addition neighbors. When the FWND scores for these two types of lexical neighbors were used as predictors in separate regression analyses, significant negative correlations between VT strength and FWND were obtained for both the stimulus substitution-neighborhoods [$r=-0.733$, $F(28)=32.6$, $p<0.0001$] and addition neighborhoods [$r=-0.460$, $F(28)=7.54$, $p<0.02$]. However, the substitution- and addition-neighborhood FWND scores for this set of stimuli were also correlated [$r=0.398$, $F(28)=5.26$, $p<0.03$], so first-order partial correlation coefficients were computed to better determine the individual contributions of the different neighbor types to VT inhibition. When variance in common between substitution- and addition-neighborhood FWND values was partialled out, the correlation between VT strength and substitution-neighborhood FWND remained highly significant [$pr=-0.675$, $t(27)=-4.76$, $p<0.0001$] while that involving addition-neighborhood FWND was nonsignificant [$pr=-0.27$, $t(27)=-1.46$, $p>0.15$]. These results obtained with the VT-strength measure converge with those from listeners' initial VT reports, which showed that the relative frequencies of specific phoneme transformations were strongly correlated with the lexical substitution neighborhoods of the stimuli, but not with their addition neighborhoods. This disparity can also be seen in the types of forms reported in this study. About 44% of addition transformations yielded non-neighbors of the stimuli, either through multiple phoneme additions or through addition combined with a phoneme substitution. In contrast, only about 18% of substitutions yielded non-neighbors, and in most instances these transformations involved a single phoneme substitution combined with one or more phoneme additions. Hence, while substitution transformations conform strongly to the similarity neighborhood of the stimulus, addition transformations do not.

At present it is not clear whether the null effects obtained for addition-neighborhood composition in this study reflect a general lack of salience of addition neighbors in speech processing. It is possible that addition neighbors do effectively compete with substitution neighbors under normal listening conditions but become ineffective when stimuli are repeated for an extended period. It should be noted that the VT paradigm does appear to be inappropriate for the study of possible deletion-neighborhood effects: Pitt and Shoaf (2001, 2002) have shown that word repetition, especially with a very brief interstimulus interval (ISI) , evokes the general auditory process of stream segregation (Bregman, 1990), which can cause spectrally distinct consonants such as stops (especially unvoiced stops) to split from the remainder of an utterance and be heard as a separate background stream. This segregation effect can also be incomplete

(Pitt and Shoaf, 2002), so that a portion of stop-consonant aspiration may remain grouped with a following vowel, resulting in the substitution of /h/ (e.g., /pi/ heard as /hi/). In the present study, deletion transformations were reported infrequently (6% of initial changes), probably due to the 250 ms ISI separating stimulus repetitions. Consistent with Pitt and Shoaf (2002), simple deletion responses occurred most frequently for the voiceless stops, and most often when they were syllable final (VC stimuli). Moreover, about 1% of transformations involved substitution of /h/ for the stop consonants, all in syllable-initial position and most unvoiced.

Deletion transformations may be too heavily confounded with auditory streaming to provide useful information about speech-specific processing. However, substitution transformations are clearly very sensitive to a lexical level of analysis: In the present study, stimulus variation in substitution-neighborhood composition accounted for nearly 70% of the variance in phonetic substitution transformations, and 50% of the variance in overall VT illusion time. These values compare favorably with those obtained using other psycholinguistic tasks, in which neighborhood FWND, the most potent stimulus variable, seldom accounts for more than 20% of performance variance (e.g., Luce and Pisoni, 1998).

The large inhibitory effect of neighborhood density on VTs was not restricted to the illusion-time measure, but also was observed for the number of different forms reported by individual listeners. Forms decreased in a strong linear fashion with increasing FWND [$r = -0.74$, $F(22) = 28.5$, $p < 0.0001$]. Interestingly, a paper by Yin and MacKay (1992, cited by MacKay *et al.*, 1993) reported the opposite result for a set of 20 words varying in neighborhood density, which suggests that the dominance of neighborhood inhibition observed with simple biphone stimuli in the present study may give way to dominance by activated competitors—producing a greater range of transformations—when stimulus complexity is increased. This possibility, which is consistent with the present finding that neighborhood effects upon VTs are chiefly manifested through substitution transformations, will be examined in further planned work that will systematically increase syllabic complexity, beginning with CVCs composed of the same stops and vowels used for the biphones of this study.

Acknowledgments

This work was supported in part by Grant No. R01DC000208 from the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

References and links

[1]For a discussion of repetition-induced verbal summation occurring independently of adaptation, see Warren *et al.* (1996).

[2]Verbal transformations involving multiple phonemic changes have been found to decrease in frequency with the age of the listener, from childhood through old age (Warren and Warren, 1966), indicating that stimulus neighborhoods become phonetically more restricted throughout the lifespan.

Bashford, J. A., Jr., Warren, R. M., and Lenz, P. W. (**2006**). "Polling the effective neighborhoods of spoken words with the verbal transformation effect," J. Acoust. Soc. Am. **119**, EL55–EL59.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound*. (MIT Press, Cambridge, MA).

Goldinger, S. D., Luce, P. A., and Pisoni, D. B. (**1989**). "Priming lexical neighbors of spoken words: Effects of competition and inhibition," J. Mem. Lang. **28**, 501–518.

Landauer, T., and Streeter, L. A. (**1973**). "Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition," J. Verbal Learn. Verbal Behav. **12**, 119–131.

Luce, P. A., Goldinger, S. D., Auer, E. T. J., and Vitevitch, M. S. (**2000**). "Phonetic priming, neighborhood activation, and PARSYN," Percept. Psychophys. **62**, 615–662.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.

MacKay, D. G., Wulf, G., Yin, C., and Abrams, L. (**1993**). "Relations between word perception and production: New theory and data on the verbal transformation effect," J. Mem. Lang. **32**, 624–646.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cognit Psychol. **18**, 1–86.

Norris, D. (**1994**). "Shortlist: A connectionist model of continuous speech recognition," Cognition **52**, 189–234.
Pitt, M. A., and Shoaf, L. (**2002**). "Linking verbal transformations to their causes," J. Exp. Psychol. Hum. Percept. Perform. **28**, 150–162.
Pitt, M. A., and Shoaf, L. (**2001**). "The source of lexical bias in the verbal transformation effect," Lang. Cognit. Processes **16**, 715–721.
Vitevitch, M. S., and Luce, P. A. (**1998**). "When words compete: Levels of processing in perception of spoken words," Psychol. Sci. **9**, 325–329.
Vitevitch, M. S., and Luce, P. A. (**1999**). "Probabilistic phonotactics and neighborhood activation in spoken word recognition," J. Mem. Lang. **40**, 374–408.
Vitevitch, M. S., and Luce, P. A. (**2004**). "A Web-based interface to calculate phonotactic probability for words and nonwords in English," Behav. Res. Methods Instrum. Comput. **36**, 481–487.
Warren, R. M. (**1961**). "Illusory changes of distinct speech upon repetition—the verbal transformation effect," Br. J. Psychol. **52**, 249–258.
Warren, R. M. (**1996**). "Auditory illusions and perceptual processing of speech," in *Principles of Experimental Phonetics*, by N. J. Lass (Ed.). (Mosby, St. Louis), pp. 435–466.
Warren, R. M., Healy, E. W., and Chalikia, M. H. (**1996**). "The vowel-sequence illusion: Intrasubject stability and intersubject agreement of syllabic forms," J. Acoust. Soc. Am. **100**, 2452–2461.
Warren, R. M., and Warren, R. P. (**1966**). "A comparison of speech perception in childhood, maturity, and old age by means of the verbal transformation effect," J. Verbal Learn. Verbal Behav. **5**, 142–146.
Yin, C., and MacKay, D. G. (**1992**). "Auditory illusions and aging: Transmission of priming in the verbal transformation paradigm," Poster presented at the *Cognitive Aging Conference*, Atlanta.

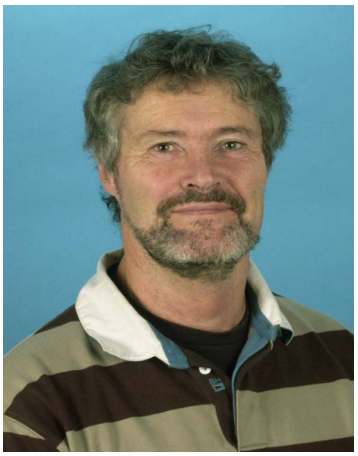
ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America

Dani Byrd—For research on the relation of linguistic structures to the temporal realization of speech
Photo by donmilici@earthlink.net



Brian R. Glasberg—For characterization of frequency selectivity and modeling of loudness

Brad H. Story—For advancing the understanding of the human vocal tract

Jennell Vick named recipient of the 2007 Stetson Scholarship

ASA member Jennell Vick was selected the recipient of the 2007-08 Raymond H. Stetson Scholarship in Phonetics and Speech Production. Jennell is a graduate student in the Department of Speech and Hearing Sciences at the University of Washington in Seattle.

She received a B.S. in Hearing and Speech Sciences from Ohio University and an M.A. in Communication Sciences from Case Western Reserve University.

The Stetson Scholarship, which was established in 1998, honors the memory of Professor Raymond H. Stetson, a pioneer investigator in phonetics and speech science. Its purpose is to facilitate the research efforts of promising graduate students and postgraduates. The Scholarship includes a $4,000.00 USD stipend for one academic year.

Applications for the award are due in March each year. For further information about the award, please contact the Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org/fellowships.html

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.

2008

29 June–4 July	Joint Meeting of the Acoustical Society of America, European Acoustics Association and the Acoustical Society of France, Paris, France) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
28 July–1 Aug	9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609, Tel. 860-572-0680; Web: www.icben.org. E-mail icben2008@att.net
10–14 Nov	156th Meeting of the Acoustical Society of America, Miami, FL [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].

2009

18–22 May	157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

 0001-4966/2008/123(3)/1211/2/$23.00

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

Volumes 1–10, 1929–1938: JASA, and Contemporary Literature, 1937–1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; Nonmembers $10

Volumes 11–20, 1939–1948: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 395. Out of Print

Volumes 21–30, 1949–1958: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75

Volumes 31–35, 1959–1963: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90

Volumes 36–44, 1964–1968: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print.

Volumes 36–44, 1964–1968: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print

Volumes 45–54, 1969–1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound)

Volumes 55–64, 1974–1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound)

Volumes 65–74, 1979–1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound)

Volumes 75–84, 1984–1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound)

Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound)

Volumes 95–104, 1994–1998: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632, Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound)

Volumes 105–114, 1999–2003: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 616, Price: ASA members $50; Nonmembers $90 (paperbound)

ACOUSTICAL STANDARDS NEWS

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

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

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

Standards Meetings Calendar–National

•30 June 2008

ASA Committee on Standards (ASACOS) and ASACOS Steering will meet in Paris in conjunction with the 155th meeting of the Acoustical Society of America.

Note that there will be NO standards committee meetings in Paris in June 2008.

•10–14 November 2008

Meetings of the National Standards Committees S1, Acoustics; S2, Mechanical Vibration and Shock; S3, Bioacoustics; S3/SC1, Animal Bioacoustics; and S12, Noise, and the U.S. TAGs to ISO/TC 43, ISO/TC 43/SC 1, ISO/TC 108 and its subcommittees, will be held in conjunction with the 156th meeting of the Acoustical Society of America in Miami, Florida.

Specific meeting dates and times will be announced soon.

Standards Meetings Calendar—International

19–23 May 2008-Stockholm, Sweden
•IEC/TC 29 Electroacoustics
26–30 May 2008-Boras, Sweden
•ISO/TC43 Acoustics, and ISO/TC 43/SC 1 Noise
27–31 May 2008-Kyoto, Japan
•ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines
3–7 November 2008-St. Louis, MO
•ISO/TC 108, Mechanical vibration, shock and condition monitoring
•ISO/TC 108/SC 3, Use and calibration of vibration and shock measuring instruments
•ISO/TC 108/SC 6, Vibration and shock generating systems

Photos from New Orleans

All the Standards Committees, the TAGs, and many of the Working Groups met in New Orleans. Here are photos from just a few of them with apologies to those whose meetings are not represented here.

Fig. 1. S3/WG 87 Manikins.

Fig. 2. The first meeting of S3/SC 1 Animal Bioacoustics.

 0001-4966/2008/123(3)/1213/4/$23.00

Fig. 3. S12/WG 46 relocatable classrooms.

Fig. 4. ASACOS Steering Group.

Fig. 5. Accredited Standards Committee S1, Acoustics.

Fig. 6. ASACOS.

Fig. 7. ASACOS #2.

Standards News from the United States

Call for Members in Standards Committees

The Acoustical Society of America Standards Secretariat is seeking organizational members in the categories of User Interest, Government Interest, Trade Association, and/or General Interest for Accredited Standards Committees S1–Acoustics, S2–Mechanical Vibration and Shock, S3–Bioacoustics, S3/SC1–Animal Bioacoustics, and S12–Noise, as well as for participation in the U.S. TAGS to IEC/TC 29–Electroacoustics, ISO/TC 43–Acoustics, ISO/TC 43/SC 1–Noise, and ISO/TC 108 and all of its subcommittees.

If your company or organization is interested in becoming a member of any of these committees, please contact Susan Blaeser, Standards Manager, ASA Standards Secretariat, (631) 390–0215 or sblaeser@aip.org for additional information.

American National Standards Call for Comment on Proposals Listed

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

ASME (American Society of Mechanical Engineers)

Reaffirmations

BSR/ASME Y32.18-1972 (R200x), Symbols for Mechanical and Acoustical Elements as Used in Schematic Diagrams (reaffirmation of ANSI/ASME Y32.18-1972 (R2003))

Presents standard symbols and definitions that may be used in constructing schematic diagrams for mechanical and acoustical systems whose performances are describable by finite sets of scalar variables (29 January 2008)

ATIS (Alliance for Telecommunications Industry Solutions)

Revisions

BSR/ATIS 0300255-200x, In Service, Nonintrusive, Measurement Device (INMD)—Methodology for Applying INMD Measurements to Customer Opinion Models (revision of ANSI T1.255-2003)

Allows INMD measurements to be used to evaluate the performance of telecommunications connections and services and to detect speech level, noise, and echo anomalies on telecommunications connections. (11 February 2008)

SCTE (Society of Cable Telecommunications Engineers)

New Standards

BSR/SCTE 24-23-200x, BV32 Speech Codec Specification for Voice over IP Applications in Cable Telephony (new standard)

Contains the description of the BV32 speech codec1. BV32 compresses 16-kHz sampled wideband speech to a bit rate of 32 kb/s (kilobits per second) by employing a speech coding algorithm called Two-Stage Noise Feedback Coding (TSNFC), developed by Broadcom. (11 February 2008)

UL (Underwriters Laboratories, Inc.)

Revisions

BSR/UL 217-200x, Single and Multiple Station Smoke Alarms (revision of ANSI/UL 217-2006)

Proposes the following changes: Voice messages in alarm signals; smoke box air circulation fans; jarring test; sound output measurement; temporary alarm silencing; battery operational temperature ranges for RV and marine alarms; installation instructions; and situations where smoke alarms may not be effective. (17 December 2007)

Final Actions on American National Standards

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

ASA (ASC S3) (Acoustical Society of America)

Reaffirmations

ANSI S3.46-1997 (R2007), Methods of Measurement of Real-Ear Performance Characteristics of Hearing Aids (reaffirmation of ANSI S3.46-1997 (R2002))

ASA (ASC S12) (Acoustical Society of America)

Revisions

ANSI/ASA S12.9-Part 5-2007, Quantities and Procedures for Description and Measurement of Environmental Sound—Part 5: Sound Level Descriptors for Determination of Compatible Land Use (revision of ANSI S12.9-Part 5-1998 (R2003))

Withdrawals

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

SCTE (Society of Cable Telecommunications Engineers)

New Standards

ANSI/SCTE 24-22-2007, iLBCv2.0 Speech Codec Specification for Voice over IP Applications in Cable Telephony (new standard)

Project Initiation Notification System (PINS)

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

The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

ATIS (Alliance for Telecommunications Industry Solutions)

BSR/ATIS 0300255-200x, In-Service, Nonintrusive Measurement Device (INMD)—Methodology for Applying INMD Measurements to Customer Opinion Models (revision of ANSI T1.255-2003) This document is intended for use as the North American standard for mapping measurements made with In-Service, Nonintrusive Measurement Devices (INMDs) to the parameters used in customer opinion models for voice services. Project Need: To allow INMD measurements to be used to evaluate the performance of telecommunications connections and services and to detect speech level, noise, and echo anomalies on telecommunications connections. Stakeholders: Telecommunications Industry.

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

BSR C63.4-200x, Measurement of Radio-Noise Emissions from Low-Voltage Electrical and Electronic Equipment in the Range of 9 kHz to 40 GHz (revision of ANSI C63.4-2003)

Revises ANSI C63.4 to address the following specific issues and other issues as may be raised by the revision task group. (A) Add to Figure 2 (LISN) a tabular list in addition to the graph; (B) Clarify paragraphs 11.1.3(e) and (f) on exercise of displays; (C) In section 12.2.1, replace "5 V peak-to-peak VITS signal supplied through the video input port." with "the highest input voltage where video is still processed or present, or visible on the TV screen without error, should be recorded and tested"; (D) LISN calibration procedures; and (E) Definition of cable loss as a function of temperature. Project Need: To revise ANSI C63.4 to address issues raised by the revision task group. Stakeholders: EMC test laboratories, EMC test equipment manufacturers (software designers).

BSR C63.19 Amendment-200x, Methods of Measurement of Compatibility between Wireless Communications Devices and Hearing Aids (addenda to ANSI C63.19-2007)

Supplements and revises portions of ANSI C63.19-2007 to deal with the topics listed in the project description. Project Need: To cover range between 698 MHz to 6 GHz and to revise the methods of measurement. Stakeholders: Manufacturers of cellular phones and hearing aids, service providers, hearing aid wearers, regulators.

TIA (Telecommunications Industry Association)

BSR/TIA 855-A-200x, Telecommunications—Telephone Terminal Equipment—Stutter Dial Tone Detection Device Performance Requirements (revision of ANSI/TIA 855-2001)

Revises ANSI/TIA/EIA-855-2001 to update the noise immunity requirements, provide missing tolerances for several requirements, and to clarify several ambiguous requirements. Project Need: To update the noise immunity requirements, to provide missing tolerances for several requirements, and to clarify several ambiguous requirements. Stakeholders: Telecommunications Industry Association.

Standards News from Abroad

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

Newly Published ISO and IEC Standards

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

ISO Standards

BASES FOR DESIGN OF STRUCTURES (TC 98)

ISO 10137:2007, Bases for design of structures—Serviceability of buildings and walkways against vibrations

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 2017-2:2007, Mechanical vibration and shock—Resilient mounting systems—Part 2: Technical information to be exchanged for the application of vibration isolation associated with railway systems

ISO 8041/Cor1:2007, Human response to vibration—Measuring instrumentation—Corrigendum

PLASTICS (TC 61)

ISO 6721-9/Amd1:2007, Plastics—Determination of dynamic mechanical properties—Part 9: Tensile vibration—Sonic-pulse propagation method—Amendment 1: Precision

ROLLING BEARINGS (TC 4)

ISO 15242-4:2007, Rolling bearings—Measuring methods for vibration—Part 4: Radial cylindrical roller bearings with cylindrical bore and outside surface

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

ISO 23600:2007, Assistive products for persons with vision impairments and persons with vision and hearing impairments—Acoustic and tactile signals for pedestrian traffic lights

ISO Technical Reports

ACOUSTICS (TC 43)

ISO/TR 25417:2007, Acoustics—Definitions of basic quantities and terms

IEC Standards

ELECTROACOUSTICS (TC 29)

IEC 60318-6 Ed. 1.0 b:2007, Electroacoustics—Simulators of human head and ear—Part 6: Mechanical coupler for the measurement on bone vibrators

PERFORMANCE OF HOUSEHOLD ELECTRICAL APPLIANCES (TC 59)

IEC 60704-2-14 Ed. 1.0 b:2007, Household and similar electrical appliances—Test code for the determination of airborne acoustical noise —Part 2-14: Particular requirements for refrigerators, frozen-food storage cabinets and food freezers

IEC 62552 Ed. 1.0 b:2007, Household refrigerating appliances—Characteristics and test methods

ENVIRONMENTAL CONDITIONS, CLASSIFICATION AND METHODS OF TEST (TC 104)

IEC 60068-2-6 Ed. 7.0 b:2007, Environmental testing—Part 2-6: Tests—Test Fc: Vibration (sinusoidal)

IEC Technical Specifications

ULTRASONICS (TC 87)

IEC/TS 61949 Ed. 1.0 en:2007, Ultrasonics—Field characterization—In situ exposure estimation in finite-amplitude ultrasonic beams

ISO and IEC Draft International Standards

ISO

ACOUSTICS (TC 43)

ISO/DIS 11201, Acoustics—Noise emitted by machinery and equipment—Measurement of emission sound pressure levels at a work station and at other specified positions in an essentially free field over a reflecting plane with negligible environmental corrections (8 February 2008)

ISO/DIS 11202, Acoustics—Noise emitted by machinery and equipment—Measurement of emission sound pressure levels at a work station and at other specified positions applying approximate environmental corrections (8 February 2008)

ISO/DIS 11204, Acoustics—Noise emitted by machinery and equipment—Measurement of emission sound pressure levels at a work station and at other specified positions applying accurate environmental corrections (8 February 2008)

ISO/DIS 17201-3, Acoustics—Noise from shooting ranges—Part 3: Guidelines for sound propagation calculations (16 February 2008)

IEC Draft Standards

104/448/FDIS, IEC 60068-2-27 Ed. 4.0: Environmental testing—Part 2-27: Tests—Test Ea and guidance: Shock (1 February 2008)

46A/875/FDIS, IEC 61196-1-325: Coaxial communication cables—Part 1-325: Mechanical test methods—Aeolian vibration (11 January 2008)

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.

Ultrasonic Processes and Machines: Dynamics, Control and Applications

Vladimir K. Astashev and Vladimir I. Babitsky

Springer-Verlag, Berlin and Heidelberg, 2007. 332 pp. price $169.00 (hardcover), ISBN: 978-3-540-72060-7.

Ultrasonic machining is an unconventional mechanical process used for low-rate material removal in otherwise hard to do cases. It is distinctly different from other specialized material removal processes, such as electrical discharge and electrochemical machining, in that its application is not limited by the electrical or chemical characteristics of the workpiece material. Ultrasonic machining can be used on both conductive and nonconductive materials, especially those with low ductility and high hardness. Following Paul Langévin's groundbreaking work, the history of ultrasonic machining began with a paper by Robert Williams Wood and Alfred Lee Loomis in 1927 and took off in earnest after the Second World War. Pioneering work was conducted in the former Soviet Union by the authors of this volume following the footsteps of other prominent members of the Russian school including O. V. Abramov, O. I. Babikov, A. I. Markov, L. G. Merkulov, V. P. Severdenko, and others. Some of this seminal work was previously unavailable in English for the expanding international community of researchers, graduate students, and industrial practitioners working in ultrasonics. The Springer series entitled *Foundations of Engineering Mechanics* publishes scientific monographs and graduate-level textbooks on various topics of applied mechanics. A unique feature of this series is that it presents selected works of Russian and Eastern European scientists that had been previously unpublished in the West. The volume entitled *Ultrasonic Processes and Machines* was expertly translated from Russian into English by Karima Khusnutdinova.

The physical processes involved in ultrasonic machining include a number of specific effects. Peculiar behavior is exhibited by materials in the presence of high-intensity ultrasonic vibration, for example, dry friction in the region of contact between two compressed surfaces transforms into viscous friction and displays substantially decreased yield strength. Chapter 1 presents quantitative models of the ultrasonic process based upon independently measurable macroscopic parameters such as the compressive force and cutting speed of the machine, the elastic modulus, yield strength, and friction coefficient of the material, and the ultrasonic vibration velocity of the tool. The authors illustrate that macroscopic phenomenological models can describe surprisingly well the ultrasonic machining process without microscopic details of the local material/acousto-mechanical interaction. Chapter 2 focuses on the particular dynamic characteristics of ultrasonic machines with special emphasis placed on the most fundamental longitudinal and bending tool vibrations and the operation of magnetostrictive and piezoelectric actuators. Chapter 3 addresses the problem of nonlinear interaction between the tool and the workpiece, especially in the presence of contact gaps that lead to impact phenomena, and the complex influence of nonlinearity on the resonant behavior of the whole system. The efficiency of ultrasonic machining crucially depends on the ultrasonic vibration velocity being larger than the cutting speed. At a cutting speed of 10 m/s, this threshold can be achieved above 20 kHz using ultrasonic displacements in excess of 100 μm, wherein lies the main challenge of ultrasonic machining. Even at driving power levels as high as a few kilowatts, adaptive resonant tuning is required to maintain the necessary vibration amplitude in the presence of highly variable process parameters. Chapter 4 discusses the sensitivity of machining efficiency to nonlinear loading variations and the operation of automatic frequency control systems using different feedback strategies. Finally, Chapter 5 presents numerous practical examples of ultrasonically assisted machining including cutting, turning, and drilling.

Is this volume a valuable edition to the existing extensive literature on high-power applications of ultrasonics? Considering the applied nature of the topic, the analysis is a little too mathematical, and the presentation is heavily loaded by algebraic derivations. The reader probably will be put off by the oddity of the format, for example, the mostly low-quality figures and the annoying lack of figure captions that makes it particularly difficult to search for the relevant details in the surrounding text. There are numerous typographic errors and inconsistencies throughout the book, and in some sections the text seems to refer to equation numbers in an earlier publication that are completely different from the present one. About 40% of the cited references are published only in Russian and are 40 years old on the average with the latest one published back in 1987, i.e., 20 years ago. Although parts of the book, especially Chapter 5, have been obviously updated by more recent references from the international literature, these new results are only barely touched upon in the text without in-depth discussion, and the extensive reference list at the end of the book is still less than adequate for current researchers. All in all, the book is a hard read, but a worthwhile addition to the literature that can be recommended to anybody who is specializing in the physical effects and practical applications of ultrasonics. Unquestionably, Astashev and Babitsky's new volume reads like an old book in a new edition that still shows its age, but it also reveals the authors' distinctive insight that could greatly benefit a new generation of readers.

PETER B. NAGY
Department of Aerospace Engineering
and Engineering Mechanics
University of Cincinnati
Cincinnati, Ohio 45221-0070

 0001-4966/2008/123(3)/1217/1/$23.00

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
GEOFFREY EDELMANN, *Naval Research Laboratory, Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,214,871

43.38.Fx TRANSMITTER FOR SOUND RECORDING OF AN ELECTRIC SIGNAL FROM AN ACOUSTIC DRUM

Jürgen Hasenmaier, 89542 Herbrechtingen, Germany
8 May 2007 (Class 84/723); filed in Germany 8 July 2003

The inventor proposes a design for using a piezoelectric sensor that can be clamped onto the rim of a drum. Particular attention is given to the shape of the sensor since the author believes that a sector of a circle is the preferred shape to eliminate the proximity effect of rolls near the rim. Of course, no theory or measurement is given to demonstrate why this might be the case. At least one of the figures was taken directly from European Patent No. EP 0542706 ("An acoustic drum transmitter and a holder therefor"). A much earlier but still relevent reference is U.S. Patent No. 5,449,964.—MK

7,266,040

43.30.Gv ACTIVE SONAR SIMULATION

Ronald Frances Norwood and Cathy Christensen Matthews, assignors to United States of America as represented by the Secretary of the Navy
4 September 2007 (Class 367/13); filed 10 August 2005

For the purpose of training sonar operators, a system is described that simulates realistic sounding reverberation from an underwater acoustic source in a given environment. The real-time simulations are not exact. The training system makes various visual displays of the modeled noise.—GFE

7,266,042

43.30.Vh MULTISTAGE MAXIMUM LIKELIHOOD TARGET ESTIMATOR

Kurt Gent and Kenneth J. McPhillips, assignors to The United States of America as represented by the Secretary of the Navy
4 September 2007 (Class 367/118); filed 31 March 2006

An algorithm is described that provides range and bearing estimates for target acquisition by removing time estimate errors and tracking with a coarse grid search followed with Gauss–Newton and then Cartesian coordinate maximum-likelihood estimation.—GFE

7,266,044

43.30.Wi METHOD AND APPARATUS FOR ACOUSTIC SOURCE TRACKING USING A HORIZONTAL LINE ARRAY

Tsih C. Yang, assignor to United States of America represented by the Secretary of the Navy
4 September 2007 (Class 367/124); filed 17 December 2004

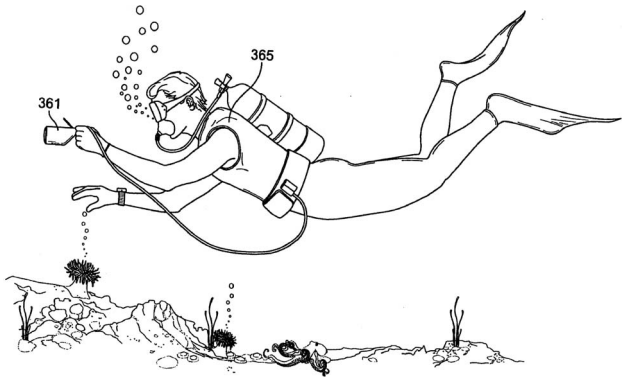
Adaptive array processing is sensitive to target motion when the number of samples is limited; a method to compensate for motion in the beam domain is described.—GFE

7,272,075

43.30.Vh PERSONAL SONAR SYSTEM

Matthew Pope, Los Angeles, California
18 September 2007 (Class 367/131); filed 10 October 2006

With the fear of sharks the motivation, this document describes a surfboard or scuba mounted system that somehow detects only large animals



and warns the water sports enthusiast. It also, more tractably, warns him when he drifts away from his friends.—GFE

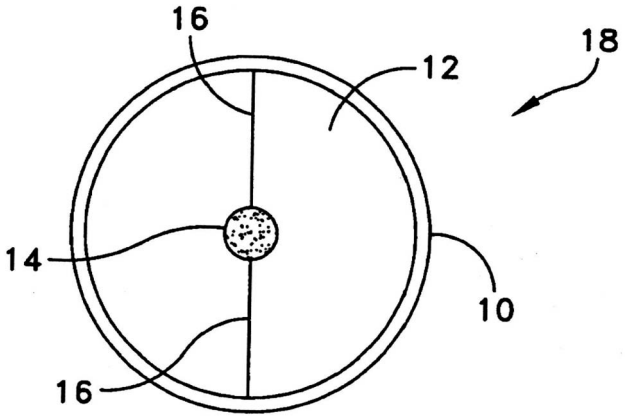
 0001-4966/2008/123(3)/1219/18/$23.00

7,266,046

43.30.Yj MINIATURE LOW FREQUENCY ACOUSTIC TRANSMITTER

Anthony A. Ruffa, assignor to The United States of America as represented by the Secretary of the Navy
4 September 2007 (Class 367/171); filed 26 June 2006

Compact low-frequency transducer **18** is comprised of air-filled balloon **14** (other gases can be used to obtain different frequencies), which is held by tethers **16** in water **12** filled pressure vessel **10** (although other fluids can be used). The walls of pressure vessel **10** need to be on the order of several radii away from miniature balloon **14**. The system is driven into resonance by using transducers that are mounted to the pressure vessel or by



using a piston to drive fluid into and out of the pressure vessel. The resonant system thus formed is said be very effective at low frequencies, as the effective stiffness of the system is that of the air bubble while its effective mass is that of the surrounding water. An open system is also described. The patent is clear and to the point with mathematical formulas included in the detailed description.—NAS

7,269,096

43.30.Wi IDENTIFICATION AND LOCATION OF AN OBJECT VIA PASSIVE ACOUSTIC DETECTION

Rhonda L. Millikin, assignor to Her Majesty the Queen in Right of Canada as represented by the Minister of National Defense
11 September 2007 (Class 367/127); filed 24 February 2003

Localization via time difference of arrival by any other name is still apparently easy to submit as novel.—GFE

7,275,436

43.35.Yb METHOD AND APPARATUS FOR MEASURING FILM THICKNESS AND FILM THICKNESS GROWTH

Scott F Grimshaw, assignor to Tangidyne Corporation
2 October 2007 (Class 73/579); filed 21 October 2005

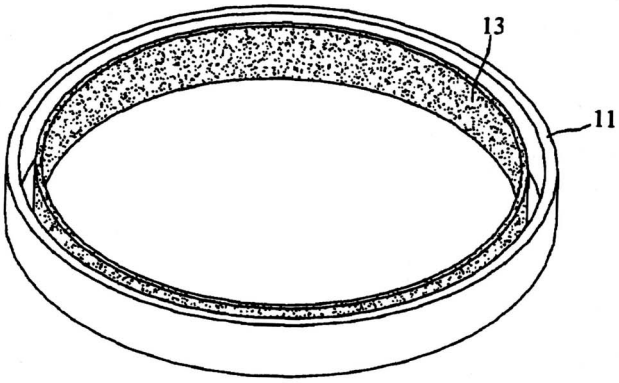
This patent discloses the use of IT-cut quartz crystals for film thickness monitoring in a vacuum chamber environment. Usually, an AT-cut quartz crystal is used in thin-film deposition systems as a process monitor, wherein the oscillation frequency of the crystal drops as the mass (thickness) of the film being deposited increases. In order to maintain gauging accuracy, the AT-cut crystals must be cooled to 20 °C–25 °C in the chamber. The authors disclose that IT-cut crystals can operate stably at temperatures of 100 °C with consequently simpler temperature regulation. A second advantage of this high-temperature operation is that the stability of the oscillation frequency with high-stress films like magnesium fluoride is increased. Why this is so is not explained.—JAH

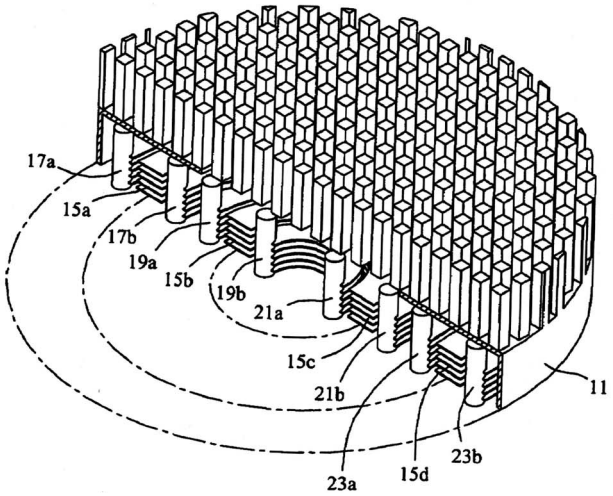
7,278,271

43.35.Ud RADIAL HIGH ENERGY ACOUSTIC DEVICE AND THE APPLIED THERMOACOUSTIC DEVICE

Ya-Wen Chou and Ming Shan Jeng, assignors to Industrial Technology Research Institute
9 October 2007 (Class 62/6); filed in Taiwan 27 December 2004

This very short patent explains that a thermoacoustic device operates by immersing a heat exchange stack in a strong acoustic standing wave, which causes heat to be transferred from one end of the stack to the other. In one type of prior art device, a transducer is placed at one end of a cylindrical resonance tube. In another type, radial standing waves are generated by a





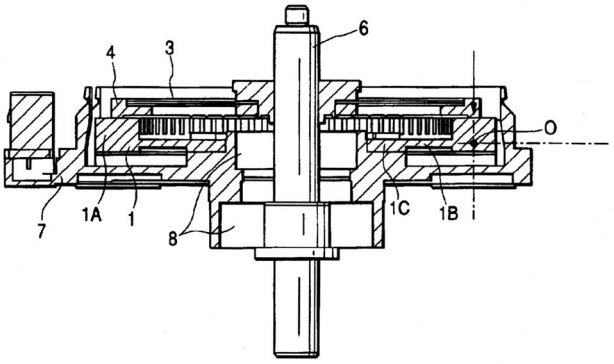
small transducer at the center of a shallow drum-shaped cavity. The patent asserts that greater acoustic efficiency can be achieved by employing a ring-shaped transducer **13** near the outer boundary **11**, thus increasing cooling capacity without increasing electrical power consumption.—GLA

7,245,058

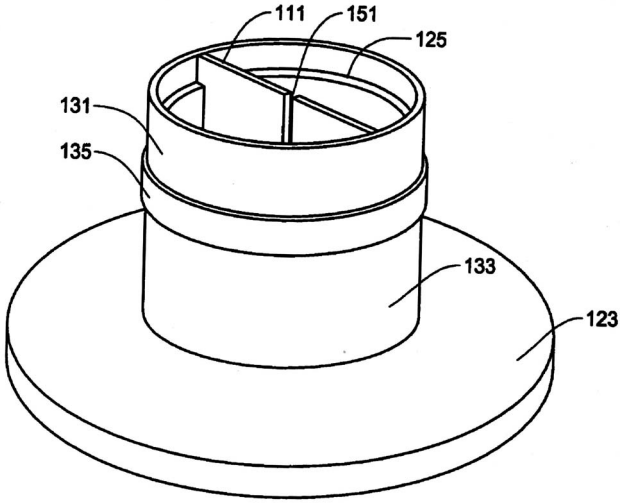
43.35.Zc VIBRATION WAVE DRIVING APPARATUS AND METHOD OF SETTING SHAPE OF SUPPORT MEMBER SUPPORTING ELASTIC MEMBER FORMING VIBRATION MEMBER OF VIBRATION WAVE DRIVING APPARATUS

Kosuke Fujimoto et al., assignors to Canon Kabushiki Kaisha
17 July 2007 (Class 310/317); filed in Japan 7 December 2001

An ultrasonic vibrating wave motor producing very smooth turning motion that is not prone to abnormal wear or squeaks due to single point contact is claimed. Rotation on axis **6** is achieved as follows: Stator **1** is driven in ninth order out of plane bending mode by piezoelectric transducers



attached to its bottom side. Rotor **4**, pressed to stator **1** by a spring, is caused to move via a fifth order torsional mode of stator vibration. Stator dimensions are proportioned to have these two modes at near the same frequency. Single point contact, said to cause excess wear, is minimized by driving the plate at a frequency just below its modal resonance frequency, e.g., at 35 kHz when its modal resonance is 45 kHz.—AJC



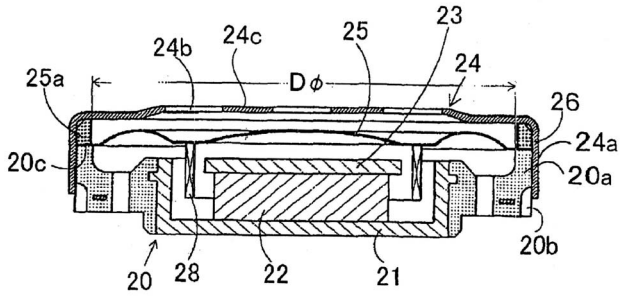
is provided to maintain the air gap in the magnetic circuit. Heat is conducted from ring **131** to heat dissipation plate **111**. Both pole piece **125** and heat dissipation plate **11** have slots to reduce electric currents therein.—NAS

7,266,213

43.38.Dv COMPACT SPEAKER WITH A PROTECTIVE COVER

Masahito Furuya, assignor to Citizen Electronics Company, Limited
4 September 2007 (Class 381/398); filed in Japan 30 October 2001

Protective cover **24** extends over the peripheral projector **20a** of case **20**. Diaphragm **25b** is held between hold ring **26** in protector **24** and a



peripheral projection **20a** of case **20** at peripheral portion **25a**. The cover **24** is secured to case **20** by an adhesive.—NAS

7,272,238

43.38.Dv LOUDSPEAKER HAVING COOLING SYSTEM

Jason Kemmerer and Robert True, assignors to Alpine Electronics, Incorporated
18 September 2007 (Class 381/414); filed 12 October 2004

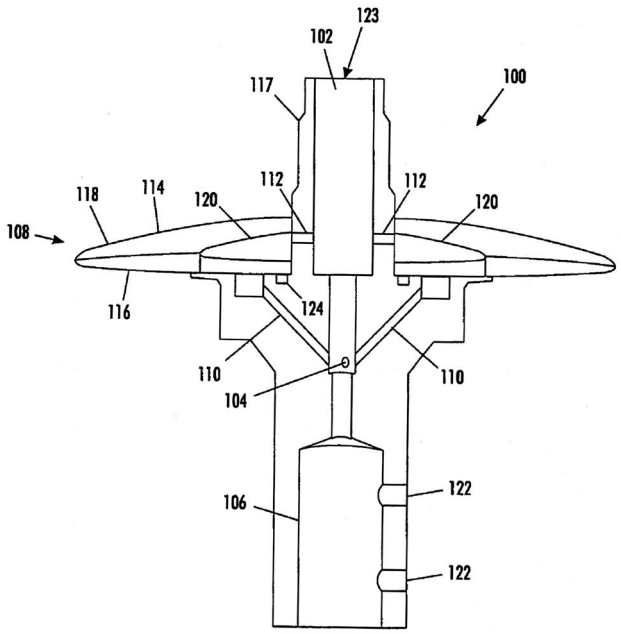
To reduce distortion, electrically and thermally conductive but non-magnetic shorting rings **131** and **133** are positioned on pole piece **125**. Steel ring **135**, which has magnetic properties that can interact with the voice coil,

7,244,377

43.38.Fx ACOUSTIC ENHANCEMENT OF PARTICLE FABRICATION BY SPINNING

Eric J. Shrader, assignor to Palo Alto Research Center Incorporated
17 July 2007 (Class 264/8); filed 21 October 2005

A particle production spinning apparatus is claimed where spinning disk **108** is also vibrating via electroactuator material **120**. The vibration



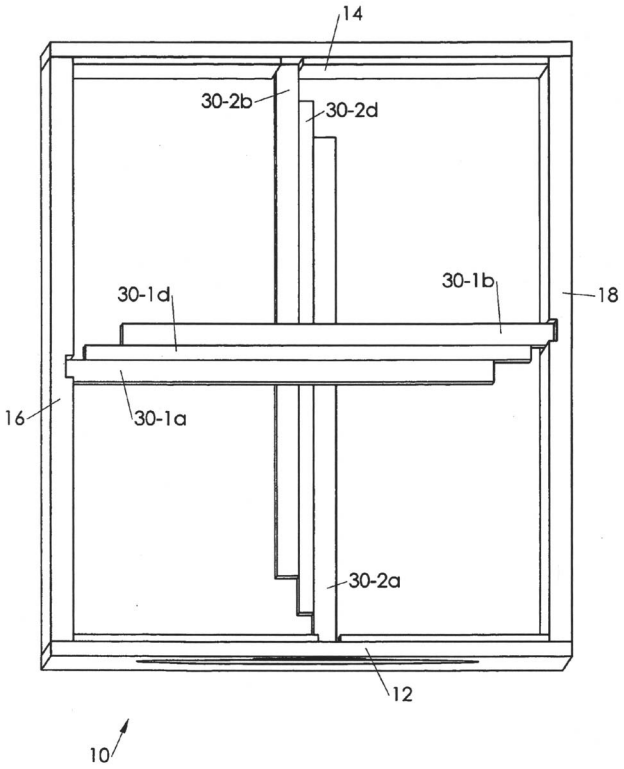
mode shape is unspecified. When vibrated at 6500–7500 Hz, particle size distribution produced favors 150 μm in diameter.—AJC

7,270,215

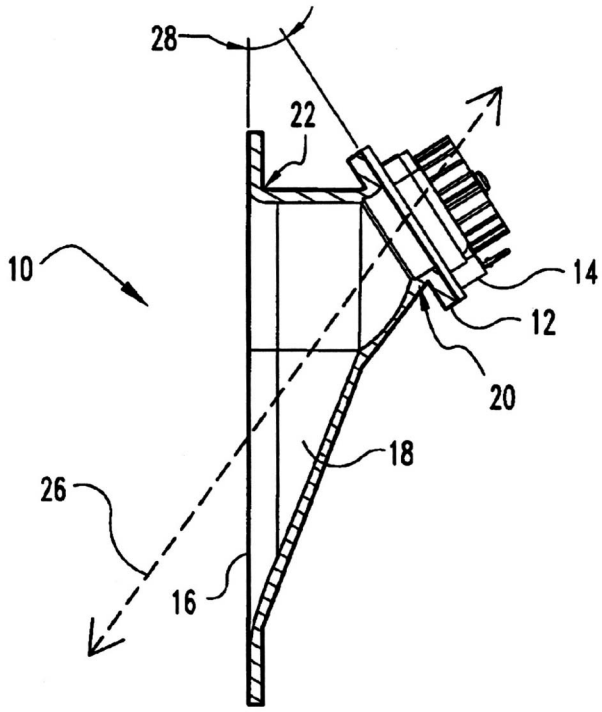
43.38.Ja LOUDSPEAKER ENCLOSURE WITH DAMPING MATERIAL LAMINATED WITHIN INTERNAL SHEARING BRACE

Enrique M. Stiles, assignor to STEP Technologies Incorporated
18 September 2007 (Class 181/151); filed 15 April 2005

Loudspeaker enclosures can vibrate since the transducers within them vibrate, and the enclosure vibrations can be detrimental to the overall sound emanating from the loudspeaker system. The cabinet walls can still vibrate even though the cabinet is fitted with rigid internal bracing. The patent presents a scheme in which the stiff internal bracing members **30-1**, **2 a** and



b are laminated to damping members **30-1**, **2 d**. The stiff members are therefore connected to just one side, and do not provide a short circuit between two sides of enclosure **10**. The damping member is said to help damp the vibrations of the cabinet.—NAS



theater surround sound, although a number of other applications come to mind. In short, this appears to be a well-designed device that meets its performance goals. However, the patent claims are broadly written and do not describe any specific geometry. Instead they postulate a throat plane that forms an angle with a mouth plane, the two being connected by an elongated duct "…such that sound waves generated by said transducer propagate unreflected down the elongated duct." It seems to me that thisdescription applies equally well to a planar-face, dual-horn tweeter manufactured by University Loudspeakers around 1950.—GLA

7,275,621

43.38.Ja SKEW HORN FOR A LOUDSPEAKER

Rogelio Delgado, Jr., assignor to Klipsch, LLC
2 October 2007 (Class 181/192); filed 18 January 2005

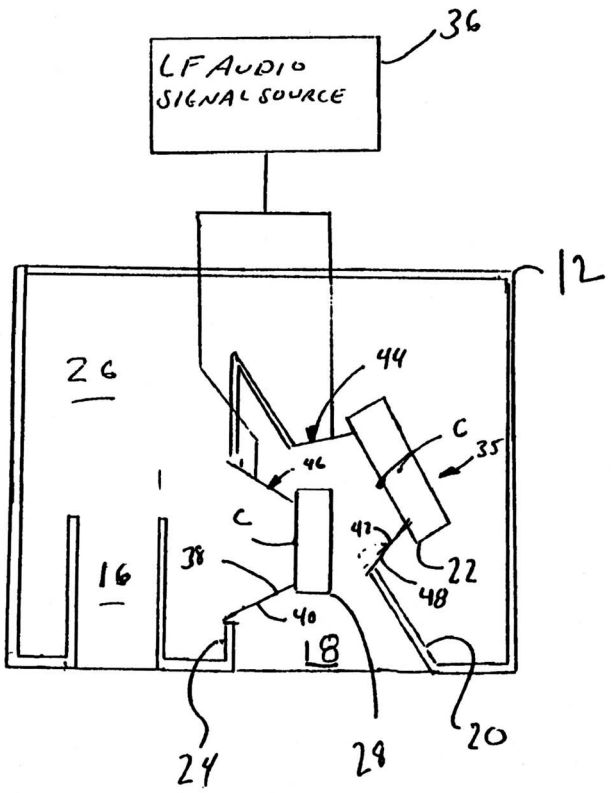
The skewed horn shown here in cross section has a flat mouth that can be flush-mounted in a wall or baffle. The patent includes polar response plots showing that a symmetrical pair of horns can deliver wide horizontal coverage over a range of two octaves or more. The intended usage is home

7,277,552

43.38.Ja INCREASED LF SPECTRUM POWER DENSITY LOUDSPEAKER SYSTEM

Curtis E. Graber, Woodburn, Indiana
2 October 2007 (Class 381/182); filed 9 August 2004

As can be seen from the illustration, the invention is basically an electrovoice "manifold" loudspeaker system with one woofer inverted. It is known that such a reversed pair of woofers reduces even-order distortion, but this feature is not mentioned in the patent. Instead, the comparatively small radiating area is exploited in a double-pair variant configured as a unidirectional low frequency loudspeaker. According to the patent, "The invention achieves high efficiency… in the second embodiment by dynamic control of the lobe of radiated energy by adjustment of the relative phase of a drive signal supplied (to) each of two transducer assemblies." There



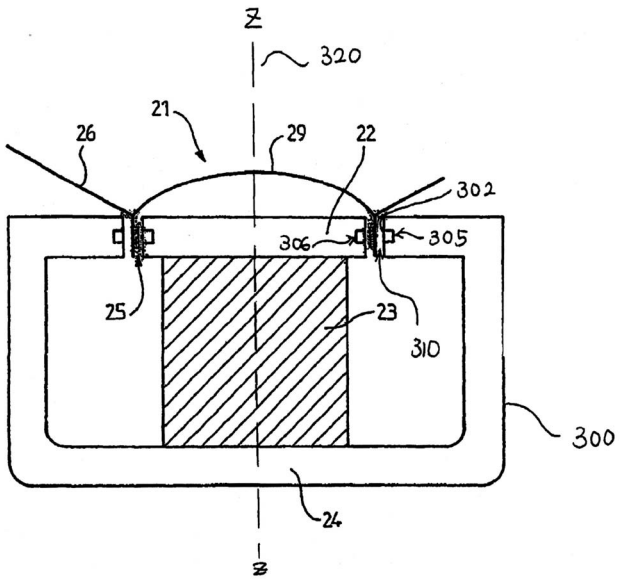
seems to be some confusion between pattern control and relative efficiency, which is actually quite low for any gradient pair.—GLA

7,283,642

43.38.Ja MOVING-COIL ELECTRODYNAMIC MOTOR FOR A LOUDSPEAKER, LOUDSPEAKER, AND POLE PIECE

Gilles Milot and Francois Malbos, assignors to Harman International Industries, Incorporated
16 October 2007 (Class 381/400); filed in France 13 February 2002

Here is my best guess as to what has been patented after reading all 24 patent claims twice: the magnetic gap of a loudspeaker includes at least one groove **305**, **306** in either the top plate, the pole piece, or both, and the



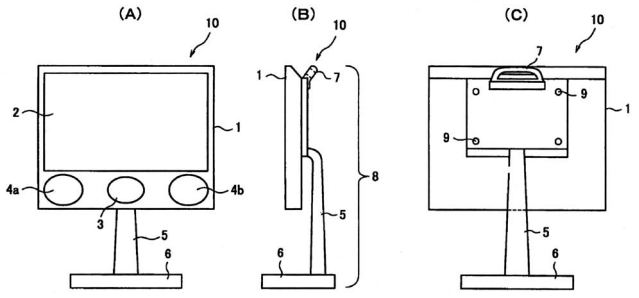
groove(s) may or may not be fitted with a conducting ring. That is also as much information as I was able to glean from five pages of descriptive text.—GLA

7,280,665

43.38.Ja IMAGE DISPLAY DEVICE WITH BUILT-IN LOUDSPEAKERS

Ichiro Tamura and Kohji Ohtsuka, assignors to Sharp Kabushiki Kaisha
9 October 2007 (Class 381/333); filed in Japan 15 March 2002

Flat panel TV receivers allow very little space for loudspeakers. In this design, left and right speakers **4a** and **4b** are conventionally mounted below the screen. Common woofer **3** is loaded by a small rear chamber with a vent tube that exits from the bottom of the cabinet. Thus, "…sound output from



the duct outlet **12** can be constantly absorbed by and reflected from the base surface **6**…without being affected by the flooring material," all of which makes no sense but adds one more patent to Sharp's portfolio.—GLA

7,261,182

43.38.Lc WIDE BAND SOUND DIFFUSER WITH SELF REGULATED LOW FREQUENCY ABSORPTION AND METHODS OF MOUNTING

Liviu Nikolae Zainea, Athens, Greece
28 August 2007 (Class 181/293); filed in Greece 21 May 2002

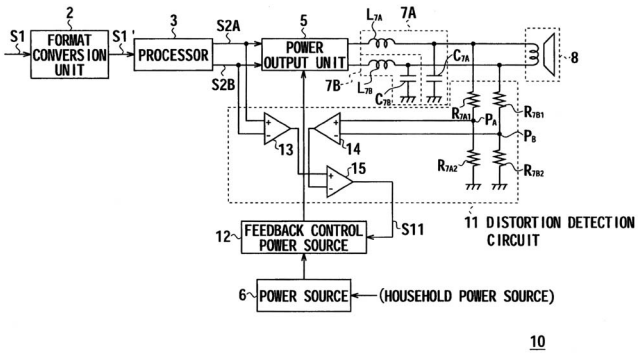
An improved acoustic diffuser is described that diffuses mid to high audible frequencies and absorbs frequencies lower than 250 Hz.—GFE

7,276,964

43.38.Lc PWM POWER AMPLIFIER AND METHOD FOR CONTROLLING THE SAME

Chitoshi Ishikawa, assignor to Sony Corporation
2 October 2007 (Class 330/10); filed in Japan 17 November 2004

Class D (switching) audio amplifiers have received considerable attention in recent years, both for high-power and low-power applications. Their high efficiency allows high-power amplifiers to be relatively lightweight and small. It also allows cellular phones and CD players to operate with very low current drain. However, conventional analog negative feedback cannot be used, and prior art includes a number of techniques for monitoring and reducing distortion. This patent describes a balanced PWM design in which



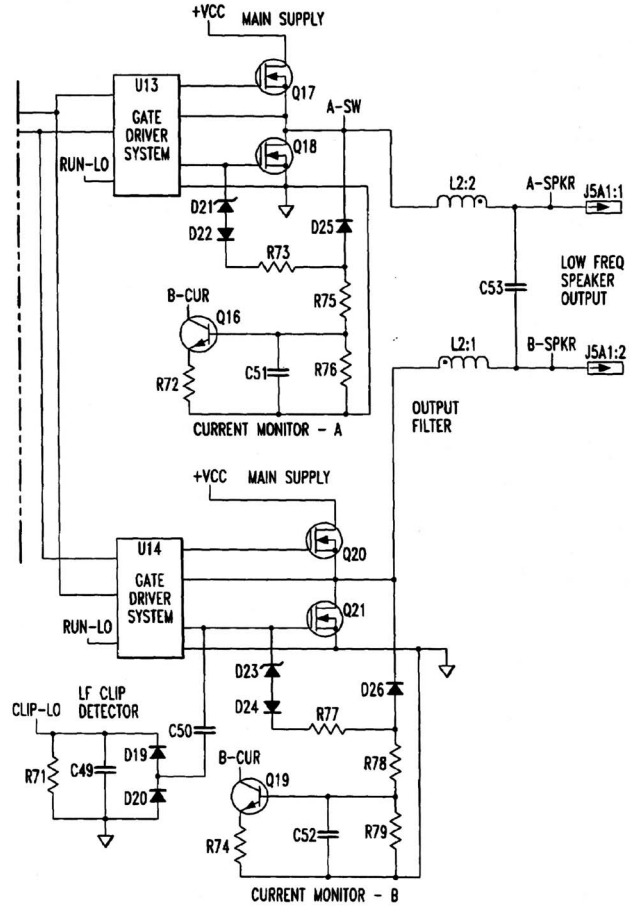
an analog signal is converted to pulse width modulation in conversion unit **2**. Processor **3** then derives separate positive and negative pulses, which are amplified by power unit **5**. Loudspeaker **8** is driven through lowpass filters **7A** and **7B**. Opamp **14** derives a difference value from the push-pull output signal and opamp **13** derives a difference value from the PWM signals at the input to the power unit. (Although not shown, the signals to **13** are also lowpass filtered to provide a compatible reference). Any difference between the two values is sensed by opamp **15** and used to compensate the power source accordingly.—GLA

7,279,967

43.38.Lc MULTI-CHANNEL, MULTI-POWER CLASS D AMPLIFIER WITH REGULATED POWER SUPPLY

Patrick H. Quilter, assignor to QSC Audio Products, Incorporated
9 October 2007 (Class 330/10); filed 12 January 2006

This patent goes into extensive detail about an improved Class D audio amplifier design. Separate high-frequency and low-frequency sections allow switching frequencies and output configurations to be optimized for individual frequency bands. The amplifier includes suppression of on-off transients, overload protection, and minimal switching losses. Six pages of text are devoted to the power supply design, which is described as an envelope-rectified, active-clamp ("ERAC") scheme. According to the patent, the advantages of this amplifier and power supply include, "…increased acoustic output, reduced distortion, wider frequency range, smaller and lighter speaker housings, and lower cost."—GLA

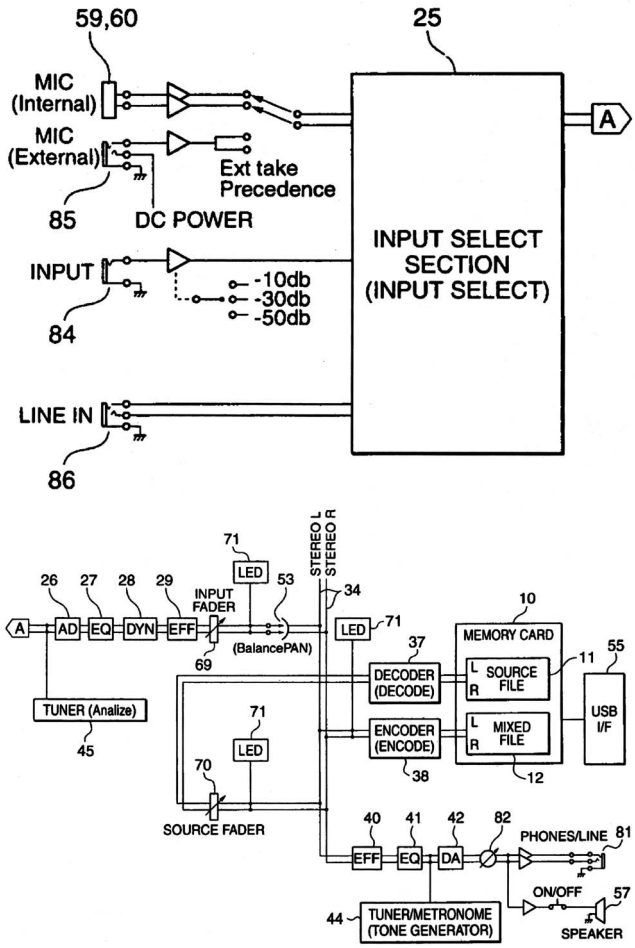


7,212,469

43.38.Md MIXING RECORDER, CONTROL METHOD THEREFOR, AND COMPUTER-READABLE MEDIUM INCLUDING PROGRAM FOR IMPLEMENTING THE CONTROL METHOD

Seiji Hirade et al., assignors to Yamaha Corporation
1 May 2007 (Class 369/4); filed in Japan 5 November 2002

There is nothing amazing about this straightforward description of a Yamaha overdubbing mixer (the "sound sketcher"). As shown, the mixer includes equalization **27**, dynamic processing **28**, and delay effects (like reverb) **29**. Note how the input has an analysis block **45** and a click track generator (metronome) on the output. For further details about the user interface and software internals, see U.S. patent 7,119,267.—MK



7,206,414

43.38.Md METHOD AND DEVICE FOR SELECTING A SOUND ALGORITHM

Donald Schulz, assignor to Grundig Multimedia B.V.
17 April 2007 (Class 381/56); filed in Germany 29 September 2001

Although missing many details, this patent tries to choose different sound processing algorithms depending on the specific signal parameters, particularly the dynamics, spectra, spectral flatness, and spectral peaks. The specific use of this disclosure is for discrimination of music and film channels in a television viewer.—MK

7,214,868

43.38.Md ACOUSTIC SIGNAL PROCESSING APPARATUS AND METHOD, SIGNAL RECORDING APPARATUS AND METHOD AND PROGRAM

Mototsugu Abe et al., assignors to Sony Corporation
8 May 2007 (Class 84/600); filed in Japan 12 December 2002

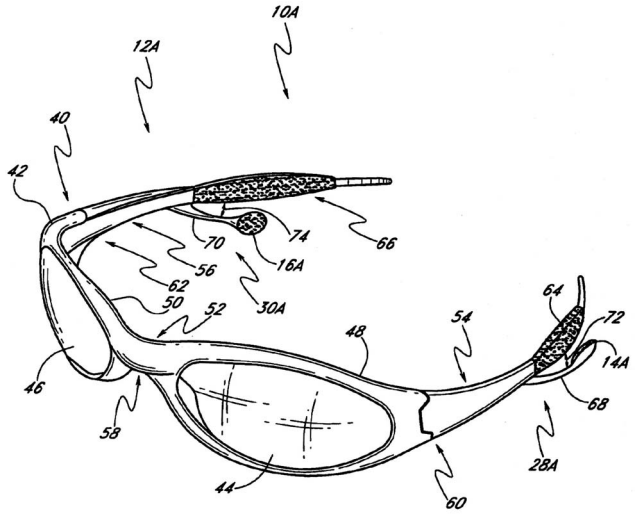
At issue is the automatic detection of interesting events in an audio track where "interesting" includes events like baseball hits and the like, which generate significant audience noise. The invention proposes using weighted power spectra and time duration calculations (although other possible parameters are mentioned in passing). Further use of discrimination and classification methods (such as the horribly mistranslated "Baizes" instead of Bayesian!) are proposed for event detection. There is nothing here that has not already been described in the information retrieval literature.—MK

7,264,350

43.38.Md MULTIDIRECTIONAL ADJUSTMENT DEVICES FOR SPEAKER MOUNTS FOR EYEGLASS WITH MP3 PLAYER

James Jannard et al., assignors to Oakley, Incorporated
4 September 2007 (Class 351/158); filed 12 October 2004

A means of mounting miniature loudspeakers **14A**, **16A** to an eyeglass frame **12A** is described along with a means of using wireless devices to



connect to the eyeglass assembly via built-in electronic packages **64**, **66**.—NAS

7,267,070

43.38.Pf DEVICE FOR CONTROLLING STEERING OF A TOWED UNDERWATER OBJECT

Yann Le Page and Frederic Schom, assignors to Cybernetix
11 September 2007 (Class 114/245); filed in France 18 May 2004

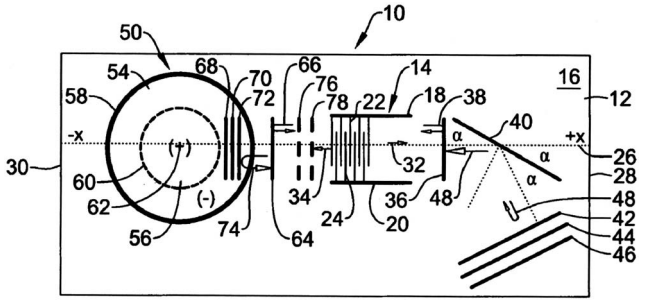
A steering fin designed for ocean towed acoustic arrays that allows inclination control is described.—GFE

7,243,547

43.38.Rh MEMS SAW SENSOR

Cornel P. Cobianu et al., assignors to Honeywell International Incorporated
17 July 2007 (Class 73/579); filed 13 October 2004

An interrogated MEMS temperature and pressure sensor for harsh environments is claimed where ultrasound surface acoustic wave (SAW) propagation on axis **26** senses pressure and temperature on axis **a**. Pressure is sensed on diaphragm **50** formed by etching away the back side of substrate **12**, leaving rigid support at circle **58**. A pressure force on the back side of sensor **10** will create tension on surface **62** and compression on surface



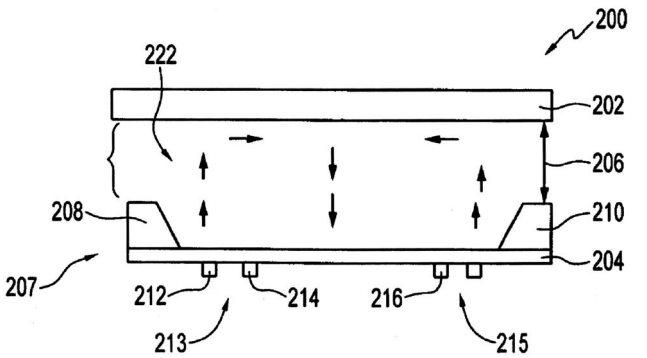
ring **54**. SAW **34**, reflected by bars **68–72** located on ring **54** will shift in phase according to said pressure or force. Temperature is sensed by reflector **40** diverting SAW energy at an angle in a direction that produces a phase change with temperature. Bars **42–46** reflect wave **48** back to IDT **22–24**. An antenna (not shown) can receive energy from a remote interrogator for wireless send-read operation.—AJC

7,243,549

43.38.Rh ACOUSTIC WAVE LUBRICITY SENSOR

James ZT Liu et al., assignors to Honeywell International Incorporated
17 July 2007 (Class 73/592); filed 19 May 2005

A SAW fuel lubricity sensor is claimed where diesel fuel passes between plate **204** vibrating in a bending mode and plate **202** vibrating in a



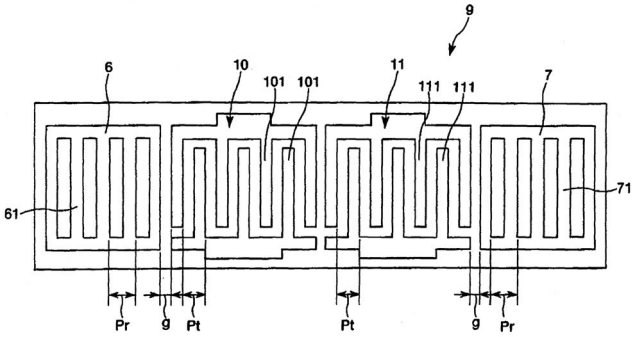
face shear mode. The change of amplitude and phase of either mode is claimed to be a measure of "lubricity" required for long life of diesel engine fuel metering components.—AJC

7,245,193

43.38.Rh SURFACE ACOUSTIC WAVE ELEMENT AND ELECTRONIC EQUIPMENT PROVIDED WITH THE ELEMENT

Tsukasa Funasaka, assignor to Seiko Epson Corporation
17 July 2007 (Class 333/193); filed in Japan 29 August 2003

It is claimed that the impedance characteristic and the insertion loss of surface acoustic wave (SAW) resonator **9** is improved when the ratio of the reflector conducting finger **10–101**, **11–111** spacing ("pitch") Pr to the transmit-receive finger **6–61**, **7–71** pitch Pt is in the range Pr/Pt=0.7 to



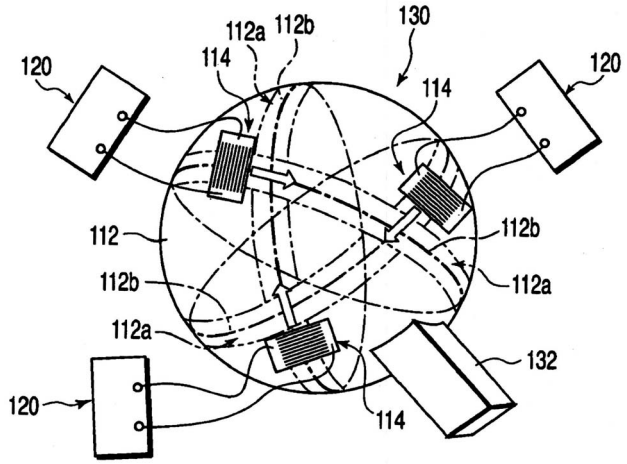
0.9995. It is further claimed that the preferred thickness of the piezoelectric layer be equal to or less than twice the SAW wavelength. It is also claimed that an intermediate layer will provide better capability to adjust impedance and to operate at higher frequencies.—AJC

7,247,969

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE AND ENVIRONMENTAL DIFFERENCE DETECTING APPARATUS USING THE SURFACE ACOUSTIC WAVE DEVICE

Noritaka Nakaso and Kazushi Yamanaka, assignors to Toppan Printing Company, Limited
24 July 2007 (Class 310/313 R); filed in Japan 19 September 2003

A spherical environmental difference detector **130** is claimed where several surface waves propagate on separate circumferential paths



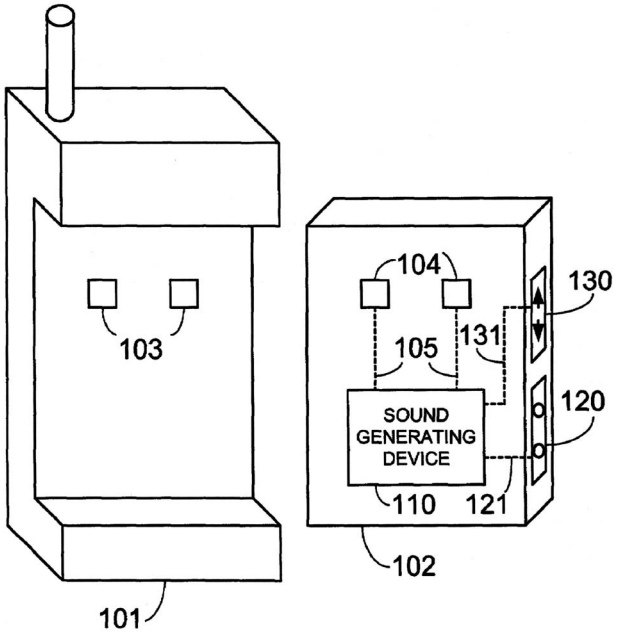
112a-112b-112c on sphere **112**. The presence of a different environment, e.g., raindrops, will attenuate and otherwise alter SAW amplitude and phase.—AJC

7,277,734

43.38.Si DEVICE, SYSTEM, AND METHOD FOR AUGMENTING CELLULAR TELEPHONE AUDIO SIGNALS

Mark Kirkpatrick, assignor to AT&T BLS Intellectual Property, Incorporated
2 October 2007 (Class 455/567); filed 28 September 2001

This patent describes a cellular phone add-on, "…device, system, and method that enables a user to easily choose any type of audio sounds to be used by a cellular telephone and allow such audio sounds to be designated as



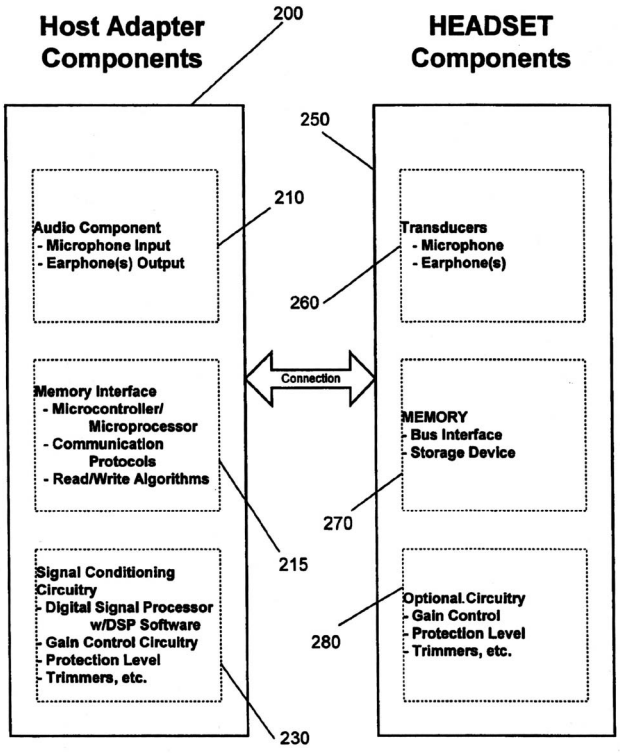
the incoming telephone call audio alert signal." It is reassuring to learn that genuine audio sounds are generated, rather than the bogus nonaudio sounds produced by ordinary cell phones.—GLA

7,283,635

43.38.Si HEADSET WITH MEMORY

Arthur G. Anderson and Robert M. Khamashta, assignors to Plantronics, Incorporated
16 October 2007 (Class 381/74); filed 9 December 1999

Information about measured performance, preferred equalization curves, and the like are stored in a memory chip located in an otherwise conventional headset. A host adapter communicates with the chip and makes



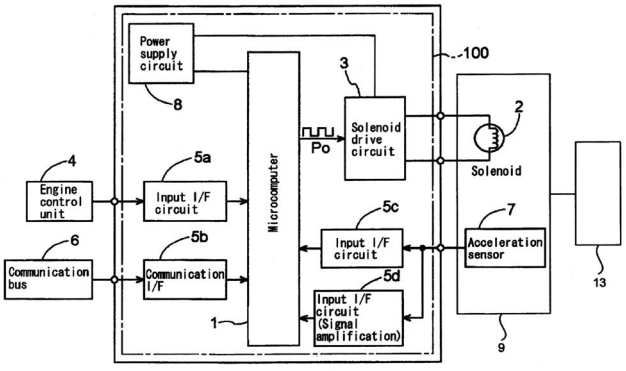
corresponding electronic adjustments as required. This dual-location scheme allows a single host adapter to control any headset that may be plugged in. It also allows operating parameters to be identified and changed at any time.—GLA

7,281,522

43.38.Yn ACCELERATION SENSOR STATUS DETECTING APPARATUS

Kenichi Sato et al., assignors to Aisin Seiko Kabushiki Kaisha
16 October 2007 (Class 123/492); filed 28 February 2007

This patent deals with active vibration control. Solenoid **2** is a part of engine mount **9**. Its drive circuitry uses a signal from sensor **7** to produce antiphase forces of appropriate frequency and amplitude. But suppose that the wiring between the control unit and the sensor is intermittent, or that the



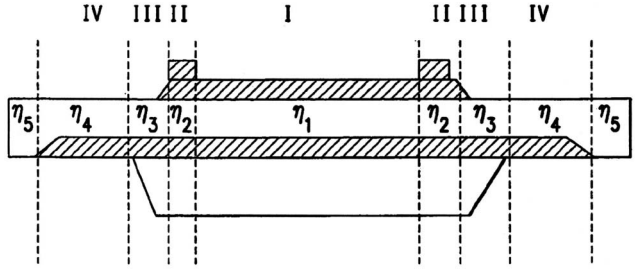
sensor itself suddenly conks out? A relatively simple scheme is described that maintains a running comparison between the slope of the current signal and that of an earlier sample. If the deviation exceeds acceptable limits, an error state is declared. The patent does not concern itself with other possible problems in the solenoid, the solenoid drive circuit, the communication bus, etc.—GLA

7,280,007

43.40.Dx THIN FILM BULK ACOUSTIC RESONATOR WITH A MASS LOADED PERIMETER

Hongjun Feng et al., assignors to Avago Technologies General IP (Singapore) Pte. Limited
9 October 2007 (Class 333/187); filed 15 November 2004

This patent discloses the construction and design details of a film bulk acoustic resonator (FBAR) that uses a raised-perimeter electrode ring to enhance the energy trapping and increase resonator Q. It is claimed that this



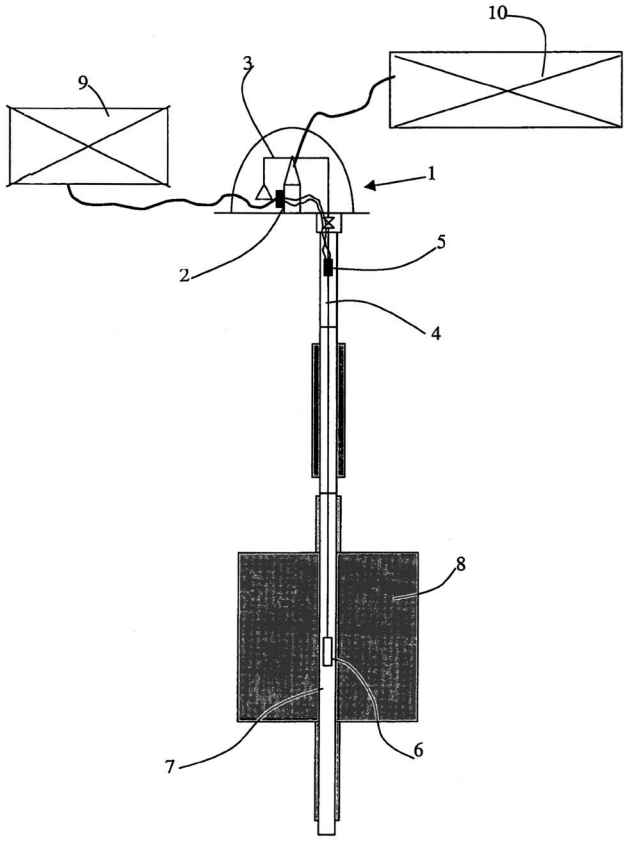
is accomplished by reducing the leakage of lateral (Lamb wave) modes in the body of the FBAR. Several before and after examples are given of the efficacy of this design, and the description is clear and easy to read.—JAH

7,246,516

43.40.Le DEVICE FOR COUPLING THERMOGRAVIMETRIC ANALYSES AND ACOUSTIC EMISSION MEASUREMENTS

Francois Ropital and Philippe Dascotte, assignors to Institut Francais du Petrole
24 July 2007 (Class 73/86); filed in France 20 August 2004

An acoustic emission sensor **5-2-9** application accompanying thermogravimetric analysis (TGA) of suspended specimen **6** is claimed. An



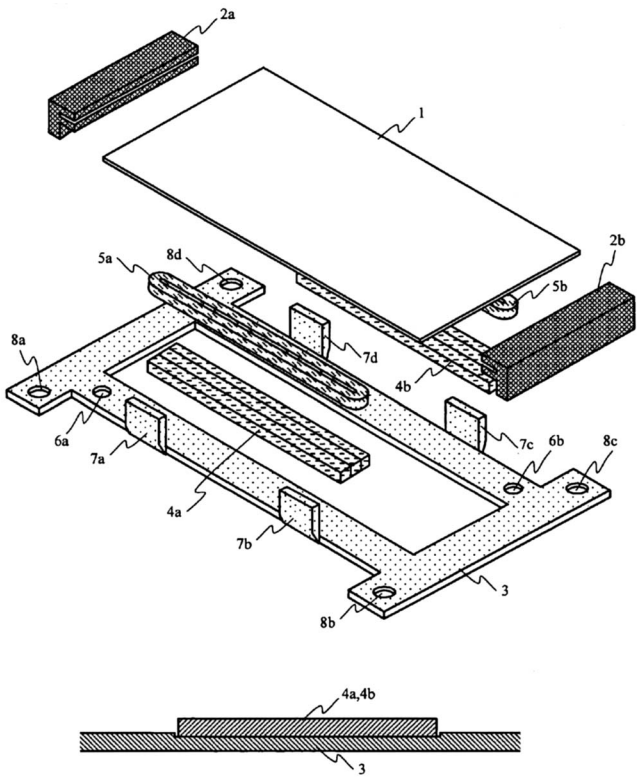
ultrasound receiving transducer **5** is incorporated in suspension **4** of specimen **6** out of environment **8**, which can include high temperatures and acid atmospheres.—AJC

7,245,048

43.40.Rj TACTILE PANEL

Noboru Fujii et al., assignors to Fujitsu Component Limited
17 July 2007 (Class 310/12); filed in Japan 29 July 2004

A tactile (haptic) panel capable of creating or countering vibration forces on contacting objects or fluids is claimed where panel **1** is attached to



coils **5a** and **5b** adjacent to permanent magnets **4a** and **4b**, supported by fixed yoke **3**. Resilient pads **2a** and **2b** support the coil and pad with minimal resistance to motion or large displacement of panel **1**.—AJC

7,261,274

43.40.Tm VIBRATION ISOLATION SYSTEM FOR DAGGER MOUNTED EQUIPMENT

Richard D. Vatsaas et al., assignors to Lockheed Martin Corporation
28 August 2007 (Class 248/635); filed 6 May 2003

A "dagger pin" arrangement permits one to mount a device blindly in a pocket type compartment where there is little clearance around the device. A dagger pin is a short rod with a rounded or conical end that is fixed to a surface in the pocket and mates with a hole in a "dagger plate" attached to the device. A dagger pin restrains motion of the device in directions perpendicular to the pin's axis; a dagger-pin-mounted device is provided with additional constraints to restrict its motion in the direction parallel to the pins' axes. The present patent describes dagger plates and additional constraining fasteners that include resilient elements for vibration isolation.—EEU

7,264,303

43.40.Tm BODY PANEL WITH VIBRATION DAMPING MATERIAL, VIBRATION DAMPING MATERIAL COATER, AND DAMPING MATERIAL APPLICATION METHOD

Hideki Fukudome et al., assignors to Mitsubishi Jidosha Kogyo Kabushiki Kaisha
4 September 2007 (Class 296/191); filed in Japan 2 October 2001

This patent describes a production means for adding a constrained-layer viscoelastic damping treatment to an automotive body panel or the like. The viscoelastic material is sprayed onto the panel by a first fan-shaped nozzle and the constraining layer is sprayed atop the viscoelastic layer by a nearby second fan-shaped nozzle as the panel is moved along the production line.—EEU

7,267,196

43.40.Tm METHOD AND APPARATUS FOR REDUCING ACOUSTIC NOISE

Gopal P. Mathur, assignor to The Boeing Company
11 September 2007 (Class 181/208); filed 12 February 2004

This patent describes dampers in which thin films of air contained between a basic plate element and secondary closely spaced parallel plate elements are squeezed as the basic plate element vibrates. The secondary plates are resiliently attached to the primary one over only a fraction of their perimeters, so that the air in the narrow gap between the plates can be "pumped" in directions parallel to the plates.—EEU

7,273,380

43.40.Tm FLEXIBLE RING INTERCONNECTION SYSTEM

Atsuhito Noda et al., assignors to Molex Incorporated
25 September 2007 (Class 439/66); filed 25 January 2007

Multiple small rings of a conducting material are used to provide flexible connections between a substrate and a microelectronic device package. The rings are soldered to the substrate and the device package, but their sides are kept free of solder, so that some of their flexibility is maintained in the installed condition. They can accommodate elevation differences while providing vibration isolation.—EEU

7,278,437

43.40.Tm METHOD FOR CLOSING FLUID PASSAGE, AND WATER HAMMERLESS VALVE DEVICE AND WATER HAMMERLESS CLOSING DEVICE USED IN THE METHOD

Tadahiro Ohmi et al., assignors to Fujikin Incorporated
9 October 2007 (Class 137/2); filed in Japan 19 December 2002

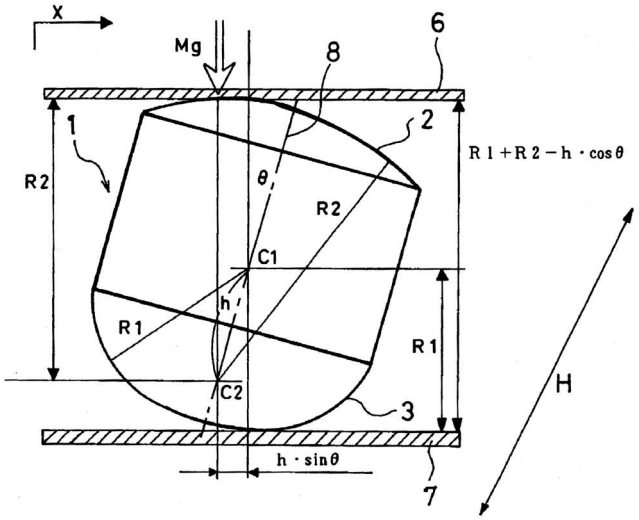
Water hammer—a sharp increase in pressure—occurs when a passage through which a fluid flows is closed abruptly. Instead of earlier means for reducing water hammer (such as closing the passage slowly, releasing the pressure via a bypass passage, or absorbing the pressure in an accumulator) the system described in this patent achieves rapid closure of a valve by moving the valve body in two steps. In the first step the valve body moves rapidly to a predetermined position short of full closure and in the second step it moves to full closure.—EEU

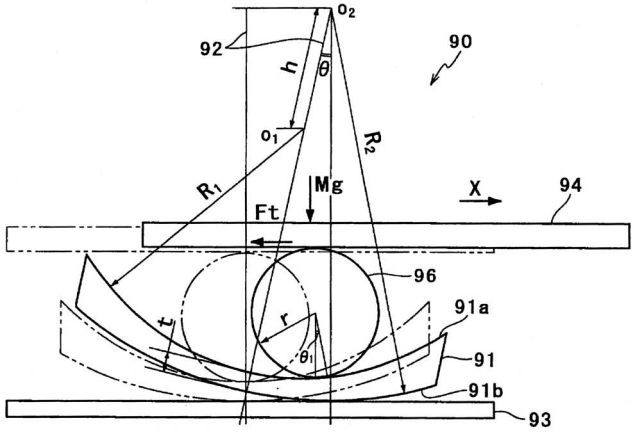
7,278,623

43.40.Tm VIBRATION CONTROL UNIT AND VIBRATION CONTROL BODY

Masashi Yasuda et al., assignors to Tokkyokiki Corporation
9 October 2007 (Class 248/562); filed in Japan 3 August 2001

Isolation of horizontal vibrations is achieved by means of rolling elements **1** placed between a bottom plate **7** and a top plate **6**. A rolling element **1** has an upper convex surface **2** with radius of curvature R2 and a lower convex surface **3** with radius of curvature R1, with the overall height of the





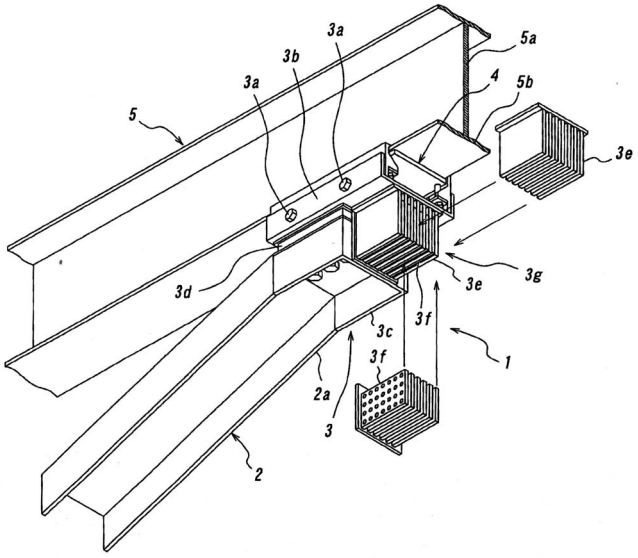
element smaller than R1+R2. The general concept may also embodied using a spherical element **96** rolling on a rocking element **91** that is concave at its top and convex at its bottom with different curvatures on its two surfaces.—EEU

7,281,287

43.40.Tm PROP-TYPE DAMPING DEVICE

Yoji Shimazaki et al., assignors to Tokai University Educational System
16 October 2007 (Class 14/14); filed in Japan 9 September 2003

This device for damping the vibrations of a bridge girder **5** uses a prop **2** that is pin-connected to a column or pier of the bridge at its lower end and



carries a damping arrangement at its upper end. The damping arrangement is clamped to the girder with the aid of bolts **3a**, **3b**, and **3c** and has a series of copper plates **3e** attached to the clamped end. A rubber pad **3d** permits the prop to slide along the length of the girder, enabling the permanent magnets **3f** that are attached to the prop to interact with the copper plates **3e** with which they are interlaced. Rubber pads between the bridge girder and the piers that support it permit some vertical motion of the girder.—EEU

7,269,995

43.40.Vn METHOD AND APPARATUS OF VIBRATION ISOLATION, IN PARTICULAR FOR ELECTRON BEAM METROLOGY TOOLS

Peter Heiland, assignor to Integrated Dynamics Engineering Gmb
18 September 2007 (Class 73/105); filed in Germany 25 September 2003

This broadly written patent relates to measurement tools in which the distance between the measuring element and the item to be measured needs to be maintained precisely in the presence of vibrations. The system described here involves sensing of the relative or absolute motions of the measuring element and the measured item and using an active control system to maintain the desired distance.—EEU

7,278,391

43.40.Vn CYLINDER DEACTIVATION TORQUE LIMIT FOR NOISE, VIBRATION, AND HARSHNESS

Kevin C. Wong et al., assignors to GM Global Technology Operations, Incorporated
9 October 2007 (Class 123/198 F); filed 11 September 2006

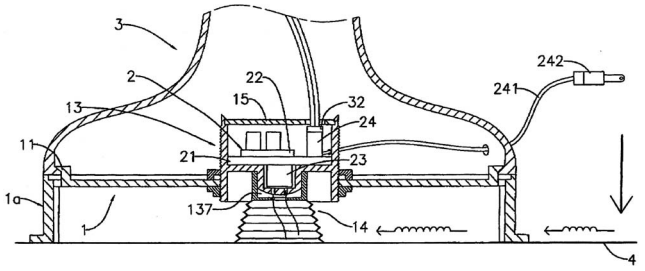
This patent addresses a system for controlling an engine to transition between a mode where all cylinders are active and a deactivated mode where some of the cylinders are not active, which may be used for the sake of fuel economy when low engine torque suffices. Transition between the two modes is determined on the basis of a noise-vibration-harshness (NVH) limit, the engine speed, the vehicle speed, and the desired torque.—EEU

7,213,943

43.40.Yq TAP SENSING LAMP SWITCH

Peng Kuo Chung and Chen Yang Chun, assignors to Fun Plus Corporation
8 May 2007 (Class 362/276); filed 31 August 2005

Rather than search for the light switch, the inventors want to use acoustic signals to turn on the lamp. As shown, the "sound collection element" **14** is connected to the microphone **23**, which is fed to the circuit



board **21** that will control the light. No mention is made of how to distinguish between heavy footsteps and light touch.—MK

7,262,834

43.40.Yq SENSOR FOR MEASURING VELOCITY OF VIBRATION USING LIGHT WAVEGUIDE

Kazuro Kageyama et al., assignors to Toudai Tlo, Incorporated
28 August 2007 (Class 356/28); filed in Japan 27 June 2001

This vibration sensor is based on the inventors' observation that the frequency of light that passes through a curved section of an optical fiber that is subjected to vibration is shifted in frequency. Light from a laser is input to an optical fiber with a curved portion via a coupler. A half mirror sends some of the light to an acoustic optical modulator, to which the light coming out of the fiber is also sent. This modulator then determines the frequency changes, whose magnitudes depend on the vibrational velocity of the curved section of the fiber.—EEU

7,263,028

43.50.Gf COMPOSITE ACOUSTIC ATTENUATION MATERIALS

Gerald B. Thomas et al., assignors to United States of America as represented by the Secretary of the Navy
28 August 2007 (Class 367/1); filed 1 October 2004

An acoustic attenuation material, according to this patent, is made of two layers. One of these consists of a matrix in which there are embedded a plurality of small particles of high and/or low acoustic impedance compared to that of the matrix material. The second layer consists of a fibrous mat in minimal contact with the first layer. The patent claims that absorption and transmission characteristics can be adjusted by appropriate selection of the materials. Performance data of certain implementations are presented in the patent, but it is not clear whether these data refer to absorption or transmission.—EEU

7,217,879

43.55.Br REVERBERATION SOUND GENERATING APPARATUS

Tsugio Ito, assignor to Yamaha Corporation
15 May 2007 (Class 84/630); filed in Japan 26 March 2003

Given two room impulse responses, how can a new impulse response be generated that is midway between the two? The inventor proposes windowing and then using the Fourier transform for each impulse response (all the preprocessing is done offline). A linear combination is made and then the inverse Fourier transform is applied, (n.b. Of course, using partioned frequency domain convolution would eliminate the inverse transform step).—MK

7,261,242

43.55.Wk SOUND INSULATION FOR ELECTRIC RELAY

François Houde, assignor to Honeywell International Incorporated
28 August 2007 (Class 236/1 C); filed 2 February 2004

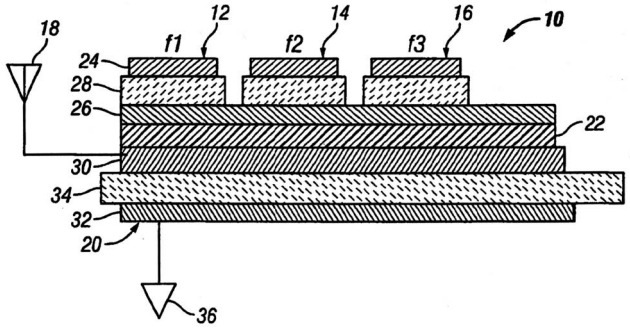
A device for dampening the audible click of an electrical relay switch is put forth.—GFE

7,276,892

43.58.Kr RESONATOR BASED SPECTRUM ANALYZER AND METHOD

William Richard Trutna, Jr. and Steven Rosenau, assignors to Avago Technologies Wireless IP (Singapore) Pte. Limited
2 October 2007 (Class 324/76.19); filed 29 April 2005

This patent discloses the concept of a MEMS spectrum analyzer based upon parallel-connected film bulk acoustic wave resonators (FBARs). The concept is not new; the only possible novelty here is in the way that the resonators are driven, which is illustrated in the figure but not described well in the text. There is no mention of the interstitial layer thickness, of whether



one wants tight or loose coupling between FBARs **12**, **14**, **16** and FBAR **20**, and no measurements of performance are given. It would seem from the lack of detail given that the authors have never actually built one of these devices, and clearly a lot more work remains to be done to make this a practical, useful device.—JAH

7,266,043

43.60.Fg METHOD FOR DIRECTIONAL LOCATION AND LOCATING SYSTEM

Hans-Ueli Roeck, assignor to Phonak AG
4 September 2007 (Class 367/124); filed 29 November 2004

A very scientific sounding patent is put forth that determines the direction of sound source by determining which (nonomnidirectional) sensor it was loudest on, just like your ears.—GFE

7,272,477

43.60.Gk VEHICLE PARKING ASSISTING SYSTEM AND METHOD

Tetsuya Oki et al., assignors to Denso Corporation
18 September 2007 (Class 701/36); filed in Japan 4 November 2005

Sonar data from a vehicle bumper array are displayed from a bird's eye view so that drivers will know where nearby objects are located.—GFE

7,266,045

43.60.Jn GUNSHOT DETECTION SENSOR WITH DISPLAY

Kevin C. Baxter and Ken Fisher, assignors to ShotSpotter, Incorporated
4 September 2007 (Class 367/128); filed 24 January 2005

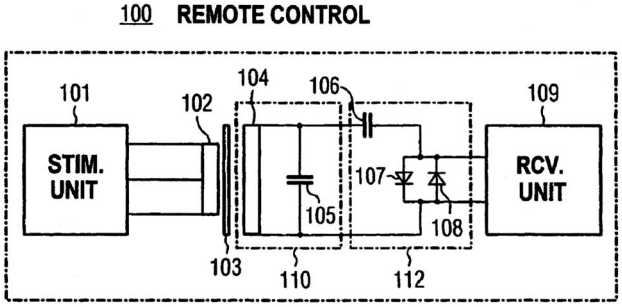
Man portable sensors are to collect acoustic data and global position in a net centric gunshot detection scheme.—GFE

7,277,553

43.66.Ts TRANSMISSION COIL SYSTEM AND REMOTE CONTROL FOR A HEARING AID

Jürgen Reithinger, assignor to Siemens Audiologische Technik GmbH
2 October 2007 (Class 381/315); filed in Germany 22 May 2003

Two loosely coupled coils with an unequal number of windings that are wound alongside each other on a coil core produce a stronger magnetic field for wireless control of hearing aids than if they were closely coupled.



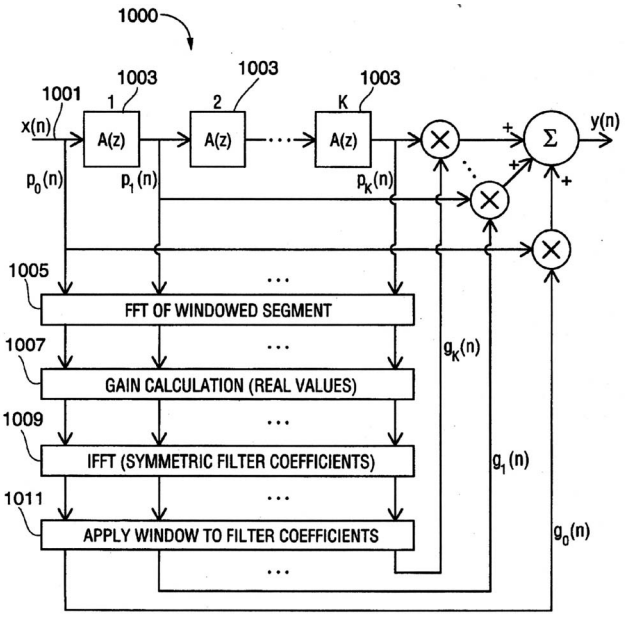
The second transmission coil is used as a part of a resonant circuit and is excited from the first coil with a phase shift. The coils may be used in either transmit or receive mode.—DAP

7,277,554

43.66.Ts DYNAMIC RANGE COMPRESSION USING DIGITAL FREQUENCY WARPING

James M. Kates, assignor to GN ReSound North America Corporation
2 October 2007 (Class 381/316); filed 13 November 2001

To more closely match human auditory latencies while achieving adequate frequency analysis resolution, a compression filter bank for a hearing aid uses cascaded all-pass sections to provide constant group delay, independent of instantaneous compression gain. After the delayed samples are windowed, a FFT or DFT is performed and first and second warped power spectra are calculated, summed, and normalized to compute frequency



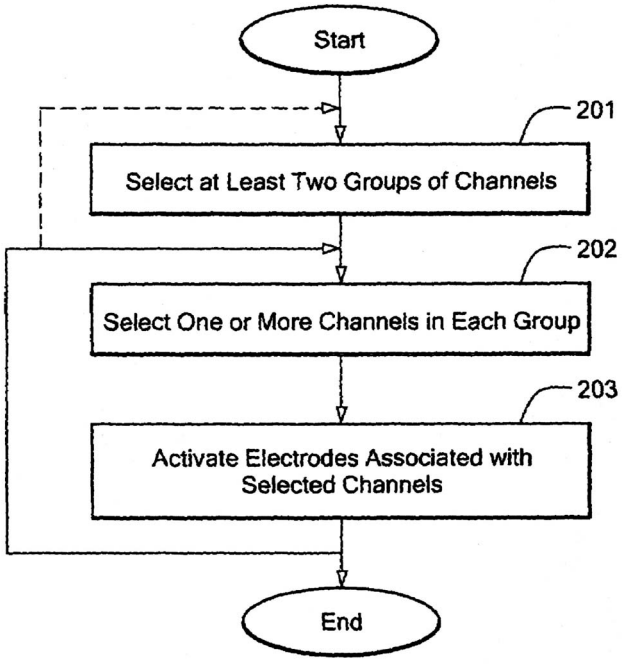
domain level estimates. Thereafter, an inverse transform back into a warped time domain produces compression gain filter coefficients that are convolved with the outputs of the all-pass delay line to produce a digital output signal.—DAP

7,283,876

43.66.Ts ELECTRICAL STIMULATION OF THE ACOUSTIC NERVE BASED ON SELECTED GROUPS

Clemens M. Zierhofer, assignor to MED-EL Elektromedizinische Geraete GmbH
16 October 2007 (Class 607/57); filed 8 March 2005

At least one channel in each of at least two groups of channels is selected for stimulation in a multichannel electrode cochlear implant. At least one of the groups contains a number of channels, each of which is



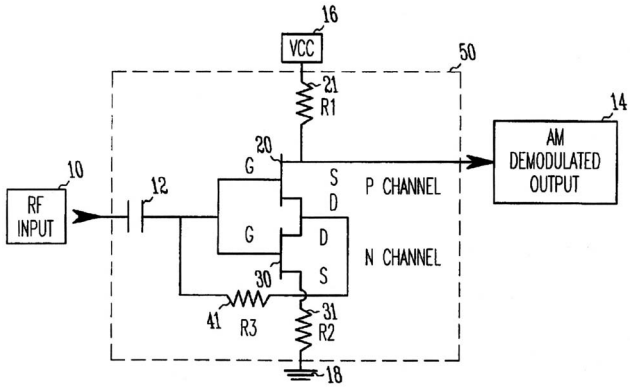
associated with a different implanted electrode. This procedure is repeated so that the selected channels and their corresponding activated electrodes vary in at least one group. The activation of electrodes associated with each channel may be sequential or simultaneous.—DAP

7,277,687

43.66.Ts LOW POWER AMPLITUDE MODULATION DETECTOR

John David Terry, assignor to Starkey Laboratories, Incorporated
2 October 2007 (Class 455/293); filed 3 December 2003

Methodology is described for implementing a low power, small size amplitude modulation detector on an integrated circuit. Both *p*-channel and



n-channel metal oxide semiconductor transistors are utilized in a feedback configuration to receive the radio frequency input.—DAP

7,283,638

43.66.Ts HEARING AID WITH ERROR PROTECTED DATA STORAGE

Casper Kruger Troelsen and Rene Mortensen, assignors to GN ReSound A/S
16 October 2007 (Class 381/312); filed in Denmark 14 November 2000

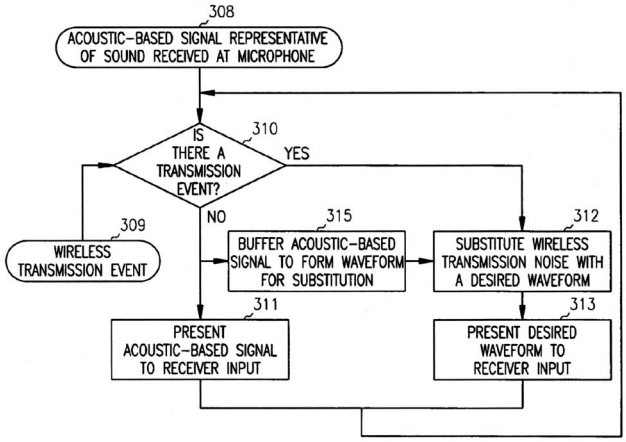
Hearing aid wearers may open their battery compartments during a data logging operation, which constitutes a power failure that may cause corrupted signal processing parameter values and other data such as volume control setting to be stored into the nonvolatile memory of the hearing aid. To prevent the storage of corrupted data, a first set of the original data is written and a second set of updated data is written at a later time into different areas of storage. Thereafter, either the first or second dataset is marked as the valid data, which is automatically loaded at power on during reboot.—DAP

7,283,639

43.66.Ts HEARING INSTRUMENT WITH DATA TRANSMISSION INTERFERENCE BLOCKING

Robert Kariniemi, assignor to Starkey Laboratories, Incorporated
16 October 2007 (Class 381/312); filed 10 March 2004

Hearing aid wearers are prevented from listening to interfering signals generated by external devices such as digital cellular phones. If an interfering signal produced by a wireless transmission is detected at the input, a



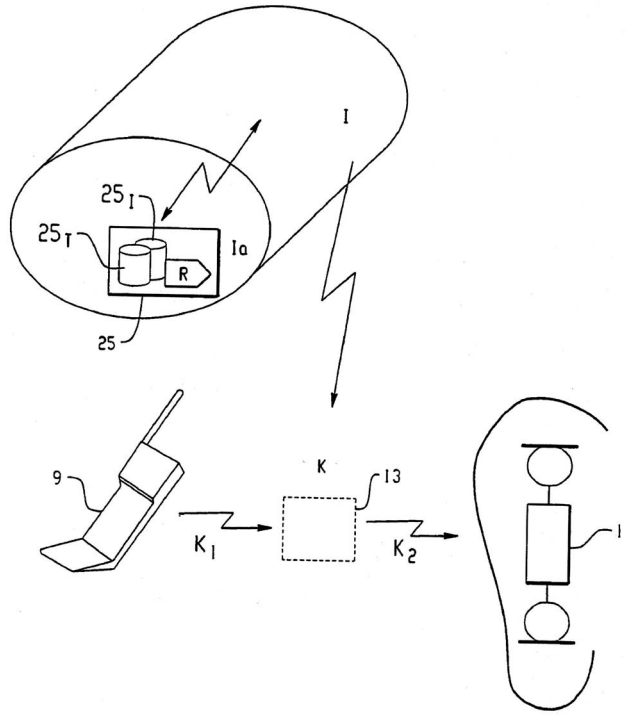
substitute wave form is sent to the hearing aid output. The substitute wave form may be generated from signals present at the input immediately prior to the interfering signal.—DAP

7,283,842

43.66.Ts FITTING-SETUP FOR HEARING DEVICE

Christian Berg, assignor to Phonak AG
16 October 2007 (Class 455/557); filed 7 November 2005

To facilitate remote adjustments and remote control of hearing aids, a cell phone is used to download fitting programs, updated software and commands to hearing aids in a monaural or a binaural fitting via a wireless communication link, and convey data about the hearing aids to a database



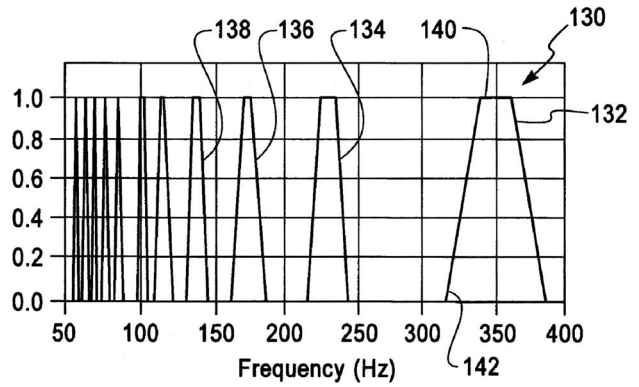
server. The type of communication link may be acoustic, infrared, or high frequency via a converter relay, and the database access may be via the Internet. The converter, which may be integrated into the cell phone or be a separate external device, communicates with the hearing aid via wireless or wired means.—DAP

7,272,551

43.72.Ar COMPUTATIONAL EFFECTIVENESS ENHANCEMENT OF FREQUENCY DOMAIN PITCH ESTIMATORS

Alexander Sorin, assignor to International Business Machines Corporation
18 September 2007 (Class 704/205); filed 24 February 2003

This speech pitch analyzer would evaluate pitch candidates based on a comparison of weighted amplitudes of the spectral lines of a Fourier transform (FFT) of each speech data frame. The FFT resolution is first improved by using a Dirichlet interpolation. In order to improve the tradeoff between definition and resolution for pitch frequencies below 90 Hz, two adjacent



FFTs are then combined with a phase shift, producing a longer window. The spectral lines are then weighted by a pattern of trapezoidal weightings specific to each pitch candidate frequency. A sequence of tests is performed next to select one of the pitch candidates for the frame. The patent presents full mathematical details for performing these steps.—DLR

7,266,493

43.72.Gy PITCH DETERMINATION BASED ON WEIGHTING OF PITCH LAG CANDIDATES

Huan-Yu Su and Yang Gao, assignors to Mindspeed Technologies, Incorporated
4 September 2007 (Class 704/207); filed 13 October 2005

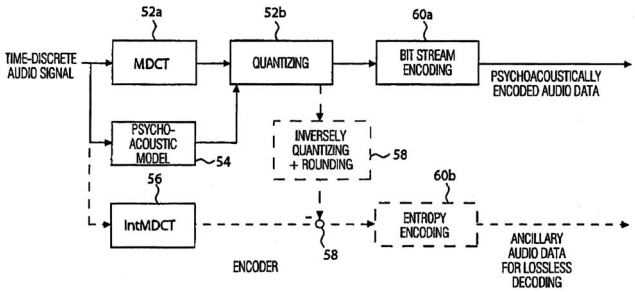
This patent presents a method for selecting a pitch lag value for encoding frames of speech following the code-excited linear prediction technique. As each preceding frame is encoded, for those frames determined to be voiced an adaptive weighting system is updated to provide a pitch estimate for the next frame. A pitch value is then determined by a combination of the current long-term linear prediction result and the adaptive weighting, taking into consideration a preference for the lower frequency value when multiple candidates are available having a common frequency multiplier, i.e., when one or more harmonics are present, the fundamental would be selected. There does not seem to be any clear statement of how this encoding system would be related to any of the many existing speech encoding standards.—DLR

7,275,036

43.72.Gy APPARATUS AND METHOD FOR CODING A TIME-DISCRETE AUDIO SIGNAL TO OBTAIN CODED AUDIO DATA AND FOR DECODING CODED AUDIO DATA

Ralf Geiger et al., assignors to Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V.
25 September 2007 (Class 704/500); filed in Germany 18 April 2002

The goal is less expensive hardware for a lossless audio signal decoder that does not require a frequency/time conversion to calculate the error



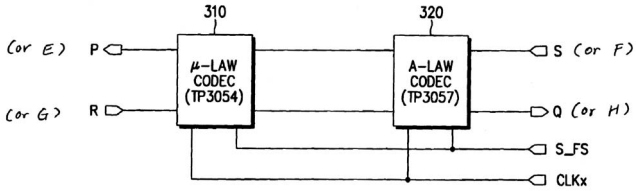
between the original signal and the coded/decoded signal. A quantizer provides spectral values of the audio signal using a psychoacoustic model and the resultant is inverse quantized and rounded. Integer spectral values are generated and compared with a combiner to the rounded spectral values to obtain difference spectral values. Coded audio data are generated, including information on the quantized values and difference values. Decoding provides similar steps in reverse to generate a temporal representation of the audio signal.—DAP

7,277,490

43.72.Gy APPARATUS FOR CONVERTING PCM SIGNAL

Chang-Rae Jeong, assignor to Samsung Electronics Company, Limited
2 October 2007 (Class 375/242); filed in Republic of Korea 11 July 1998

An interfacing protocol in a communication system converts efficiently between one or two selected channels out of a multichannel set of,



for example, A-law encoded signals, which are widely used in Europe, and μ-law encoded signals, which are widely used in the United States and Japan.—DAP

7,283,965

43.72.Gy DELIVERY AND TRANSMISSION OF DOLBY DIGITAL AC-3 OVER TELEVISION BROADCAST

James A. Michener, assignor to The DirecTV Group, Incorporated
16 October 2007 (Class 704/500); filed 30 June 1999

Described is methodology to broadcast television programs via satellite with the same audio quality that would be heard in theaters with a DVD. Program content in AC-3 format is transmitted as AES-3 signal bit streams. A predetermined time or word count at which each packet in the AES-3 bit stream is to arrive and whether a link disruption has occurred during packet transmission are used as criteria to enable packet data being sent to the output and a response indicating packet has been received successfully.—DAP

7,274,967

43.72.Ja SUPPORT OF A WAVETABLE BASED SOUND SYNTHESIS IN A MULTIPROCESSOR ENVIRONMENT

Marius Tico et al., assignors to Nokia Corporation
25 September 2007 (Class 700/94); filed 10 October 2003

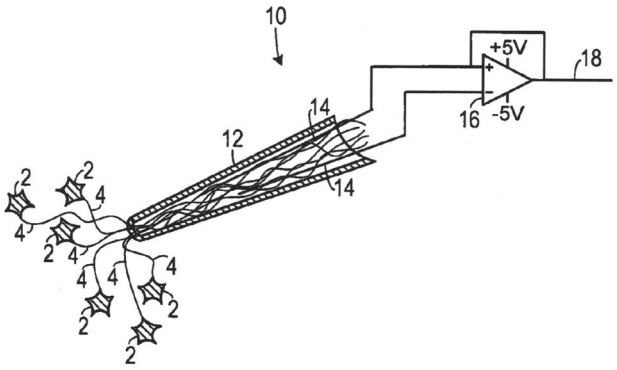
A first processor selects samples of stored wavetable data and makes them available to a second processor that generates audio frames. Selection criteria are based on a model of a pitch-shift evolution during a single frame and on the number of samples that have been used so far by a second processor to generate previous audio frames.—DAP

7,275,035

43.72.Ja SYSTEM AND METHOD FOR SPEECH GENERATION FROM BRAIN ACTIVITY

Philip R. Kennedy, assignor to Neural Signals, Incorporated
25 September 2007 (Class 704/271); filed 8 December 2004

This patent is a logical development upon the increasing levels of success in measuring and recording brain activity. Many ways have been devised for patients with various types of muscular maladies to operate a speech synthesizer. Now, at last, the synthesizer would be controlled based on a computer analysis of electrical activity recorded directly from the brain. One proposed technique for detecting, the relevant brain patterns would be a



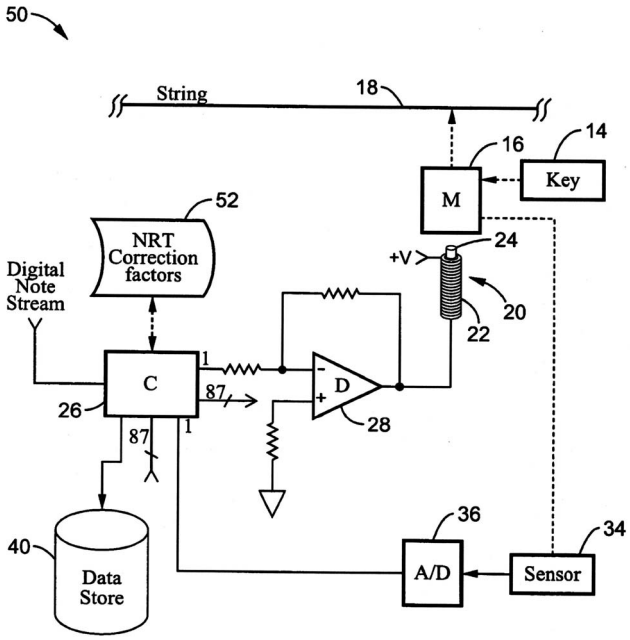
neurotrophic electrode, as shown in the figure. In this case, neurons are persuaded by various chemical means to grow dendrites up into the glass tube of the electrode. Other means are discussed in the patent. In any case, a machine learning system is trained to associate specific neural patterns with the patient's desire to speak a particular sound. The synthesizer then produces that sound.—DLR

7,217,880

43.75.Mn PREVIOUS EVENT FEEDBACK SYSTEM FOR ELECTRONIC PLAYER PIANO SYSTEMS

Mark Van Sant, assignor to Burgett, Incorporated
15 May 2007 (Class 84/645); filed 20 March 2006

A slightly different take on mechanical replay for pianos. The inventor wants to use a correction table for hammer velocity so that the playback will match historical player pianos and the like. As shown, there is a solenoid



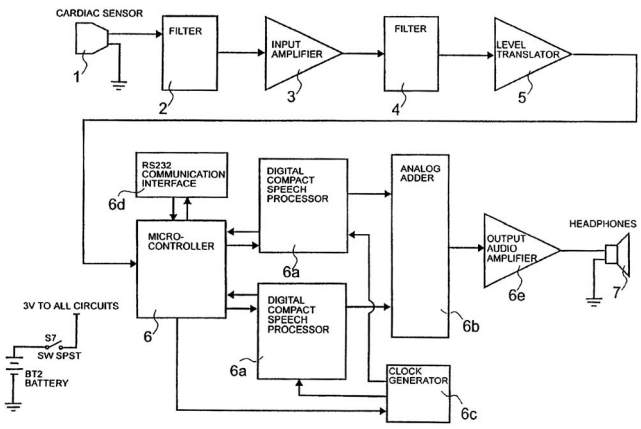
20 with coil **22** and plunger **24**. The plunger strikes the hammer 'M' **16**. Note the sensor **34** that can be recorded and then corrected via the nonreal time (NRT) factors by the controller **C**.—MK

7,207,935

43.75.St METHOD FOR PLAYING MUSIC IN REAL-TIME SYNCHRONY WITH THE HEARTBEAT AND A DEVICE FOR THE USE THEREOF

Mordechai Lipo, 96431 Jerusalem, Israel
24 April 2007 (Class 600/28); filed in Israel 21 November 1999

Here is a concept patent: the inventor claims that having music in synchrony with heartbeats is "enjoyable and relaxing." So, he proposes measuring the heart rate and somehow changing the music. Details like beat



detection algorithms and how to use that information to change the music are notably missing.—MK

7,209,153

43.75.St SYSTEM AND METHOD OF REPRESENTING PERSONAL PROFILE IN AUDITORY FORM

Barbara Lehman, Scottsdale, Arizona
24 April 2007 (Class 345/956); filed 2 March 2005

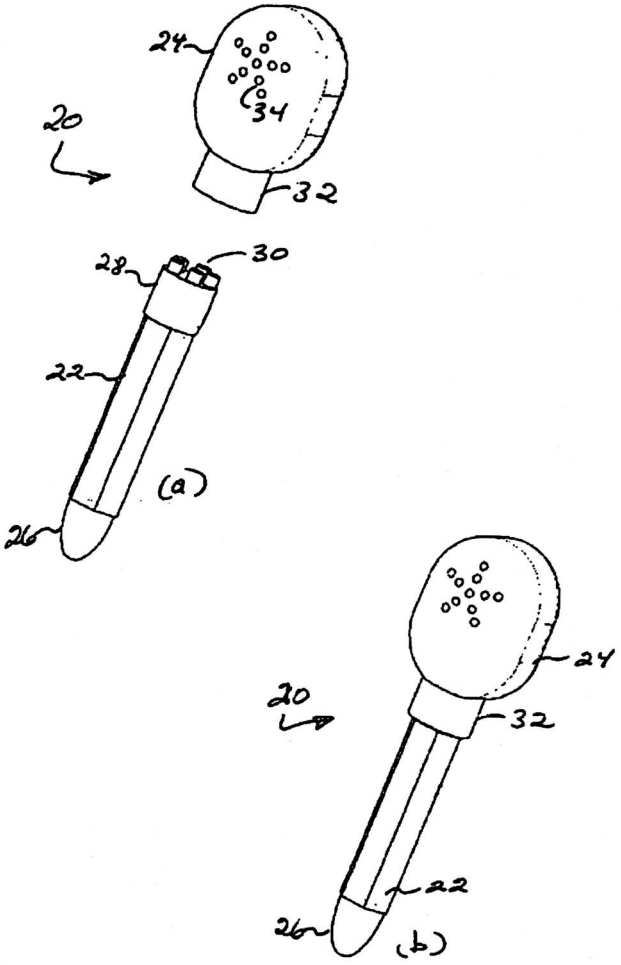
It's true! You can represent your own personality profile in music and images! Just answer this simple personality test (it will only take a minute) and your new personality profile will be generated. Naturally, details about tune generation are conspicuously missing. But lower pitches will be generated for masculine attributes. It is all true!—MK

7,204,655

43.75.Tv SOUND PRODUCING CRAYON

Henry Tsang, Richmond Hill, Ontario and Loretana Verrelli, Montreal, Quebec, both of Canada
17 April 2007 (Class 401/195); filed 21 February 2006

This is easy enough: add a ubiquitous sound chip to the top of a



crayon. But will different colors have different sounds? Synaesthesiasts would have a field day.—MK

7,207,135

43.75.Tv AUDIO FISHING LURE SYSTEM FOR ATTRACTING SPECIES SPECIFIC FISH

Jimmie D. Williams, Jr., Greenville, North Carolina
24 April 2007 (Class 43/42.31); filed 27 February 2006

If hunters can use various devices to entrap deer, elk, and moose, then why can't the fishing crowd do the same thing? Well, now they can. It is easy enough again: Store the appropriate wave form and then play it back through a waterproof transducer. You can even use a metallic (insulated) line to supply current.—MK

7,205,470

43.75.Wx PLAY DATA EDITING DEVICE AND METHOD OF EDITING PLAY DATA

Kaoru Tsukamoto et al., assignors to Oki Electric Industry Company, Limited
17 April 2007 (Class 84/609); filed in Japan 24 May 2004

Given the limited frequency range (and fidelity) of small cellular telephone loudspeakers, how is a General MIDI stream converted to a suitable output? After dividing the stream into chord, melody, and bass, the microprocessor will use "appropriate" mappings to eliminate certain sounds. It is all very vague.—MK

7,275,439

43.80.Vj PARAMETRIC ULTRASOUND IMAGING USING ANGULAR COMPOUNDING

James A. Zagzebski et al., assignors to Wisconsin Alumni Research Foundation
2 October 2007 (Class 73/625); filed 4 February 2004

Parameters such as scatterer size, spacing, density, and attenuation that characterize the structure of tissue are estimated using ultrasonic signals acquired at different angles to reduce the variance in the estimates.—RCW

7,278,968

43.80.Vj ULTRASONOGRAPH

Shin-ichiro Umemura and Ken-ichi Kawabata, assignors to Hitachi Medical Corporation
9 October 2007 (Class 600/458); filed in Japan 20 September 2001

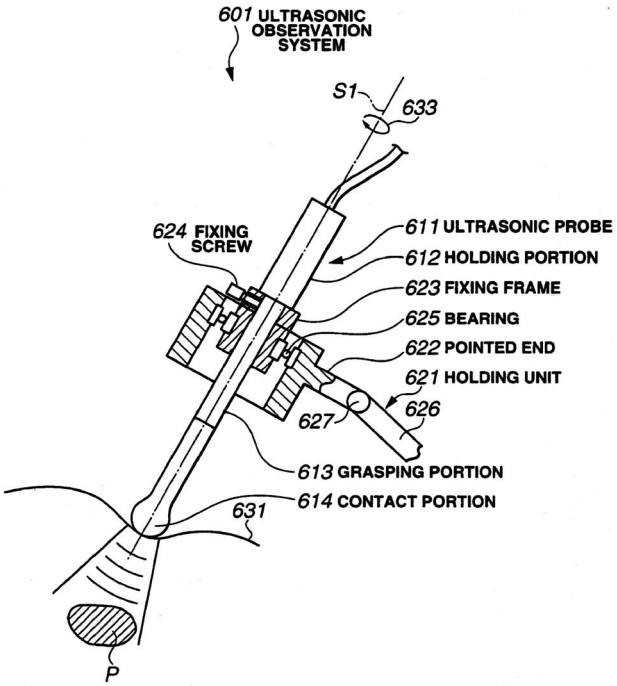
The decrease in density of a microbubble contrast agent in a given region is inhibited. This is accomplished by superimposing higher harmonics on a fundamental wave and adjusting the phase of both. The phase of the higher harmonics and the fundamental wave can also be adjusted differently to decrease the density of the microbubble contrast agent.—RCW

7,278,969

43.80.Vj ULTRASONIC OBSERVATION SYSTEM

Masaaki Ueda, assignor to Olympus Corporation
9 October 2007 (Class 600/463); filed in Japan 25 March 2002

This system has a long ultrasonic probe with transducer elements at the distal end. The probe can be arbitrarily rotated around its axis. A



mechanism to grasp and hold the probe in position is included.—RCW

7,280,694

43.80.Vj APPARATUS AND METHOD FOR IDENTIFYING AN ORGAN FROM AN INPUT ULTRASOUND IMAGE SIGNAL

Nam Chul Kim et al., assignors to Medison Company, Limited
9 October 2007 (Class 382/181); filed in Republic of Korea 23 July 2002

An organ in a human body is identified by calculating feature vectors from an ultrasonic image and then using the feature vectors to classify the organ from a database of average feature vectors and corresponding standard deviations for various organs.—RCW

7,273,455

43.80.Vj CORRECTIONS FOR WAVEFRONT ABERRATIONS IN ULTRASOUND IMAGING

Bjørn A. J. Angelsen, 7051 Trondheim, Norway et al.
25 September 2007 (Class 600/437); filed 19 July 2004

Aberration of wave fronts used in ultrasonic imaging is corrected by filters that are determined by eigenvectors of a cross-correlation matrix formed from the signals at the elements of a transducer array.—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.

Comparing turbulence models for flow through a rigid glottal model (L)

Jungsoo Suh[a] and Steven H. Frankel[b]
Maurice J. Zucrow Laboratories, School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907-2014

(Received 9 May 2007; revised 28 December 2007; accepted 2 January 2008)

Flow through a rigid model of the human vocal tract featuring a divergent glottis was numerically modeled using the Reynolds–averaged Navier–Stokes approach. A number of different turbulence models, available in a widely used commercial computational fluid dynamics code, were tested to determine their ability to capture various flow features recently observed in laboratory experiments and large eddy simulation studies. The study reveals that results from unsteady simulations employing the $k-\omega$ shear stress transport model were in much better agreement with previous measurements and predictions with regard to the ability to predict glottal jet skewing due to the Coanda effect and the intraglottal pressure distribution or related skin friction coefficient, than either steady or unsteady simulations using the Spalart–Allmaras model or any other two-equation turbulence model investigated in this study. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2836783]

PACS number(s): 43.70.Aj, 43.70.Bk [BHS] Pages: 1237–1240

I. INTRODUCTION

In human phonation, sound is generated as a result of air flowing through a time-varying orifice formed by flow-induced vibration of the human vocal folds. The quasisteady assumption is often invoked to justify substitution of unsteady flow variations by a sequence of steady flows. Although the quasisteady approximation is widely used to simplify aerodynamics between the vocal folds during phonation, complex flow physics such as turbulence transition and asymmetric flow separation due to the Coanda effect are important. Recently, Suh and Frankel[1] conducted large eddy simulation (LES) through a static glottal model, providing detailed analyses of the flow and sound field. Transition to turbulence, flow asymmetries, and the far-field sound were captured, and highlighted the prominent role of the dipole sound source associated with the unsteady intraglottal flow forces due to vortex shedding and its role in tonal sound.

Although advanced LES of vocal tract aeroacoustics can significantly improve fundamental understanding of such flows, their expense precludes their current use as a clinical tool. Hence, there is a desire to validate simpler models. Flow models based on Bernoulli's equation are limited to before the onset of flow separation and cannot explain the pressure difference across the vocal folds, boundary layer separation, or the formation of a free jet.[2] Other theoretical flow models, such as Pohlhausen or Thwaite's method, can predict the pressure and velocity in steady and unsteady flows reasonably well, but only when unsteady or viscous effects are not significant.[3,4] Despite the relatively low Reynolds number associated with laryngeal air flow ($O(10^3)$), the free jet formed just downstream of the glottis can become turbulent. Turbulence can have a dramatic influence on intraglottal pressure distribution and asymmetric flow separation due to the Coanda effect, as well as the wall shear stress. Despite the wide use of commercial computational fluid dynamics (CFD) codes, such as FLUENT, for two-dimensional laminar glottal flow,[5–10] published work focused on modeling intraglottal transition and turbulent glottal jet dynamics using such codes is limited.

In the present study, a number of different commercially available turbulence models were investigated and their capabilities to predict various flow features discussed and compared to previously published measurements and simulations.

II. PROBLEM DESCRIPTION

The geometry and mesh for the vocal tract model and glottis considered here are shown in Fig. 1. The divergent glottal shape is typical of the closing phase in human phonation. The included glottal angle was 20°. The glottal region

[a]Current address: IIHR-Hydroscience and Engineering, The University of Iowa, Iowa City, IA 52242. Electronic mail: suhj@ecn.purdue.edu
[b]Electronic mail: frankel@ecn.purdue.edu

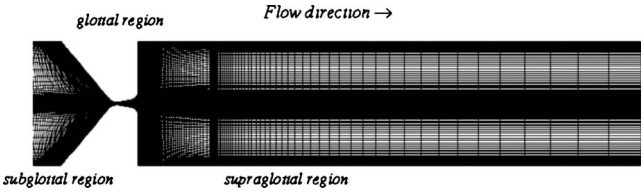


FIG. 1. Computational grid of human vocal tract model with included glottal angle of 20° (divergent glottis).

was the slit between the two vocal folds and the minimal glottal diameter (δ) was set at 0.04 cm. The vocal fold models employed here are based on the M5 geometries studied by Scherer *et al.*[5] The length of the planar channels upstream and downstream of vocal fold models are set to 12.5δ and 175δ, respectively. The computational domain includes the glottal far field, where asymmetric velocity fluctuations are usually generated,[1] to facilitate natural generation of the Coanda effect.

Two structured grids containing a total of 46,960 and 71,966 cells, respectively, were used. In the intraglottal and centerline region, the grid was refined to capture possible flow separation and complex jet flow phenomena including the transition to turbulence. The Reynolds number ($\mathrm{Re}=V_0\delta/\nu$) based on the minimum glottal diameter (δ), kinematic viscosity (ν), and reference velocity $V_0=\sqrt{s\Delta p/p}$ was 1340. The fluid properties were chosen as air (density of 1.225 $\mathrm{kg/m^3}$ and viscosity of 1.7894×10^{-5} kg/m s). The boundary conditions involved an imposed constant pressure difference (15 cm H_2O) between inlet and outlet. The inlet height and 3% were used for the hydraulic diameter and turbulence intensity at the inlet, respectively.

The commercial CFD code FLUENT was used to solve the steady and unsteady incompressible Reynolds–averaged Navier–Stokes (RANS) equations. The SIMPLE (semi-implicit method for pressure-linked equations) algorithm was used for pressure–velocity coupling. The second-order central scheme and second-order upwind scheme was used for the spatial discretization of the pressure and the other variables, respectively. The second-order implicit scheme was used for time discretization. Double-precision was used for all simulations. For the convergence criteria, $1.0-2.0\times10^{-4}$ was used. More details on the numerical method can be found in Versteeg and Malalasekera[11] and the FLUENT user's guide.[12]

III. TURBULENCE MODELS

Six different turbulence models, readily available in FLUENT, were used in the present study. The first three models are slight variations of the classical $k-\epsilon$ model. The standard $k-\epsilon$ model is a semiempirical model based on modeled transport equations for the turbulent kinetic energy (k) and its dissipation rate (ϵ). The k transport equation is derived exactly from the Navier–Stokes equations and then modeled via gradient transport. The ϵ transport equation is obtained from a dimensional analogy to the k transport equation. The standard $k-\epsilon$ model is most suitable for high Reynolds number, fully turbulent flows where the rate of dissipation and production of turbulent kinetic energy are in equilibrium.[12,13] However, this model cannot represent the suppression of turbulent mixing near the wall, which is critical for accurate skin friction predictions. Hence, it has typically been used with an additional wall treatment.[13] The Renormalization group (RNG) $k-\epsilon$ model represents an improvement for flows with strong curvature, swirl, or pressure gradient over the standard $k-\epsilon$ model, using renormalization group methods.[12] The realizable $k-\epsilon$ model was proposed to improve several deficiencies in the standard $k-\epsilon$, including the generation of negative normal stresses.[12]

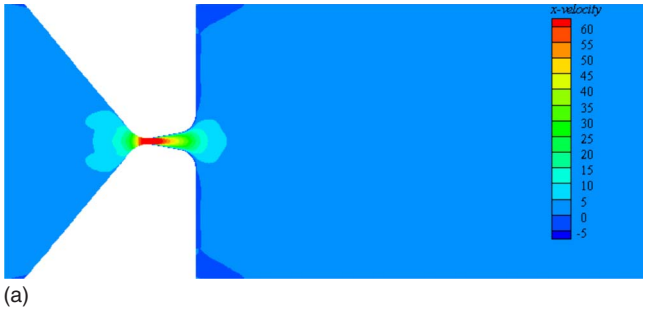
The fourth model considered herein is the low-Reynolds-number $k-\omega$ model, which is another semiempirical model based on modeled transport equations for k and its specific dissipation rate (ω), which can be thought as ϵ/k. It is reported to work well in the near-wall region without any additional wall damping.[13] However, this model has a very strong sensitivity to the free stream value of ω for the outer region of boundary layer and free shear flows.[14] The low-Reynolds-number version of this model has demonstrated good performance for wall bounded low-Reynolds-number flows and transitional flows.[12]

The fifth model is the shear-stress transport (SST) $k-\omega$ model (hereafter as the SST model). The SST model was proposed by Menter[14] to overcome two weak points of the original $k-\omega$ model. Its tendency to overpredict shear stress in adverse pressure gradient regions was fixed by imposing a bound on the shear stress ratio. Hence, the turbulent shear stress is predicted to be proportional to the k in the wake region of the boundary layer as previously observed. In addition, a blending function of the $k-\omega$ model near the wall and the $k-\epsilon$ model for the rest of the flow was used to remove strong free stream sensitivity of the original $k-\omega$ model.[13,14]

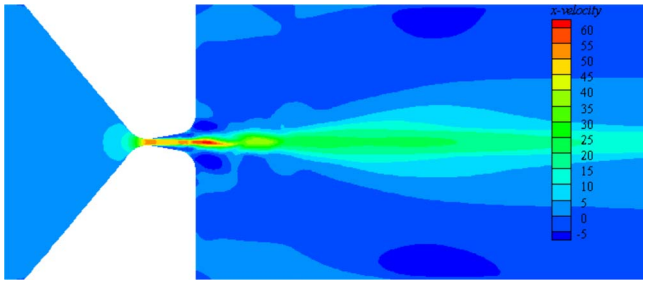
The last model used in the present work is the Spalart–Allmaras (SA) one-equation model, a popular model for aerodynamic flow simulations (hereafter as the SA model). In this model, a phenomenologically based nonlinear transport equation is solved for the eddy viscosity.[15] A system rotation and streamline curvature corrected version of the SA model was suggested,[16] but is not included in the present study as it was unavailable in most recent version of the FLUENT code used in this study.

It should be noted that all turbulence models considered in the present study employed an eddy viscosity based on the Boussinesq hypothesis, which assumes isotropic turbulence and equilibrium between Reynolds stress and mean rate of strain.[13] In the presence of an adverse pressure gradient or separated flow region, this is not strictly true.

For the near-wall treatment in turbulence modeling, the enhanced wall treatment approach was used in this study. This method requires the restriction that the near-wall mesh must be sufficiently fine to resolve the viscous sublayer (typically $y^+\approx1$). Although this approach requires significant computational resources, it is more appropriate for low-Reynolds-number flows or separated flows. In the present study, 62 grid points were placed between the two vocal folds and the first node is located 0.00018 cm from the wall. All simulation results were checked to resolve the near wall



(a)



(b)

FIG. 2. (Color online) Streamwise velocity: (a) realizable $k-\epsilon$ model and (b) $k-\omega$ SST model.

region after computation. The y^+ value of the first mesh from the RNG $k-\epsilon$ model was around 1.3, whereas the other models showed smaller values.

IV. RESULTS AND DISCUSSION

Figure 2 plots streamwise velocity contours from two representative two-dimensional steady RANS computation obtained using the realizable $k-\epsilon$ and the SST model. The other $k-\epsilon$ models showed similar results to the realizable $k-\epsilon$ model, whereas the low-Reynolds-number $k-\omega$ model and the SA model showed similar results to the SST model. Computations using the $k-\epsilon$ models failed to respond to the adverse pressure gradient and were unable to predict separation from vocal fold. Consequently, the glottal jet predicted by these models could not fully develop as seen in the SST model, the low-Reynolds-number $k-\omega$, and the SA models. Similar lackluster performance of the $k-\epsilon$ model in the presence of an adverse pressure gradient was previously reported.[17] Although the SST, low-Reynolds-number $k-\omega$, and SA models all showed slightly better predictions, steady computations using these models were unable to predict flow asymmetries and the resulting discrepancy between the flow-wall and non-flow-wall pressure distributions as observed in the previous measurements.[18] This is not surprising because the present divergent glottal shape corresponds to the large transitory stall regime,[19] which is inherently unsteady. Nevertheless, unsteady simulations with the $k-\epsilon$ model series or the SA model show almost same result as their steady counterparts.

On a more positive note, it was found that unsteady two-dimensional simulations using the SST model, which can account for the effects of an adverse pressure gradient on the turbulent shear stress, was able to predict glottal jet skewing due to the Coanda effect (see Fig. 3). As shown in

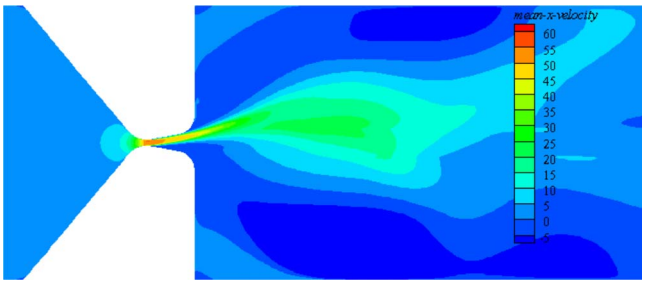


FIG. 3. (Color online) Mean streamwise velocity: unsteady $k-\omega$ SST model.

Fig. 4, the unsteady SST model was in good agreement with experimental data[18] and LES results[1] for the intraglottal pressure distribution on the nonflow-wall and flow wall up to the maximum pressure drop. However, the unsteady SST model underpredicted the pressure drop near the glottal exit due to its enhanced diffusivity. The unsteady simulation of the low-Reynolds-number $k-\omega$ model did predict flow asymmetry for one specific inlet condition, but a slight variation of the inlet condition altered the entire flow field. The strong inlet condition sensitivity for this model was previously reported[14] and hence this model is not recommended for glottal jet simulations.

Figure 5 depicts profiles of mean skin friction coefficient $C_f=\tau_w/\rho U_0^2/2$, based on the estimated maximum velocity $U_0=\sqrt{2\Delta P/\rho}$, within the glottis and immediately downstream. Until the separation point in the LES, both the unsteady computation with the SST model and the LES are in good agreement. At the minimum glottal diameter, both cases predieted similar values for the highest skin friction on both flow and nonflow walls. However, discrepancies can be noted downstream of the separation point and are most likely due to the isotropic equilibrium assumption associated with all turbulence models. Because turbulence is fully three dimensional and anisotropic when vortices continue to move along the wall after separated from it. The present divergent glottis is in the large transitory stall regime.[19] Note that the

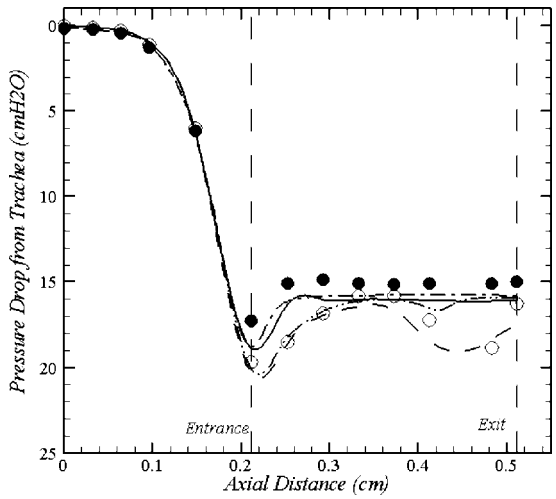


FIG. 4. Intraglottal pressure distribution: nonflow wall/experiment (●), flow wall/experiment (○), nonflow wall/LES (—), flow wall/LES (- - -), nonflow wall/unsteady SST (–··–), and flow wall/unsteady SST (–···–).

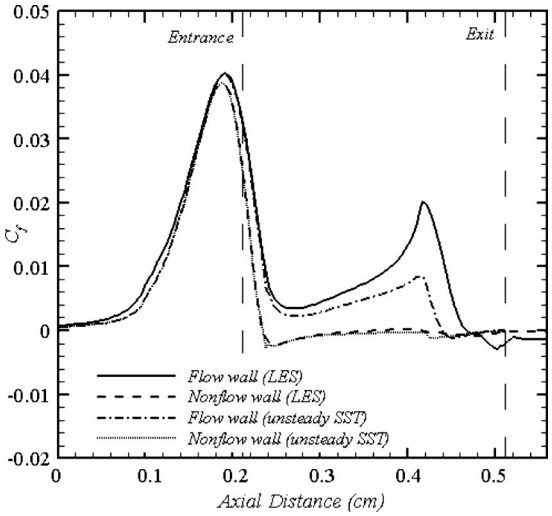


FIG. 5. Mean skin friction coefficient C_f based on the estimated maximum velocity $\sqrt{2\Delta P/\rho}$.

unsteady computation using the SST model predicts almost same value of the skin friction coefficient of nonflow wall with LES. This suggests the unsteady simulation using the SST model can predict the flow separation and its location, which are very important for phonation study, with reasonable accuracy.

V. CONCLUSIONS

RANS predictions, obtained from a commercial CFD code, comparing a number of different turbulence models for flow through a rigid model of the human vocal tract with a divergent glottis, showed that *unsteady* simulations using the $k-\omega$ SST model were in the best agreement with previously published measurements and simulations with regard to prediction of glottal flow asymmetry and jet separation as compared to all the other one- or two-equation turbulence models investigated in this study. These findings suggest that unsteady simulations using the $k-\omega$ SST model may offer an acceptable trade-off between accuracy and efficiency, which, when combined with wall motion for fluid–structure interaction studies and/or coupled to an acoustic model for aeroacoustic studies, may serve as the core flow prediction tool in a more comprehensive phonation model. Although laryngeal acoustics were not considered in the present study, the significant role of the adverse pressure gradient on the turbulent shear stress distribution and hence aerodynamic glottal flow forces, highlighted here, has important ramifications for phonation aeroacoustics through the dominant dipole sound source and warrants further study.

ACKNOWLEDGMENTS

This work was supported by Research Grant Nos. R01 DC 03577 and R01 DC 05788 from the National Institute on Deafness and other Communication Disorders, National Institutes of Health.

[1]J. Suh and S. H. Frankel, "Numerical simulation of turbulence transition and sound radiation for flow through a rigid glottal model," J. Acoust. Soc. Am. **121**, 3728–3739 (2007).
[2]A. Hirschberg and S. Rienstra, *Large-Eddy Simulation for Acoustics*, 1st ed. (Cambridge University Press, New York, 2007), pp. 24–88.
[3]X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, and Y. Auregan, "Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431 (1994).
[4]C. E. Vilain, X. Pelorson, C. Fraysse, M. Deverge, A. Hirschberg, and J. Willems, "Experimental validation of a quasi-steady theory for the flow through the glottis," J. Sound Vib. **276**, 475–490 (2004).
[5]R. C. Scherer, D. Shinwari, K. J. De Witt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh, "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630 (2001).
[6]R. C. Scherer, K. J. De Witt, and B. R. Kucinschi, "The effect of exit radii on intraglottal pressure distributions in the convergent glottis," J. Acoust. Soc. Am. **110**, 2267–2269 (2001).
[7]S. L. Thomson, L. Mongeau, and S. H. Prankel, "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700 (2005).
[8]S. Li, R. C. Scherer, M. Wan, S. Wang, and H. Wu, "The effects of glottal angle on intraglottal pressure," J. Acoust. Soc. Am. **119**, 539–548 (2006).
[9]S. Li, R. C. Scherer, M. Wan, S. Wang, and H. Wu, "Numerical study of the effects of inferior and superior vocal fold surface angles on vocal fold pressure distributions," J. Acoust. Soc. Am. **119**, 3003–3010 (2006).
[10]B. R. Kucinschi, R. C. Scherer, K. J. DeWitt, and T. T. Ng, "An experimental analysis of the pressures and flows within a driven mechanical model of phonation," J. Acoust. Soc. Am. **119**, 3011–3021 (2006).
[11]H. K. Versteeg and W. Malalasekera, *Introduction to Computational Fluid Dynamics*, 1st ed. (Longman Scientific and Technical, Essex, UK, 1995).
[12]*FLUENT 6.2 User Guide* (Fluent Inc., , Lebanon, NH, 2005).
[13]P. A. Durbin and B. A. Pettersson Reif, *Statistical Theory and Modeling for Turbulent Flows*, 1st ed. (Wiley, West Sussex, UK, 2001).
[14]F. R. Menter, "Two-equation eddy-viscosity turbulence models for engineering applications," AIAA J. **32**, 1598–1605 (1994).
[15]P. R. Spalart and S. R. Allmaras, "A one equation model for aerodynamic flows," Rech. Aerosp. **1**, 5–21 (1994).
[16]M. L. Shur, M. K. Strelets, A. K. Travin, and P. R. Spalart, "Turbulence modeling in rotating and curved channels: Assessing the Spalart-Shur correction," AIAA J. **38**, 784–792 (2000).
[17]G. Iaccarino, "Predictions of a turbulent separated flow using commercial CFD codes," J. Fluid Mech. **123**, 819–828 (2001).
[18]B. Erath, "An experimental investigation of velocity fields in diverging glottal models of the human vocal tract," Master's thesis, School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, 2005.
[19]R. W. Fox and S. J. Kline, "Flow regimes in curved subsonic diffusers," J. Fluids Eng. **84**, 303–316 (1962).

Near field Rayleigh wave on soft porous layers

N. Geebelen, L. Boeckx,[a] G. Vermeir, and W. Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Heverlee, Belgium

J. F. Allard and O. Dazel
Laboratoire d'Acoustique de l'Université du Maine, Avenue Olivier Messiaen, 72000 Le Mans, France

(Received 24 September 2007; revised 23 November 2007; accepted 12 December 2007)

Simulations performed for a typical semi-infinite reticulated plastic foam saturated by air show that, at distances less than three Rayleigh wavelengths from the area of mechanical excitation by a circular source, the normal frame velocity is close to the Rayleigh pole contribution. Simulated measurements show that a good order of magnitude estimate of the phase speed and damping can be obtained at small distances from the source. Simulations are also performed for layers of finite thickness, where the phase velocity and damping depend on frequency. They indicate that the normal frame velocity at small distances from the source is always close to the Rayleigh pole contribution and that a good order of magnitude estimate of the phase speed of the Rayleigh wave can be obtained at small distances from the source. Furthermore, simulations show that precise measurements of the damping of the Rayleigh wave need larger distances. Measurements performed on a layer of finite thickness confirm these trends. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2832326]

PACS number(s): 43.20.Gp, 43.20.Jr, 43.20.Ye [LLT] Pages: 1241–1247

I. INTRODUCTION

Soft porous media like plastic foams and glass wools are widely used in room and building acoustics and by aircraft and car manufacturers as sound and vibration absorbers. Evaluating the rigidity coefficients at audio frequencies can be necessary to predict the performances of these materials, but most of the available methods solely allow measurements restricted to the quasistatic range.[1–4] It has been shown by Feng and Johnson that for a nondamped semi-infinite porous medium saturated by a light fluid and excited mechanically, the Rayleigh wave is the detectable surface mode.[5] In case of acoustical excitation, another type of surface wave will be dominant.[6] This wave is associated with the penetration of sound into the pores and is detectable by means of a microphone instead of a solid particle velocity sensor. Measuring the phase speed and damping of the Rayleigh wave in the audible frequency range can provide information concerning the rigidity coefficients. Some general properties of the frames of these media must be taken into account. The loss angle is large, around 1/10, and the velocity of the Rayleigh waves is around 50 m/s. The available porous samples have a finite thickness, generally around 2 cm, and measurements must be performed in the medium and the high frequency range to avoid strong finite thickness effects. Simulations have been performed previously with a short excitation in the time domain. It was shown by Allard *et al.*[7] that at a distance from the source large compared to the Rayleigh wavelength, for an ordinary damped soft medium saturated by air, the vertical deformation at the surface of the medium is mainly due to the contribution of the Rayleigh pole related to the elastic frame, slightly modified by the light saturating fluid. If this last property holds at small distances from the source, measurements at small distances of the source of the speed of the Rayleigh wave should be possible and easier than at large distances due to the large damping in most soft porous media. Moreover, measurements on small samples should be possible. Simulated measurements of the damping and the speed of the Rayleigh wave created by a circular source are performed in the time domain close to the source for a typical reticulated plastic foam. The relation between the simulated measurements of the speed and the damping of the Rayleigh wave and the phase speed and damping related to the Rayleigh pole at a given frequency is detailed, including the effects of finite layer thickness.

II. FRAME DISPLACEMENT INDUCED BY A SURFACE STRESS FIELD

The porous layer is represented in Fig. 1. It is in contact with air at the upper face. The layer is semi-infinite or has a finite thickness l. The layer with a finite thickness is bonded to an impervious rigid backing at the lower face. In this case the x and z displacement components of the frame and the z displacement component of the saturating air are equal to 0 at the lower face. Simulations have been performed for two normal stress fields with a time dependence $f(t)$. The true excitation is created by a small vibrating disc, with radius R, bonded onto the porous layer. Describing this excitation by means of a stress field is difficult. Instead of using a constant stress acting on a limited surface area $r<R$, an expression is used that will yield a fast and regular decreasing of its Hankel transform. The space dependence g_1 of the stress field acting on the frame at the air–porous layer interface is given by

[a]Electronic mail: laurens.boeckx@fys.kuleuven.be

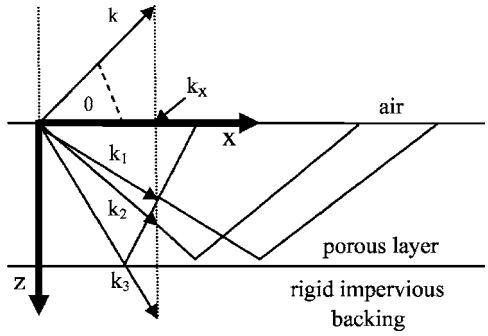


FIG. 1. The porous sample.

$$g_1(r) = (a_1/\pi)^{1/2} \exp(-a_1 r^2), \tag{1}$$

where r is the radial distance from the origin. The total force acting on the surface area $r < r_0$ is equal to $[1-\exp(-a_1 r_0^2)] \times (\pi/a_1)^{1/2}$ and the main contribution of the stress field is localized on the surface area $r < 2a_1^{1/2}$. In a first step the Biot theory,[8] with the formalism developed in Ref. 9, is used to predict the vertical displacement u_z^s of the frame at the air–porous layer interface created by a stress field $\sigma_{zz} = \exp(ik_x x - i\omega t)$ acting on the frame at the surface of the layer. The plane waves that can propagate in a porous layer bonded to a rigid impervious backing are described in Ref. 10. The model by Johnson *et al.*[11] is used to describe the viscous and the inertial interactions and the model by Lafarge[12] is used for the bulk modulus of the saturating air. Let k_1 and k_2 be the wave numbers of the two Biot compressional waves and k_3 the wave number of the Biot shear wave. Let $\varphi_j^{\pm}$, $j=1,2$, be the velocity potentials of the compressional waves with the related displacement $\mathbf{u}^s = i\nabla\varphi\omega$. $\mathbf{\Psi}^{\pm} = \mathbf{n}\varphi_3^{\pm}$ are the velocity potentials of the shear waves, $\mathbf{n}$ being the unit vector on the y axis, with the related frame displacement $\mathbf{u}^s = i\nabla \wedge \mathbf{\Psi}/\omega$. The functions $\varphi_j^{\pm}$ can be written as

$$\varphi_j^{\pm} = a_j^{\pm} \exp(\pm i\alpha_j z + i\xi x), \quad j = 1,2,3, \tag{2}$$

$$\alpha_j = (k_j^2 - \xi^2)^{1/2}. \tag{3}$$

Three frame displacement fields $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ are defined in Ref. 10 by the relations

$$\mathbf{b}_1 = \frac{i}{\omega}[\nabla(\exp(i\alpha_1 z) + r_{1,1}\exp(-i\alpha_1 z) + r_{1,2} \times \exp(-i\alpha_2 z)) + \nabla \wedge \mathbf{n} r_{1,3} \exp(-i\alpha_3 z)], \tag{4}$$

$$\mathbf{b}_2 = \frac{i}{\omega}[\nabla(r_{2,1}\exp(-i\alpha_1 z) + \exp(i\alpha_2 z) + r_{2,2} \times \exp(-i\alpha_2 z)) + \nabla \wedge \mathbf{n} r_{2,3} \exp(-i\alpha_3 z)], \tag{5}$$

$$\mathbf{b}_3 = \frac{i}{\omega}[\nabla(r_{3,1}\exp(-i\alpha_1 z) + r_{3,2}\exp(-i\alpha_2 z)) + \nabla \wedge \mathbf{n}(\exp(i\alpha_3 z) + r_{3,3}\exp(-i\alpha_3 z))]. \tag{6}$$

The coefficients r_{ij} are chosen so that each $\mathbf{b}_i$ satisfies the boundary conditions at the lower face. For a semi-infinite layer, the parameters r_{ij} are equal to 0. Let σ_{ij}^f be the stress components of the air in the porous medium and let σ_{ij}^s be the stress components of the frame. These stress components are related to forces per unit area of porous medium. At the upper face, the boundary conditions can be written, the common factor $\exp(ik_x x - i\omega t)$ being removed,

$$\phi \dot{u}_z^f + (1-\phi)\dot{u}_z^s = v_z, \tag{7}$$

$$\sigma_{zz}^f = -\phi p, \tag{8}$$

$$\sigma_{zz}^s = 1 - (1-\phi)p, \tag{9}$$

$$\sigma_{xz}^s = 0, \tag{10}$$

where ϕ is the porosity of the porous medium, $\dot{u}_z^f$ and $\dot{u}_z^s$ are the z components of the velocity of the saturating air and of the frame at the air–porous layer interface, p is the pressure, and v_z is the z velocity component in the free air close to the boundary. The pressure p and the velocity component v_z are related to a wave created in the free air at the boundary with a spatial dependence $\exp[i(xk_x - zk\cos\theta)]$, where k is the wave number in the free air. They satisfy the relation

$$v_z = -Zp/\cos\theta, \tag{11}$$

where Z is the characteristic impedance of air and θ is defined by $\sin\theta = k_x/k$. Using Eq. (11), p and v_z can be removed from the set of equations (7)–(9) that becomes

$$\sigma_{zz}^s - \frac{1-\phi}{\phi}\sigma_{zz}^f = 1, \tag{12}$$

$$(\phi\dot{u}_z^f + (1-\phi)\dot{u}_z^s) - \frac{Z}{\cos\theta}\frac{\sigma_{zz}^f}{\phi} = 0. \tag{13}$$

The frame displacement field in the porous medium is the sum $\lambda_1\mathbf{b}_1 + \lambda_2\mathbf{b}_2 + \lambda_3\mathbf{b}_3$ that satisfies Eqs. (10), (12), and (13). In Ref. 10, σ_{xy}^s, $\dot{u}_z^s$, $\dot{u}_z^f$, σ_{zz}^f, and σ_{zz}^s are given in Eqs. (A1)–(A5) by

$$\sigma_{xy}^s = M_{11}\lambda_1 + M_{12}\lambda_2 + M_{13}\lambda_{33}, \tag{14}$$

$$\dot{u}_z^s = M_{21}\lambda_1 + M_{22}\lambda_2 + M_{23}\lambda_3, \tag{15}$$

$$\dot{u}_z^f = M_{31}\lambda_1 + M_{32}\lambda_2 + M_{33}\lambda_3, \tag{16}$$

$$\sigma_{zz}^f = M_{41}\lambda_1 + M_{42}\lambda_2 + M_{43}\lambda_3, \tag{17}$$

$$\sigma_{zz}^s = M_{51}\lambda_1 + M_{52}\lambda_2 + M_{53}\lambda_3. \tag{18}$$

The following expressions for λ_1, λ_2, λ_3 are obtained:

$$\lambda_1 = \frac{B_2 M_{13} - B_3 M_{12}}{\Delta}, \tag{19}$$

$$\lambda_2 = \frac{B_3 M_{11} - B_1 M_{13}}{\Delta}, \tag{20}$$

$$\lambda_3 = \frac{B_1 M_{12} - B_2 M_{11}}{\Delta}, \tag{21}$$

where B_i, A_i, $i=1,2,3$, are given by

$$B_i = \phi M_{3i} + (1-\phi)M_{2i} - \frac{Z}{\cos\theta}\frac{M_{4i}}{\phi}, \tag{22}$$

$$A_i = M_{5i} - \frac{1-\phi}{\phi}M_{4i}, \tag{23}$$

$$\Delta = \begin{bmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ M_{11} & M_{12} & M_{13} \end{bmatrix}. \tag{24}$$

The z component of the frame velocity is given by Eq. (15) where λ_1, λ_2, λ_3 are given by Eqs. (19)–(21). The Hankel transform $\hat{g}_1$ of g_1 is given by

$$\hat{g}_1(k_r) = \int_0^\infty r g_1(r) J_0(k_r r) dr \tag{25}$$

and the z component of the frame velocity $v(r,t)$ at r is given by

$$v(r,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega \int_0^\infty dk_r \times \exp(-i\omega t) k_r \hat{g}_1(k_r) J_0(k_r r) \dot{u}(k_r). \tag{26}$$

Using $J_0(u) = 0.5[H_0^{(1)}(u) - H_0^{(1)}(e^{i\pi}u)]$ (Ref. 13, Chap. 9) and $\dot{u}(k_r) = \dot{u}(-k_r)$, Eq. (26) can be rewritten

$$v(r,t) = \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} dk_r \exp(-i\omega t) k_r \hat{g}_1(k_r) H_0^{(1)} \times (k_r r)\dot{u}(k_r). \tag{27}$$

The function $\hat{g}_1$ is given by

$$\hat{g}_1(k_r) = \exp\left(\frac{-k_r^2}{4a_1}\right)/(2\sqrt{\pi a_1}). \tag{28}$$

A zero of the determinant Δ at $k_x = k_p$ corresponds to a pole of v at $k_r = k_p$, the contribution of the pole residue for v in Eq. (27) is given by

$$v_p(r,t) = \frac{i}{2}\int_{-\infty}^{\infty} d\omega \exp(-i\omega t) k_p \hat{g}_1(k_p) H_0^{(1)}(k_p r) R(k_p). \tag{29}$$

In Eq. (29), $R(k_p)$ is given by

$$R(k_p) = M_{21}\frac{B_2 M_{13} - B_3 M_{12}}{[\partial\Delta(k_x)/\partial k_x]_{k_x=k_p}} + M_{22}\frac{B_3 M_{11} - B_1 M_{13}}{[\partial\Delta(k_x)/\partial k_x]_{k_x=k_p}} + M_{23}\frac{B_1 M_{12} - B_2 M_{11}}{[\partial\Delta(k_x)/\partial k_x]_{k_x=k_p}}. \tag{30}$$

Another spatial dependence, $g_2(x)$, has been considered for the stress field source, given by

$$g_2(x) = \cos a_2 x, \quad 0 < |x| < \frac{\pi}{2a_2} \tag{31}$$

with $a_2 = 250\pi$. Using the double Fourier transform

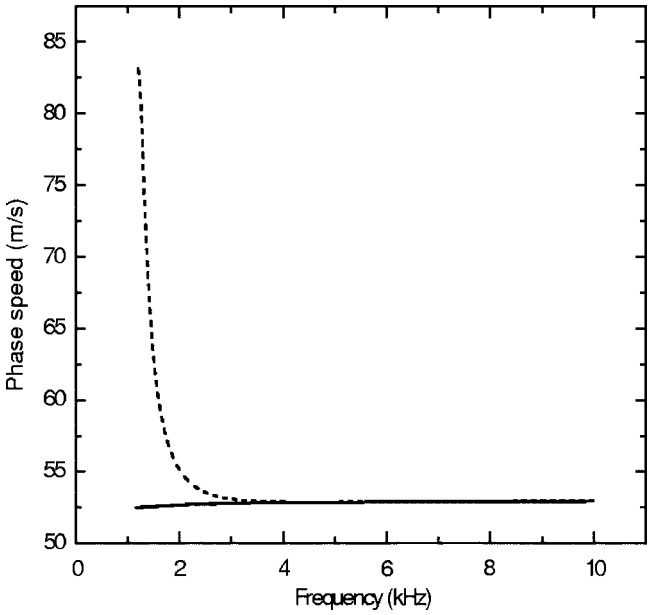


FIG. 2. Phase speed $\omega/\mathrm{Re}\,k_p$ of the modified Rayleigh wave, parameters of Table I, semi-infinite layer (—), thickness $l=2$ cm (- - -).

$$\tilde{g}_2(k_x)\tilde{f}(\omega)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dt dx \exp(i\omega t - ik_x x) g_2(x) f(t), \tag{32}$$

the z component of the frame velocity at the surface of the porous medium is given by

$$v(x) = (2\pi)^{-2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\omega dk_x \times \exp(-i\omega t + ik_x x)\tilde{g}_2(k_x)\tilde{f}(\omega)\dot{u}_z^s(k_x). \tag{33}$$

The function $\tilde{g}_2$ is given by

$$\tilde{g}_2(k_x) = \frac{2a_1}{a_1^2 - k_x^2}\cos\frac{\pi k_x}{2a_1}. \tag{34}$$

The contribution of the pole residue for v is given by

$$v_p(x,t) = \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \times \exp(-i\omega t + ik_p x)\tilde{g}_2(k_p)\tilde{f}(\omega)\dot{u}_z^s(k_p)R(k_p). \tag{35}$$

III. SIMULATIONS

A. Rayleigh pole

The Rayleigh pole is located at $k_x = k_p$ where $\dot{u}_z^s$ is infinite. The wave number component k_p is obtained by an iterative method which minimizes $1/\dot{u}_z^s$ as a function of k_x. The initial value of k_x is the wave number of the Biot shear wave. The quantity $\omega/\mathrm{Re}\,k_p$, which is the phase speed of the Rayleigh wave, is represented in Fig. 2 for a layer of the medium of Table I similar to current reticulated plastic foams with a semi-infinite thickness and a thickness $l=2$ cm, as a function of frequency, and $\mathrm{Im}\,k_p$ is represented in Fig. 3. There is a strong finite thickness effect at frequencies lower than 2 kHz. At higher frequencies, k_p is not noticeably modified by the finite thickness.

TABLE I. Parameters for the porous material.

Porosity ϕ (—)	0.99
Flow resistivity σ (N m^{-4} s)	12 000
Thermal permeability k_0' (m^2)	1.5×10^{-9}
Viscous dimension Λ (μm)	100
Thermal dimension Λ' (μm)	400
Tortuosity α_∞ (—)	1.01
Density ρ_s (kg/m^3)	24.5
Shear modulus N (kPa)	$80(1+i/8)$
Poisson ratio [ν] (—)	0.3

B. Frame velocity

Simulations have been performed of the frame velocity resulting from a normal stress field $g_1(r)$ with a time dependence $f(t)$ and with the porous medium described in Table I. The time dependence $f(t)$ is the product of a sine function at different frequencies and a Blackman–Harris window with a duration equal to ten periods. The integration in ω in Eq. (27) is replaced by a fast Fourier transform. For the spatial dependencies g_1 of the excitation, $a_1=1.725\times10^5$. The integral in k_x in Eq. (27) is performed on the real axis in the physical Riemann sheet of the complex k_r plane. If the porous layer is semi-infinite, the physical sheet is characterized by

$$\mathrm{Im}\cos\theta>0, \tag{36}$$

$$\mathrm{Im}\,\alpha_i>0,\quad i=1,2,3. \tag{37}$$

If the porous sample is semi-infinite, the initial path of integration can be replaced by the three cuts related to the three Biot waves and the cut related to the wave in air in the same sheet. The residues of the poles which are crossed when the initial path is modified must be added to the integral on the new path. The determinant Δ for a semi-infinite layer is a polynomial in k_x and there is a finite number of poles. If the porous sample has a finite thickness the three cuts related to the Biot waves disappear and the determinant Δ has an infinite number of roots. The three cuts are replaced by the contributions of an infinite number of poles. The transition between the semi-infinite thickness and the finite thickness is described by Tamir and Felsen[14] for electromagnetic waves in the presence of a dielectric slab. The z component of the velocity v of the porous frame as a function of time at $r=3$ cm calculated with Eq. (27) is represented in Fig. 4 for a semi-infinite layer and in Fig. 5 for a layer of thickness $l=2$ cm. The contribution v_p of the Rayleigh pole given by Eq. (29) is also represented in Figs. 4 and 5. The pole contribution is limited to the frequencies in the interval 0.8–10 kHz in Fig. 5. It was verified that the contribution of the frequencies out of this interval is negligible. The Ray-

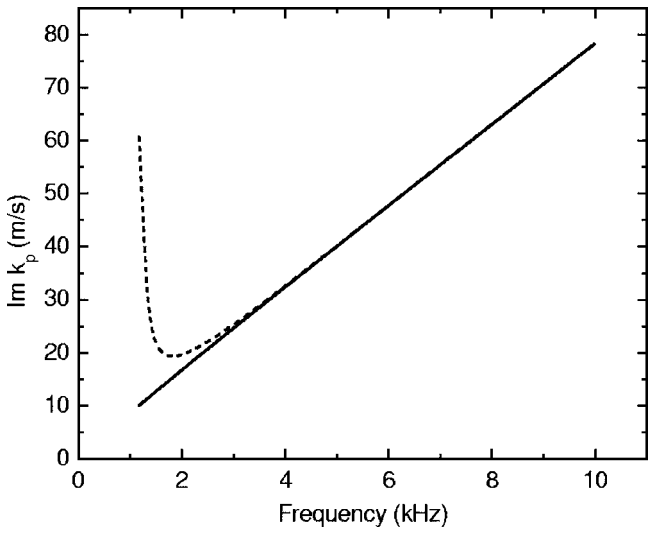


FIG. 3. Damping Im ξ_p of the modified Rayleigh wave, parameters of Table I, semi-infinite layer (—), thickness $l=2$ cm (- - -).

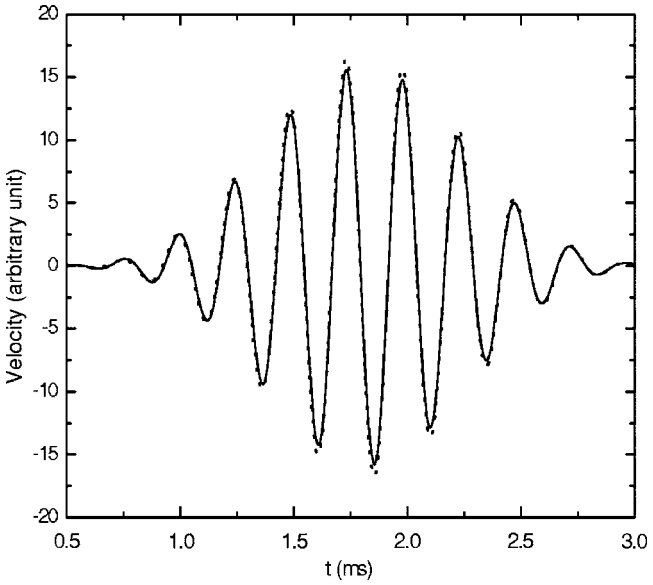


FIG. 4. Normal velocity v of the porous frame, semi-infinite layer, parameters of Table I, central frequency 4 kHz, circular source, $r=3$ cm, integral on the real k_r axis (—), pole contribution (- - -).

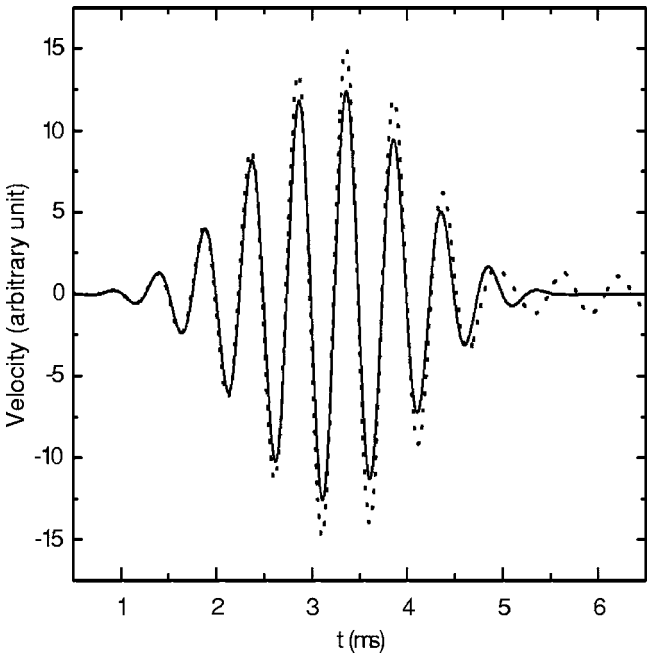


FIG. 5. Normal velocity v of the porous frame, thickness $l=2$ cm, parameters of Table I, central frequency 2 kHz, circular source, $r=3$ cm, integral on the real k_r axis (—), pole contribution (- - -).

TABLE II. Predicted $\omega/\mathrm{Re}\,k_r$ and $\mathrm{Im}\,k_r$ and simulated measurements of c_r and D at 4 kHz. Semi-infinite layer and finite thickness $l=5$ cm.

	$\omega/\mathrm{Re}\,k_p$ (m/s), c_r	$\mathrm{Im}\,k_p$ (m^{-1}), D
Prediction (semi-infinite layer) $\omega/\mathrm{Re}\,k_p$, $\mathrm{Im}\,k_p$	*52.8*	*32.4*
Simulation 3–4 cm pole c_r, D	52.6	32.5
Simulation 3–4 cm c_r, D	53.8	30.5
Simulation 15–16 cm pole c_r, D	53.2	32.0
Simulation 15–16 cm c_r,D	53.2	33.8
Simulation 3–5 cm $l=5$ cm c_r, D	53.2	36.1

leigh pole contribution is close to the total displacement in both cases, especially for the semi-infinite layer. Due to the fact that the contribution of the Rayleigh wave is dominant, an approximateevaluation of the speed and the damping of the Rayleigh wave should be possible at short distances from the source.

C. Simulated measurement of the speed and the damping of the Rayleigh wave

The intercorrelation function $I(\tau)$ given by

$$I(\tau)=\int f(t+\tau)v(r,t)dt \tag{38}$$

is successively predicted for a couple r_1, r_2 with $r_2>r_1$. Let I_1 and I_2 be the maxima of I at the distances r_1 and r_2, respectively, and let τ_1 and τ_2 be the related time lags. The speed c_r is evaluated by

$$c_r=(r_2-r_1)/(\tau_2-\tau_1). \tag{39}$$

Using the asymptotic expression of the Hankel function (Ref. 13, Chap. 9)

$$H_0^{(1)}(i)=\left(\frac{2}{\pi u}\right)^{1/2}\exp[i(u-\pi/4)], \tag{40}$$

the damping D is evaluated by

$$D=\frac{1}{r_2-r_1}\log\left(\sqrt{\frac{r_1}{r_2}}I_1/I_2\right). \tag{41}$$

The simulated measurements of c_r and D at the center frequency 4 kHz of the excitation are compared with the phase speed $\omega/\mathrm{Re}\,k_p$ and $\mathrm{Im}\,k_p$ at the same frequency in Table II for the semi-infinite layer. The simulated measurements are performed for the couple $r_1=3$ cm, $r_2=4$ cm and the couple $r_1=15$ cm, $r_2=16$ cm. In Eq. (38), I is successively evaluated with v and with v_p. The predicted values of the speeds c_r and $\omega/\mathrm{Re}\,k_p$ and of D and $\mathrm{Im}\,k_p$, respectively, are close to each other for both couples r_1, r_2. They are closer for the pole contribution than for the full contribution, and slightly closer for the couple 15–16 cm than for the couple 3–4 cm. This shows that for the chosen time dependence of the excitation at a given center frequency the Rayleigh wave speed and damping are close to the phase speed and $\mathrm{Im}\,k_p$ at the same frequency. The comparisons also show that the limitation in the precision of the evaluation is mainly due to the contributions of the cuts which decrease faster than the contribution of the pole with the distance. Simulations performed at 4 kHz for a layer of thickness $l=5$ cm present the same trends as for the semi-infinite layer. In Table III, the simulated measurements of the speed c_r and the damping D are compared with the predicted phase speeds $\omega/\mathrm{Re}\,k_p$ and $\mathrm{Im}\,k_p$ modified by the finite thickness $l=2$ cm, for the central frequencies 2, 3, and 4 kHz. The predicted values of the speeds are close to each other with a precision better than 10/100. The simulated measurement of D can be very different from $\mathrm{Im}\,k_p$. At 4 kHz, a simulated measurement of c_r and D is performed where the full evaluation of v_z is replaced by the pole contribution. With v_2 restricted to the pole contribution, in the last line of Table III, the quantities c_r and $\omega/\mathrm{Re}\,k_p$, and D and $\mathrm{Im}\,k_p$, respectively, are close to each other. This shows that the differences between D and $\mathrm{Im}\,k_p$ when the full evaluation of v is performed with Eq. (27) is not due to the dispersion of $\mathrm{Im}\,k_p$ but to the contribution of the other poles of $\dot{u}_z$. The fluctuation in amplitude due to the contribution of the other poles is much lower than the variation due to the damping when the difference r_2-r_1 is sufficiently large. This allows more precise measurements of $\mathrm{Im}\,k_p$ via the measurement of D by using larger distances r_2-r_1. The simulated measurements of D given in Table III for the couple $r_1=3$ cm, $r_2=10$ cm are close to the predicted $\mathrm{Im}\,k_p$ with an error smaller than 20/100.

TABLE III. Predicted $\omega/\mathrm{Re}\,k_p$ and $\mathrm{Im}\,k_p$ and simulated measurements of c_r and D, $l=2$ cm.

	$\omega/\mathrm{Re}\,k_p$ (m/s), c_r	$\mathrm{Im}\,k_p$(m^{-1}), D
Prediction 2 kHz $\omega/\mathrm{Re}\,k_p$, $\mathrm{Im}\,k_p$	*55.1*	*19.7*
Simulation 2 kHz, 3–4 cm c_r, D	55.5	17.2
Simulation 2 kHz, 4–5 cm c_r, D	54.3	18.9
Simulation 2 kHz, 3–10 cm c_r, D	52.4	21.6
Prediction 3 kHz $\omega/\mathrm{Re}\,k_p$, $\mathrm{Im}\,k_p$	*53.1*	*25.4*
Simulation 3 kHz, 3–4 cm c_r, D	53.2	5.2
Simulation 3 kHz, 4–5 cm c_r, D	56.0	8.4
Simulation 3 kHz, 3–10 cm c_r, D	52.4	21.6
Prediction 4 kHz $\omega/\mathrm{Re}\,k_p$, $\mathrm{Im}\,k_p$	*52.9*	*32.8*
Simulation 4 kHz, 3–4 cmc_r, D	47.7	25.8
Simulation 4 kHz, 4–5 cm c_r, D	58.1	7.2
Simulation 4 kHz, 3–10 cm c_r, D	53.4	31.5
Simulation 4 kHz, 3–4 cm, pole c_r, D	53.2	32.6

The same study has been performed with the geometry g_2 of the source. The results are very similar and are not reported.

IV. MEASUREMENTS

Measurements have been performed on a layer of thickness $l=2$ cm of a urethane foam. The surface area of the sample was 1.5×1.5 m^2, which is sufficiently large so that boundary effects do not affect the measurement of the Rayleigh wave velocity. For a melamine foam, it was even verified experimentally that the sample size could be reduced to 15×15 cm^2 without compromising the Rayleigh wave velocity. To obtain a uniform boundary condition double-faced tape was used to glue the porous layer onto a rigid impervious backing. Rayleigh waves were excited by means of a magnetic transducer putting in motion a circular plate of di-

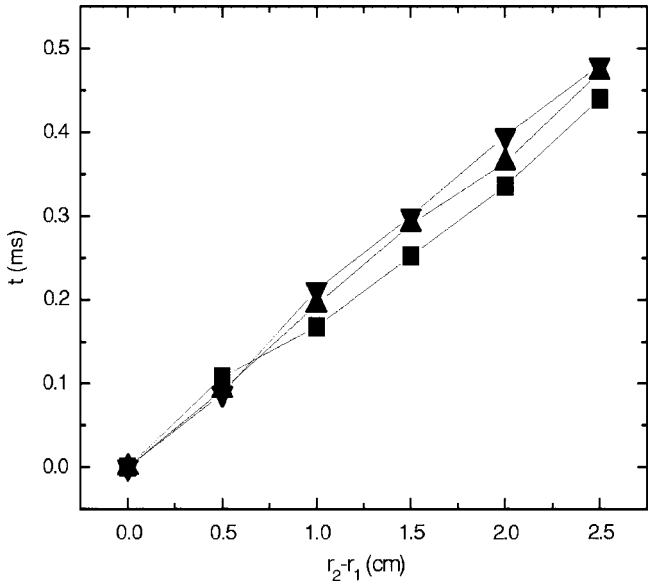


FIG. 6. Measured time $t=\tau_2-\tau_1$ vs displacement r_2-r_1, central frequency 2 kHz (■–■), 3 kHz (▼–▼), 3.5 kHz (▲–▲).

ameter 1 cm bonded on the layer. The magnetic transducer was consecutively fed with a sine burst signal of 2, 3, and 3.5 kHz, respectively. The time dependence of the signal is similar to the one in the modeling and the normal velocity of the frame is measured with a laser Doppler vibrometer. Measurement points are located on a radius through the excitation point. The radius was scanned up to 2.5 cm and with a typical step size of 5 mm. The damping D and the time of flight between two locations are evaluated from the intercorrelations like in the previous sections. The measured time of flight $t=\tau_2-\tau_1$ versus the distance r_2-r_1 is shown in Fig. 6 and the damping is shown in Fig. 7. The ratio $(r_2-r_1)/t$ is close to the speed 54 m/s with an order of magnitude of 10/100 for the dispersion. This result is in a reasonable agreement with measurements of the rigidity parameters of the same medium performed with a different method.[10] The

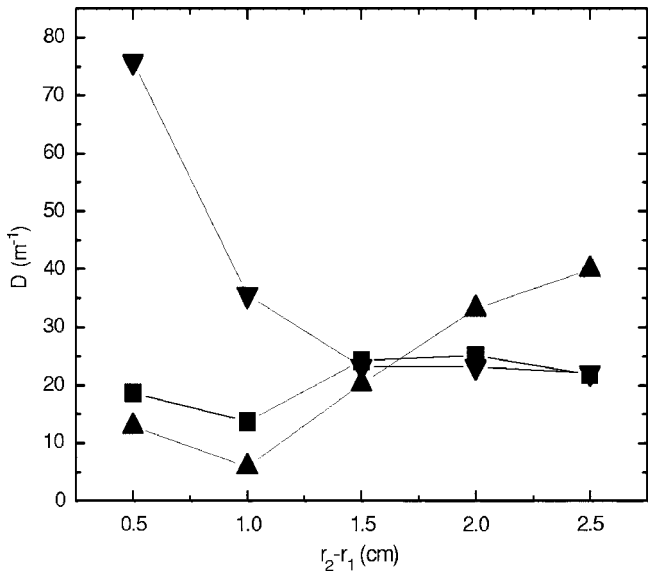


FIG. 7. Measured damping D, $r_1=3$ cm, central frequency 2 kHz (■–■), 3 kHz (▼–▼), 3.5 kHz (▲–▲).

evaluation of the speed of the Rayleigh wave can be performed at a small distance from the source but the measured damping at a given frequency presents a large dispersion and cannot be related to a precise Im k_p.

V. CONCLUSION

Predictions were performed in the frequency domain for soft and strongly damped porous media in contact with air. The wave number interface component of the pole of the Rayleigh plane wave was predicted as a function of frequency for a layer having a small thickness and a semi-infinite layer. The dependence of the phase velocity on the frequency due to the finite thickness was demonstrated. Simulations in the time domain were performed close to a circular source, showing that the residue related to the Rayleigh pole is the dominant contribution to the normal velocity of the frame for the semi-infinite layer and the layer of finite thickness. As a consequence, the speed of the Rayleigh wave and the damping can be evaluated at a small distance from the source for the semi-infinite layer, allowing an evaluation of the phase speed and the imaginary part of the Rayleigh pole. For the layer of finite thickness, the speed evaluation is also possible, but the contributions of the other poles create a fluctuation in the amplitude of the normal velocity which does not allow a precise evaluation of the damping at small distances from the source. Measurements are performed on a layer of small thickness. These measurements confirm that the evaluation of the speed is possible at small distances from the source.

ACKNOWLEDGMENT

L.B. is thankful for financial support obtained through a Research Mandatory of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

[1]T. Pritz, "Measurement methods of complex Poisson's ratio of viscoelastic materials," Appl. Acoust. **60**, 279–292 (2000).
[2]M. Melon, M. Mariez, C. Ayrault, and S. Sahraoui, "Acoustical and mechanical characterization of anisotropie open-cell foams," J. Acoust. Soc. Am. **104**, 2622–2627 (1998).
[3]J. Park, "Measurement of the frame acoustic properties of porous and granular materials," J. Acoust. Soc. Am. **118**, 3483–3490 (2005).
[4]V. Tarnow, "Dynamic measurements of the elastic constants of glass wool," J. Acoust. Soc. Am. **118**, 3672–3678 (2005).
[5]S. Feng and D. L. Johnson, "High-frequency acoustic properties of a fluid/porous solid interface. I. New surface mode. II. The 2D reflection Green's function," J. Acoust. Soc. Am. **74**, 906–924 (1983).
[6]T. L. Richards and K. Attenborough, "Solid particle motion resulting from a point source above a poro-elastic half space," J. Acoust. Soc. Am. **86**, 1085–1092 (1989).
[7]J. F. Allard, G. Jansens, G. Vermeir, and W. Lauriks, "Frame-borne surface waves in air-saturated porous media," J. Acoust. Soc. Am. **111**, 690–696 (2002).
[8]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[9]J. F. Allard, *Propagation of Sound in Porous Media, Modelling Sound Absorbing Materials* (Elsevier, London, 1993).
[10]J. F. Allard, B. Brouard, N. Atalla, and S. Ghinet, "Excitation of soft porous frames resonances and evaluation of the rigidity coefficients," J.

Acoust. Soc. Am. **121**, 78–84 (2007).
[11]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[12]D. Lafarge, P. Lemarinier, J. F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[13]M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[14]T. Tamir and L. B. Felsen, "On lateral waves in slab configurations and their relation to other wave types," IEEE Trans. Antennas Propag. **13**, 410–422 (1965).

Attenuation and scattering of axisymmetrical modes in a fluid-filled round pipe with internally rough walls

German A. Maximov[a] and Eugenii V. Podjachev
Moscow Engineering Physics Institute, Kashirskoe Shosse 31, Moscow, 115409 Russian Federation

Kirill V. Horoshenkov[b]
School of Engineering, Design and Technology, University of Bradford, Bradford, West Yorkshire, BD7 1DP United Kingdom

(Received 19 June 2007; accepted 10 December 2007)

The attenuation of axisymmetric eigenmodes in a cylindrical, elastic, fluid-filled waveguide with a statistically rough elastic wall is studied. It is shown that small perturbation theory can be used to relate explicitly the statistical characteristics of the internal wall surface roughness of an elastic pipe to the attenuation and scattering coefficients of the acoustic modes in the filling fluid. Analytical expressions for modal attenuation coefficients are obtained. The analysis of the frequency dependent attenuation coefficients and the ratio between the roughness correlation length and the inner radius of the pipe is made for different correlation functions of the roughness. It is shown that two scale parameters control the overall behavior of the modal attenuation coefficients. These are the ratios of the roughness correlation length and the inner pipe radius to the acoustic wavelength. The numerical results for sound propagation in a pipe and in a borehole with statistically rough, elastic walls are obtained and discussed. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2831932]

PACS number(s): 43.20.Mv, 43.20.Bi, 43.20.Hq, 43.30.Hw, 43.20.Tb [JHG] Pages: 1248–1259

I. INTRODUCTION

Pipelines are used widely to convey fluids and gases in the petrochemical, water, and energy sectors. The quality of the inner pipe wall can deteriorate rapidly because of chemical reactions, wall material erosion, thermal cracking, and sedimentation processes. This leads to increased hydrodynamic drag in the pipe, reduced hydraulic capacity and potential structural failures. In the majority of cases direct visual quality inspection of pipes is difficult or impossible because of operational, safety and access issues. As a result, there is a clear need for quick, inexpensive and accurate methods for the characterization of the boundary conditions in pipes. In this respect, the use of Stoneley and Lamb waves which can propagate long distances along the fluid–solid interface of a buried, fluid filled pipe appears a very attractive noninvasive boundary characterization technique. The frequency dependent phase velocity and attenuation (dispersion characteristics) of these modes are sensitive to the wall thickness and material properties and can be measured *in situ* using remote sensors to provide a basis for the inversion problem.

A considerable number of methods have been developed to predict the dispersion characteristics of Lamb and Stoneley waves.[1–8] The attenuation of Lamb waves is primarily determined by the viscosity of the fluid contents of the pipe, finite damping in the pipe material and the presence of the surrounding medium, and the attenuation can be used to monitor the condition of the inner surface of the pipe wall.[9–14] It has been shown that Lamb waves interact strongly with structural defects in pipes[15–20] which result in increased modal attenuation, reflection and cross-modal energy exchange.[20–24] The extra modal attenuation and cross-modal energy exchange are particularly pronounced in the presence of structural and surface impedance discontinuities and wall thickness variations.[25–29] Some additional attenuation can also be observed if the inner surface of the pipe is no longer smooth because of the process of corrosion and/or sedimentation. In order to consider this effect it is common to express the extra attenuation as a function of the statistical parameters of the boundary roughness (e.g., Refs. 30 and 31). The problem of sound propagation in a pipe with a periodically varying wall has been originally considered by Rayleigh.[32] However, none of the previous works seems to account for the variation in the wall thickness due to the presence of roughness or to include the elastic properties of the pipe wall. So far these effects have been only considered in application to Sholte–Stoneley wave scattering by flat rough, impedance boundaries (e.g. Refs. 33–38) and have not been included explicitly in the existing models for modal propagation in elastic cylindrical waveguides. In terms of the practical application of the acoustic technique for the diagnosis of the quality of pipelines and boreholes one can argue that measuring the wall thickness variation and the change in the elastic properties of the pipe wall material are essential for the early detection of quality degradation due to corrosion, wall cracking, and sedimentation.

This work presents an improved method for the prediction of the modal attenuation in a cylindrical, elastic waveguide with statistically rough internal wall surfaces. The method is based on the hypothesis that the roughness scale is small in comparison with the acoustic wavelength. It is shown that small perturbation theory can be used to relate

[a]Electronic mail: maximov@dpt39.mephi.ru
[b]Electronic mail: k.horoshenkov@bradford.ac.uk

 0001-4966/2008/123(3)/1248/12/$23.00

explicitly the statistical characteristics of the internal wall surface roughness of an elastic pipe to the attenuation and scattering coefficients of the acoustic modes in the filling fluid. Analytical results presented here are obtained for a realistic section of concrete pipe surrounded by a vacuum. The latter assumption is valid in many practical cases if the compressional and shear sound speeds in the surrounding fluid are relatively small in comparison with that measured in the pipe wall material and the filling fluid.[29,39–41]

II. THEORETICAL FORMULATION

We consider an infinitely long, fluid-filled, elastic waveguide of circular cross section, with a statistically rough inner wall (see Fig. 1). It is assumed that the pipe is suspended in vacuum. It is also assumed that the roughness is cylindrically symmetrical and its height distribution is described by a statistical function $\eta(z)$ so that the actual inner radius of the pipe is given by $r=R_1+\eta(z)$, where R_1 is the mean inner radius of the pipe. The mean wall roughness $\sigma=\sqrt{\langle \eta^2(z)\rangle}$ is small in comparison with the acoustics wavelength λ, i.e., $\sigma/\lambda \ll 1$. The outer wall of the pipe is smooth and the outer radius of the pipe is R_2. It is suggested that the coordinate z runs along the central axis of the pipe and that a monochromatic source is positioned in the center of the pipe (see Fig. 1).

It is proposed to represent the acoustic field in the fluid phase in terms of the displacement scalar potential φ_f. Accordingly, the acoustic field in the elastic wall can be presented in terms of the scalar φ and vector $\vec{\psi}$ potentials which correspond to the longitudinal and shear components of the displacement field, i.e., $\vec{U}=\text{grad}\,\varphi+\text{rot}\,\vec{\psi}$. Because of the axial symmetry of the problem the vector potential can be fully described with a single azimuthal component of the displacement potential $\vec{\psi}=(0,\psi,0)$, which implies that only the radially polarized acoustic waves are considered in this formulation.

The displacement potential components should satisfy the classical wave equations for the acoustic field in the pipe fluid and in elastic pipe wall. In the cylindrical coordinate system (r,θ,z) these wave equations can be written in the form of Helmholtz equations after the double Fourier transform by time t and axial variable z [the correspondence notation is $f(r,k,\omega)=\int_{-\infty}^{\infty}dz e^{ikz}\int_{-\infty}^{+\infty}dt e^{-i\omega t}f(r,z,t)$]

$$\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{r}-\nu_f^2\right)\varphi_f(r,k,\omega)=\frac{A(\omega)}{2\pi r}\delta(r),$$

$$\left(\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{r}-\nu_l^2\right)\varphi(r,k,\omega)=0,$$

$$\left(\frac{\partial}{\partial r}\frac{1}{r}\frac{\partial}{\partial r}r-\nu_s^2\right)\psi(r,k,\omega)=0, \tag{1}$$

where $\nu_j=\sqrt{k^2-k_j^2}$ for $|k|\geqslant\omega/c_j$, $\nu_j=i\sqrt{k_j^2-k^2}$ for $|k|\leqslant\omega/c_j$, and $A(\omega)$ is the source strength. The indices $j=l$, s, and f are used to denote the wave numbers/sound speed of the longitudinal and shear waves in the elastic medium of the pipe waveguide wall and the wave number/sound speed in the fluid phase, respectively. There is a singularity at the source position, $r=0$ which is expressed in Eq. (1) with the delta-function $\delta(r)$.

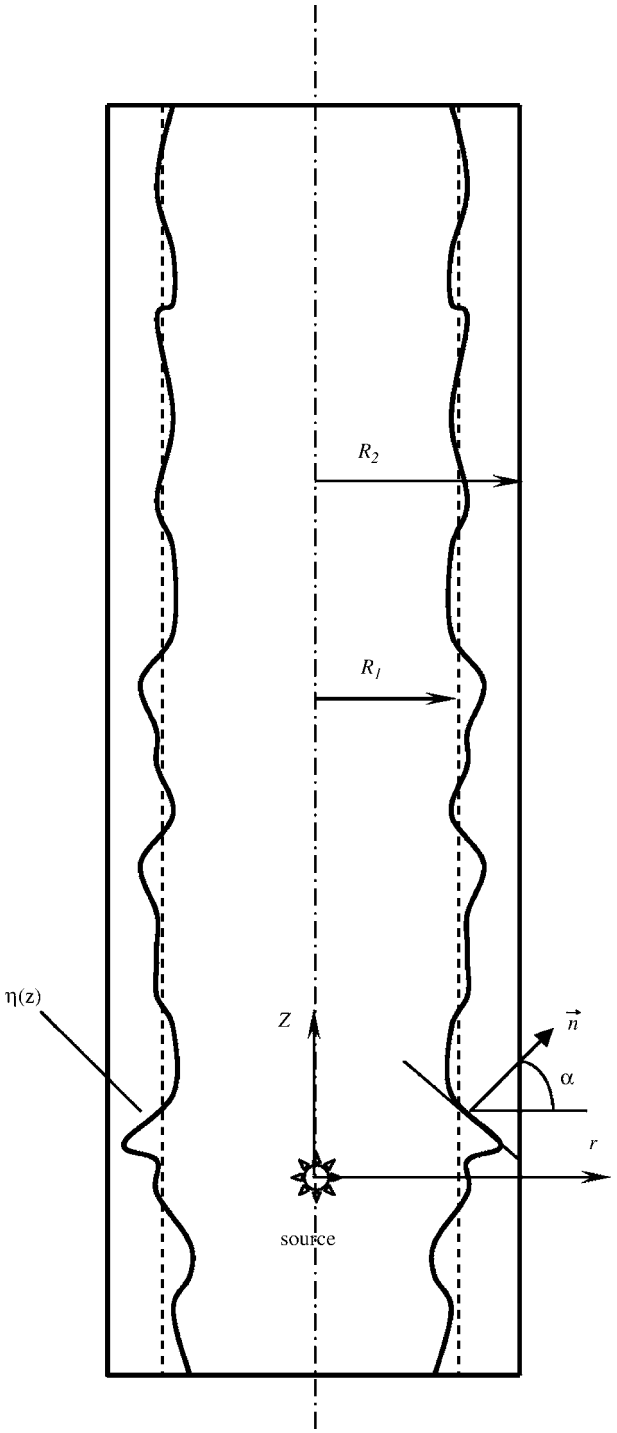


FIG. 1. Sound propagation in a pipe with rough, elastic walls.

The boundary conditions assume the equality of the pressures and stresses acting on the rough inner wall of the pipe

$$n_r(\sigma_{rr}n_r+\sigma_{rz}n_z)+n_z(\sigma_{zr}n_r+\sigma_{zz}n_z)=-P,$$

$$\tau_r(\sigma_{rr}n_r+\sigma_{rz}n_z)+\tau_z(\sigma_{zr}n_r+\sigma_{zz}n_z)=0, \tag{2a}$$

and the radial displacements

$$u_{fr}n_r + u_{fz}n_z = u_r n_r + u_z n_z \tag{2b}$$

at $r=R_1+\eta(z)$. In the previous expressions σ_{ij} is the stress tensor, P is the fluid pressure in the waveguide, u_{fi} and u_i are the displacement components in the fluid and in the wall, respectively. The radial and axial components of the displacement vector in the fluid and in the wall are defined using the indices r and z, respectively. $\vec{n}$ and $\vec{\tau}$ are the normal, $\vec{n}=(\cos\alpha,\sin\alpha)$, and tangential, $\vec{\tau}=(-\sin\alpha,\cos\alpha)$, vectors to the waveguide inner wall. α is the angle between the coordinate-dependent vector $\vec{n}$ and the normal to the mean surface vector $\langle\vec{n}\rangle$ (see Fig. 1).

The boundary conditions on the outer wall, $r=R_2$, are pressure-release, i.e.,

$$\sigma_{rr}=0, \quad \sigma_{rz}=0. \tag{3}$$

Because of the condition $\sigma/\lambda \ll 1$, the boundary conditions can be expanded using the Taylor series at $r=R_1$. Using first-order approximation theory, neglecting the second-order terms in the expansion it can be shown that the boundary conditions on the inner wall for the displacement potential in a rough cylindrical waveguide can be expressed in the following form:

$$\begin{aligned}
2ik\frac{\partial\psi}{\partial r} - \Omega_s\phi + \frac{2}{R}\frac{\partial\phi}{\partial r} - \frac{\rho_f}{\rho}k_s^2\phi_f &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk'\,\eta(k') \\
&\times\left[-2i\tilde{k}\frac{\partial^2\tilde{\psi}}{\partial r^2} + \tilde{\Omega}_s\frac{\partial\tilde{\phi}}{\partial r} - \frac{2}{R}\frac{\partial^2\tilde{\phi}}{\partial r^2} + \frac{2}{R^2}\frac{\partial\tilde{\phi}}{\partial r}\right. \\
&\left. + \frac{\rho_f}{\rho}k_s^2\frac{\partial\tilde{\phi}_f}{\partial r}\right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\varphi_f}{\partial r} - \frac{\partial\varphi}{\partial r} - ik\psi &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk'\,\eta(k')\left[\left(\frac{\partial^2\tilde{\varphi}}{\partial r^2} - i\tilde{k}\frac{\partial\tilde{\psi}}{\partial r}\right.\right. \\
&\left. - \frac{\partial^2\tilde{\varphi}_f}{\partial r^2}\right) - \left(i\tilde{k}\tilde{\varphi} + \frac{1}{R}\tilde{\psi} + \frac{\partial\tilde{\psi}}{\partial r}\right. \\
&\left.\left. - i\tilde{k}\tilde{\varphi}_f\right)ik'\right],
\end{aligned} \tag{4}$$

where functions with a tilde depend on argument $\tilde{k}=k-k'$, and the following notations are introduced:

$$\Omega_s = 2k^2 - k_s^2, \quad \tilde{\Omega}_s = 2\tilde{k}^2 - k_s^2, \quad \tilde{\nu}_{js}^2 = \tilde{k}^2 - k_j^2, \quad j=l,s,f,$$

$$k_s=\omega/c_s, \quad k_l=\omega/c_l.$$

In the previous expressions $\eta(k)=\int_{-\infty}^{+\infty}dz e^{-ikz}\eta(z)$ is the Fourier transform of the range-dependent statistical roughness function. Expressions (4) include explicitly the characteristics of the wall roughness and the effect of the elastic wall on the displacement potential in fluid filling the pipe. The details of the derivation of expressions (4) and the expressions for the outer wall boundary conditions are provided in Appendix A.

The expressions for the displacement potential which satisfy the wave equations (1) can be written in the following form:

$$\begin{aligned}
\varphi_f(r,k,\omega) &= -\frac{A}{2\pi}K_0(\nu_f r) + C_1 I_0(\nu_f r), \\
\varphi(r,k,\omega) &= C_2 K_0(\nu_l r) + C_4 I_0(\nu_l r), \\
\psi(r,k,\omega) &= C_3 K_1(\nu_s r) + C_5 I_1(\nu_s r),
\end{aligned} \tag{5}$$

where $I_n(\nu_i r), K_n(\nu_i r)$ are the modified Bessel and MacDonald functions of order n, respectively, and C_i, $i=1,\ldots,5$ are some arbitrary constants which need to be chosen to satisfy the inner [$r=R_1$, expression (4)] and the outer wall [$r=R_2$, expression (A3)] boundary conditions. Applying the boundary conditions to the previous form for the displacement potential [Eq. (5)] one can obtain a system of integral equations for the five coefficients, C_i, $i=1,\ldots,5$, which can be presented in the following matrix form:

$$\mathbf{L}_0(k)\vec{C}(k) = \mathbf{L}_1^*(k,k-k')\vec{\mathbf{C}}(k') + \vec{\mathbf{Q}}^*(k), \tag{6}$$

where

$$\vec{\mathbf{C}}(k) = (C_1, C_2, C_3, C_4, C_5,)^T$$

is the coefficient vector,

$$\mathbf{L}_0(k) = \begin{pmatrix}
\frac{\rho_f}{\rho}k_s^2 I1_0^f & \Omega_s K1_0^l + \frac{2\nu_l}{R_1}K1_0^l & 2ik\left(\nu_s K1_0^s + \frac{1}{R_1}K1_1^s\right) & \Omega_s I1_0^l - \frac{2\nu_l}{R_1}I1_1^l & 2ik\left(-\nu_s I1_0^s + \frac{1}{R_1}I1_1^s\right) \\
0 & -2ik\nu_l K1_1^l & \Omega_s K1_1^s & 2ik\nu_l I1_1^l & \Omega_s I1_1^s \\
\nu_f I1_1^f & \nu_l K1_1^l & ikK1_1^s & -\nu_l I1_1^l & ikI1_1^s \\
0 & \Omega_s K2_0^l + \frac{2\nu_l}{R_2}K2_0^l & 2ik\left(\nu_s K2_0^s + \frac{1}{R_2}K2_1^s\right) & \Omega_s I2_0^l - \frac{2\nu_l}{R_2}I2_1^l & 2ik\left(-\nu_s I2_0^s + \frac{1}{R_2}I2_1^s\right) \\
0 & -2ik\nu_l K2_1^l & \Omega_s K2_1^s & 2ik\nu_l I2_1^l & \Omega_s I2_1^s
\end{pmatrix} \tag{7}$$

is the matrix operator for the unperturbed (smooth wall) case and $\mathbf{L}_1^*(k,\tilde{k})$:

$$\mathbf{L}_1^*(k,\tilde{k}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk'\,\eta(k')\mathbf{L}_1(k,\tilde{k}),$$

$$\mathbf{L}_1(k,\tilde{k}) = \begin{pmatrix} L_{11} & L_{12} & L_{13} & L_{14} & L_{15} \\ L_{21} & L_{22} & L_{23} & L_{24} & L_{25} \\ L_{31} & L_{32} & L_{33} & L_{34} & L_{35} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

are the perturbation matrix operators which account for the effect of finite wall roughness. In the earlier expressions the following notations are used:

$$K_n(\nu_i R_j) = Kj_n^i, \quad I_n(\nu_i R_j) = Ij_n^i, \quad n = 0,1, \quad i = f,l,s, \quad j = 1,2.$$

The expressions for the components of the above operators are presented in Appendix B.

Integral matrix equation (6) can be solved using the mean field method which is detailed, for example, in Ref. 42. Here we present the final equation for the mean vector of coefficients $\langle\vec{\mathbf{C}}\rangle$,

$$[\mathbf{L}_0(k) - \sigma^2\mathbf{L}_2(k)]\langle\vec{\mathbf{C}}(k)\rangle = \sigma^2\vec{\mathbf{Q}}(k), \tag{8}$$

where σ^2 is the dispersion for the statistical roughness height and

$$\mathbf{L}_2(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk'\,W(k')\mathbf{L}^2(k,\tilde{k}),$$

$$\mathbf{L}^2(k,\tilde{k}) = \mathbf{L}_1(k,\tilde{k})\mathbf{L}_0^{-1}(\tilde{k})\mathbf{L}_1(\tilde{k},k),$$

$$\vec{\mathbf{Q}}(k) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} dk'\,W(k')\mathbf{L}_1(k,\tilde{k})\mathbf{L}_0^{-1}(\tilde{k})\vec{\mathbf{Q}}'(\tilde{k}) + \langle\vec{\mathbf{Q}}^*(k)\rangle.$$

In the previous expression the function $W(k')$ is used to denote the Fourier transform of the correlation function of the statistical roughness $W(z)$, where $\sigma^2 W(z-z') = \langle\eta(z)\,\eta(z')\rangle$. The solutions which satisfy the homogeneous matrix equation (8) $[\vec{\mathbf{Q}}(k) = \mathbf{0}]$ correspond to the eigenmodes, which can propagate in the fluid filling the rough waveguide. The conditions for the existence of particular eigenmodes are controlled by the solution of the following dispersion equation:

$$\det[\mathbf{L}_0(k) - \sigma^2\mathbf{L}_2(k)] = 0. \tag{9}$$

Because $\sigma/\lambda \ll 1$ the previous expression can be reduced to

$$\det\mathbf{L}_0(k) - \sigma^2 H(k) = 0, \tag{10}$$

where

$$H(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk'\,W(k')\frac{M(k,\tilde{k})}{\det\mathbf{L}_0(\tilde{k})},$$

$$M(k,\tilde{k}) = \sum_{i,j=1}^{3} L_{ij}^2(k,\tilde{k})\,\mathrm{minor}(\mathbf{L}_0(k)_{ij}). \tag{11}$$

The solution of the dispersion equation (10) can be obtained by solving the classical problem for sound propagation in a smooth waveguide and incorporating the method of small perturbations. In this way the wave numbers for sound propagation in a waveguide with a statistically rough inner wall can be expressed as a superposition of the solutions for the sound field in the smooth waveguide and small perturbation terms due to the presence of roughness. The proposed solution should be sufficiently general to allow a number of eigenmodes propagating in the fluid-filled waveguide to be considered. This case is different from a more simple problem when a single Rayleigh wave is allowed to propagate in an empty borehole in an elastic medium.[42] The sound speed and attenuation of these modes can be determined from the complex eigennumbers

$$k_i = k_i^0 + \delta k_i, \tag{12}$$

where the wave numbers k_i^0 are the solutions of the unperturbed form of Eq. (10), i.e., $\det L_0(k) = 0$, and δk_i is the wave number perturbation due to the presence of roughness. If $i = 0$, then $k_0 = k_{\mathrm{St}}$ which corresponds to the case of the Stoneley wave with wave number k_{St}. If $i \geqslant 1$, then the sound field is represented as a set of Lamb waves with $k_i = k_{i,L}$.

The explicit solution of the equation $\det\mathbf{L}_0(k) = 0$ is rather cumbersome [see Eq. (7)]. However, in the particular case $R_2 \to \infty$ the expression for the operator $\mathbf{L}_0$ is reduced to

$$\mathbf{L}_0(k) = \begin{pmatrix} \frac{\rho_f}{\rho}k_s^2 I1_0^f & \Omega_s K1_0^l + \frac{2\nu_l}{R_1}K1_0^l & 2ik\left(\nu_s K1_0^s + \frac{1}{R_1}K1_1^s\right) & 0 & 0 \\ 0 & -2ik\nu_l K1_1^l & \Omega_s K1_1^s & 0 & 0 \\ \nu_f I1_1^f & \nu_l K1_1^l & ikK1_1^s & 0 & 0 \\ 0 & 0 & 0 & \Omega_s & -2ik\nu_s 0 \\ 0 & 0 & 0 & 2ik\nu_l & \Omega_s \end{pmatrix}. \tag{13}$$

The determinant of the previous matrix can be decomposed as a product of two separate determinants: $\det \mathbf{L}_0 = \det \mathbf{L}_0^0 \det \mathbf{L}_0^1$. The first determinant defines the Rayleigh wave

$$\det \mathbf{L}_0^0 = \Omega_s^2 - 4k^2 \nu_l \nu_s, \tag{14}$$

which propagates in the inner elastic space which surrounds the fluid. The second determinant is given by the following expression:

$$\det \mathbf{L}_0^1(k) = \left(\Omega_s^2 K_0^l K_1^s - 4k^2 \nu_l \nu_s K_1^l K_0^s - \frac{2}{R} \nu_l k_s^2 K_1^l K_0^s \right) \nu_f I_1^l + \frac{\rho_f}{\rho} \nu_l k_s^4 K_1^l K_1^s I_0^f. \tag{15}$$

The form of dispersion equation (15) is well-known.[43,44] This equation depends upon k^2 and, at a given frequency, it has a finite number of roots on the real axis. The fundamental mode, the Stoneley wave, can exist at any frequency of sound,[44] whereas the other, the so-called higher Lamb modes can only exist above the critical cutoff frequencies ω_{0i}.[46] The dispersion curve for the Stoneley wave differs qualitatively from the higher Lamb modes and lies below the dispersion curve for any other high-order mode. The phase velocity of the Stoneley wave increases monotonously with frequency from the lower limit of the tube wave propagation velocity $c_t = c_f / \sqrt{1 + (\rho_f / \rho)(c_f / c_s)^2}$ (Ref. 44) and approaches to the upper limit of Sholte–Stoneley wave propagation velocity for a flat boundary at high frequencies. In the vicinity of the cutoff frequency the phase velocities of higher-order Lamb waves approach the velocity of the shear waves in an elastic space $c_i^0(\omega \to \omega_{0i}) = c_s$ and further for higher frequencies it decreases approaching asymptotically to the sound speed in a fluid $c_i^0(\omega \to \infty) = c_f$. Two particular roots of the dispersion equation $k = \pm k_i^0(\omega)$, $k_i^0 = \omega / c_i^0(\omega)$ correspond to the eigenmodes which propagate in opposite directions along the axis of the waveguide with the frequency-dependent phase velocity of $c_i^0(\omega)$.

If the wall thickness is finite the behavior of the dispersion curves for the sound speed is similar to the previous case, but the cutoff frequency for the first Lamb mode is equal to zero as illustrated in Fig. 2. As a result, in the low frequency regime there are two propagating eigenmodes: (i) the so-called α-mode, which resembles in behavior the Stoneley wave in a circular cylindrical waveguide surrounded by an semiinfinite elastic medium (e.g., the behavior of a Stoneley wave in a borehole); (ii) the so-called β-mode [in the terminology presented in Ref. 18 it is the mode $L(0,1)$], which at low frequencies propagates at a phase velocity slightly below that for the longitudinal waves in the elastic wall of the waveguide and at higher frequencies asymptotically approaches the sound speed in the fluid.

Substituting Eq. (12) in Eq. (11) and expanding the operators $\mathbf{L}_0(k)$ and $H(k)$ results around $k = k_i^0$ we obtain

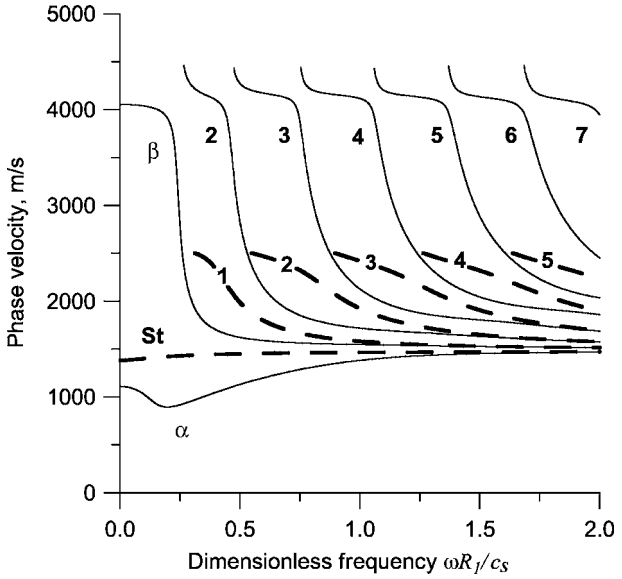


FIG. 2. Phase velocity as a function of dimensionless frequency for the first seven axially symmetrical modes which propagate in a pipe and borehole with smooth walls. The principal modes in the pipe are marked as α and β. The wall and fluid pipe parameters are described in the text. The analogous modes in a borehole with the same parameters are shown by dashed lines. Symbol St corresponds to the Stoneley wave.

$$\left. \frac{\partial \det \mathbf{L}_0}{\partial k} \right|_{k=k_i^0} \delta k_i - \sigma^2 H(k_i^0) - \sigma^2 \left. \frac{\partial H(k)}{\partial k} \right|_{k=k_i^0} \delta k_i = 0. \tag{16}$$

The last term in Eq. (16) is the second order of magnitude in comparison with the other terms and can be neglected. As a result, the expression for the correction to the wave number has the following form:

$$\delta k_i = \sigma^2 H(k_i^0) \Big/ \left. \frac{\partial \det \mathbf{L}_0(k)}{\partial k} \right|_{k_i^0}. \tag{17}$$

The real and imaginary parts of the wave number determine the phase velocity $c_i(\omega)$ and modal attenuation $\alpha_i(\omega)$, i.e.,

$$k_i(\omega) = \frac{\omega}{c_i(\omega)} + i\alpha_i(\omega). \tag{18}$$

Because the acoustic normal modes in a smooth waveguide are dispersive, but nonattenuating, it is of interest to study the imaginary part of the wave number correction term, which is responsible for the modal attenuation due to the finite scale of roughness

$$\alpha_i(\omega) = \mathrm{Im}(\delta k_i) = \mathrm{Im}\left(\sigma^2 H(k_i^0) \Big/ \left. \frac{\partial \det \mathbf{L}_0(k)}{\partial k} \right|_{k_i^0} \right). \tag{19}$$

It is convenient to rewrite Eq. (17) in nondimensional variables by introducing the dimensionless wave number x: $k = k_s x$, where $k_s = \omega / c_s$ is the wave number for the shear wave in the wall of the waveguide. Similarly, substituting for the dimensionless wave number in the operators $L_0(k)$ and $L_1^*(k, \tilde{k})$ and taking the coefficients $K1_1^l$, $K1_1^s$, $I1_0^f$, $I2_0^l$, $I2_0^s$,

out of the brackets, one can express the modal attenuation coefficient [Eq. (19)] in the following form:

$$\alpha_i(\omega) = \mathrm{Im}\left(\sigma^2 k_s^4 \frac{H(x_i^0)}{\left.\frac{\partial \det \mathbf{L}_0(x)}{\partial x}\right|_{x_i^0}} \right) \tag{20}$$

where $k_i^0 = k_s x_i^0$ ($x_i^0 = c_s / c_i^0(\omega)$). The expressions for the other operators which appear in Eq. (20) are provided in Appendix B.

Equation (20) suggests that the modal attenuation coefficient can be linked explicitly to the correlation function $W(k')$ for the wall roughness of the waveguide

$$\alpha_i(\omega) = \frac{\sigma^2 k_s^4}{2\pi \left(\partial \det \mathbf{L}_0(x)/\partial x\right)|_{x_i^0}} \int_{+\infty}^{-\infty} dx' W(k') \times M(x_i^0, \tilde{x}) / \det \mathbf{L}_0(\tilde{x}), \tag{21}$$

where $M(x,\tilde{x}) = \Sigma_{i,j=1}^{5} L_{ij}^2(x,\tilde{x}) \mathrm{minor}\, \mathbf{L}_0(x)_{ij}$ is the modal scattered amplitude, $\mathrm{minor}(L_0(x)_{ij})$ is the minor of ijth element of the matrix $\mathbf{L}_0$, $\tilde{x} = x_i^0 - x$ and x_i^0 are dimensionless solutions of the dispersion equation for the case of the waveguide with smooth walls.

The imaginary part of the correction to the wave number in Eq. (12) relates to the following behavior of the integral in Eq. (11): (i) poles which correspond to the Stoneley wave at $\tilde{k} = \pm k_{\mathrm{St}}$ and higher order Lamb waves $\tilde{k} = \pm k_i^0$ and (ii) contributions from the branch cuts at the points $|k| \leqslant \omega / c_l$ and $|k| \leqslant \omega / c_s$ associated with the scattering of eigenmodes into the compressional and shear waves. Therefore, the total modal attenuation,

$$\alpha_i = \sum_j \alpha_{ij} + \alpha_{is} + \alpha_{il}, \tag{22}$$

can be represented as a superposition of the partial attenuation coefficients associated with the mode scattering into secondary eigenmodes (α_{ij}), and also into shear (α_{is}) and longitudinal (α_{il}) head waves.

Let us first consider the contribution to the attenuation factor due to scattering into other eigenmodes. This contribution can be evaluated analytically by application of the residue theorem to integral (21)

$$\alpha_{ij}(\omega) = \sigma^2 k_s^4 \frac{\{W(k_i^0 - k_j^0) M(x_i^0, x_j^0) + W(k_i^0 + k_j^0) M(x_i^0, -x_j^0)\}}{\left.\frac{\partial \det \mathbf{L}_0(x)}{\partial x}\right|_{x_i^0} \left.\frac{\partial \det \mathbf{L}_0(x)}{\partial x}\right|_{x_j^0}} \tag{23}$$

The first term in the numerator of Eq. (23) describes scattering of mode i into mode j both of which propagate in the same direction (forward cross-mode scattering). The second term in the numerator corresponds to scattering of mode i into mode j, which propagates in the opposite direction (cross-mode backscattering).

As is seen from Eq. (23), for the case of the mode scattering into the same eigenmode ($i=j$: $k_j^0 = k_i^0$) the argument of the Fourier spectrum of the correlation function in the first term equals to zero. Hence for correlation functions with decreasing spectra the first term will dominate at large correlation length (correlation length a is determined as a space scale where the correlation function is essentially different from zero) and the principal contribution to the attenuation coefficient will come from forward scattering. In this case the attenuation coefficient will not depend on the form of the correlation function. For other cases the attenuation coefficient will be dependent on the spectrum of the correlation function.

Further, we consider the behavior of the second and third terms in Eq. (22), which correspond to shear and compressional head waves, respectively. The associated contribution to the attenuation factor is given by the following integral:

$$\alpha_{is}(\omega) + \alpha_{il}(\omega) = \frac{\sigma^2}{2\pi} \frac{k_s^4}{\left.\frac{\partial \det \mathbf{L}_0(x)}{\partial x}\right|_{x_i^0}} \mathrm{Im} \int_{x_i^0 - 1}^{x_i^0 + 1} dx' W(k_s x') \times \frac{M(x_i^0, x_i^0 - x')}{\det \mathbf{L}_0(x_i^0 - x')}. \tag{24}$$

In the region of $x' \in \lfloor c_s/c_i^0 - 1; c_s/c_i^0 - c_s/c_l \rfloor \cup \lfloor c_s/c_i^0 + c_s/c_l; c_s/c_i^0 + 1 \rfloor$ the attenuation is due to scattering of the mode i into shear head waves which propagate in the elastic medium. In the the region $x' \in \lfloor c_s/c_i^0 - c_s/c_l; c_s/c_i^0 + c_s/c_l \rfloor$ the attenuation is primarily due to modal scattering into head compressional waves.

The form of Eqs. (23) and (24) suggests that the behavior of the modal attenuation coefficient is controlled by the roughness correlation function, $W(k)$, scattering strength coefficient, $M(x,\tilde{x})$, and the first derivative, $\partial \det \mathbf{L}_0(x)/\partial x$, taken along the phase curves. The spectrum of the roughness correlation function depends on $ka = \omega a / c_s$, and the scattering amplitude depends on the parameters $kR_1 = \omega R_1 / c_s$ $kR_2 = \omega R_2 / c_s$. Therefore, the two scale ratios, ka and kR_1 ($kR_1 \sim kR_2$), control the overall behavior of the modal attenuation, whereby the forward scattering strength of mode i is only controlled by the scale $kR_1 \sim kR_2$. However, if we take into account that the phase curves are more closely grouped at high frequencies then this behavior will be specific wholly for forward partial attenuation coefficients. Thus it is convenient to carry out the analysis of partial attenuation coefficients due to forward and backward scattering at the two different scales $k_s R_1$ and $k_s a$, respectively.

It is convenient for comparison to normalise the partial attenuation coefficients. Considering that the partial attenuation coefficients are proportional to $\sigma^2 k_s^4$ and the roughness correlation spectrum $W(k) \propto a$ [see Eqs. (23) and (24)], it is sensible to adopt the normalization factor of $(\sigma^2 a / R_1^4)$, which has the dimension of per meter. In this case, the amplitude of the forward-scattering strength becomes independent of the form of the roughness correlation function and the correlation length. We note that the proposed choice of the

normalization factor is not universal and its choice depends on the type of the correlation function and scattering type.

III. RESULTS

The earlier expressions have been used to predict the phase velocity and overall attenuation for several modes propagating in a water-filled concrete pipe with the following characteristics: longitudinal wave speed in concrete $c_l = 4500$ m/s, shear wave speed in concrete, $c_s = 2500$ m/s, density of concrete, $\rho = 2000$ kg/m^3, sound speed in water, $c_f = 1500$ m/s, density of water, $\rho_f = 1000$ kg/m^3, inner pipe radius, $R_1 = 0.5$ m and outer pipe radius, $R_2 = 0.6$ m. Figure 2 presents the frequency dependent phase velocity of several eigenmodes propagating in the pipe with smooth, elastic walls. Figure 2 also presents the results for the equivalent borehole for which $R_2 \to \infty$.

The results illustrate a case of modal transformation when the Stoneley wave in the borehole with semiinfinite walls is transformed in the so-called α-mode in the pipe with finite wall thickness.[42,45,46] The first Lamb mode in the borehole is transformed in the so-called β-mode in the pipe which does not have a cut-on frequency and can be generated at very low frequencies of sound (see Fig. 2). The higher-order modes in the borehole undergo direct transformation into the corresponding higher-order Lamb modes in the pipe. Detailed analysis of the behavior of these modes can be found in Refs. 12–14.

In the higher frequency limit the behavior of the eigenmodes in the pipe and in the borehole are asymptotic. The phase velocity of the Stoneley wave in the borehole in the higher frequency range asymptotically approaches the wave speed in the fluid and remains greater than the phase velocity of the α-mode in the pipe. The major difference in the modal behavior between the borehole and the pipe is that the phase velocity of the higher-order modes at the cut-on frequencies in the pipe is equal to infinity, fastly drops to the longitudinal speed of sound in the wall material and reduces with the frequency to the sound speed in the fluid, whereas the phase velocity of these modes in the borehole at the cut-on frequencies is equal to the speed of a shear wave in the surrounding medium. Therefore, the cut-on frequencies in the pipe and in the borehole are determined wholly by the longitudinal and shear wave speeds in the wall material, respectively.

Figure 3 presents the frequency dependence of the partial attenuation coefficients [Eq. (24)] in the water filled pipe. For comparison, Fig. 3 also presents the partial attenuation coefficients for the Stoneley wave in the borehole. The following roughness correlation function was used to obtain this set of results:

$$W(x) = e^{-x^2/a^2}, \quad W(k) = a\sqrt{\pi}e^{-(ka)^2/4}. \tag{25}$$

The data for the partial attenuation coefficients are normalized by the factor $\sigma^2 a/R_1^4$ and presented for a range of ratios of the roughness correlation length to the inner pipe radius, a/R_1. The modal attenuation due to the forward- and backscattering is controlled by the roughness correlation length and by the inner pipe radius. Therefore, the frequency scales in these graphs are different and normalized by a/c_s and R_1/c_s, respectively. The results presented in Fig. 3 show that the modal attenuation due to the wall roughness is a combination of the forward- and backscattering of the α-mode into itself and in other higher-order modes. In the case of the wall roughness with a small correlation length, i.e., $a/R_1 \ll 1$, the backscattering is the dominant mechanism of attenuation. As the ratio a/R_1 increases the modal attenuation due to forward-scattering becomes more pronounced and dominates for $a/R_1 \cong 1$. The results also suggest that for $a/R_1 > 0.1$ the attenuation due to backscattering is more pronounced in the low frequency range whereas the attenuation due to forward-scattering is more pronounced at the higher frequencies of sound.

In general, the effect of the wall roughness on the acoustic attenuation due to cross-modal scattering in the pipe is greater than that in the borehole. It is interesting to note that there is a strong resemblance in the spectra for the modal backscattering coefficients in the borehole and in the pipe. However, the spectra for the forward-scattering coefficients for these two cases are markedly different, particularly in the low frequency regime. Figure 4 replots the data shown in Fig. 3 using the logarithmic scale to illustrate the functional dependence of the attenuation coefficients. The results presented Fig. 4 suggest that in the low frequency regime in the pipe the attenuation of the α-mode in itself, and in the β-mode due to backscattering by the wall roughness, can be described by a quadratic function of the frequency. A similar behavior is observed in the case of the Stoneley wave in the borehole. The behavior of the attenuation due to the forward-scattering of these modes shown in Fig. 4 is considerably different between the cases of the pipe and borehole. The behavior of the forward-scattering attenuation of the α-mode is also quadratic and by several orders of magnitude exceeds the analogous attenuation of the Stoneley wave in the borehole.

IV. CONCLUSIONS

A set of analytical expressions for the partial attenuation coefficients of axisymmetric eigenmodes in a cylindrical, elastic, fluid filled waveguide with a statistically rough elastic wall has been derived using small perturbation theory. These expressions relate explicitly the mean roughness height, standard deviation, and correlation length with the forward- and backscattering modal attenuation coefficients of the acoustic modes in the filling fluid. An analysis of the frequency-dependent attenuation coefficients and the ratio between the roughness correlation length and the inner radius of the pipe is made for different correlation functions of roughness. It is shown that two scale parameters control the overall behavior of the modal attenuation coefficients. These are the a/λ and R_1/λ, where a is the roughness correction length, R_1 is the inner pipe radius, and λ is the acoustic wavelength.

Numerical results for sound propagation in a pipe and in a borehole with statistically rough, elastic walls have been obtained and discussed. It has been shown that in the case of a wall roughness with a small correlation length, i.e., a/R_1

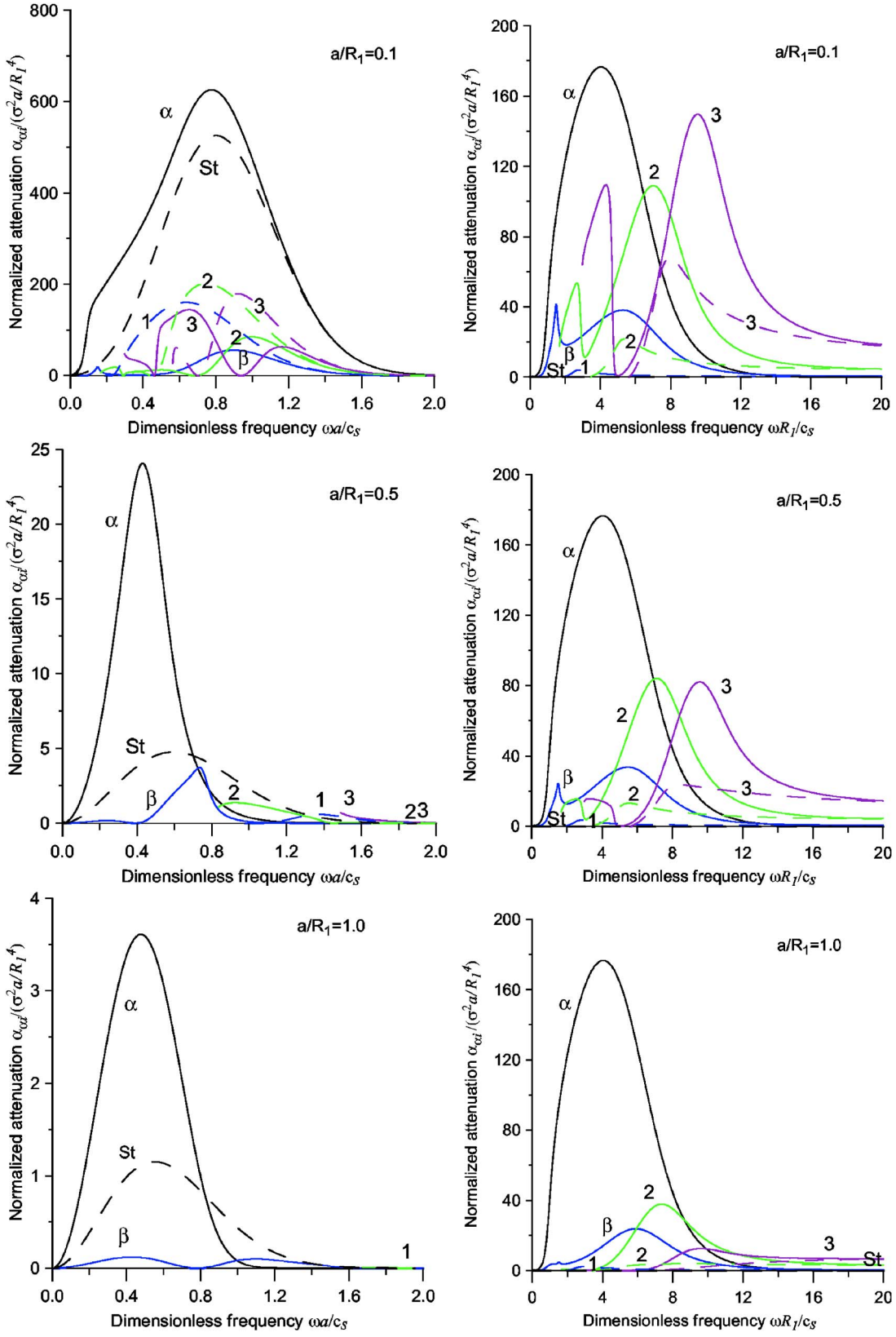


FIG. 3. (Color online) The overall normalized partial attenuation coefficients due to α-mode scattering into itself and other higher modes plotted against dimensionless frequency for different ratios of correlation length to inner radius of the waveguide, a/R_1. The graphs to the left correspond to attenuations due to forward scattering processes, those to the right are for backward scattering processes. The principal modes in the pipe are marked as α and β. The higher modes are marked by numbers. For comparison the partial attenuation coefficients for analogous modes in a borehole with the same parameters are shown by dashed lines. Symbol St corresponds to the Stoneley wave.

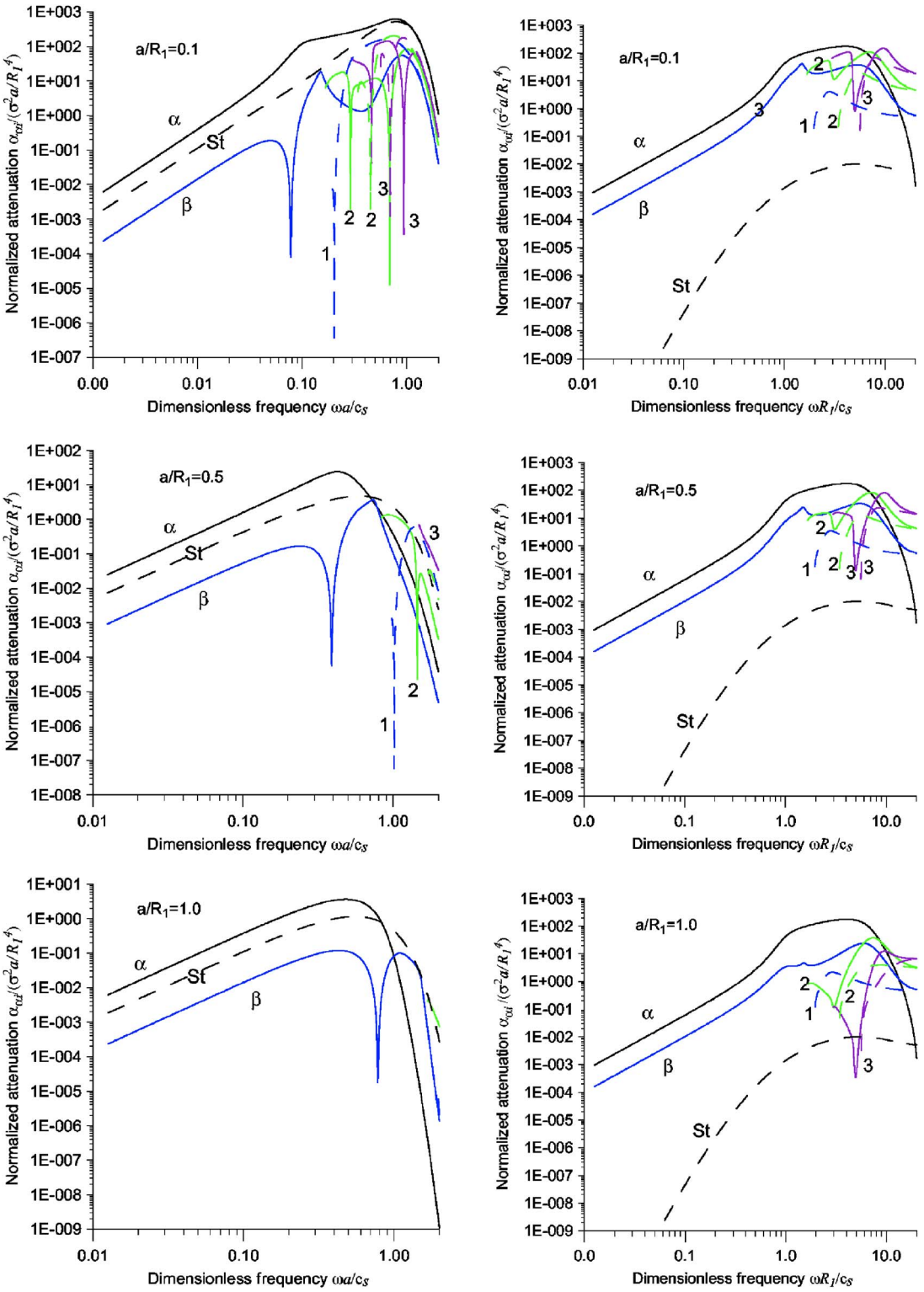


FIG. 4. (Color online) The results from Fig. 3 replotted on a logarithmic scale.

$\ll 1$, backscattering is the dominant mechanism of modal attenuation. The forward-scattering mechanism of modal attenuation is dominant for $a/R_1 \cong 1$. For $a/R_1 > 0.1$ the backscattering attenuation can be more pronounced in the low frequency regime, whereas the forward-scattering attenuation is more pronounced at the higher frequencies of sound. It has been observed that the effect of the wall roughness on the acoustic attenuation due to cross-modal scattering in the pipe is greater than that in the borehole and that there is a resemblance between the spectra of the backscattering attenuation coefficients. The behavior of the forward-scattering attenuation coefficient in the pipe and in the borehole is considerably different across the considered frequency range.

The results obtained in this work can be used to develop a method for the noninvasive inspection of the wall quality in pipes and boreholes from attenuation data. The predicted

modal frequency-dependent attenuation behavior is linked strongly to the statistical parameters of the wall roughness and could thus be used to determine the roughness mean height and dominant correlation length in practical situations.

ACKNOWLEDGMENTS

The authors are grateful to the UK's Engineering and Physical Sciences Research Council (EPSRC Grant No. EP/D031311/1) for support of the staff exchange between the University of Bradford (UK) and the Moscow Engineering Physics Institute (Russian Federation).

APPENDIX A: BOUNDARY CONDITIONS ON THE INNER SURFACE OF A ROUGH PIPE

The expressions for the local normal and tangential vectors to the pipe surface can be presented in the following forms:

$$\vec{n} = (\cos\alpha, \sin\alpha) = \left(1, -\frac{\partial\eta(z)}{\partial z}\right) \Bigg/ \sqrt{1+\left(\frac{\partial\eta(z)}{\partial z}\right)^2},$$

$$\vec{\tau} = (-\sin\alpha, \cos\alpha) = \left(-\frac{\partial\eta(z)}{\partial z}, 1\right) \Bigg/ \sqrt{1+\left(\frac{\partial\eta(z)}{\partial z}\right)^2}. \tag{A1}$$

Here α is the angle between the local normal $\vec{n}$ and the mean normal direction to the inner wall (see Fig. 1). The previous equations can be reduced to

$$(\sigma_{zr} n_r + \sigma_{zz} n_z)|_{R+\eta(z)} = -P\sin\alpha$$

$$(\sigma_{rr} n_r + \sigma_{rz} n_z)|_{R+\eta(z)} = -P\cos\alpha. \tag{A2}$$

The components of the stress and displacement tensor can be expressed via the displacement potentials for the longitudinal and shear waves written in a system of cylindrical coordinates

$$\sigma_{rr} = 2\mu\left(\frac{\partial^2\varphi}{\partial r^2} - \frac{\partial^2\psi}{\partial r\partial z}\right) + \frac{\lambda}{c_l^2}\frac{\partial^2\varphi}{\partial t^2},$$

$$\sigma_{rz} = \mu\left(2\frac{\partial^2\varphi}{\partial r\partial z} + \frac{1}{c_s^2}\frac{\partial^2\psi}{\partial t^2} - 2\frac{\partial^2\psi}{\partial z^2}\right),$$

$$\sigma_{zz} = 2\mu\left(\frac{\partial^2\varphi}{\partial z^2} + \frac{1}{r}\frac{\partial\psi}{\partial z} + \frac{\partial^2\psi}{\partial r\partial z}\right) + \frac{\lambda}{c_l^2}\frac{\partial^2\varphi}{\partial t^2},$$

$$u_r = \frac{\partial\varphi}{\partial r} - \frac{\partial\psi}{\partial z}, \quad u_z = \frac{\partial\varphi}{\partial z} + \frac{1}{r}\frac{\partial r\psi}{\partial r}, \tag{A3}$$

where μ, λ are the Lame coefficients and second derivatives by radius are expressed by use of wave equations (1), for the potentials. The pressure and displacement potential in the fluid are related by the expression

$$P = -\rho_f\frac{\partial^2\varphi_f}{\partial t^2}. \tag{A4}$$

Making use of Eqs. (A1)–(A4) we can present the boundary conditions (2a) and (2b) in the following form:

$$2\frac{\partial^2\varphi}{\partial r\partial z} - 2\frac{\partial^2\psi}{\partial z^2} + \frac{1}{c_s^2}\frac{\partial^2\psi}{\partial t^2} = -\left[\frac{\mu+\lambda}{c_l^2\mu}\frac{\partial^2\varphi}{\partial t^2} + \frac{1}{r}\left(\frac{\partial\psi}{\partial z} - \frac{\partial\varphi}{\partial r}\right) - \frac{\rho_f}{\rho c_s^2}\frac{\partial^2\varphi_f}{\partial t^2}\right]\sin 2\alpha,$$

$$2\frac{\partial^2\psi}{\partial r\partial z} + 2\frac{\partial^2\varphi}{\partial z^2} + \frac{2}{r}\frac{\partial\varphi}{\partial r} - \frac{1}{c_s^2}\frac{\partial^2\varphi}{\partial t^2} + \frac{\rho_f}{\rho c_s^2}\frac{\partial^2\varphi_f}{\partial t^2} = -2\left[\frac{\lambda+\mu}{\mu c_l^2}\frac{\partial^2\varphi}{\partial t^2} + \frac{1}{r}\left(\frac{\partial\psi}{\partial z} - \frac{\partial\varphi}{\partial r}\right) - \frac{\rho_f}{\rho c_s^2}\frac{\partial^2\varphi_f}{\partial t^2}\right]\sin^2\alpha,$$

$$\frac{\partial\varphi_f}{\partial r}\cos\alpha + \frac{\partial\varphi_f}{\partial z}\sin\alpha = \left(\frac{\partial\varphi}{\partial r} - \frac{\partial\psi}{\partial z}\right)\cos\alpha + \left(\frac{\partial\varphi}{\partial z} + \frac{1}{r}\psi + \frac{\partial\psi}{\partial r}\right)\sin\alpha. \tag{A5}$$

Assuming that the roughness is small, i.e., $|\eta(z)| \ll \lambda$, and applying the Taylor series expansion near the mean internal radius $r=R_1$ we obtain

$$f(r=R_1+\eta(z)) = f(r=R_1) + \left.\frac{\partial f}{\partial r}\right|_{r=R_1}\eta(z) + \cdots.$$

Neglecting any second order terms with respect to $\eta(z)$ in Eq. (A5) we obtain

$$2\frac{\partial^2\varphi}{\partial r\partial z} - 2\frac{\partial^2\psi}{\partial z^2} + \frac{1}{c_s^2}\frac{\partial^2\psi}{\partial t^2} = \left(-2\frac{\partial^3\varphi}{\partial r^2\partial z} + 2\frac{\partial^3\psi}{\partial r\partial z^2} - \frac{1}{c_s^2}\frac{\partial^3\psi}{\partial t^2\partial r}\right)\eta + 2\left[\frac{\mu+\lambda}{c_l^2\mu}\frac{\partial^2\varphi}{\partial t^2} + \frac{1}{R}\left(\frac{\partial\psi}{\partial z} - \frac{\partial\varphi}{\partial r}\right) - \frac{\rho_f}{\rho c_s^2}\frac{\partial^2\varphi_f}{\partial t^2}\right]\frac{\partial\eta}{\partial z},$$

$$2\frac{\partial^2\psi}{\partial r\partial z} + 2\frac{\partial^2\varphi}{\partial z^2} + \frac{2}{R}\frac{\partial\varphi}{\partial r} - \frac{1}{c_s^2}\frac{\partial^2\varphi}{\partial t^2} - \frac{\rho_f}{\rho c_s^2}\frac{\partial^2\varphi_f}{\partial t^2} = \left(-2\frac{\partial^3\psi}{\partial r^2\partial z} - 2\frac{\partial^3\varphi}{\partial r\partial z^2} - \frac{2}{R}\frac{\partial^2\varphi}{\partial r^2} + \frac{1}{c_s^2}\frac{\partial^3\varphi}{\partial t^2\partial r} + \frac{\rho_f}{\rho c_s^2}\frac{\partial^3\varphi}{\partial t^2\partial r}\right)\eta,$$

$$\frac{\partial\varphi_f}{\partial r} - \frac{\partial\varphi}{\partial r} + \frac{\partial\psi}{\partial z} = \left(\frac{\partial^2\varphi}{\partial r^2} - \frac{\partial^2\psi}{\partial r\partial z} - \frac{\partial^2\varphi_f}{\partial r^2}\right)\eta - \left(\frac{\partial\varphi}{\partial z} + \frac{1}{R}\psi + \frac{\partial\psi}{\partial r} - \frac{\partial\varphi_f}{\partial z}\right)\frac{\partial\eta}{\partial z}. \tag{A6}$$

Applying the time and spatial Fourier transform along the z-axis to Eq. (A6) yields

$$2ik\frac{\partial\phi}{\partial r}+\Omega_s\psi=\frac{1}{2\pi}\int_{-\infty}^{+\infty}dk'\,\eta(k')\Bigg[-2i\tilde{k}\frac{\partial^2\tilde{\phi}}{\partial r^2}-\tilde{\Omega}_s\frac{\partial\tilde{\psi}}{\partial r}+\left(-(k_s^2-k_l^2)\tilde{\phi}+\frac{1}{R}\left(i\tilde{k}\tilde{\psi}-\frac{\partial\tilde{\phi}}{\partial r}\right)+\frac{\rho_f}{\rho}k_s^2\tilde{\varphi}_f\right)2ik'\Bigg],$$

$$2ik\frac{\partial\psi}{\partial r}-\Omega_s\phi+\frac{2}{R}\frac{\partial\phi}{\partial r}-\frac{\rho_f}{\rho}k_s^2\phi_f=\frac{1}{2\pi}\int_{-\infty}^{+\infty}dk'\,\eta(k')\Bigg[-2i\tilde{k}\frac{\partial^2\tilde{\psi}}{\partial r^2}+\tilde{\Omega}_s\frac{\partial\tilde{\phi}}{\partial r}-\frac{2}{R}\frac{\partial^2\tilde{\phi}}{\partial r^2}+\frac{2}{R^2}\frac{\partial\tilde{\phi}}{\partial r}+\frac{\rho_f}{\rho}k_s^2\frac{\partial\tilde{\phi}_f}{\partial r}\Bigg],$$

$$\frac{\partial\varphi_f}{\partial r}-\frac{\partial\varphi}{\partial r}-ik\psi=\frac{1}{2\pi}\int_{-\infty}^{+\infty}dk'\,\eta(k')\Bigg[\left(\frac{\partial^2\tilde{\varphi}}{\partial r^2}-i\tilde{k}\frac{\partial\tilde{\psi}}{\partial r}-\frac{\partial^2\tilde{\varphi}_f}{\partial r^2}\right)-\left(i\tilde{k}\tilde{\varphi}+\frac{1}{R}\tilde{\psi}+\frac{\partial\tilde{\psi}}{\partial r}-i\tilde{k}\tilde{\varphi}_f\right)ik'\Bigg], \tag{A7}$$

where

$$\tilde{k}=k-k',\quad \Omega_s=2k^2-k_s^2,\quad \tilde{\Omega}_s=2\tilde{k}^2-k_s^2,$$

$$\tilde{\nu}_{js}^2=\tilde{k}^2-k_j^2,\quad j=l,s,f,\quad k_s=\omega/c_s,\quad k_l=\omega/c_l,$$

and $\eta(k)=\int_{-\infty}^{+\infty}dx e^{-ikx}\eta(x)$ is the spatial Fourier transform of the boundary roughness.

APPENDIX B: COMPONENTS OF MATRIX OPERATOR $L_1(k,\tilde{k})$

The matrix operator $\mathbf{L}_1(k,\tilde{k})$ has the following components:

$$L_{11}=-2ik'\frac{\rho_f}{\rho}k_s^2\tilde{I}_0^f,$$

$$L_{12}=2i\tilde{k}\tilde{\nu}_l\tilde{I}_l^1+2ik'\left((k_l^2-k_s^2)\tilde{I}_0^l-\frac{1}{R_1}\tilde{\nu}_l\tilde{I}_1^l\right),$$

$$L_{13}=-2i\tilde{k}\tilde{\nu}_l\tilde{K}_l^1+2ik'\left((k_l^2-k_s^2)\tilde{K}_0^l+\frac{1}{R_1}\tilde{\nu}_l\tilde{K}_1^l\right),$$

$$L_{14}=\tilde{\Omega}_s\tilde{K}_s^1-\frac{2}{R_1}k'\tilde{k}\tilde{K}_1^s$$

$$L_{15}=-\tilde{\Omega}_s\tilde{I}_s^1-\frac{2k'\tilde{k}}{R_1}\tilde{I}_1^s,$$

$$L_{21}=\frac{-\rho_f}{\rho}k_s^2\tilde{\nu}_f\tilde{I}_1^f,\quad L_{22}=-\tilde{\nu}_l\left(\tilde{\Omega}_s\tilde{I}_1^l+\frac{2}{R_1}\left(\frac{1}{R_1}\tilde{I}_1^l-\tilde{I}_l^1\right)\right),$$

$$L_{23}=\tilde{\nu}_l\left(\tilde{\Omega}_s\tilde{K}_1^l+\frac{2}{R_1}\left(\tilde{K}_l^1+\frac{1}{R_1}\tilde{K}_1^l\right)\right),$$

$$L_{24}=2i\tilde{k}\left(\tilde{K}_1^s\left(\tilde{\nu}_s^2+\frac{2}{R_1^2}\right)+\tilde{K}_0^s\frac{\tilde{\nu}_s}{R_1}\right),$$

$$L_{25}=2i\tilde{k}\left(\tilde{I}_1^s\left(\tilde{\nu}_s^2+\frac{2}{R_1^2}\right)-\frac{\tilde{\nu}_s}{R_1}\tilde{I}_0^s\right),$$

$$L_{31}=-(\tilde{I}_f^1\tilde{\nu}_f+k'\tilde{k}\tilde{I}_0^f),\quad L_{32}=\tilde{\nu}_l\tilde{I}_l^1+k'\tilde{k}\tilde{I}_0^l,$$

$$L_{33}=\tilde{\nu}_l\tilde{K}_l^1+k'\tilde{k}\tilde{K}_0^l,$$

$$L_{34}=i(\tilde{k}\tilde{K}_s^1+k'\tilde{K}_0^s\tilde{\nu}_s),\quad L_{35}=-i(\tilde{k}\tilde{I}_s^1+k'\tilde{\nu}_s\tilde{I}_0^s).$$

In the previous expressions the following notations are used to denote the modified Bessel functions of first and second kind:

$$I_n(\nu_iR_j)=Ij_n^i,\quad K_n(\nu_iR_j)=Kj_n^i,\quad n=0,1,$$

$$i=f,l,s,\quad j=1,2.$$

[1] J. Blitz and G. Simpson, "Ultrasonic methods of non-destructive testing," (Chapman and Hall, London, 1996).
[2] W. Mohr and P. Holler, "On inspection of thin walled tubes for transverse and longitudinal flaws by guided ultrasonic waves," IEEE Trans. Sonics Ultrason. **23**, 369–374 (1976).
[3] M. G. Silk and K. P. Bainton, "The propagation in metal tubing of ultrasonic wave modes equivalent to Lamb waves," Ultrasonics **17**, 11–19 (1979).
[4] W. Bottger, H. Schneider, and W. Wiengarten, "Prototype EMAT system for tube inspection with guided ultrasonic waves," Nucl. Eng. Des. **102**, 356–376 (1987).
[5] J. L. Rose, J. J. Ditri, A. Pilarski, K. Rajana, and F. T. Carr, "A guided wave inspection technique for nuclear steam generator tubing," NDT & E Int. **27**, 307–330 (1994).
[6] J. L. Rose, Y. Cho, and J. L. Ditri, "Cylindrical guided wave leakage due to liquid loading," in *Review of Progress in QNDE*, edited by D. O. Thopson and D. E. Chimenti (Plenum, New York, 1994), Vol. **13**, pp. 259–266.
[7] R. Long, T. Vogt, M. J. S. Lowe, and P. Cawley, "Measurement of acoustic properties of near-surface soils using an ultrasonic waveguide," Geophysics **69**, 460–465 (2004).
[8] F. B. Cegla, P. Cawley, and M. J. S. Lowe, "Material property measurement using the quasi-Scholte mode—A waveguide sensor," J. Acoust. Soc. Am. **117**, 1098–1107 (2005).
[9] D. N. Alleyne and P. Cawley, "The interaction of Lamb waves with defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[10] D. N. Alleyne and P. Cawley, "Optimisation of Lamb wave inspection techniques," NDT & E Int. **25**, 11–22 (1992).
[11] J. J. Ditri, "Utilization of guided elastic waves for the characterization of circumferential cracks in hollow cylinders," J. Acoust. Soc. Am. **96**, 3769–3775 (1994).
[12] M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "Defect detection in pipes using guided waves," Ultrasonics **36**, 147–154 (1998).
[13] A. Demma, P. Cawley, M. J. S. Lowe, and A. G. Roosenbrand, "The reflection of the fundamental torsional mode from cracks and notches in pipes," J. Acoust. Soc. Am. **114**, 611–625 (2003).
[14] D. N. Alleyne and P. Cawley, "The long range detection of corrosion in pipes using Lamb waves," in *Review of Progress in QNDE*, edited by D. O. Thopson and D. E. Chimenti (Plenum, New York, 1995), Vol. **14**, pp.

2073–2080.
[15]D. N. Alleyne, M. J. S. Lowe, and P. Cawley, "The reflection of guided waves from circumferential notches in pipes," ASME J. Appl. Mech. **65**, 635–641 (1998).
[16]M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "Mode conversion of guided waves by defects in pipes," in *Review of Progress in QNDE*, edited by D. O. Thopson and D. E. Chimenti (Plenum, New York, 1997), Vol. **16**, pp. 1261–1268.
[17]M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "The mode conversion of a guided waves by a part circumferential notch in a pipe," ASME J. Appl. Mech. **65**, 649–656 (1998).
[18]P. Cawley, M. J. S. Lowe, F. Simonetti, C. Chevalier, and A. G. Roosenbrand, "The variation of the reflection coefficient of extensional guided waves in pipes from defects as a function of defect depth, axial extent, circumferential extent and frequency," Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci. **216**, 1131–1143 (2002).
[19]T. Vogt, M. J. S. Lowe, and P. Cawley, "The scattering of guided waves in partly embedded cylindrical structures," J. Acoust. Soc. Am. **113**, 1258–1272 (2003).
[20]N. Amir, H. Matzner, and S. Shtrikman, "Acoustics of a flanged cylindrical pipe using singular basis functions," J. Acoust. Soc. Am. **107**, 714–724 (2000).
[21]D. N. Alleyne and P. Cawley, "The effect of discontinuities on the long range propagation of Lamb waves in pipes," Proc. I. Mech. E., Part E: J. Process. Mech. Eng. **210**, 217–226 (1996).
[22]D. N. Alleyne and P. Cawley, "Long range propagation of Lamb waves in chemical plant pipework," Mater. Eval. **55**, 504–508 (1997).
[23]J. M. Muggleton and M. J. Brennan, "Axisymmetric wave propagation in buried, fluid-filled pipes: Effects of wall discontinuities," J. Sound Vib. **281**, 849–867 (2005).
[24]Y. Gao, M. J. Brennan, P. F. Joseph, J. M. Muggleton, and O. Hunaidi, "A model of the correlation function of leak noise in buried plastic pipes," J. Sound Vib. **277**, 133–148 (2004).
[25]Y. Gao, M. J. Brennan, and P. F. Joseph, "A comparison of time delay estimators for the detection of leak noise signals in plastic water distribution pipes," J. Sound Vib. **292**, 552–570 (2006).
[26]J. M. Muggleton and M. J. Brennan, "Leak noise propagation and attenuation in submerged plastic water pipes," J. Sound Vib. **278**, 527–537 (2004).
[27]V. N. R. Rao and J. K. Vandiver, "Acoustics of fluid filled boreholes with pipe: Guided propagation and radiation," J. Acoust. Soc. Am. **105**, 3057–3066 (1999).
[28]M. P. Horne and R. Hansen, "Sound propagation in pipe containing a liquid of comparable acoustic impedance," J. Acoust. Soc. Am. **71**, 1400–1405 (1982).
[29]M. J. Moloney and D. L. Hatten, "Acoustic quality factor and energy losses in cylindrical pipes," Am. J. Phys. **69**, 311–314 (2001).
[30]C. Aristegui, M. J. S. Lowe, and P. Cawley, "Guided waves in fluid-filed pipes surrounded by different fluids," Ultrasonics **39**, 367–375 (2001).
[31]R. Long, P. Cawley, and M. J. S. Lowe, "Acoustic wave propagation in buried iron pipes," Proc. R. Soc. London, Ser. A **459**, 2749–2770 (2003).
[32]R. Long, M. J. S. Lowe, and P. Cawley, "Attenuation characteristics of the fundamental modes that propagate in buried iron pipes," Ultrasonics **41**, 509–519 (2003).
[33]F. G. Bass and I. M. Fuks, *Wave Scattering from Statistically Rough. Surfaces* (Nauka, Moscow, 1972), p. 424 (in Russian).
[34]L. M. Brekhovskikh, "Wave diffraction by a rough surface," Sov. Phys. JETF Parts 1 and 2 **23**, 289–304 (1952).
[35]S. Rayleigh, *Theory of Sound*, 2nd ed. (Dover, New York, 1976).
[36]F. Gilbert and L. Knopoff, "Seismic scattering from topographic irregularities," J. Geophys. Res. **65**, 3437–3444 (1960).
[37]A. D. Lapin, "Scattering of fluid-borne surface waves on a rough rigid interface," Akust. Zh. **15**, 387–392 (1969) (in Russian).
[38]E. I. Urazakov and L. A. Falcovskii, "On the propagation of the Rayleigh wave along a rough boundary," Sov. Phys. JETF **63**, 2297–2302 (1972) (in Russian).
[39]A. G. Eguiluz and A. A. Maradudin, "Frequency shift and attenuation length of a Rayleigh wave due to surface roughness," Phys. Rev. B **28**, 728–747 (1983).
[40]N. E. Glass, R. London, and A. A. Maradudin, "Propagation of Rayleigh surface waves across a large-amplitude grating," Phys. Rev. B **24**, 6843–6841 (1981).
[41]V. V. Kosachev and A. Shchegrov, "Dispersion and attenuation of surface acoustic waves of various polarizations on a stress-free randomly rough surface of solid," Ann. Phys. (N.Y.) **240**, 225–265 (1995).
[42]G. A. Maximov, E. Ortega, E. V. Pod'yachev, and M. R. Chillemi, "Rayleigh wave attenuation due to the scattering by two-dimensional irregularities of the walls of an empty borehole," Akust. Zh. **50**, 682–692 (2004) (in Russian); [also in Acoust. Phys. **50**, 585–595 (2004), in English].
[43]E. V. Pod'yachev, G. A. Maximov, and E. Ortega, "Attenuation of a Stoneley wave and higher modes in fluid-filled borehole due to their scattering by two-dimensional irregularities of the walls," Proceedings of the XIII Session of the Russian Acoustical Society, Moscow, 25–29 August 2003, Vol. **1**, pp. 245–248 (in Russian).
[44]G. A. Maximov, E. Ortega, and E. V. Pod'yachev, "Attenuation of a Stoneley wave and higher Lamb modes due to the scattering by two-dimensional irregularities of the walls of fluid-filled borehole," Akust. Zh. **53**, 20–37 (2007) (in Russian); [also Acoust. Phys. **53**, 16–32 (2007), in English].
[45]M. A. Biot, "Propagation of elastic waves in a cylindrical borehole containing a fluid," J. Appl. Phys. **23**, 997–1005 (1952).
[46]G. I. Petrashen', L. A. Molotkov, and P. V. Krauklis, *Waves in Layered Homogeneous Isotropic Elastic Media* (Leningrad, Nauka, 1985), Vol. **2**, p. 302 (in Russian).

Low frequency wind noise contributions in measurement microphones

Richard Raspet, Jiao Yu, and Jeremy Webster[a)]
National Center for Physical Acoustics and the Department of Physics and Astronomy, University of Mississippi, University, Mississippi 38677

(Received 23 July 2007; revised 12 December 2007; accepted 13 December 2007)

In a previous paper [R. Raspet, *et al.*, J. Acoust. Soc. Am. **119**, 834–843 (2006)], a method was introduced to predict upper and lower bounds for wind noise measured in spherical wind-screens from the measured incident velocity spectra. That paper was restricted in that the predictions were only valid within the inertial range of the incident turbulence, and the data were from a measurement not specifically designed to test the predictions. This paper extends the previous predictions into the source region of the atmospheric wind turbulence, and compares the predictions to measurements made with a large range of wind-screen sizes. Predictions for the turbulence–turbulence interaction pressure spectrum as well as the stagnation pressure fluctuation spectrum are calculated from a form fit to the velocity fluctuation spectrum. While the predictions for turbulence–turbulence interaction agree well with measurements made within large (1.0 m) wind-screens, and the stagnation pressure predictions agree well with unscreened gridded microphone measurements, the mean shear–turbulence interaction spectra do not consistently appear in measurements.
 [DOI: 10.1121/1.2832329]

PACS number(s): 43.28.Dm, 43.28.Ra, 43.28.Vd, 43.50.Rq [GCL] Pages: 1260–1269

I. INTRODUCTION

In an earlier paper[1] we argued that wind noise measurements near the earth's surface should be bound between the stagnation pressure measured on a smooth object in the flow and by the turbulence–turbulence interaction pressure as calculated by Batchelor.[2] The stagnation pressure fluctuation and turbulence–turbulence interaction pressure fluctuation spectra in the inertial range were then predicted based on the measured wind velocity turbulence spectrum and compared to measurements of wind noise. The calculations in the earlier paper were limited to the inertial range of the turbulence and it was assumed that the mean shear–turbulence interaction pressure was negligible. However, extrapolation of the inertial range results from George *et al.*[3] indicates that the mean shear–turbulence interaction pressures could dominate at low wave numbers. In this paper we extend the three predictions into the source region of the turbulent flow so that the results may be applied to infrasonic measurements.

The turbulence–turbulence and mean shear–turbulence interaction pressures are intrinsic to the turbulent flow. These contributions are the pressure fluctuations that would be measured if we could construct a perfect wind-screen, which could eliminate all pressure fluctuations due to local interactions of the turbulence with the surface of the wind-screen. The sum of the turbulence–turbulence and mean shear–turbulence interaction pressures represents a lower limit on the wind noise reduction that can be achieved by a compact wind-screen. The two intrinsic pressure contributions have scales which are comparable to the velocity turbulence scales. At low frequency, the correlation length of the intrinsic pressure fluctuations is much larger than a typical wind-screen.

The stagnation pressure is thought to represent the maximum pressure fluctuation generated when a bluff object, the microphone or microphone wind-screen combination, is placed in a turbulent flow and interacts with the incident flow field. The measurements of Ref. 1 showed that the fluctuating pressure levels measured by an unscreened gridded microphone in the flow were well predicted by the calculated stagnation pressure term in the inertial range of the turbulence. Current theories of wind noise reduction by spherical wind-screens assume that the dominant pressure fluctuation is generated locally at the surface of the wind-screen.[4,5] The pressure measured by a microphone at the center of the wind-screen is predicted to be the pressure fluctuation averaged over the entire surface of the wind-screen.[5,6] The principle mechanism of wind noise reduction is the variation in phase of the locally generated pressure fluctuations. At low frequencies, the phase distribution is assumed to be the steady state pressure distribution, while at higher frequencies the correlation length is inversely proportional to the frequency.[4] As the wind-screen size is increased, the averaging area is increased and the wind noise reduction increases. The basic assumption in either case is that the scales of the locally generated pressure fluctuations are much smaller than the scales of the same frequency turbulent flow away from the wind-screen. The stagnation pressure and locally generated pressure fluctuations have much higher amplitudes than the intrinsic pressure fluctuations in the inertial range.

In this paper we have measured the pressure fluctuation levels in the bare gridded microphone and in a series of different size spherical screens. If the scales of the intrinsic and locally generated pressure are correctly estimated above,

[a)]Electronic mail: jwebster@olemiss.edu.

then increasing the size and area of the wind-screen will have a large effect in reducing the average level of the locally generated pressure fluctuations and a much smaller effect on the large scale intrinsic pressure fluctuations. If the surface contributions are reduced sufficiently by increasing the wind-screen size, then the turbulence–turbulence and mean shear–turbulence pressure fluctuations may become the dominant contribution to the measured wind noise.

Much of the theoretical work in this paper is based on that presented by George *et al.*[3] George *et al.* compared predictions of the turbulence–turbulence interaction pressure and mean shear–turbulence interaction pressure based on turbulent velocity measurements to pressure measurements made by a streamlined probe in a laboratory turbulent shear flow. We will apply their theories and our calculation of the stagnation pressure spectrum to study the wind noise contributions on bare, gridded, and a series of different size screened microphones in outdoor turbulent wind. In Ref. 1 our conclusions were limited because the largest wind-screen was not mounted at the same height as the other microphones and the ultrasonic anemometer used to measure the fluctuating wind velocity components. In addition, the acoustic transmission characteristics of the largest screen were not measured so there were uncertainties in interpreting the measurements.

The predictions are compared to wind noise measurements using laboratory grade, low frequency microphones (0.07 Hz) deployed in a variety of bare and screened arrangements at 1.0 m above the ground. The wind-screens are all nominally spherical. The smaller screens are standard open cell plastic foam while the larger (60 and 100 cm) are constructed of fiberglass wool.

II. THEORY

A. A spectral model for turbulence

George *et al.*[3] use the von Karman spectrum to model the turbulence in a free jet. The one-dimensional spectrum in the direction of flow is given by

$$F_{11}^{1}(k_1) = \frac{18}{55}\frac{\alpha\epsilon^{2/3}\lambda^{5/3}}{[1+(k_1\lambda)^2]^{5/6}}, \tag{1}$$

where α is Kolmogorov's constant taken as 1.5, ϵ is the mean energy dissipation per unit mass of fluid, λ is a length parameter determining the location of the transition from the source region to the inertial range, and is related to the longitudinal integral scale l as

$$\lambda = 0.96l, \tag{2}$$

18/55 is a constant relating the three-dimensional energy spectrum to the one-dimensional spectrum and is an appropriate value for a one-sided spectrum with the wave number in the direction of flow k_1 ranging from 0 to $+\infty$. The von Karman spectrum is appropriate for purely mechanical turbulence.

Kaimal[7] and Højstrup[8] have developed models for velocity spectra measured outdoors where the turbulence arises from convective heating terms in addition to the mechanical terms. In this case the spectra depend on additional factors such as the roughness length, the stability of the atmosphere, and the inversion height. Højstrup[9] has also developed models for unstable conditions downstream from changes in roughness and heat flux which may be more appropriate for our measurement site at Clegg Field—Oxford University Airport.

Our velocity spectra were taken over 15 min periods under unstable, unsteady conditions. The 15 min period was chosen as appropriate for short-term acoustic measurements while still being long enough to examine low frequency contributions to the wind velocity and infrasound spectra.

Figure 1 displays the power spectral density of the wind velocity on the average wind direction during four measurement periods. We have not observed separate convective and mechanical contributions to the spectra as described by Højstrup's model. This is due to our relatively short sample times and possibly terrain and vegetation factors. Although Clegg Field is relatively open, there are terrain changes within 100 m of the measurement site.

All of our measured data sets displayed an inertial range with an approximately $k^{-5/3}$ slope and a near constant source region. We found that our spectra could be fit to the curve

$$F_{11}^{1}(k_1) = \frac{C}{[1+(k_1\lambda)^2]^{5/6}}. \tag{3}$$

C and λ are now fit parameters which are used to best match Eq. (3) to the measured velocity spectra, and cannot be interpreted in terms of Kolmogorov's constant and the mean dissipation. k_1 is calculated from the measured frequency and average wind velocity in the direction of flow using Reynold's frozen turbulence hypothesis: $k_1 = 2\pi f/U$. Equation (3) matches the form of the von Karman spectra, but the constants are determined empirically and are not those of von Karman.

This choice is convenient since the calculations by George *et al.* of turbulence–turbulence and mean shear–turbulence interaction spectra from the fit to the velocity spectrum can be used to analyze our wind noise spectra with minimal modifications.

Throughout this paper we have retained C and λ as the fit parameters to our data. Integration of Eq. (3) yields a relationship between the mean square velocity fluctuation in the direction of flow and the fit parameters

$$\overline{u_1^2} = 2.103\frac{C}{\lambda}. \tag{4}$$

This relation is useful for relating the results of this paper to other atmospheric conditions.

B. Turbulence–turbulence and mean shear–turbulence interaction pressure

George *et al.*[3] develop calculations for the turbulence–turbulence interaction pressure spectrum and the mean shear–turbulence interaction pressure spectrum for the von Karman spectrum. Both calculations are based on the Navier–Stokes equation applied to incompressible flow,

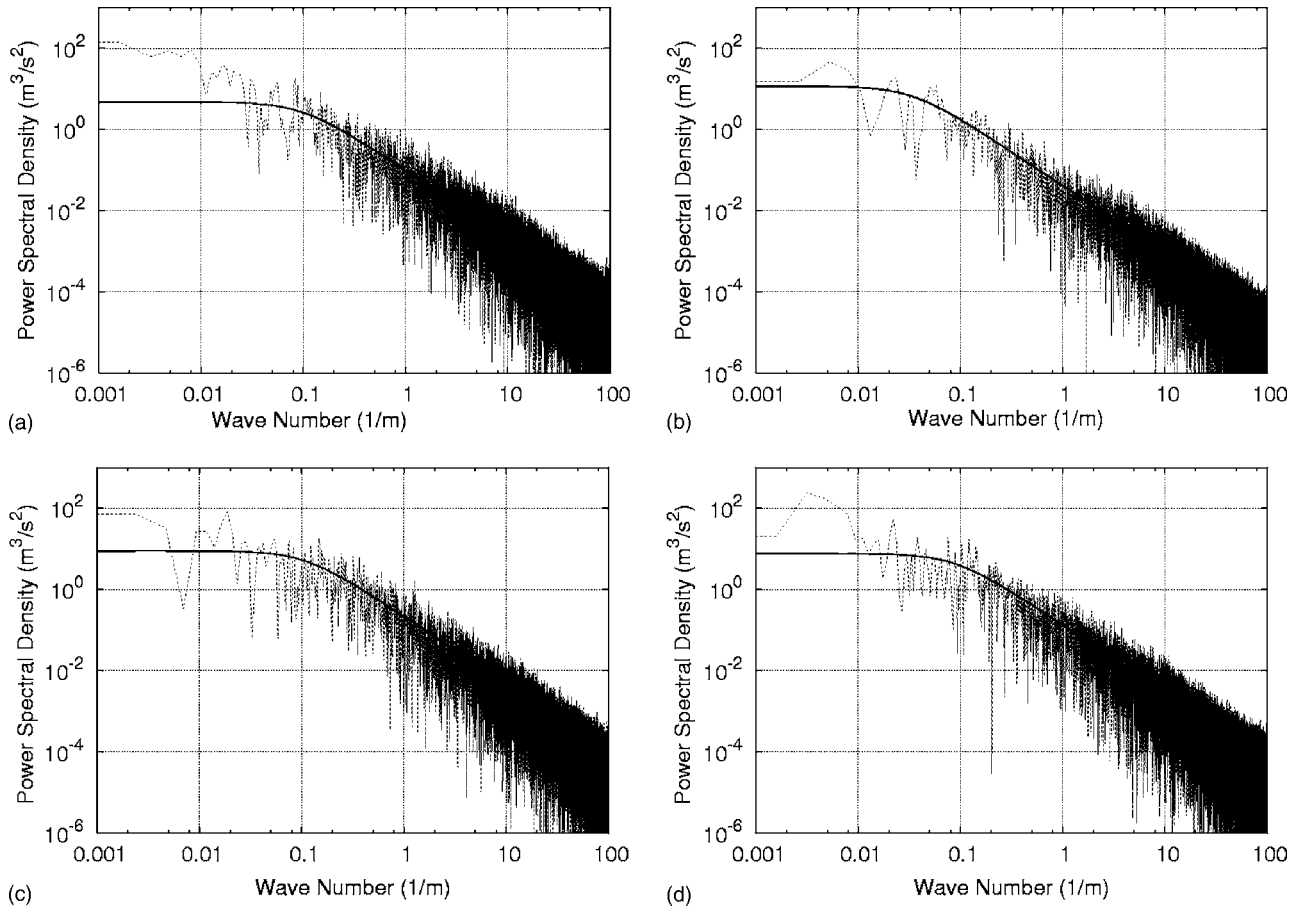


FIG. 1. Measured velocity spectra with fit spectra. The velocity spectra are used to predict the pressure spectra displayed in the corresponding figures: (a) Fig. 6, (b) Fig. 7, (c) Fig. 8, and (d) Fig. 9.

$$\frac{1}{\rho}\nabla^2 p = -\frac{\partial^2 u_i u_j}{\partial x_i \, \partial x_j}. \tag{5}$$

First the Green's function solution is used to form an integral equation for the pressure fluctuations in terms of derivatives of the average velocity and the fluctuation velocity components. The average velocity is restricted to a unidirectional flow with constant mean shear. We assume that the shear is changing slowly enough at our measurement height so that this will be valid for our measurements. Next the cross correlation of the pressure at two points is formed in terms of the cross correlations between velocity components.

Then the three-dimensional Fourier transform of the equation is performed resulting in

$$\frac{1}{\rho^2}F_{pp}(\mathbf{k}) = 4K^2\left[\frac{k_1^2}{k^4}F_{2,2}(\mathbf{k})\right] + 2iK\left[\frac{k_1 k_l k_m}{k^4}(F^*_{2,lm}(\mathbf{k}) - F_{2,lm}(\mathbf{k}))\right] + \frac{k_i k_j k_l k_m}{k^4}F_{ij,lm}(\mathbf{k}). \tag{6}$$

K is the mean shear of the flow in m/s/m, and $F_{2,2}$, $F_{2,lm}$, and $F_{ij,lm}$ are the three-dimensional Fourier transforms of the velocity correlation functions:

$$B_{2,2}(\mathbf{r}) = \overline{u_2(\mathbf{y})u_2(\mathbf{y}+\mathbf{r})},$$

$$B_{2,lm}(\mathbf{r}) = \overline{u_2(\mathbf{y})u_l(\mathbf{y}+\mathbf{r})u_m(\mathbf{y}+\mathbf{r})},$$

$$B_{ij,lm}(\mathbf{r}) = \overline{u_i(\mathbf{y})u_j(\mathbf{y})u_l(\mathbf{y}+\mathbf{r})u_m(\mathbf{y}+\mathbf{r})} - \overline{u_i(\mathbf{y})u_j(\mathbf{y})u_l(\mathbf{y}+\mathbf{r})u_m(\mathbf{y}+\mathbf{r})}. \tag{7}$$

The mean flow is in the 1 direction and the gradient of the mean flow is in the 2 direction. In Eq. (6) $F_{pp}(\mathbf{k})$ is the three-dimensional Fourier transform of the pressure fluctuations, the first term on the right-hand side is the second moment mean shear–turbulence interaction contribution, the second term is the third moment mean shear–turbulence interaction term, and the last term is the turbulence–turbulence interaction terms. George notes that the second term is zero for isotropic turbulence. We have assumed that its contribution is negligible compared to the second moment term, although the justification for this is weaker for anisotropic atmospheric turbulence at low wave numbers than for George's laboratory measurements.

In the absence of detailed information we follow George and assume that the general forms for the velocity spectra are related like isotropic turbulence and that the relations derived by Batchelor for isotropic homogeneous turbulence may be used to simplify the above-presented expressions.

Kraichnan[10] has calculated the effect of anisotropy on the pressure field in homogeneous turbulence. He found that anisotropy produced only small reductions in the predicted pressure fluctuation fields. If we assume his analysis is ap-

plicable to the turbulence fields of this study, the predictions here represent a small overestimation at low wave numbers.

The velocity spectra $F_{i,j}(\mathbf{k})$ are given in terms of the three-dimensional velocity spectral function $E(k)$,

$$F_{i,j}(\mathbf{k}) = \frac{E(k)}{4\pi k^2}\left[\delta_{ij} - \frac{k_i k_j}{k^2}\right], \tag{8}$$

where $E(k)$ is the spectral energy density per unit mass,

$$E(k) = \frac{1}{2}\iint F_{i,i}(\mathbf{k})d\sigma(\mathbf{k}). \tag{9}$$

The integration in Eq. (9) is performed over the area of the spherical shell of radius k in wave number space.

The last step in the analysis forms the one-dimensional spectrum along the direction of flow by integrating the three-dimensional spectrum over the k_2, k_3 plane.

For isotropic turbulence with a power spectral density given by Eq. (3), the energy spectrum is given by

$$E(k) = \frac{55}{18}(k\lambda)^4 \frac{C}{[1+(k\lambda)^2]^{17/6}}. \tag{10}$$

The mean shear–turbulence interaction pressure spectrum in the direction of flow is

$$F^1_{ppm}(k_1) = K^2 2\rho^2 \frac{55}{18} k_1^2 C\lambda^4 [J_1 + (k_1\lambda)^2 J_2], \tag{11}$$

where J_1 and J_2 can be expressed in terms of the nondimensional variable $\chi = k\lambda$:

$$J_1 = \int_{\chi_1}^{\infty} \frac{d\chi}{\chi[1+\chi^2]^{17/6}}, \tag{12a}$$

$$J_2 = \int_{\chi_1}^{\infty} \frac{d\chi}{\chi^3[1+\chi^2]^{17/6}}. \tag{12b}$$

K is the gradient of the average wind speed at the observation height. C and λ are determined by fits to the measured velocity spectrum in the average wind direction.

We did not measure the wind velocity profile to determine K. In Chap. 6 of Ref. 11, Panofsky and Dutton present models for estimating the wind profile as a function of wind speed, surface roughness, and atmospheric observations. The Monin–Obukhov length is estimated from observation of surface conditions, time of day, cloud cover, and wind speed. For our unstable conditions, we found that estimating the slope of the wind velocity profile using the simple logarithm profile provides an upper limit to the value of the derivative at our 1.0 m observation height and that the more complicated Businger–Dyer formula produced at most a 15% reduction in the gradient. With the experimental uncertainties and nonideal site, use of the more complicated formula was not justified. In addition, earlier measurements of the wind velocity profile over a smooth flat plate at Clegg Field were well fit with the logarithmic profile.[12]

The formula for K at 1.0 m is derived from the logarithmic profile:

$$K = \frac{U(z)}{z\ln(z/z_0)}, \quad z = 1.0, \tag{13}$$

where z_0 is the roughness length taken as 0.01 m for mown grass surfaces.[11]

The turbulence–turbulence interaction pressure spectrum is given by

$$F^1_{ppt}(k_1) = \frac{2^{-17/6}}{\lambda}\left(\frac{55}{18}C_\rho\right)^2 \int_{\chi_1}^{\infty} \chi^{7/3} d\chi \int_0^{\infty} \frac{I(a)y^{19/6}dy}{[1+\chi^2 y^2]^{17/6}}, \tag{14}$$

where

$$I(a) = \left[\frac{186\,624}{5005}a^2 - \frac{1728}{143}\right][(a+1)^{1/6} - (a-1)^{1/6}] - \frac{31\,104}{5005}a[(a+1)^{1/6} + (a-1)^{1/6}], \tag{15}$$

and

$$a = \frac{1+(k\lambda)^2(1+y^2)}{2y(k\lambda)^2}. \tag{16}$$

For numerical evaluation, the integrals in Eq. (14) can be performed once to give the integrals as a function of χ_1. The value of F^1_{ppt} is then evaluated as a function of k_1 using the relation $k_1 = \chi_1/\lambda$ in terms of the fit parameter λ.

C. Stagnation interaction pressure

The source equation for pressure in an incompressible flow can be derived in an alternative form by taking the divergence of the Navier-Stokes equation in vector form and applying the continuity equation.[13]

$$\frac{1}{\rho}\nabla^2 p = \nabla^2\left(\frac{1}{2}V^2\right) - \nabla\cdot[\mathbf{V}\times\nabla\times\mathbf{V}]. \tag{17}$$

If wind impinges on a bluff body, such as a wind-screen, the dominant term under some conditions will be the change in the magnitude $\rho V^2/2$ from the free stream to the front surface of the bluff body. If the second term is neglected, Eq. (17) can be integrated to give Bernoulli's equation. We will discuss the validity of this approximation in regard to experimental results in Sec. IV.

Bernoulli's equation,

$$P(t) = \frac{1}{2}\rho V^2(t), \tag{18}$$

can be expanded in terms of the average wind velocity U and the fluctuating components u_i:

$$P(t) = \frac{1}{2}\rho U^2 + \rho U u_1 + \frac{1}{2}\rho u_i u_i. \tag{19}$$

To derive the pressure spectral density in terms of the fit to the measured average velocity and one-dimensional spectrum of the wind velocity fluctuation in the direction of flow, we follow George's method of calculation.

First, the mean pressure is calculated:

$$\overline{P(t)} = \frac{1}{2}\rho U^2 + \rho U\overline{u_1} + \frac{1}{2}\rho\overline{u_i u_i}. \tag{20}$$

Equation (20) can be simplified by assuming isotropic turbulence so that $\overline{u^2}=\overline{u_i^2}$ for any i. Since the mean of the fluctuation velocity in the flow direction ($\overline{u_1}$) is zero, Eq. (20) can be written

$$\overline{P(t)} = \frac{1}{2}\rho U^2 + \frac{3}{2}\rho\overline{u^2}. \tag{21}$$

This is subtracted from Eq. (19) to give an equation for the fluctuation pressure alone,

$$p(t) = \rho U u_1 + \frac{1}{2}\rho u_i u_i - \frac{3}{2}\rho\overline{u^2}. \tag{22}$$

Next, the spatial correlation function is formed:

$$\overline{\frac{pp'}{\rho^2}} = \overline{\left(Uu_1 + \frac{1}{2}u_i u_i - \frac{3}{2}\overline{u^2}\right)\left(Uu_1' + \frac{1}{2}u_j' u_j' - \frac{3}{2}\overline{u'^2}\right)}, \tag{23}$$

where the unprimed quantities are evaluated at the origin and the primed quantities at position $\mathbf{r}$. Using the homogeneity condition, this becomes

$$\overline{\frac{pp'}{\rho^2}} = U^2\overline{u_1 u_1'} + \frac{1}{4}\overline{(u_i u_i - 3\overline{u^2})(u_j' u_j' - 3\overline{u'^2})}. \tag{24}$$

This relation can be expressed in terms of the cross correlations of the velocity components $B_{1,1}$ and $B_{ii,jj}$ and the cross correlation of the pressures B_{pp}:

$$\frac{B_{pp}}{\rho^2} = U^2 B_{1,1}(\mathbf{r}) + \frac{1}{4}B_{ii,jj}(\mathbf{r}). \tag{25}$$

The spectral densities of the pressure and velocity moments are three-dimensional Fourier transforms of the cross correlations:

$$\frac{F_{pps}(\mathbf{k})}{\rho^2} = U^2 F_{1,1}(\mathbf{k}) + \frac{1}{4}F_{ii,jj}(\mathbf{k}), \tag{26}$$

where $F_{1,1}$ is the spectrum of the u_1 component of the velocity.

The fourth-order moments are calculated by assuming they can be related to the second-order moments as if they were Gaussian,[2,3]

$$F_{ij,lm}(\mathbf{k}) = \int F_{i,l}(\mathbf{k}-\mathbf{k}')F_{j,m}(\mathbf{k}')d^3\mathbf{k}' + \int F_{i,m}(\mathbf{k} - \mathbf{k}')F_{j,l}(\mathbf{k}')d^3\mathbf{k}'. \tag{27}$$

Assuming isotropy, the three-dimensional velocity spectra in terms of the velocity spectrum function $E(k)$ are given by Eq. (8). Combining Eqs. (8) and (27) and substituting into Eq. (26) gives

$$\frac{F_{pps}(\mathbf{k})}{\rho^2} = U^2 F_{1,1}(\mathbf{k}) + \frac{1}{32\pi^2}\int \frac{E(|\mathbf{k}-\mathbf{k}'|)E(\mathbf{k}')}{|\mathbf{k}-\mathbf{k}'|^4 k'^2}[|\mathbf{k} - \mathbf{k}'|^2 + (|\mathbf{k}|\cos\theta - |\mathbf{k}'|)^2]d^3\mathbf{k}', \tag{28}$$

where θ is the angle between $\mathbf{k}$ and $\mathbf{k}'$. If we apply spherical coordinates, $d^3\mathbf{k}'=2\pi k'^2\cos\theta dk'$, we can write

$$\frac{F_{pps}(\mathbf{k})}{\rho^2} = U^2 F_{1,1}(\mathbf{k}) + \frac{1}{16\pi} \times \left[\int_0^\infty\int_{-1}^1 \frac{E(|\mathbf{k}-\mathbf{k}'|)E(\mathbf{k}')(k^2+2k'^2)}{|\mathbf{k}-\mathbf{k}'|^4}dxdk' - \int_0^\infty\int_{-1}^1 \frac{E(|\mathbf{k}-\mathbf{k}'|)E(\mathbf{k}')4kk'x}{|\mathbf{k}-\mathbf{k}'|^4}dxdk' + \int_0^\infty\int_{-1}^1 \frac{E(|\mathbf{k}-\mathbf{k}'|)E(\mathbf{k}')k^2x^2}{|\mathbf{k}-\mathbf{k}'|^4}dxdk'\right], \tag{29}$$

where $x=\cos\theta$.

Equation (10) is used for the energy spectrum, and the variable a defined as in Eq. (16), and y is defined as $y=k'/k$. Then F_{pps} can be written as the sum of $U^2F_{1,1}$ and three double integrals in x and y. After integrating out the angular dependence:

$$\frac{F_{pps}(\mathbf{k})}{\rho^2} = U^2 F_{1,1}(\mathbf{k}) + 2^{-29/6}\left(\frac{55}{18}C\right)^2\frac{\lambda^{7/3}k^{4/3}}{\pi} \times \left[\int_0^\infty \frac{y^{7/6}}{4[1+(k\lambda)^2y^2]^{17/6}}[I_1(a)+I_3(a)]dy + \int_0^\infty \frac{y^{19/6}}{2[1+(k\lambda)^2y^2]^{17/6}}I_1(a)dy - \int_0^\infty \frac{y^{13/6}}{[1+(k\lambda)^2y^2]^{17/6}}I_2(a)dy\right], \tag{30}$$

where

$$I_1(a) = -\frac{6}{11}[(a+1)^{-11/6} - (a-1)^{-11/6}], \tag{31}$$

$$I_2(a) = \frac{6}{5}[(a+1)^{-5/6} - (a-1)^{-5/6}] + a\left[-\frac{6}{11}[(a+1)^{-11/6} - (a-1)^{-11/6}]\right], \tag{32}$$

$$I_3(a) = a^2\left[-\frac{6}{11}[(a+1)^{-11/6} - (a-1)^{-11/6}]\right] - 2a\left[-\frac{6}{5}[(a+1)^{-5/6} - (a-1)^{-5/6}]\right] + 6[(a+1)^{1/6} - (a-1)^{1/6}]. \tag{33}$$

To compare to measurements, an expression for the one-dimensional spectrum must be generated from the three-dimensional spectrum F_{pps}. This is accomplished by integrating over the k_2 and k_3 variables,

$$F_{pps}^1(k_1) = \iint_{-\infty}^{\infty} F_{pps}(\mathbf{k})dk_2dk_3. \tag{34}$$

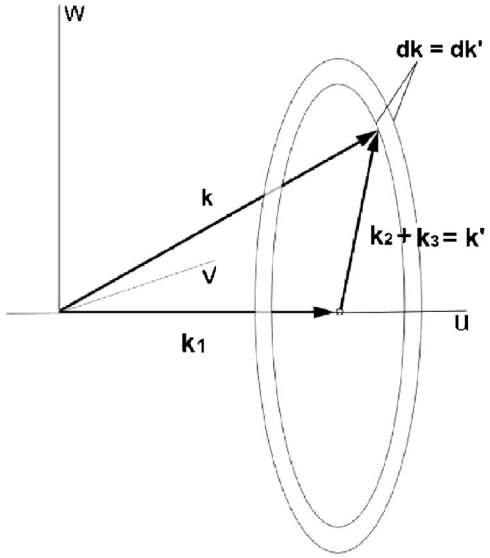


FIG. 2. The geometry used to simplify the integral in Eq. (30). Integration in the k_2, k_3 plane is replaced by integration over k from k_1 to ∞ for each value of k_1.

The double integral over the k_2, k_3 plane can be rewritten as a single integral by considering the geometry shown in Fig. 2. Since the turbulence is isotropic, the integral in the fixed k_1 plane can be done in polar integrals. Examining Fig. 2 shows

$$k^2 = k_1^2 + k'^2, \tag{35}$$

and

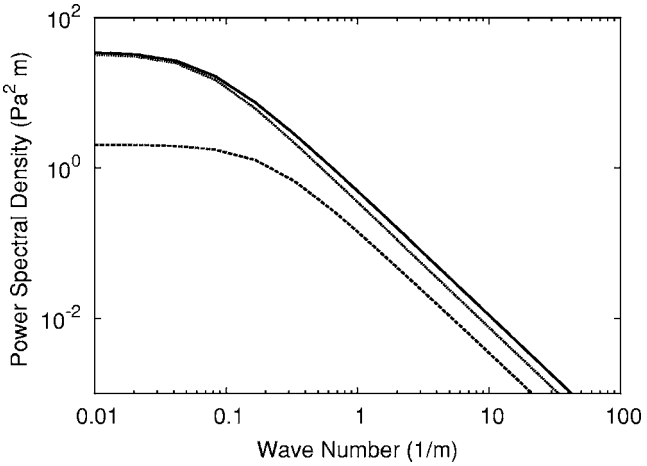


FIG. 3. Comparison of terms in the stagnation equation [Eq. (39)]. The dotted line is the first term, the dashed line is the second term, and the solid line is their sum. For small wave numbers, the stagnation spectrum can be estimated using the (much easier to calculate) first term.

$$2k'dk' = 2kdk, \tag{36}$$

where k is the total wave number vector magnitude, and k' is the polar magnitude. It follows that the integration over the plane can be written as

$$F_{pps}^1(k_1) = 2\pi \int_{k_1}^{\infty} F_{pps}(\mathbf{k}) k dk, \tag{37}$$

The stagnation interaction spectrum is now given by

$$F_{pps}^1(k_1) = \rho^2 U^2 F_{1,1}(k_1) + \left(\frac{55}{18}\rho C\right)^2 \frac{\lambda^{7/3}}{2^{29/6}} \int_{k_1}^{\infty} k^{7/3} dk \int_0^{\infty} \frac{y^{7/6}(I_1(a) + I_3(a)) + 2y^{19/6} I_1(a) - 4y^{13/6} I_2(a)}{[1 + (k\lambda)^2 y^2]^{17/6}} dy, \tag{38}$$

where $F_{1,1}(k_1)$is the one-sided measured velocity spectrum in the mean flow direction, and $F_{pps}^1(k_1)$ is the one-sided stagnation pressure spectrum in the direction of flow.

The calculation of the second term in Eq. (38) can be simplified by performing a change of variables as described in Sec. II B:

$$F_{pps}^1(k_1) = \rho^2 U^2 F_{1,1}(k_1) + \left(\frac{55}{18}\rho C\right)^2 \frac{1}{2^{29/6}\lambda} \int_{\chi_1}^{\infty} \chi^{7/3} dx \int_0^{\infty} \frac{y^{7/6}(I_1(a) + I_3(a)) + 2y^{19/6} I_1(a) - 4y^{13/6} I_2(a)}{[1 + \chi^2 y^2]^{17/6}} dy. \tag{39}$$

As before, the integrals in Eq. (39) can be calculated once and used for each measurement set by scaling with C and λ.

The result of Eq. (39) is shown in Fig. 3 along with the first and second terms shown separately. In the source region below 0.1 m^{-1}, the much simpler $\rho U u$ term is an accurate estimation of the pressure spectrum.

As a check, the turbulence–turbulence prediction and the stagnation pressure prediction were compared to the corresponding inertial range predictions from Ref. 1. Figure 4 shows that the two sets of predictions agree very well in the inertial range, and diverge as expected in the source region.

III. PREDICTIONS AND FITS

Since the predictions all have the common form of power law behavior at high and low wave number with a transition scale corresponding to λ of Eq. (3), they all can be fit with the forms analogous to Eq. (3). We provide the fits for discussion of the nature of the solutions and for use by others. In this section we have set $\rho = 1.2$ kg/m^3.

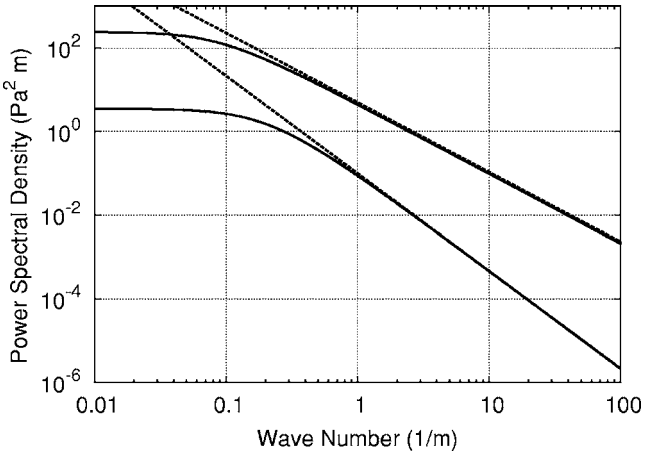


FIG. 4. Comparison of predictions for stagnation pressures and turbulence–turbulence interaction pressures from Ref. 1 (dotted lines) with corresponding predictions from this paper (solid lines).

A. Mean shear–turbulence

$$F^1_{ppm}(k_1) = 7.380CK^2 \frac{\lambda^2 (k_1\lambda)^{5/3}}{[1 + 1.622(k_1\lambda)^2]^{8/3}}. \quad (40)$$

At low wave numbers the mean shear–turbulence interaction spectrum increases as $k^{5/3}$, at high wave numbers it decays as $k^{-11/3}$. The equivalent scale factor determining the transition range is

$$\lambda_M = 1.273\lambda. \quad (41)$$

The peak of the mean shear–turbulence spectrum occurs at slightly lower wave numbers than the transition from the source to the inertial range of the velocity spectrum.

B. Turbulence–turbulence

$$F^1_{ppt}(k_1) = 0.811 \frac{C^2}{\lambda} \frac{1}{[1 + 0.1792(k_1\lambda)^2]^{7/6}}. \quad (42)$$

The turbulence–turbulence pressure spectrum is constant at low wave numbers and decays as $k^{-7/3}$ power at high wave numbers. The transition scale is given by

$$\lambda_T = 0.4233\lambda, \quad (43)$$

so the transition to the source region occurs at significantly higher wave numbers than for the velocity spectrum.

C. Stagnation pressure

$$F^1_{pps}(k_1) = \frac{1.44U^2C}{[1 + (k_1\lambda)^2]^{5/6}} + \frac{1.451C^2}{\lambda[1 + 0.1129(k_1\lambda)^2]^{5/6}}. \quad (44)$$

The first term is just $1.44U^2$ times the velocity spectrum and follows the behavior of the velocity spectrum. The second term has a constant source region at low wave number and a $k^{-5/3}$ decay at high wave number. The transition in the second term is determined by the scale

$$\lambda_S = 0.3360\lambda, \quad (45)$$

so the transition from the source region to the inertial region occurs at a higher wave number than the velocity transition.

For our measurements near the earth's surface in unstable conditions, the second term in the stagnation pressure only contributes in the inertial range, but its contribution is significant in that region.

D. Relative values of the terms

For our measurements, the mean shear–turbulence interaction term is predicted to be negligible except for a peak in the transition region. Under less turbulent conditions the coefficient C could be much smaller and the predicted mean shear–turbulence interaction pressure could become dominant over a larger range. The leading term in the stagnation pressure would also increase in relation to the other terms under less turbulent, higher average wind conditions.

IV. MEASUREMENT AND ANALYSIS

The experimental data presented here were taken at Clegg Field—Oxford University Airport in Oxford, MS.

Wind velocity measurements were taken with a Gill Instruments R3A-100 Ultrasonic Research Anemometer mounted approximately 1.0 m from the ground. This anemometer has an internal sampling rate of 100 Hz and can measure down to 0 Hz. It outputs data in the form of voltages proportional to the x, y, and z components of the wind velocity.

Four wind-screens were used in the measurements: two spherical open cell polyurethane foam screens with diameters 9 and 18 cm, and two fiberglass screens with diameters 60 and 100 cm. The fiberglass wind-screens were constructed by removing the paper backing from home insulation type fiberglass and rolling it into balls. They were approximately spherical. The acoustic transmission characteristics of the fiberglass screens were measured and the screens were determined to be acoustically transparent up to at least 100 Hz.

Pressure measurements were taken at 1.0 m above ground using Brüel & Kjær type 4193 1/2 in. microphones powered by Nexus brand conditioning amplifiers. The frequency responses for these microphones drop off below 0.07 Hz, but the low frequency cutoff of the pressure data is 0.1 Hz, which is set by the high pass filter in the Nexus. All data reported herein were taken with the microphones oriented vertically.

All of the sensors were connected to a National Instruments AD/DA data acquisition card controlled by a program written in LABVIEW. Unless otherwise noted, each data run lasted 900 s and was taken at a sample rate of 200 samples/s. After acquisition, all data analysis was done within MATLAB.

Power spectra were generated in the same manner as in Ref. 1. Each data set was broken into nonoverlapping blocks. Each block was detrended and windowed before its power spectral density (PSD) was calculated. Finally the average of the block PSDs was calculated and converted from frequency to wave number space. Pressure spectra were broken into blocks of size 4096 points as a compromise between good averaging and good resolution, while the velocity spec-

TABLE I. Screen size, U, U_{rms} and fit parameters λ and C for various graphs.

Figures	Screen diameters	U(m/s)	U_{rms}	λ	C
1(a), 6	Grid and 0.09 m	4.35	1.40	11.96	7.16
1(b), 7	Grid and 0.18 m	2.65	1.08	5.38	1.32
1(c), 8	Grid and 1.0 m	2.98	1.49	9.33	8.96
1(d), 9	0.60 and 1.0 m	4.41	1.61	11.30	7.60

tra were generated with all 180,000 points with no averaging in order to retain the low wave number components.

The windowing of the data blocks was done using a Hamming window. This choice is in accordance with the recommendation by Kaimal and Kristensen,[14] who noted that the spectra derived from short, untapered time series consistently overestimated the power spectra generated from much longer samples. This overestimation indicates that an appropriate windowing function should be employed, especially when the length of the time series is small compared to the period of the oscillations being measured. Furthermore, it is important to choose a windowing function with spectral slope greater than the spectral slope of the data. Doing otherwise leads to an overestimation of the power spectrum at low frequencies. Since the slopes of the velocity and stagnation spectra are −5/3 in the inertial subrange, the Hamming window, with a spectral slope of −2, was shown by Kaimal and Kristensen to be an appropriate choice when analyzing outdoor wind spectra. Differences in the spectra with and without the use of the Hamming window were minor.

The velocity data from each measurement were fit using a least chi-squared approach to Eq. (3) with λ and C as fit parameters. Since the spectra obey a −5/3 power law in the inertial range, the values at high wave numbers can be six orders of magnitude smaller than those at low wave numbers. In this paper, each data point was multiplied by its corresponding wave number k. The adjusted data were then fit to Eq. (3), also multiplied by k:

$$kF_{11}^{1}(k) = \frac{Ck}{[1 + (k\lambda)^2]^{5/6}}. \tag{46}$$

When the fits are plotted on the full frequency range of the data, the spectral fluctuations at high wave numbers make it difficult to determine how well the fits match the data. If the spectra are averaged over shorter blocks of data, agreement between the data and the inertial range fit is achieved. In meteorology, $kF(k)$ is often plotted with a log–log scale in order to emphasize the peak contribution to $\langle U_{\text{rms}}^2 \rangle$.

Figure 1 shows example plots of the velocity spectra taken simultaneously with the pressure data plotted below. No averaging was performed in order to retain the low frequency portion of the spectrum. Table I lists the overall mean wind speed along with the rms values of the turbulent fluctuations for each run.

Initially, stagnation measurements with unscreened microphones were performed with the protective grid removed. This resulted in measured pressure spectra which were considerably greater than the stagnation pressure. Figure 5 shows the spectra measured simultaneously by a gridded and a bare microphone along with the predicted stagnation spectrum. As discussed in Sec. III C, the stagnation pressure prediction neglects the contribution due to the term $\nabla \cdot [\mathbf{V} \times \nabla \times \mathbf{V}]$. In particular, the interaction of turbulence with thin shear layers formed on the microphone diaphragm is the probable source of the large wind noise contribution of the bare microphone. The grid displaces shear layers away from the sensing element. This interaction is similar to the mean shear–turbulence interaction pressure, but involves different assumptions than used to develop Eq. (11). For this reason all unscreened microphone measurements are done with the microphone oriented vertically with the protective grid on. The hypothesis of Ref. 1 that the wind noise on a bluff body or bare microphone is well approximated by the calculated stagnation pressure must be changed to apply only to gridded microphones.

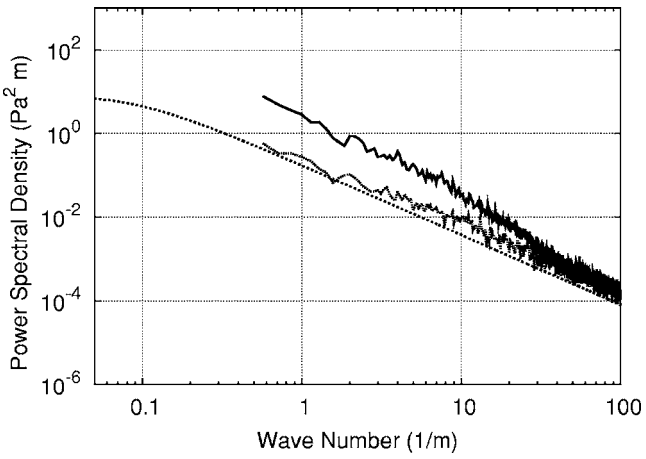


FIG. 5. Microphone with grid attached (small dots) compared to without grid (solid line). This clearly shows how the pressures measured by an absolutely bare microphone overshoot the predicted stagnation pressures. The theory line is the predicted stagnation spectrum.

Informal comparisons of the wind noise levels of gridded microphones pointed perpendicular to the wind and into the wind displayed only small differences. The same comparison with bare microphones produced significant differences in the levels.

A. Unscreened, gridded microphone data: Stagnation pressure calculation

All of the unscreened gridded microphone data presented in Figs. 6–8 are well represented by the stagnation pressure calculation. The slopes are similar and all data are roughly within 3 dB of the prediction. The acoustic measurements do not extend far enough into the source region to determine if this contribution follows the predicted curve at very low frequencies.

B. Wind-screened microphone data

In Figs. 6–9 we display the wind noise measured in wind-screens 9, 18, 60, and 100 cm in diameter. The wind noise spectral level for the 9 cm screen (Fig. 6) is significantly less than the unscreened gridded microphone but significantly above the turbulence–turbulence or mean shear–turbulence levels. Doubling the size of the wind-screen to

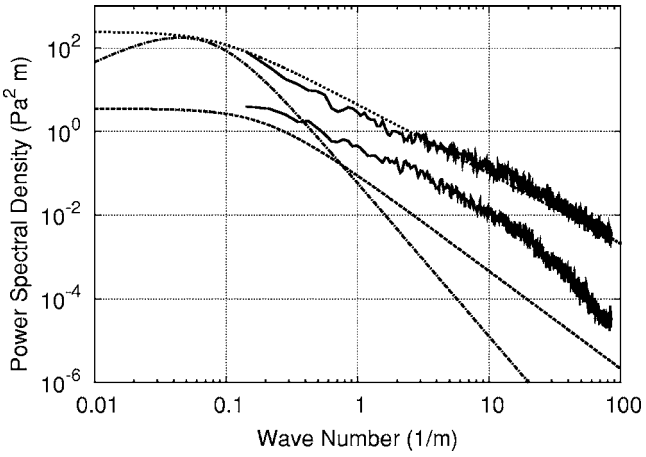


FIG. 6. Gridded mike and 9 cm foam wind-screen with wind noise contributions. Theory lines are: Stagnation (dotted lines), turbulence–turbulence (dashed line), and mean shear (dash-dot lines).

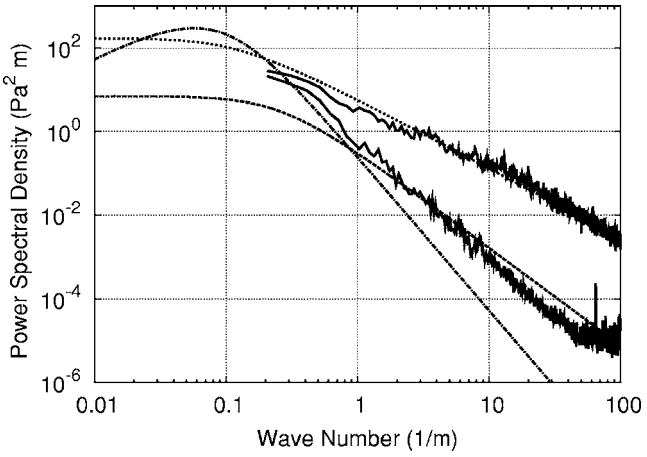


FIG. 8. Gridded mike and 1 m fiberglass wind-screen with wind noise contributions. Theory lines are: Stagnation (dotted lines), turbulence–turbulence (dashed line), and mean shear (dash-dot lines).

18 cm (Fig. 7) produces even more reduction of the wind noise levels, consistent with the model of area averaging of the locally generated pressure fluctuations. The measured levels are still much larger than the predicted intrinsic contributions. Both of these screens display a transition from lower spectral slope at low frequencies, to steeper spectral slope at higher frequencies as observed by van den Berg[4] and Morgan.[6]

To investigate whether an even larger wind-screen could reduce the surface generated pressure fluctuations to the level of the intrinsic turbulence contributions, a 100-cm-diam, nominally spherical wind-screen, was constructed of loose fiberglass and tested. The results of this measurement are displayed in Fig. 8. The 100 cm wind-screen reduces the total wind noise measured to levels comparable to the predictions for the sum of the intrinsic pressure fluctuation contributions. In addition, the behavior of the spectrum differs from that of the smaller wind-screens. The spectrum displays an initial high slope at very low frequencies, then displays a slope slightly steeper than the turbulence–turbulence prediction without a further transition. This result is consistent with the idea that a large enough wind-screen can reduce the surface generated local pressure fluctuations to the point that the intrinsic terms begin to contribute to the total measured wind noise. A 100 cm wind-screen is large enough so that some averaging of even the larger scale intrinsic pressure fluctuation is likely, and this could be contributing to the increased attenuation at higher frequencies.

A 60 cm wind-screen was constructed of fiberglass and compared to the 100 cm wind-screen (Fig. 9). The wind noise levels measured in the 60 cm screen are slightly higher than those measured in the 100 cm screen and again, the slope is slightly steeper than the turbulence–turbulence interaction spectrum.

These results are not definitive in identifying the measured pressure fluctuations in the 100 and 60 cm wind-screens as the turbulence–turbulence interaction pressure but do confirm that the stagnation pressure calculation and the turbulence–turbulence interaction pressure prediction do represent upper and lower bounds for wind noise measurements in outdoor flow for compact wind-screens.

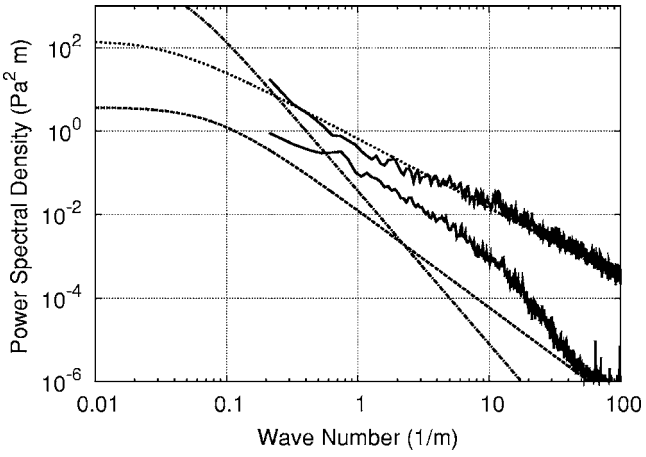


FIG. 7. Gridded mike and 18 cm foam wind-screen with wind noise contributions. Theory lines are: Stagnation (dotted lines), turbulence–turbulence (dashed line), and mean shear (dash-dot lines).

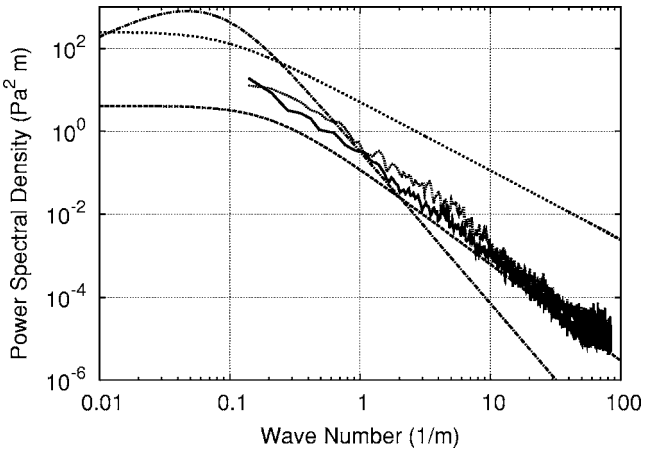


FIG. 9. Comparison of 60 cm wind-screen (small dots) and 100 cm wind-screen (solid). Theory lines are: Stagnation (dotted lines), turbulence–turbulence (dashed line), and mean shear (dash-dot lines). The two spectra are nearly indistinguishable.

C. Mean shear–turbulence interaction pressures

The predicted mean shear–turbulence interaction pressures fluctuation level is larger than or similar to the stagnation pressure spectrum at low wave numbers ($k<0.2$ m^{-1}) and larger than the turbulence–turbulence interaction pressure spectrum at moderately low wave numbers ($k<1.0$ m^{-1}). The predicted levels display a peak at low wave numbers, then decay proportional to k^2. The mean shear–turbulence prediction is large at low wave numbers but is not orders of magnitude larger than the turbulence–turbulence contributions in the source range as extrapolations of the inertial range results indicate. The measured levels for the gridded and wind-screened microphones do not consistently indicate the presence of this predicted contribution. The small improvement in the prediction by the use of the Businger–Dyer formula for the velocity profile does not significantly affect this conclusion. We do note that the statistics at these very low wave numbers are poor. Measurements on an acoustically meaningful time scale (15 min) are extremely short on a meteorological time scale, but the shorter measurement periods are necessary to reflect the changing acoustic environment.

Kraichnan[15] has predicted that curvature of the gradient of the average flow velocity will reduce the pressure fluctuations observed on a flat surface below a turbulent boundary layer. Perhaps the curvature of the velocity gradient plays a role in reducing the mean shear–turbulence pressure contribution.

V. CONCLUSIONS

The primary goal of this paper was to extend the turbulence–turbulence and stagnation pressure predictions from the measured velocity fluctuation spectra developed in Ref. 1 into the source range of atmospheric wind turbulence. This was accomplished by applying a turbulence–turbulence interaction theory developed by George *et al.*[3] for laboratory measurements to measurements made outdoors, and by developing a prediction for stagnation pressure levels measured by an unscreened, gridded sensor. The turbulence–turbulence interaction prediction agrees well with measurements made within large (60 and 100 cm) wind-screens. The stagnation prediction agrees well with measurements made by a gridded, unscreened microphone.

An additional goal was to incorporate the mean shear–turbulence interaction pressures predicted by George *et al.* into the intrinsic pressure prediction for outdoor measurements. Our predictions indicated that this interaction should generate significant pressures within the source region and that the pressure contributions are of the same order of magnitude as the turbulence–turbulence interaction pressure in the source region. However, no consistent evidence of this interaction pressure was identified in our pressure fluctuation measurements.

ACKNOWLEDGMENTS

We wish to thank Keith Wilson, Gerald Lauchle, and G. P. van den Berg for valuable discussions and for sharing their ideas about wind noise and wind-screening.

[1] R. Raspet, J. Webster, and K. Dillion, "Framework for wind noise studies," J. Acoust. Soc. Am. **119**, 834–843 (2006).
[2] G. K. Batchelor, "Pressure fluctuations in isotropic turbulence," Proc. Cambridge Philos. Soc. **47**, 359–374 (1951).
[3] W. K. George, P. D. Beuther, and R. E. A. Arndt, "Pressure spectra in turbulent free shear flows," J. Fluid Mech. **148**, 155–191 (1984).
[4] G. P. van den Berg, "Wind-induced noise in a screened microphone," J. Acoust. Soc. Am. **119**, 824–833 (2005).
[5] Z. C. Zheng and B. K. Tan, "Reynolds number effects on flow/acoustic mechanisms in spherical windscreens," J. Acoust. Soc. Am. **113**, 161–166 (2003).
[6] S. Morgan, "An investigation of the sources and attenuation of wind noise in measurement microphones," Ph.D. thesis, University of Mississippi, University, MS, 1993.
[7] J. C. Kaimal, "Turbulence spectra, length scales and structure parameters in the stable surface layer," Boundary-Layer Meteorol. **4**, 289–309 (1973).
[8] J. Højstrup, "Velocity spectra in the unstable planetary boundary layer," J. Atmos. Sci. **39**, 2239–2248 (1982).
[9] J. Højstrup, "A simple model for the adjustment of velocity spectra in unstable conditions downstream of an abrupt change in roughness and heat flux," Boundary-Layer Meteorol. **21**, 341–356 (1981).
[10] R. H. Kraichnan, "Pressure field within homogeneous anisotropic turbulence," J. Acoust. Soc. Am. **28**, 64–72 (1955).
[11] H. A. Panofsky and J. A. Dutton, *Atmospheric Turbulence, Models and Methods for Engineering Applications* (Wiley, New York, 1984).
[12] K. Dillion, "An investigation of wind noise over a flat plate," Master's thesis, University of Mississippi, University, MS, 2005.
[13] A. L. Fetter and J. Walecka, *Theoretical Mechanics of Particles and Continua* (McGraw Hill, New York, 1980).
[14] J. C. Kaimal and L. Kristensen, "Time series tapering for short data samples," Boundary-Layer Meteorol. **57**, 187–194 (1991).
[15] R. H. Kraichnan, "Pressure fluctuations in turbulent flow over a flat plate," J. Acoust. Soc. Am. **28**, 378–390 (1955).

Range-dependent waveguide scattering model calibrated for bottom reverberation in a continental shelf environment

Ameya Galinde, Ninos Donabed, and Mark Andrews
Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115

Sunwoong Lee and Nicholas C. Makris
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Purnima Ratilal
Department of Electrical and Computer Engineering, Northeastern University, Boston, Massachusetts 02115

(Received 3 September 2007; revised 13 December 2007; accepted 14 December 2007)

An analytic model is developed for scattering from random inhomogeneities in range-dependent ocean waveguides using the Rayleigh–Born approximation to Green's theorem. The expected scattered intensity depends on statistical moments of fractional changes in compressibility and density, which scatter as monopoles and dipoles, respectively, and the coherence volume of the inhomogeneities. The model is calibrated for ocean bottom scattering using data acquired by instantaneous wide-area ocean acoustic waveguide remote sensing (OAWRS) and geophysical surveys of the ONR Geoclutter Program. The scattering strength of the seafloor on the New Jersey shelf, a typical continental shelf environment, is found to depend on wave number k, medium coherence volume V_c, and seabed depth penetration factor F_p following a $10\log_{10}(F_pV_ck^4)$ dependence. A computationally efficient numerical approach is developed to rapidly compute bottom reverberation over wide areas using the parabolic equation by exploiting correlation between monopole and dipole scattering terms and introducing seafloor depth penetration factors. An approach is also developed for distinguishing moving clutter from statistically stationary background reverberation by tracking temporal and spatial fluctuations in OAWRS intensity images. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2832509]

PACS number(s): 43.30.Gv, 43.30.Pc, 43.30.Ft [RCG] Pages: 1270–1281

I. INTRODUCTION

In this paper, we develop an analytic model with numerical implementation for reverberation in range-dependent ocean environments and bistatic source–receiver geometries. The model is derived from first principles using the acoustic wave equation for inhomogeneous media and by application of Green's theorem.

The model takes into account the full three-dimensional (3D) scattering interaction of the acoustic wavefield with volume or surface inhomogeneities. For efficiency, the model is implemented in terms of scattering from the spatially varying resolution footprint of the sonar, typically determined by beamforming and temporal matched filtering. Statistical moments of the randomly scattered field are expressed in terms of statistical moments of fractional changes in medium compressibility and density, and the waveguide Green function. Variations in compressibility and density lead to monopole and dipole terms, respectively. As the waveguide Green function is employed, the model directly incorporates attenuation in the sea bottom and scattering of evanescent waves by inhomogeneities.

The model is calibrated for ocean bottom reverberation with data acquired by instantaneous ocean acoustic waveguide remote sensing (OAWRS)[1,2] during the 2003 Main Acoustic Experiment and geophysical surveys of the ONR Geoclutter Program. The data were collected in the New Jersey Strataform,[3] south of Long Island, NY, a typical continental shelf environment with minimal bathymetric relief. The model is integrated to bathymetric databases to enable charting of predicted reverberation in geographic space. An approach is also developed for distinguishing moving clutter from statistically stationary background reverberation by tracking temporal and spatial fluctuations in OAWRS intensity images.

Analysis with the theory and data indicates that (a) both random fluctuations in seabed compressibility and density are important contributors to the scattered intensity and consequently reverberation, (b) the scattering strength of the seafloor on the New Jersey shelf, a typical continental shelf environment, is found to depend on wave number k, medium coherence volume V_c, and seabed depth penetration factor F_p following a $10\log_{10}(F_pV_ck^4)$ dependence, where each term is a function of frequency.

A computationally efficient numerical approach is developed here to rapidly compute bottom reverberation over wide areas using the parabolic equation[4] by exploiting correlation between monopole and dipole scattering terms and introducing seafloor depth penetration factors. Without innovations of this kind, bottom reverberation models based on

 0001-4966/2008/123(3)/1270/12/$23.00

volume scattering can be computationally intensive in range-dependent environments with bistatic source–receiver geometries because dipole scattering involves multidimensional spatial derivatives of the *complex* waveguide Green function. These typically must be computed numerically using large and dense 3D matrices.

A number of half-space models for seafloor scattering have used the Rayleigh–Born approach for local waterborne backscattering investigations.[5–8] Some monostatic range-independent models have used approximations similar to Rayleigh–Born.[8–11] Previous range-dependent waveguide reverberation models using the parabolic equation have only included monopole scattering with empirical models.[12–14]

II. ANALYTIC FORMULATION

Here we develop an analytic model for the scattered field from random medium inhomogeneities by application of Green's theorem. Assume the medium compressibility and density vary randomly in space, following a stationary random process within the local sonar resolution footprint.

Let the origin of the coordinate system be placed at the air–water interface with the positive z-axis pointing downward. Let the coordinates be defined by $\mathbf{r}_0=(x_0,y_0,z_0)$ for the source, $\mathbf{r}=(x,y,z)$ for the receiver, and $\mathbf{r}_t=(x_t,y_t,z_t)$ for the centroid of an inhomogeneity. Spatial cylindrical (ρ,ϕ,z) and spherical systems (r,θ,ϕ) are defined by $x=r\sin\theta\cos\phi$, $y=r\sin\theta\sin\phi$, $z=r\cos\theta$ and $\rho^2=x^2+y^2$. The acoustic wave number k is given by the angular frequency $\omega=2\pi f$ divided by the sound speed c.

A. Time harmonic scattered field moments

In the presence of volume inhomogeneities, the time-harmonic acoustic field $\Phi_t(\mathbf{r}_t,f)$ satisfies an inhomogeneous Helmholtz equation,[15]

$$\nabla^2\Phi_t(\mathbf{r}_t,f)+k^2\Phi_t(\mathbf{r}_t,f) = -k^2\Gamma_\kappa(\mathbf{r}_t)\Phi_t(\mathbf{r}_t,f)-\nabla\cdot[\Gamma_d(\mathbf{r}_t)\nabla\Phi_t(\mathbf{r}_t,f)], \quad (1)$$

where Γ_κ is the fractional change in compressibility,

$$\Gamma_\kappa(\mathbf{r}_t)=\frac{\kappa(\mathbf{r}_t)-\bar{\kappa}}{\bar{\kappa}}, \quad (2)$$

given compressibility $\kappa=1/dc^2$, and Γ_d is fractional change in density,

$$\Gamma_d(\mathbf{r}_t)=\frac{d(\mathbf{r}_t)-\bar{d}}{d(\mathbf{r}_t)}, \quad (3)$$

where $\bar{\kappa}$ and $\bar{d}$ are the mean compressibility and density in the region, respectively.

Then, by application of Green's theorem, given a source at $\mathbf{r}_0$ and a receiver at $\mathbf{r}$, the time-harmonic scattered field $\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)$ from inhomogeneities within the sonar resolution footprint V_s centered at $\mathbf{r}_s$ is[15]

$$\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)=\int\int\int_{V_s}[k^2\Gamma_\kappa(\mathbf{r}_t)\Phi_t(\mathbf{r}_t,f)G(\mathbf{r}|\mathbf{r}_t,f) + \Gamma_d(\mathbf{r}_t)\nabla\Phi_t(\mathbf{r}_t,f)\cdot\nabla G(\mathbf{r}|\mathbf{r}_t,f)]dV_t. \quad (4)$$

Here, $\Phi_t(\mathbf{r}_t,f)$ is the total field at the location of the inhomogeneity, and $G(\mathbf{r}|\mathbf{r}_t,f)$ is the medium's Green function. Elemental spatial variations in compressibility lead to monopole scattering, while those in density lead to dipole scattering. Density fluctuations contribute to both monopole and dipole scattering while sound speed fluctuations contribute only to monopole scattering.[15] The dipole term depends on spatial gradients of the total field and those of the Green function.

The total field at any location is a sum of the incident and the locally scattered fields,

$$\Phi_t(\mathbf{r}_t,f)=\Phi_i(\mathbf{r}_t|\mathbf{r}_0,f)+\Phi_s(\mathbf{r}_t,f), \quad (5)$$

where $\Phi_i(\mathbf{r}_t|\mathbf{r}_0,f)=(4\pi)^2G(\mathbf{r}_t|\mathbf{r}_0,f)$. The $(4\pi)^2$ normalization factor yields the convenient source level of 0 dB re 1 μPa at 1 m. For small local perturbations in medium compressibility and density, the total field at the inhomogeneity can be approximated by the incident field in Eqs. (5) and (4). This is the first-order Rayleigh–Born approximation to Green's theorem. The randomly scattered field at the receiver can then be expressed as

$$\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f) = 4\pi^2\int\int\int_{V_s}[k^2\Gamma_\kappa(\mathbf{r}_t)G(\mathbf{r}_t|\mathbf{r}_0,f)G(\mathbf{r}|\mathbf{r}_t,f) + \Gamma_d(\mathbf{r}_t)\nabla G(\mathbf{r}_t|\mathbf{r}_0)\cdot\nabla G(\mathbf{r}|\mathbf{r}_t,f)]dV_t, \quad (6)$$

which is a single-scatter approximation.

The mean scattered field from the inhomogeneities is

$$\langle\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)\rangle = (4\pi)^2\int\int\int_{V_s}[k^2\langle\Gamma_\kappa(\mathbf{r}_t)\rangle\langle G(\mathbf{r}_t|\mathbf{r}_0,f)G(\mathbf{r}|\mathbf{r}_t,f)\rangle + \langle\Gamma_d(\mathbf{r}_t)\rangle\langle\nabla G(\mathbf{r}_t|\mathbf{r}_0,f)\cdot\nabla G(\mathbf{r}|\mathbf{r}_t,f)\rangle]dV_t. \quad (7)$$

Random variables Γ_κ and Γ_d, are assumed to be independent of possible random fluctuations in the waveguide Green function.[16]

The second moment of the scattered field is

$$\begin{aligned}\langle|\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)|^2\rangle &= \langle\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)\Phi_s^*(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)\rangle \\ &= \langle(4\pi)^4\int\int\int_{V_s}[k^2\Gamma_\kappa(\mathbf{r}_t)G(\mathbf{r}_t|\mathbf{r}_0,f)G(\mathbf{r}|\mathbf{r}_t,f) \\ &\quad + \Gamma_d(\mathbf{r}_t)\nabla G(\mathbf{r}_t|\mathbf{r}_0,f)\cdot\nabla G(\mathbf{r}|\mathbf{r}_t,f)]dV_t \\ &\quad \times\int\int\int_{V_s'}[k^2\Gamma_\kappa(\mathbf{r}_t')G^*(\mathbf{r}_t'|\mathbf{r}_0,t)G^*(\mathbf{r}|\mathbf{r}_t',f) \\ &\quad + \Gamma_d(\mathbf{r}_t')\nabla G^*(\mathbf{r}_t'|\mathbf{r}_0,f)\cdot\nabla G^*(\mathbf{r}|\mathbf{r}_t',f)]dV_t'\rangle. \end{aligned} \quad (8)$$

Equation (8) requires evaluation at two spatial locations $\mathbf{r}_t$

and $\mathbf{r}_t'$, which requires knowledge of the spatial correlation function of volume inhomogeneities. A variety of different forms have been assumed for this correlation function in the case of seabed inhomogenieties. They include Gaussian,[8] exponential,[17] as well as power law[7,8,10] correlation functions.

To capture the essential physics in a manner conducive to analysis, we assume that inhomogeneities are uncorrelated when the two points lie outside the coherence volume V_c and are fully correlated within this volume. This leads, for example, to

$$\begin{aligned}\langle \Gamma_k(\mathbf{r}_t)\Gamma_k(\mathbf{r}_t')\rangle &= V_c(\mathbf{r}_s,z_t)[\langle \Gamma_k^2(\mathbf{r}_t)\rangle - |\langle \Gamma_k(\mathbf{r}_t)\rangle|^2]\delta(\mathbf{r}_t-\mathbf{r}_t') \\ &\quad + \langle \Gamma_k(\mathbf{r}_t)\rangle\langle \Gamma_k(\mathbf{r}_t')\rangle \\ &= V_c(\mathbf{r}_s,z_t)\mathrm{Var}(\Gamma_k(\mathbf{r}_t))\,\delta(\mathbf{r}_t-\mathbf{r}_t') + \langle \Gamma_k(\mathbf{r}_t)\rangle\langle \Gamma_k(\mathbf{r}_t')\rangle. \end{aligned} \tag{9}$$

Such delta function correlations are often used in theories of wave propagation in random media.[16,18] The coherence volume can be a function of horizontal position and depth. For example, overburden pressure in sediments leads to greater compaction with depth which affects coherence volume.[19]

Substituting equations similar to Eq. (9) into Eq. (8), leads to the second moment of the scattered field,

$$\begin{aligned}\langle |\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)|^2\rangle &= |\langle \Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)\rangle|^2 + \mathrm{Var}(\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)), \end{aligned} \tag{10}$$

proportional to the total intensity, which is a sum of coherent and incoherent components. The relative importance of each of these terms within the sonar resolution footprint can be estimated by analysis similar to that in Appendix A of Ref. 16.

The coherent term is the square of the mean scattered field, the square of Eq. (7). The incoherent term is the variance of the scattered field,

$$\begin{aligned}&\mathrm{Var}(\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)) \\ &= (4\pi)^4 \int\int\int_{V_s} V_c(\mathbf{r}_s,z_t) \\ &\quad \times [k^4\mathrm{Var}(\Gamma_k(\mathbf{r}_t))\langle |G(\mathbf{r}_t|\mathbf{r}_0,f)|^2|G(\mathbf{r}|\mathbf{r}_t,f)|^2\rangle \\ &\quad + \mathrm{Var}(\Gamma_d(\mathbf{r}_t))\langle |\nabla G(\mathbf{r}_t|\mathbf{r}_0,f)\cdot\nabla G(\mathbf{r}|\mathbf{r}_t,f)|^2\rangle \\ &\quad + k^2\mathrm{Cov}(\Gamma_k,\Gamma_d)\langle 2\Re\{G(\mathbf{r}_t|\mathbf{r}_0,f)G(\mathbf{r}|\mathbf{r}_t,f) \\ &\quad \times \nabla G^*(\mathbf{r}_t|\mathbf{r}_0,f)\cdot\nabla G^*(\mathbf{r}|\mathbf{r}_t,f)\}\rangle]dV_t. \end{aligned} \tag{11}$$

Measured reverberation from the sea bottom is often dominated by the incoherent term of Eq. (10). Exceptions occur, for example, when many scatterers with large impedence contrast, such as layers of gravel or rock on a sandy seafloor, fall within the sonar resolution footprint.

The approach presented may also be applied to model scattering from irregular interfaces between two media with differing sound speeds and densities. The volume integrals in Eqs. (7) and (11) for this case would be replaced with surface integrals over the interface and the coherence volume V_c replaced by the coherence area A_c.[20]

B. Broadband scattered field moments

For a potentially broadband source function, analytic expressions are provided for the statistical moments of the scattered field from inhomogeneities within the sonar resolution footprint. Let $q(t)$ be the source wave form with Fourier transform $Q(f)$. The expected time-dependent scattered field $\Psi(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,t)$ is

$$\langle \Psi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,t)\rangle = \int_{-\infty}^{\infty} Q(f)\langle \Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)\rangle e^{-j2\pi ft}df, \tag{12}$$

by inverse Fourier transform of Eq. (7), weighted by the source spectrum.

From Parseval's theorem, the mean scattered energy over the resolution footprint is proportional to[21]

$$\bar{E}(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0) = \int |Q(f)|^2\langle |\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)|^2\rangle df, \tag{13}$$

where $\langle |\Phi_s(\mathbf{r}_s|\mathbf{r},\mathbf{r}_0,f)|^2\rangle$ is obtained from Eq. (10).

Sea bottom reverberation is typically dominated by the incoherent intensity. In this case, only the variance term in Eq. (11) contributes to the total intensity in the right-hand side of Eq. (13).

III. APPLICATION TO SEAFLOOR REVERBERATION WITH CALIBRATION

In this section, we apply the scattering model developed in Sec. II to seafloor reverberation. Geophysical parameters of the model are calibrated with measured data.

A. OAWRS images of the ocean environment

Here we use OAWRS images obtained during the ONR-sponsored Main Acoustic Experiment 2003[1] on the U.S. East Coast continental shelf. A bistatic OAWRS system comprising of a vertical source array and a horizontally towed receiving array was deployed in the New Jersey Strataform located 200 km south of Long Island. The source array excited acoustic waveguides modes that propagated tens of kilometers in the continental shelf environment. Transmissions consisted of 1 s duration Tukey-shaded linear frequency modulated pulses in three different frequency bands; 390–440, 875–975, and 1250–1400 Hz. The transmissions were repeated every 50 s over the approximately 83 min duration it took to tow the receiving array along a 10 km track line at constant tow speed of 2 m/s. Data were collected over 5–6 tracks each day. Here we focus our analysis on data in the 390–440 Hz band as this was the wave form that was most frequently used for imaging.

Instantaneous OAWRS images of the ocean environment over wide areas spanning thousands of square kilometers are formed from each transmission. The arrivals are charted in range by two-way travel time and matched filter analysis, and in bearing by beamforming.[1,2] Scattered fields from objects at identical horizontal position but different

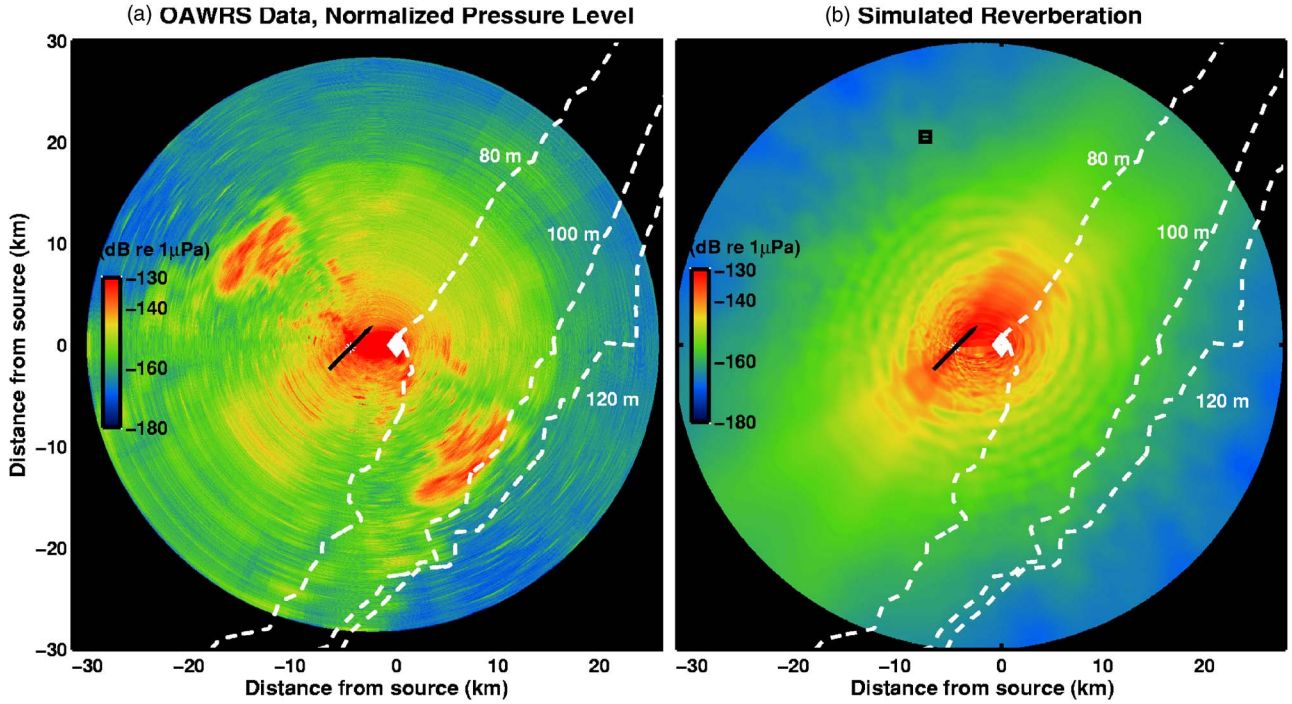


FIG. 1. (Color online) (a) Instantaneous OAWRS image of the ocean environment on the New Jersey continental shelf. The image is averaged over five consecutive transmissions centered at 09:32 EDT, 14 May 2003. The imaging diameter of the system was set to roughly 60 km (40 s imaging time). The moored source (the white diamond) operating at 390–440 Hz is the coordinate origin at 39.0563N, 73.0365W. The towed horizontal receiving array moving along 51°E heading (indicated by black line) has 2.6° azimuthal resolution at array broadside and 15 m range resolution. The positive vertical axis points north. Depth contours are indicated by dashed lines. The image has been corrected for source level. (b) The simulated bottom reverberation corresponding to the data in (a), calculated using the computationally efficient approach at the center frequency 415 Hz with a typical observed sound speed profile. The small black 1 km by 1 km box shown near the upper left-hand corner indicates the location used for calibrating the reverberation model. This is the area shown in Fig. 6.

depths are integrated by the system in the resulting two-dimensional (2D) horizontal OAWRS images. The inherent left–right ambiguity in charting returns with a horizontal line array receiver is evident in some of the OAWRS images that will be presented here.

The resolution footprint of the OAWRS system is a function of both range and azimuth. The receiving array has a cross-range resolution given by $\rho\beta(\phi)$ where the angular resolution $\beta(\phi)$ is given by $\beta(\phi)\approx 1.44(\lambda/L\cos\phi)$ for broadside $(\phi=0)$ through angles near endfire $(\phi=\pi/2)$, where λ is the wavelength of the signal and L is the length of the array. At endfire, the angular resolution is given by $\beta(\phi=\pi/2)\approx 2.8\sqrt{\lambda/L}$. A Hanning spatial window was applied to the receiving array in the beamforming to reduce the levels of the sidelobes. The range resolution is $\Delta\rho=c/2B$ after matched filtering, where B is the signal bandwidth.[12]

The OAWRS images show scattered returns from large and densely populated fish groups, small or diffuse groups of fish, some discrete geologic features of the seafloor, as well as continuous diffuse returns from the seafloor. A typical image is shown in Fig. 1(a).[1] Trends in the images include the decay of scattered intensity with increasing bistatic range due to acoustic spreading loss and attenuation in the ocean waveguide, and spatial variation in cross-range resolution which leads to spatially varying blurring.

B. Estimating scattering level of the environment from OAWRS data

A challenge in analyzing sonar imagery is to distinguish diffuse background reverberation from discrete returns due to targets or clutter. The discrete returns may arise from surface or volume inhomogeneities such as marine creatures, geologic features, or man-made objects. It is often not known in advance from where in the water column the targets or clutter features arise.

To distinguish diffuse background reverberation from clutter and to compare the scattering level of returns from different ranges, we devise a two step approach. First, we detrend an OAWRS intensity image by correcting for mean transmission over the water-column, source level, and the spatially dependent resolution footprint of the OAWRS system. This leads to an estimate of what we call the environmental scattering level (ESL) at a given horizontal location. Next we compute statistics of ESL variations over space and time, and use these to discriminate diffuse background from clutter. We then investigate either the diffuse background or the clutter. Here we focus on the background.

Let $\overline{\mathrm{Lp}(\boldsymbol{\rho}_s|\mathbf{r},\mathbf{r}_0,t)}$ be $10\log_{10}$ of the OAWRS pressure squared matched filtered output centered at horizontal location $\boldsymbol{\rho}_s$ and time t, which may include some spatial and temporal averaging in pressure squared to reduce the variance. Then the $\mathrm{ESL}(\boldsymbol{\rho}_s,t)$ at the given location and time is obtained by

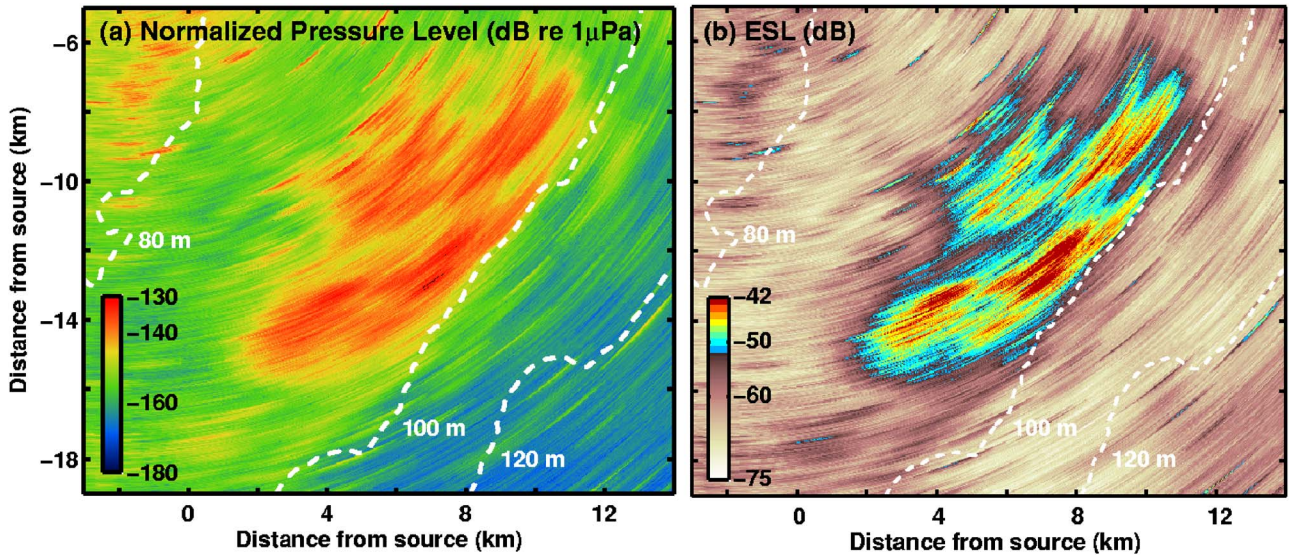


FIG. 2. (Color online) (a) A zoomed-in version of the OAWRS image shown in Fig. 1(a) in the south-east region containing a massive fish shoal. (b) Corresponding ESL image. Higher ESL levels are obtained in the region of dense fish shoal and lower levels correspond to the background. The transmission loss used for correcting this image is shown in Fig. 3(c).

$$\mathrm{ESL}(\boldsymbol{\rho}_s,t) = \overline{\mathrm{Lp}(\boldsymbol{\rho}_s|\mathbf{r},\mathbf{r}_0,t)} - \mathrm{SL}(\mathbf{r}_0) - \overline{\mathrm{TL}(\boldsymbol{\rho}_s|\mathbf{r}_0)} - \overline{\mathrm{TL}(\mathbf{r}|\boldsymbol{\rho}_s)} - 10\log_{10} A(\boldsymbol{\rho}_s), \tag{14}$$

where $\overline{\mathrm{TL}(\boldsymbol{\rho}_s|\mathbf{r}_0)}$ and $\overline{\mathrm{TL}(\mathbf{r}|\boldsymbol{\rho}_s)}$ are $10\log_{10}$ of the average of the antilog of transmission loss over the water column. The former is for transmission from source location to environmental location $\boldsymbol{\rho}_s$ and the latter is from this environmental location to the receiver. The spatially varying resolution footprint of the receiving array has area $A(\boldsymbol{\rho}_s)=[\beta(\phi)/2\pi][\pi(\rho+\Delta\rho)^2-\pi\rho^2]\simeq\rho\beta(\phi)\Delta\rho$, and $\mathrm{SL}(\mathbf{r}_0)$ is the source level. Environmental scattering level is then $10\log_{10}$ of a ratio and has units simply of deciBels, like scattering strength.

Here we use a statistical propagation model based on the range-dependent parabolic equation[4] calibrated[22] against one-way transmission data from the same set of measurements to estimate the mean transmission loss over the water-column and obtain $\overline{\mathrm{TL}(\boldsymbol{\rho}_s|\mathbf{r}_0)}$ and $\overline{\mathrm{TL}(\mathbf{r}|\boldsymbol{\rho}_s)}$. Effects such as internal waves, eddies and turbulence cause fluctuations in ocean sound speed and density structure. These lead to scintillation in ocean acoustic waveguide propagation caused by changes in the multimodal interference structure. In this environment, the acoustic waveguide modes were found experimentally to become sufficiently randomized beyond roughly 2 km in range from the source[22] that the *expected* acoustic intensity varies only gradually with increasing range and does not exhibit the strong periodic modal interference structure in range and depth present in nonrandom waveguides.[16,20] The antilog of the mean transmission, $\overline{\mathrm{TL}(\boldsymbol{\rho}_s|\mathbf{r}_0)}$ and $\overline{\mathrm{TL}(\mathbf{r}|\boldsymbol{\rho}_s)}$ are proportional to the expected acoustic power transmitted through the waveguide at any given range and azimuth.

Figures 2(a) and 2(b) are OAWRS images of the mean matched filtered scattered intensity and the corresponding ESL of the ocean environment on 14 May 2003 after incoherently averaging intensity over five consecutive transmissions (pings) centered at time instance 9:32 Eastern Daylight Time (EDT). The ESL image in Fig. 2(b) is derived from Eq. (14) where images corresponding to the right-hand side of Eq. (14) appear in Figs. 2(a) and 3. The mean transmission losses shown in Figs. 3(a) and 3(b) are in good agreement with measurements or models of transmission losses reported in Refs. 22–24 for the New Jersey Strataform.

In order to distinguish the statistically stationary background ESL from that due to moving biological clutter, we average N consecutive OAWRS images, which already were formed from five-ping running averages, to obtain a mean OAWRS image and a standard deviation OAWRS image over a wide spatial area. Figures 4(a) and 4(b) show the mean and standard deviation of $N=361$ ESL images derived from data acquired within the frequency band from 390 to 440 Hz on 14 May 2003.[25]

We find in our 2003 OAWRS images that (a) high mean and high standard deviation characterizes regions where fish are regularly passing through, (b) high mean and low standard deviation characterizes regions where fish consistently remain, (c) low mean and high standard deviation characterizes regions where fish pass through intermittently, and (d) low mean and low standard deviation characterizes regions of diffuse reverberation from the seafloor. This can be seen in Fig. 5(a) where ESL is plotted as a function of sequential ping number for various spatial locations from Fig. 4(b). In particular, ESL in regions of dense fish population is roughly 20–25 dB higher than in regions where diffuse seafloor reverberation is the dominant scattering mechanism. Note that the above-mentioned (b) could also be caused by static geologic features but was ruled out in this experiment as data from the area shown in Fig. 4, six days prior, showed the scattered levels returning to those of the background when the massive fish shoal was absent.

For estimating the background ESL due to the seafloor, we focus analysis on regions with upslope bathymetry relative to the incident wave that had ambiguous regions that were downslope. This is to help eliminate the effect of left–right ambiguity, as seafloor scattering has been shown to be

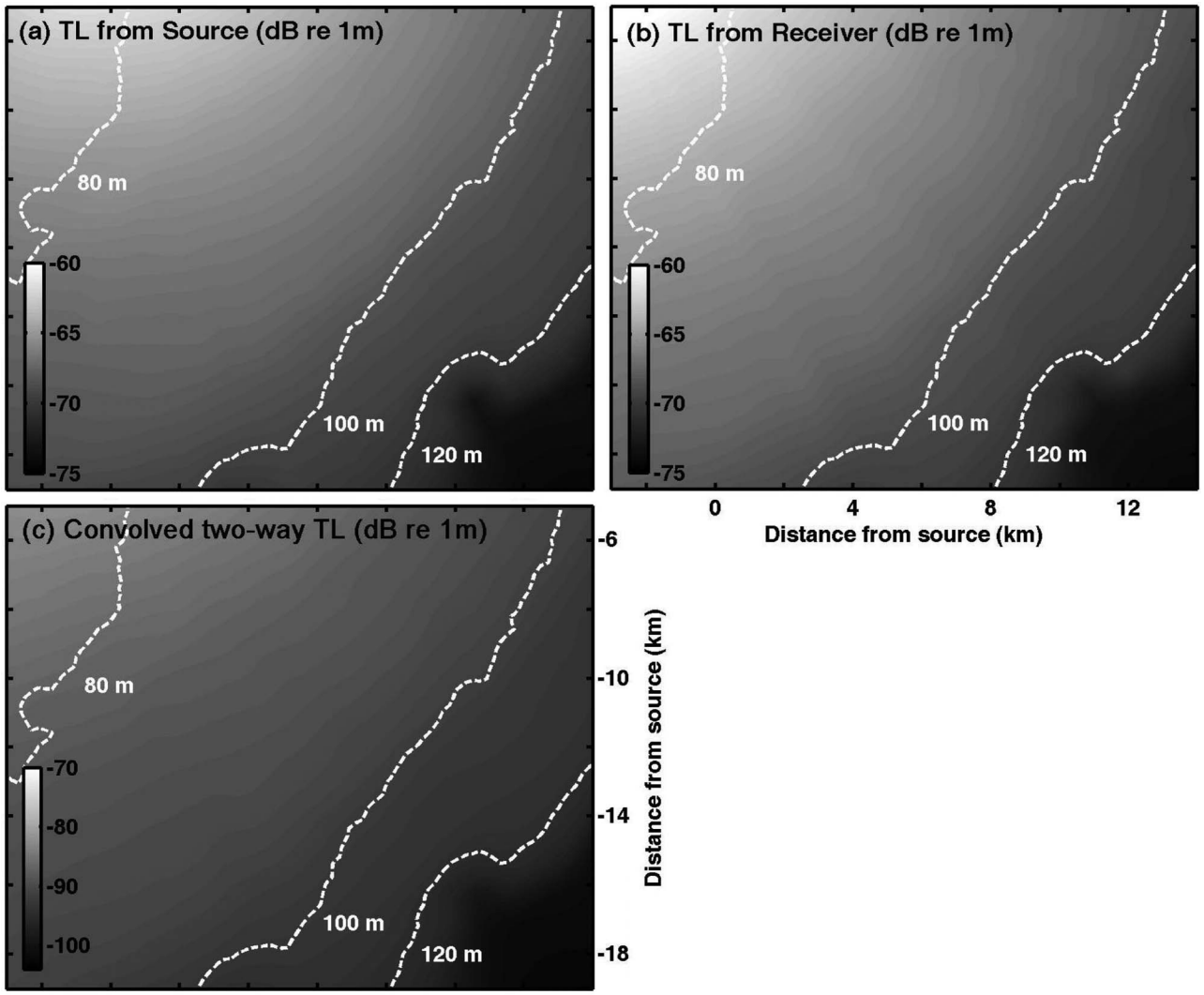


FIG. 3. (a) $\overline{\mathrm{TL}(\boldsymbol{\rho}_s|\mathbf{r}_0)}$ and (b) $\overline{\mathrm{TL}(\mathbf{r}|\boldsymbol{\rho}_s)}$ are 10 $\log_{10}$ of the mean antilog transmission losses over the water column from source and receiver respectively to the area shown for the 390–440 Hz band. (c) The quantities in (a) and (b) blurred with the spatially dependent resolution footprint of the receiving array.

more intense when bathymetry is upward sloping in the direction of the incident wave, and weaker when bathymetry is downward sloping,[13,14,26–29] as expected from physical principles. Temporal variations in ESL levels for various spatial locations with upslope bathymetry are shown in Fig. 5(b). The area shown in Fig. 4 consists of down sloping bathymetry but is shown to enable comparison of background ESL levels with those from fish shoal.

The mean seafloor ESL obtained here (and tabulated in the forthcoming Table II) can now be used to calibrate the reverberation model presented in Sec. II for seafloor scattering in the Strataform area. The standard deviation of ESL for seafloor scattering has been found by temporal and spatial averaging to be about 2 dB. This standard deviation will be useful in determining the accuracy of parameter estimates needed to calibrate the reverberation model.

C. Estimating geologic parameters necessary for reverberation model

The geological parameters necessary to implement the reverberation model defined in Eqs. (7) and (11) are estimated in the Appendix. In particular, the mean, standard deviation, and correlation of Γ_κ and Γ_d are estimated from the statistical moments of sound speed and density provided in Refs. 30, 31, and 3. The results, tabulated in Table I, are

TABLE I. Statistical geologic properties of the New Jersey Strataform used to calibrate the bottom reverberation model.

Statistical parameters	Average values over the region
$\bar{c}$ (m/s)	1700
σ_c (m/s)	35.5
$\mu^2=\sigma_c^2/\bar{c}^2$	4.6889e−04
$\bar{d}$ (g/cm^3)	1.90
σ_d (g/cm^3)	0.16
$\eta^2=\sigma_d^2/\bar{d}^2$	6.2e−3
$\langle\Gamma_\kappa\rangle$	0
$\langle\Gamma_\kappa^2\rangle$	0.0083
Var(Γ_κ)	0.0083
$\langle\Gamma_d\rangle$	−0.0063
$\langle\Gamma_d^2\rangle$	0.0066
Var(Γ_d)	0.0065
Cov(Γ_κ,Γ_d)	−0.0065

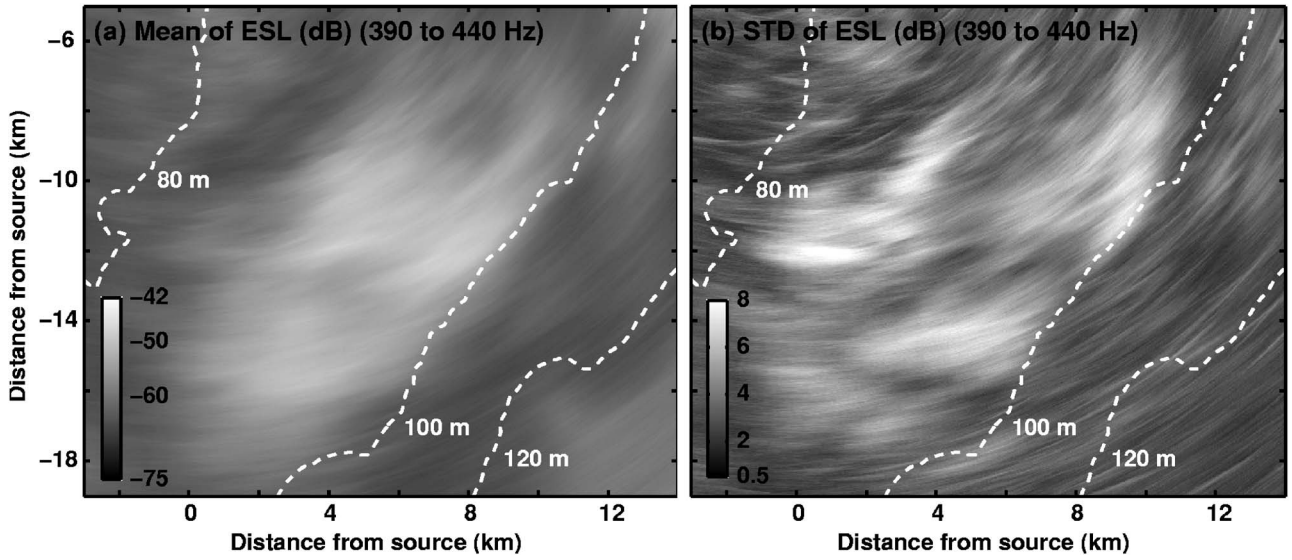


FIG. 4. Temporal and spatial variations in ESL for the area in Fig. 2 for data in the 390–440 Hz band collected on 14 May 2003. (a) Mean and (b) standard deviation of 361 ESL images formed from transmissions over a measurement time period ranging from 09:10 to 17:20 EDT.

consistent with those found in Ref. 7 which are based on measurements from ten different shallow and deep water sites around the world.

From Table I we observe that $\langle\Gamma_\kappa\rangle^2$ and $\langle\Gamma_d\rangle^2$ are at least 2 orders of magnitude smaller than $\mathrm{Var}(\Gamma_\kappa)$ and $\mathrm{Var}(\Gamma_d)$. The coherent term in Eq. (10) will therefore be negligible in comparison to the incoherent term. The total reverberation intensity can then be approximated as the variance of the scattered field using Eqs. (13) and (11).

D. Maximum likelihood estimator for coherence volume V_c

To implement the reveberation model in Eqs. (13) and (11) at a given location, an estimate is needed for the coherence volume V_c of the random seabed inhomogeneities. An expression for the maximum likelihood estimator (MLE) for the coherence volume V_c in terms of the measured reverberation data and model is provided here.

The instantaneous scattered field measured at the receiver can be modeled as a circular complex Gaussian random variable.[32] The log transform of the matched filtered intensity within the resolution footprint centered at $\boldsymbol{\rho}_s$ then obeys a Gaussian probability density when the time-bandwidth product is sufficiently large,[32] as does the ESL,

$$p(\mathrm{ESL}(\boldsymbol{\rho}_s)|V_c) = \left(\frac{\mu}{2\pi}\right)^{1/2} \exp\left(-\frac{\mu}{2}[\mathrm{ESL}(\boldsymbol{\rho}_s) - \langle \mathrm{ESL}(\boldsymbol{\rho}_s, V_c)\rangle]^2\right), \tag{15}$$

where μ is the time–bandwidth product, which for the cur-

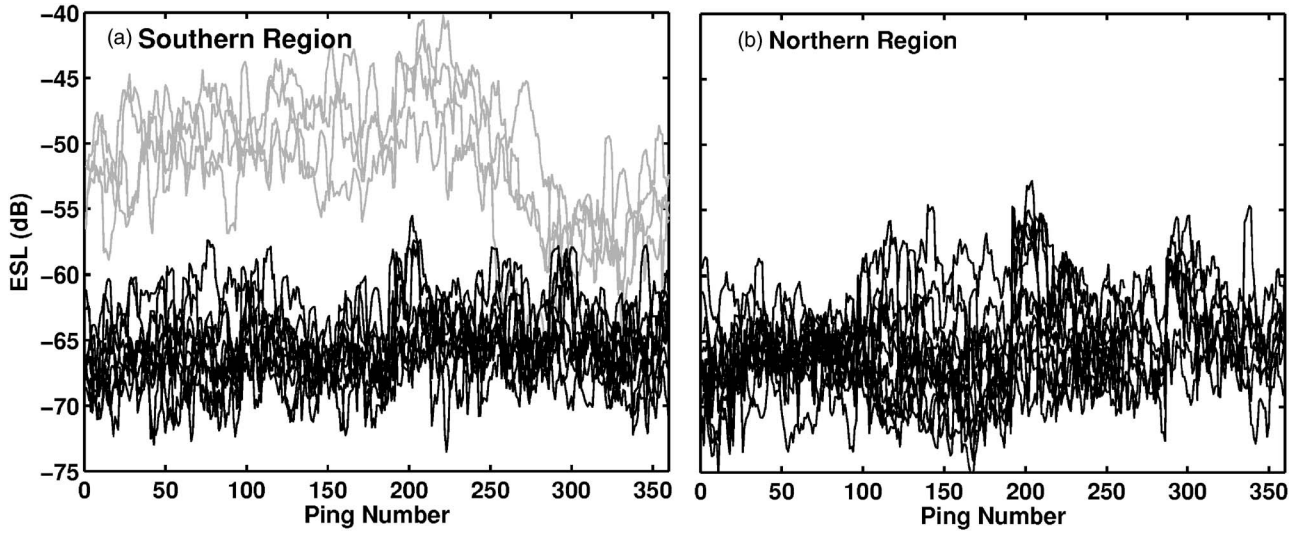


FIG. 5. (a) Temporal variation in ESL at various spatial locations as a function of transmission number or ping, for the image in Fig. 4. Gray lines in (a) are ESL at locations with persistently strong scattered returns from fish shoals. Black lines indicate ESL at locations consistent with the background. (b) Temporal variations in ESL for the region with upslope bathymetry in the vicinity of the small box shown in Fig. 1(b).

rent processing is greater than 10. This is because the matched filtered intensity data are averaged over two independent resolution cells and five independent pings.

As μ is large, the bias due to log transformation, defined in Eq. (22) of Ref. 32, is negligible. We can then make the approximation

$$\langle \mathrm{ESL}(\boldsymbol{\rho}_s, V_c)\rangle = 10\log_{10}(W(\boldsymbol{\rho}_s|\mathbf{r},\mathbf{r}_0)) + 10\log_{10}(V_c) - \langle \mathrm{TL}(\boldsymbol{\rho}_s|\mathbf{r}_0)\rangle - \langle \mathrm{TL}(\mathbf{r}|\boldsymbol{\rho}_s)\rangle, \quad (16)$$

where we have introduced the new variable

$$10\log_{10}(W(\boldsymbol{\rho}_s|\mathbf{r},\mathbf{r}_0)) = 10\log_{10}[\bar{E}(\boldsymbol{\rho}_s|\mathbf{r},\mathbf{r}_0)] - 10\log_{10}(V_c), \quad (17)$$

because it does not depend on V_c. Here, $\bar{E}(\boldsymbol{\rho}_s|\mathbf{r},\mathbf{r}_0)$ is defined in Eq. (13), and is proportional to the total energy returned from the resolution footprint of range extent Δr. It is advantageous to express our intensity data in logarithmic units as it transforms signal-dependent noise into additive signal-independent noise.[32]

The ESL of the bottom measured in N different resolution footprints of OAWRS imagery for $s=1,2,3,\ldots,N$ can be assumed to be independent and identically distributed. Their joint probability density function is

$$p(\mathrm{ESL}(\boldsymbol{\rho}_1),\mathrm{ESL}(\boldsymbol{\rho}_2),\ldots,\mathrm{ESL}(\boldsymbol{\rho}_N)|V_c) = \prod_{s=1}^{N}\left(\frac{\mu}{2\pi}\right)^{1/2}\exp\left(-\frac{\mu}{2}[\mathrm{ESL}(\boldsymbol{\rho}_s)-\langle \mathrm{ESL}(\boldsymbol{\rho}_s,V_c)\rangle]^2\right). \quad (18)$$

The maximum likelihood estimator $\hat{V}_c$ for the coherence volume V_c is given by

$$\frac{d}{dV_c}\{\ln[p(\mathrm{ESL}(\boldsymbol{\rho}_1),\mathrm{ESL}(\boldsymbol{\rho}_2),\ldots,\mathrm{ESL}(\boldsymbol{\rho}_N)|V_c)]\}\Big|_{V_c=\hat{V}_c} = 0. \quad (19)$$

Expanding Eq. (19) leads to

$$10\log_{10}(\hat{V}_c) = \frac{1}{N}\sum_{s=1}^{N}\mathrm{ESL}(\boldsymbol{\rho}_s) - \frac{1}{N}\sum_{s=1}^{N}\{10\log_{10}(W(\mathbf{r}|\mathbf{r}_0,\boldsymbol{\rho}_s)) - \mathrm{TL}(\boldsymbol{\rho}_s|\mathbf{r}_0) - \mathrm{TL}(\mathbf{r}|\boldsymbol{\rho}_s)\} \quad (20)$$

the desired estimator.

E. Reverberation model calibration

Implementation of the full reverberation model in Eq. (11) for range-dependent waveguides involves numerically calculating 3D spatial derivatives of the complex waveguide Green function. This leads to large 3D matrices that makes the implementation cumbersome and reduces computational speed. For instance at 400 Hz, the spatial sampling required is approximately 0.2 m in range, 0.5 m in depth, and roughly 3° in azimuth for computational accuracy.

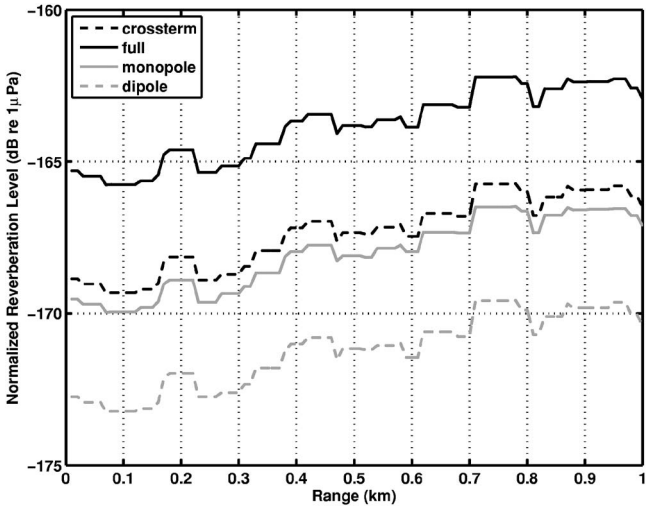


FIG. 6. Full bistatic, 390–440 Hz broadband, depth-integrated bottom reverberation model simulation for a horizontal transect across the 1 km by 1 km area shown in Fig. 1(b). The monopole term, dipole term, and cross-term contributions to the reverberation intensity are plotted for comparison.

To avoid this problem, we calibrate the full reverberation model using only a small area, 1 km by 1 km, shown in Fig. 1(b). The coherence volume V_c is then determined with the approach outlined in Sec. III D. We provide a method for implementing the reverberation model that is computationally efficient over wide areas in the next section.

The scattered intensity of Eq. (11) is a sum of (1) a purely monopole term dependent on the variance of fractional change in compressibility, (2) a purely dipole term dependent on the variance of fractional change in density, and (3) the covariance between the two. The contribution of these three terms for a horizontal transect through the small area is shown in Fig. 6.

The MLE of the coherence volume $\hat{V}_c$ is summarized in Table II. We assume the inhomogeneites are isotropic, so that $V_c=\frac{4}{3}\pi l_c^3$, where l_c is the coherence radius. The coherence lengths in x, y, and z directions are $l_x=l_y=l_z=2l_c$. The acoustically determined correlation length obtained here is comparable to the geologic correlation depth of between 0.3 and 3 m from analysis of core data.[33]

As the background ESL has a standard deviation of ±2 dB, the estimate $10\log_{10}\hat{V}_c$ also has the same standard deviation. This leads to a correlation length in the range from 15 to 21 cm. There could also be uncertainties in our estimates of the moments of Γ_κ and Γ_d. The variations in $\mathrm{Var}(\Gamma_\kappa)$ needed to match the data over the range of uncertainty in the correlation length occupies roughly a factor of 2.

TABLE II. Estimates for the New Jersey Strataform. The standard deviation for ESL and SS is ±2 dB.

f (Hz)	ESL (dB)	V_c (m^3)	$2l_c$ (m)	f_d (dB)	f_p (m)	SS (dB)
390–440	−67.5	0.0030	0.18	4.2	0.46	−37

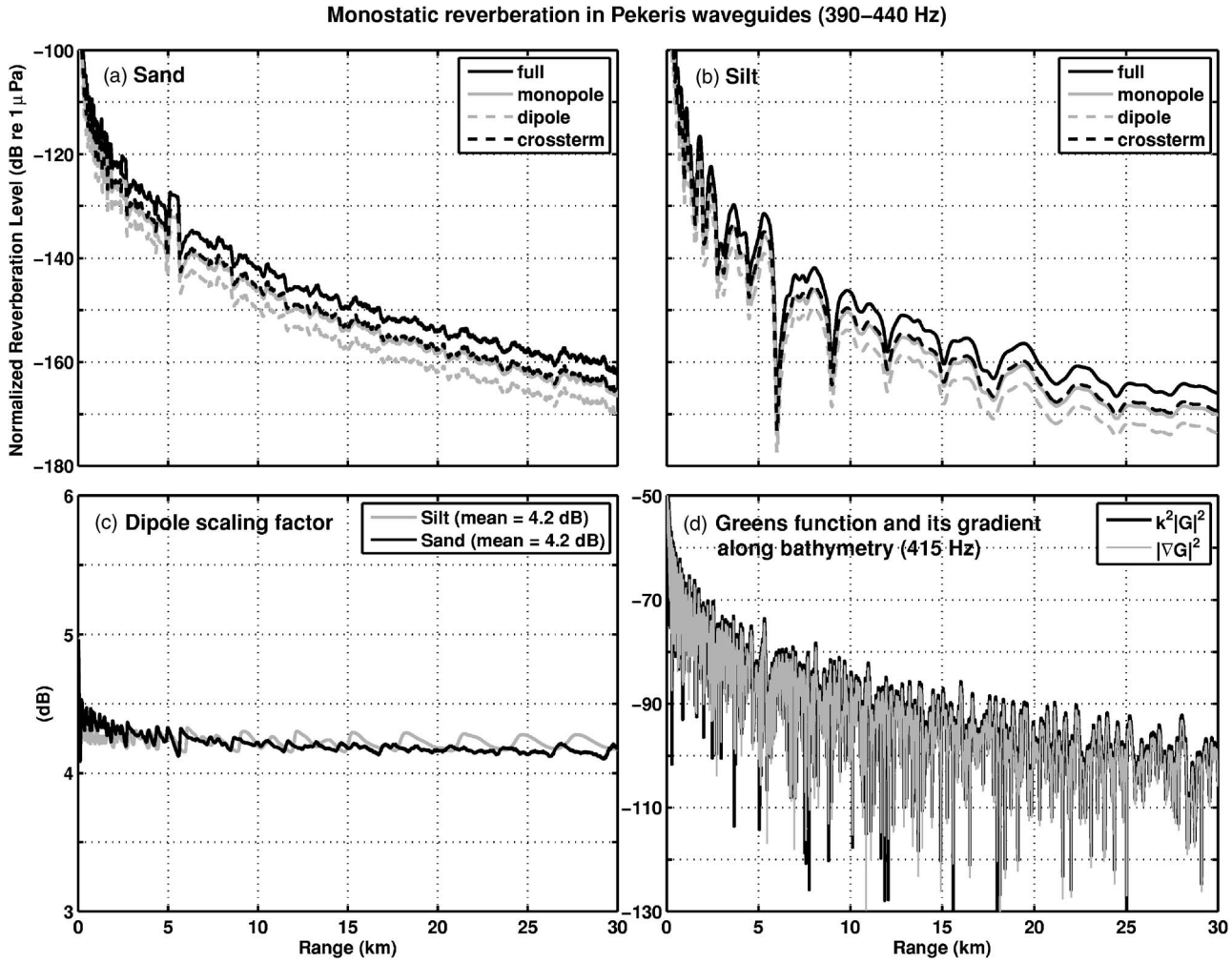


FIG. 7. Broadband 390–440 Hz, monostatic reverberation in Pekeris waveguides with (a) sand and (b) silt bottom for a 0 dB re 1 μPa at 1 m source level comprising of a single element at 50 m depth. The receiving array parameters are the same as those used in the experiment in Fig. 1(a). The Green function for Pekeris waveguides calculated using the normal mode code, Kraken (Ref. 36). (c) The dipole scaling factor equal to the ratio between the full term and the monopole term plotted in deciBels remain fairly constant to within ± 0.1 dB and is independent of the bottom type. (d) Comparison of the magnitudes of the Greens function and its spatial derivative at the water–sediment interface at 415 Hz.

Here we ignored a potential degradation due to matched filtering the data in an ocean waveguide while calibrating them with the reverberation model. This degradation is typically small, less than 3 dB for the ranges of interest.[22,34]

IV. COMPUTATIONALLY EFFICIENT IMPLEMENTATION

When working on a field experiment, it is often instructive to rapidly predict seabed reverberation over wide areas, to aid in experimental design. A computationally efficient approach for computing seabed reverberation in range dependent waveguides over wide areas is developed here. It is based on the introduction of a dipole scaling factor and a depth-penetration factor that simplify the numerical implementation.

We find it a good approximation to take the dipole and cross terms in the reverberant intensity of Eq. (11) to be proportional to the monopole term. This is illustrated for horizontally stratified Pekeris sand and silt waveguides in Figs. 7(a) and 7(b) with collocated source and receiver, and for the bistatic scenario in a range dependent waveguide in Fig. 6. The total intensity can be estimated from the monopole term alone by multiplication with a constant factor, F_d, which we call the dipole scaling factor. This can be seen by noting in Fig. 7(c) that the ratio of the total intensity to the monopole term is a constant, with $10 \log_{10} F_d \approx 4.2$ dB. This approximation is a consequence of the fact that $|\nabla G|$ is proportional to $k|G|$, as shown in Fig. 7(c), which means that the dipole and monopole terms are proportional. Without this approximation, it is currently impractical to compute reverberation over wide areas because of the tremendous computational effort needed to determine the derivatives of the Green function in the dipole term.

The depth integral of Eq. (11) for volume scattering can be approximated by the seafloor surface contribution multiplied by a depth penetration factor, F_p. The depth penetration factor provides the equivalent depth over which acoustic intensity in the seabed is of similar magnitude as that on the seafloor, before it decays rapidly with depth due to sediment attenuation. This factor is simply the ratio of the volume integral to the seafloor contribution, which are shown in Figs. 8(a) and 8(b), with corresponding F_p ratio in Fig. 8(c), which is found to be dependent on bottom type, roughly

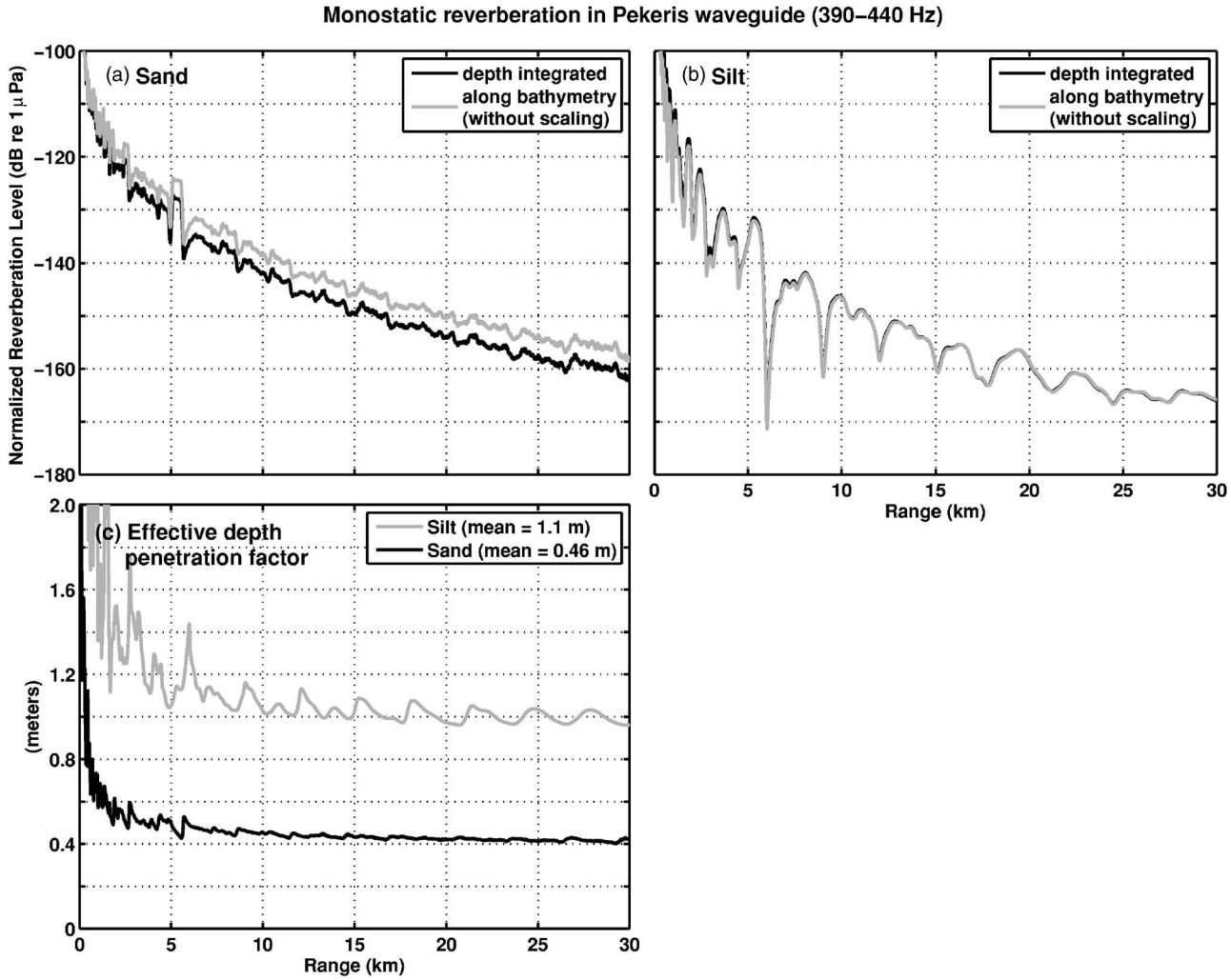


FIG. 8. Comparison of the 390–440 Hz broadband depth-integrated bottom reverberation intensity with the broadband intensity at the water–sediment interface in the Pekeris (a) sand and (b) silt waveguides. (c) The effective penetration depth in meters plotted for Pekeris sand and silt waveguides.

0.5 m for sand and 1 m for silt. The depth penetration factor is also a function of frequency, decreasing with increasing frequency.

Another approximation that can greatly increase computational efficiency is to calculate the scattered field at just the center frequency of the broadband signal. Figure 9 compares the full broadband depth integrated monostatic reveberation to the range-averaged scattered field at the center frequency with the two approximations introduced earlier applied to the result. The latter single frequency range-averaged approximation provides a good estimate to the broadband result.

The approximations made in this section require only the Green function magnitude on the seabed surface which can be stored in sparse 2D matrices rather than dense 3D matrices required for the full calculation. The reverberation intensity is accurate to within ± 0.5 dB for ranges greater than 3 km and to within 3 dB for ranges between 1 and 3 km. Figure 1(b) illustrates this approach applied to model reverberation over wide areas on the New Jersey continental shelf corresponding to the OAWRS data Fig. 1(a). The reverberation levels from the simulation are comparable to the background levels in the data, especially in areas adjacent to the massive fish shoal.

V. SCATTERING STRENGTH

Here we derive the scattering strength (SS) of the bottom on the New Jersey continental shelf using the model and calibration results. As the seabed inhomogeneities have coherence lengths that are smaller than the wavelength,[35] we can apply the sonar equation to approximate SS of the bottom from Eq. (11) with the approximations introduced in Sec. IV as

$$\mathrm{SS} \approx 10 \log_{10}[F_d F_p V_c k^4 \mathrm{Var}(\Gamma_k)]. \tag{21}$$

The SS of the sandy bottom on the New Jersey Strataform is tabulated in Table II. Analysis with a normal mode propagation model[36] indicates that the first 10 modes are important contributors to the acoustic field at a range of 20 km from the source where the calibration was done. At this range, the depth-averaged intensity of mode 10 is about 1/5 the depth-averaged intensity of mode 1. The grazing angles for modes 1–10 range from 2° to 12° and the derived SS level of roughly -37 ± 2 dB are determined by these grazing angles after long range propagation in the waveguide with the effect of multimodal coupling in the scattering process.

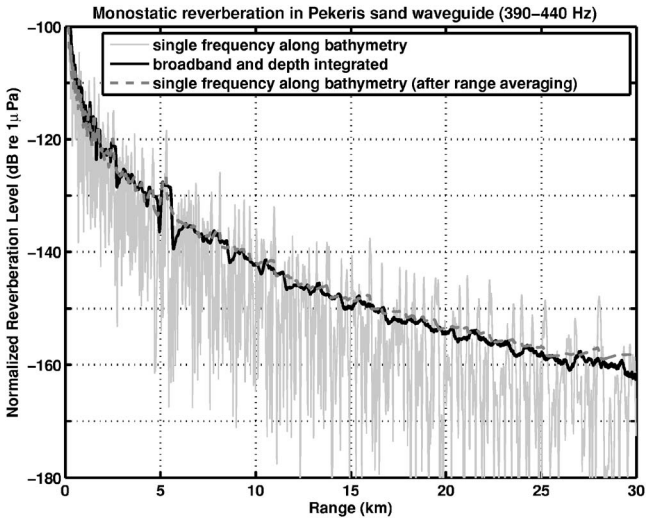


FIG. 9. Full 390–440 Hz broadband depth-integrated monostatic bottom reverberation compared to simulated result at center frequency 415 Hz calculated using efficient approach with the dipole scaling and depth penetration factors.

For reverberation measured at 390–440 Hz bandwidth considered in this paper, volume scattering is expected to be the dominant mechanism at all grazing angles for a sandy seabed as shown in Figs. 10 and 11 of Ref. 5. Our estimate agrees well with those of other ocean environments with sandy bottoms in Refs. 5 and 10, suggesting low surface roughness at this site and the consequent importance of evanescent energy in generating low-angle volume backscatter via subcritical ensonification.

The overall frequency dependence of SS in Eq. (21) is determined by the wave number k, which is directly proportional to frequency, the depth penetration factor f_p that decreases with increasing frequency, and potentially the coherence volume sampled by the acoustic wave V_c. The coherence volume may depend on frequency as short wavelength acoustic fields are more sensitive to small scale fluctuations in medium properties than long wavelength acoustic waves. This acoustically determined coherence volume may differ from the geologic coherence volume that is determined purely from spatial variations in medium properties.

VI. CONCLUSION

An analytic model for 3D, bistatic scattering from medium inhomogeneities is developed from first principles by application of Green's theorem. Statistical moments of the scattered field are expressed in terms of statistical moments of medium compressibility and density fluctuations. The model is applied to seabed reverberation and optimally calibrated with both OAWRS and geological data on the New Jersey continental shelf. Analysis with the model indicates that (1) seabed reverberation is incoherent, and (2) scattering strength varies with frequency depending on wavenumber k, medium coherence volume V_c and seabed depth penetration factor F_p following a $10\log_{10}(F_pV_ck^4)$ dependence.

An efficient numerical approach is also developed for rapidly computing seabed reverberation over wide areas for bistatic sonar systems in range-dependent ocean waveguides. It exploits the correlation between monopole and dipole scattering terms and the limited penetration of acoustic fields in the seabed. The model handles the scattering of evanescent waves in the seabed. Finally, an approach for distinguishing the statistically stationary background reverberation from the scattered fields of moving targets in sonar data by tracking the temporal and spatial evolution of the returns is presented.

ACKNOWLEDGMENTS

This work was sponsored by the Office of Naval Research and the Sloan Foundation with administrative support from the NSF Gordon Center for Subsurface Sensing and Imaging Systems.

APPENDIX: DERIVATION OF STATISTICS OF GEOACOUSTIC PARAMETERS

The seabed on the New Jersey Strataform is predominantly comprised of sand. It has been mapped by geophysical surveys with analysis provided in Refs. 3, 30, and 31. From these, we gathered the statistical moments of seabed sound speed and density fluctuations and list them in Table I. We assume that sound speed $c_t(\boldsymbol{\rho}_t) \sim N(\bar{c}, \sigma_c)$ and density $d_t(\boldsymbol{\rho}_t) \sim N(\bar{d}, \sigma_d)$ obey Gaussian random processes that are locally stationary in a given area. These statistical moments can vary between areas to account for potential changes in sediment type. We assume that sound speed and density are uncorrelated, which is a valid assumption given the large scatter in the data in Fig. 10(b) of Ref. 30. Despite this assumption, Γ_κ and Γ_d can still be correlated.

From Eqs. (2) and (3), the means are given by

$$\langle \Gamma_\kappa(\mathbf{r}_t) \rangle = 0 \tag{A1}$$

and

$$\langle \Gamma_d(\mathbf{r}_t) \rangle = 1 - \bar{d}\left\langle \frac{1}{d_t(\mathbf{r}_t)} \right\rangle, \tag{A2}$$

where $\bar{d}$ and $\bar{c}$ as defined before are the mean density and sound speed within the sonar resolution footprint, respectively. The second moments are given by

$$\langle \Gamma_d^2(\mathbf{r}_t) \rangle = 1 - 2\bar{d}\left\langle \frac{1}{d_t(\mathbf{r}_t)} \right\rangle + \bar{d}^2\left\langle \frac{1}{d_t^2(\mathbf{r}_t)} \right\rangle \tag{A3}$$

and

$$\langle \Gamma_\kappa^2(\mathbf{r}_t) \rangle = \frac{\left\langle \frac{1}{d_t^2(\mathbf{r}_t)} \right\rangle \left\langle \frac{1}{c_t^4(\mathbf{r}_t)} \right\rangle}{\left\langle \frac{1}{d_t(\mathbf{r}_t)} \right\rangle^2 \left\langle \frac{1}{c_t^2(\mathbf{r}_t)} \right\rangle^2} - 1, \tag{A4}$$

the variances by

$$\mathrm{Var}(\Gamma_d(\mathbf{r}_t)) = \langle \Gamma_d^2(\mathbf{r}_t) \rangle - \langle \Gamma_d(\mathbf{r}_t) \rangle^2 \tag{A5}$$

and

$$\mathrm{Var}(\Gamma_\kappa(\mathbf{r}_t)) = \langle \Gamma_\kappa^2(\mathbf{r}_t) \rangle - \langle \Gamma_\kappa(\mathbf{r}_t) \rangle^2, \tag{A6}$$

and their covariance by

$$\mathrm{Cov}(\Gamma_\kappa,\Gamma_d) = \left\langle \left(1 - \frac{\bar{d}}{d_t(\mathbf{r}_t)}\right)\left(\frac{\kappa_t(\mathbf{r}_t)}{\bar{\kappa}} - 1\right)\right\rangle$$
$$= \bar{d}\left\langle \frac{1}{d_t(\mathbf{r}_t)}\right\rangle - \frac{\bar{d}}{\bar{\kappa}}\left\langle \frac{1}{d_t^2(\mathbf{r}_t)c_t^2(\mathbf{r}_t)}\right\rangle, \quad \text{(A7)}$$

where

$$\bar{\kappa} = \left\langle \frac{1}{d_t(\mathbf{r}_t)}\right\rangle\left\langle \frac{1}{c_t^2(\mathbf{r}_t)}\right\rangle. \quad \text{(A8)}$$

Table I provides a summary of these calculations for the New Jersey Strataform.

[1]N. C. Makris, P. Ratilal, D. Symonds, S. Jagannathan, S. Lee, and R. Nero, "Fish population and behavior revealed by instantaneous continental-shelf-scale imaging," Science **311**, 660–663 (2006).

[2]P. Ratilal, Y. Lai, D. T. Symonds, L. A. Ruhlmann, J. Goff, C. W. Holland, J. R. Preston, E. K. Scheer, M. T. Garr, and N. C. Makris, "Long range acoustic imaging of the Continental Shelf Environment: The Acoustic Clutter Reconnaisance Experiment 2001," J. Acoust. Soc. Am. **117**, 1977–1998 (2005).

[3]J. A. Goff, D. J. P. Swift, C. S. Duncan, L. A. Mayer, and J. Hughes-Clark, "High resolution swath sonar investigation of sand ridge, dune and ribbon morphology in the offshore environment of the New Jersey margin," Mar. Geol. **161**, 307–337 (1999).

[4]M. D. Collins, "A split-step Pad solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).

[5]P. D. Mourad and D. R. Jackson, "A model/data comparison for low-frequency bottom backscatter," J. Acoust. Soc. Am. **94**, 344–358 (1993).

[6]P. C. Hines, "Theoretical model of in-plane scatter from a smooth sediment seabed," J. Acoust. Soc. Am. **99**, 836–844 (1996).

[7]T. Yamamoto, "Acoustic scattering in the ocean from velocity and density fluctuations in the sediments," J. Acoust. Soc. Am. **99**, 866–879 (1996).

[8]B. H. Tracey and H. Schmidt, "A self-consistent theory for seabed volume scattering," J. Acoust. Soc. Am. **106**, 2524–2534 (1999).

[9]A. Ivakin, "A unified approach to volume and roughness scattering," J. Acoust. Soc. Am. **103**, 827–837 (1998).

[10]D. Li, G. V. Frisk, and D. Tang, "Modeling of bottom backscattering from three-dimensional volume inhomogeneities and comparisons with experimental data," J. Acoust. Soc. Am. **109**, 1384–1397 (2001).

[11]K. D. LePage, "Spectral integral representations of volume scattering in sediments in layered waveguides," J. Acoust. Soc. Am. **108**, 1557–1567 (2000).

[12]N. C. Makris, L. Z. Avelino, and R. Menis, "Deterministic reverberation from ocean ridges," J. Acoust. Soc. Am. **97**, 3547–3574 (1995).

[13]N. C. Makris, C. S. Chia, and L. T. Fialkowski, "The bi-azimuthal scattering distribution of an abyssal hill," J. Acoust. Soc. Am. **106**, 2491–2512 (1999).

[14]C. S. Chia, N. C. Makris, and L. T. Fialkowski, "A comparison of bistatic scattering from two geologically distinct abyssal hills," J. Acoust. Soc. Am. **108**, 2053–2070 (2000).

[15]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986), Chap. 8.

[16]P. Ratilal and N. C. Makris, "Mean and covariance of the forward field propagated through a stratified ocean waveguide with three-dimensional random inhomogeneities," J. Acoust. Soc. Am. **118**, 3532–3559 (2005).

[17]J. A. Ogilvy, *Theory of Wave Scattering from Random Rough Surfaces* (IOP, Bristol, 1991), Chap. 2.

[18]*Laser Beam Propagation in the Atmosphere*, edited by J. A. Ogilvy (Springer, Berlin, 1978), p. 60.

[19]P. M. Shearer, *Introduction to Seismology* (Cambridge University Press, New York, 1999), Chap. 1.

[20]T. Chen, P. Ratilal, and N. C. Makris, "Mean and variance of the forward field propagated through three-dimensional random internal waves in a continental-shelf waveguide," J. Acoust. Soc. Am. **118**, 3560–3574 (2005).

[21]N. C. Makris and P. Ratilal, "A unified model for reverberation and submerged object scattering in a stratified ocean waveguide," J. Acoust. Soc. Am. **109**, 909–941 (2001).

[22]M. Andrews, T. Chen, S. Lee, N. C. Makris, and P. Ratilal, "Statistics of the Parseval sum, matched filter, and instantaneous single-frequency transmissions in an ocean waveguide," J. Acoust. Soc. Am. **120**, 3220 (2006).

[23]A. K. Rogers, T. Yamamoto, and W. Carey, "Experimental investigation of sediment effect on acoustic wave propagation in the shallow ocean," J. Acoust. Soc. Am. **93**, 1747–1761 (1993).

[24]J. R. Preston and D. A. Abraham, "Non-Rayleigh reverberation characteristics near 400 Hz observed on the New Jersey Shelf," IEEE J. Ocean. Eng. **29**, 215–235 (2004).

[25]N. Donabed, "Development of a calibrated bistatic Rayleigh-Born bottom reverberation model for range-dependent Continental Shelf environments," Master's of Science thesis, Northeastern University, 2006.

[26]P. Ratilal, S. Lee, Y. Lai, T. Chen, D. Symonds, N. Donabed, and N. C. Makris, "Range-dependent 3D scattering and reverberation in the shelf environment from biology, geology, and oceanography," J. Acoust. Soc. Am. **117**, 2611 (2005).

[27]P. Ratilal and N. C. Makris, "Unified model for 3-D scattering and forward propagation in a stratified ocean waveguide with random seabed inhomogeneities," J. Acoust. Soc. Am. **116**, 2527 (2004).

[28]P. Ratilal, S. Lee, and N. C. Makris, "Range-dependent reverberation modeling with the parabolic equation," J. Acoust. Soc. Am. **114**, 2302 (2003).

[29]S. Lee, P. Ratilal, and N. C. Makris, "Explaining extended linear features observed in remote sonar images of the New Jersey continental shelf break during Acoustic Clutter Experiments in 2001 and 2003," J. Acoust. Soc. Am. **115**, 2618 (2004).

[30]J. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Golick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).

[31]J. A. Goff, H. C. Olsen, and C. S. Duncan, "Correlation of sidescan backscatter intensity with grain-size distribution of shelf sediments, New Jersey margin," Geo-Mar. Lett. **20**, 43–49 (2000).

[32]N. C. Makris, "The effect of saturated transmission of scintillation on ocean acoustic intensity measurements," J. Acoust. Soc. Am. **100**, 769–783 (1996).

[33]J. A. Goff, J. A. Austin, S. Gulick, S. Nordfjord, B. Christensen, C. Sommerfield, H. Olson, and C. Alexander, "Recent and modern marine erosion on the New Jersey outer shelf," Mar. Geol. **216**, 275–296 (2005).

[34]G. Bar-Yehoshua, "Quantifying the effect of dispersion in continental-shelf sound propagation," Master's thesis, MIT, Cambridge, MA, 2002.

[35]P. Ratilal, Y. Lai, and N. C. Makris, "Validity of the sonar equation and Babinet's Principle for scattering in a stratified medium," J. Acoust. Soc. Am. **112**, 1797–1816 (2002).

[36]M. B. Porter and E. L. Reiss, "A numerical method for ocean acoustic normal modes," J. Acoust. Soc. Am. **76**, 244–252 (1984).

Bottom profiling by correlating beam-steered noise sequences

Chris H. Harrison[a)]
NURC, Viale San Bartolomeo, 400, 19126 La Spezia, Italy

Martin Siderius[b)]
HLS Research Inc., 3366 North Torrey Pines Court, La Jolla, California 92037, USA

(Received 24 April 2007; revised 30 November 2007; accepted 18 December 2007)

It has already been established that by cross-correlating ambient noise time series received on the upward and downward steered beams of a drifting vertical array one can obtain a subbottom layer profile. Strictly, the time differential of the cross correlation is the impulse response of the seabed. Here it is shown theoretically and by simulation that completely uncorrelated surface noise results in a layer profile with predictable amplitudes proportional to those of an equivalent echo sounder at the same depth as the array. The phenomenon is simulated by representing the sound sources as multiple random time sequences emitted from random locations in a horizontal plane above a vertical array and then accounting for the travel times of the direct and bottom reflected paths. A well-defined correlation spike is seen at the depth corresponding to the bottom reflection despite the fact that the sound sources contain no structure whatsoever. The effects of using simultaneously steered upward and downward conical beams with a tilted or faceted seabed and multiple layers are also investigated by simulation. Experimental profiles are obtained using two different vertical arrays in smooth and rough bottom sites in the Mediterranean. Correlation peak amplitudes follow the theory and simulations closely.
[DOI: 10.1121/1.2835416]

PACS number(s): 43.30.Pc, 43.30.Ma, 43.30.Nb, 43.30.Re [DRD] Pages: 1282–1296

I. INTRODUCTION

Ocean noise, or just noise in general, can be viewed in many ways. Traditionally, ocean noise is treated as a nuisance, distinguished only by having a spectrum, directionality, and related properties such as spatial and temporal coherence (Urick, 1975). It may also be regarded as chaotic with a prediction horizon confined to a few samples (Frison *et al.*, 1996). Alternatively, one may view ambient sound sources as a complex issue in itself. Experimental work has been done on breaking wave statistics (Ding and Farmer, 1994), and on the influence of white caps in noise production (Cato, 2000) and their spatial and temporal distribution (Melville and Matusov, 2002), and detailed statistical models of breaking wave noise have been built (Finette and Heitmeyer, 1996).

This paper concentrates on using the more broad band, featureless wind noise as a tool to infer something about geoacoustic properties rather than about the noise itself or its sources. Buckingham and Jones (1987) were able to extract the seabed's critical angle from vertical coherence measurements. Recent developments in underwater acoustics suggest that the noise may contain substantially more detailed information than one would think. From the noise (power) directionality alone measured with a vertical array it is possible to determine the seabed's reflection coefficient as a function of angle and frequency (Harrison and Simons, 2002), and with a drifting array one can obtain a (relative depth) subbottom profile (Harrison, 2004). The latter method, which relies on spectral factorization, was explored further by Harrison (2005). One can make use of the spatial coherence of the noise by cross correlating the time series from separated hydrophones to obtain the Green's function from one hydrophone to the other (Roux and Kuperman, 2004). Theory was treated by Roux *et al.* (2005), and the time required for the cross correlation to converge was treated by Sabra *et al.* (2005). Sabra *et al.* (2004) proposed array element localization as an application. Siderius *et al.* (2006) extended this approach to the domain of subbottom profiling by cross correlating the up- and downsteered beam time series from a drifting vertical array. The aim of the current paper is to develop a quantitative formula for the steered beam correlation amplitude in terms of depth, reflection properties, bandwidth, and so on, and to check it by simulation and by reference to experimental results.

To provide a clear demonstration that no special surface coherence properties are required, all the physical noise mechanisms are deliberately stripped out, and the physics is generalized by postulating an environment with many point sources distributed randomly, but uniformly in a horizontal plane, all emitting random time sequences uniformly in angle (disregarding any surface interference effects). Beneath this is a directional receiver, and beneath that a reflecting seabed as shown in Fig. 1. The aim is then to demonstrate the quantitative behavior of the normalized cross correlation $C(\tau)$ between two steered beam time series g_1, g_2,

[a)]Electronic mail: harrison@nurc.nato.int.
[b)]Electronic mail: siderius@hlsresearch.com.

 0001-4966/2008/123(3)/1282/15/$23.00

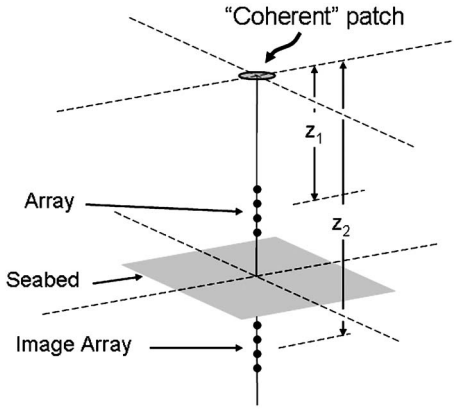


FIG. 1. General geometry showing a coherent patch of noise sources on the sea surface with the actual vertical array and its image reflected in the seabed.

$$C(\tau) = \frac{\int g_1(t) g_2(t+\tau) dt}{\sqrt{\int g_1^2(t) dt \int g_2^2(t) dt}} \tag{1}$$

by estimating the numerator (i.e., the unnormalized cross correlation) and the denominator (i.e., the standard deviations) separately. The values near, and away from, a correlation peak are derived in the Appendix, treating the receiver as a point from which emanates an upward and a downward beam. This yields formulas, which are summarized in Sec. II, in terms of sample rate, bandwidth, and array size.

It is stressed that although the normalization in Eq. (1) seems like a fairly obvious choice, it is by no means the only choice. If, for instance, the g_2 in the denominator were swapped for g_1 the result would be a quantity rather close to a time domain representation of $\overline{g_1(\omega) g_2(\omega)} / \overline{|g_1(\omega)|^2}$, which is a coherent version of the ratio of the downward to upward beam spectral powers, i.e., $|\overline{g_2(\omega)} / \overline{g_1(\omega)}|^2$, as used to determine reflection coefficient by Harrison and Simons (2002). This suggests that in the time domain alternative normalizations could be used to determine, for instance, absolute reflection coefficients, though this will not be pursued here.

In Sec. III the same calculation is approached as a simulation, retaining the processing algorithms already used on experimental data (Siderius *et al.*, 2006). Noise files are simulated for each hydrophone of a vertical array using random number sequences added and shifted in time according to their position relative to the array. The latter approach underscores the "emergent" property of the correlation, since the sources' sequences are entirely incoherent and do not contain any identifiable "clicks" or "splashes." This approach is used to investigate processing techniques, the effect of seabed reflection coherence, bottom tilt, and angle resolution of individual scatterers.

Finally in Sec. IV these findings are applied to experimental data from three vertical array drift experiments in the Mediterranean, two over smooth seabed and two over rough.

II. THEORY

The relationship between the time derivative of the noise correlation function and the time domain Green's functions between hydrophones is well established (Weaver and Lobkis, 2004; Roux *et al.*, 2005) and it is not the main interest here. Instead the paper attempts to construct and justify a function, or functions, and their normalizations that predict something useful about the seabed other than travel times.

A. Peak amplitude formula

In the Appendix it is shown that the numerator (actually peak value) of Eq. (1) can be expressed in terms of the depths of the receiver and its image in the seabed z_1, z_2, the autocorrelation function of the sound sources $C_s(\tau)$, the speed of sound c, and a constant K_0 as [Eq. (A10)]

$$\int g_1(t) g_2(t+\tau) dt = K_0 \frac{c}{(z_2 - z_1)} \int_{\tau - \tau_z}^{\infty} C_s(\tau') d\tau' . \tag{2}$$

The peak value of the time differential of this quantity is derived through the discrete difference Δ operator, the sample frequency f_s, and the reflection coefficient $R(\theta)$ as [Eq. (A14)]

$$\max\left[\Delta \left\{ \int g_1(t) g_2(t+\tau) dt \right\} \right] = K_0 R(0) \frac{c}{f_s (z_2 - z_1)}, \tag{3}$$

where "max" means the maximum of the absolute value multiplied by sign, remembering that the impulse could be negative. The denominator of Eq. (1) can be expressed in terms of the same constant K_0, the number of hydrophones M, a numerical constant β dependent on noise coherence and shading, as [Eq. (A22)]

$$\sqrt{\int g_1^2(t) dt \int g_2^2(t) dt} = K_0 |R(0)| \frac{\beta}{M}. \tag{4}$$

So the final peak value of the cross correlation is [Eq. (A25)]

$$\max[\Delta\{C(\tau)\}] = \frac{2L \operatorname{sign}(R)}{(z_2 - z_1) \gamma \beta}, \tag{5}$$

where L is the array length and γ is the ratio of sample frequency to design frequency for the array: $\gamma = f_s / f_0$.

The main dependence of the peak correlation height is on array length and its separation from the seabed. Although the peak depends on the sign of the reflection coefficient it does not depend on the reflection strength. A simple explanation for this can be seen through the three areas shown in Fig. 2. The numerator of Eq. (1) depends on noise received in the small area in which the times of arrival differ by a few samples ($\sim 1/f_s$). Applying the Fresnel approximation to the geometry shown in Fig. 1 one can see that this area is of order $2\pi(c/f_s) z_1 z_2 / (z_2 - z_1)$. In contrast, the denominator of Eq. (1) depends on noise received in the larger areas illuminated by the upward and downward endfire beams. Since the endfire beam width is roughly $\sqrt{2/M}$, these areas are, respectively, $2\pi z_1^2 / M$ and $2\pi z_2^2 / M$, and their geometric mean is $2\pi z_1 z_2 / M$. The order of magnitude peak value depends on the ratio of the small area to this geometric mean area, i.e., $Mc / \{f_s (z_2 - z_1)\}$, and this clearly reduces to the quantity evaluated in Eq. (5).

As estimated in the fourth section of the Appendix, for a 32-element array 30 m above the seabed with 12 kHz sam-

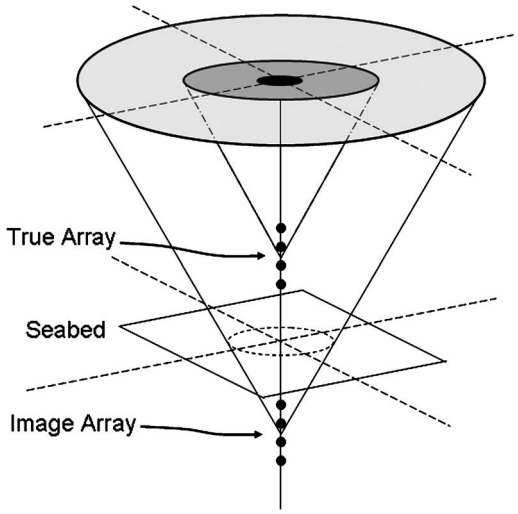


FIG. 2. The three important areas of noise sources on which the normalized cross correlation depends. The numerator of Eq. (1) depends on the small coherent area; the upbeam and downbeam parts of the denominator (standard deviations) depend, respectively, on the intermediate and largest areas.

pling, 4166.7 Hz design frequency (0.18 m hydrophone spacing), and minimal filtering, a peak height of ~0.016 is expected.

Interestingly, because of the normalization, the magnitude does not depend on the reflection coefficient R or frequency band particularly. However it will be seen in Sec. IV that there are geographic variations of "echo" strength in the experimental data. So an explanation will be sought. An important point is that the correlation peak as estimated here is an "emergent" property (see, e.g., Sabra *et al.*, 2005) that does not depend on any source coherence, i.e., clicks or splashes.

The above-presented explanations and the derivation in the Appendix assume the noise to have a flat frequency response. It is shown in the Appendix that as long as the near zero frequencies are excluded there is another processing option using a Hilbert transform with, or without, the time differentiation. The price paid for this more robust solution is loss of information on sign of the reflection, which translates to an inability to distinguish between increase and decrease of acoustic impedance.

Generally, if one is only interested in the timing of the layering rather than its exact impulse response there are more options. Since taking the time derivative is equivalent to multiplying by frequency in the frequency domain, and the noise has its own initial spectrum, there is some freedom in the prefiltering of the signal and in postfiltering the time-domain correlation function. According to Weaver and Lobkis (2004) "the spectral power density of the diffuse field may be thought of as a filter and the source of some distortion between the time derivative of the Green's function and the field-field correlation function." Indeed it is straightforward to demonstrate numerically that, for instance, multiplying the cross spectral density by frequency (postfiltering) is identical to multiplying the two initial noise signals by the square root of frequency (prefiltering). Thus no generality is lost in the definition, Eq. (1), if the time series of the up- and downsteered beams $[g(t)]$ are allowed already to be filtered.

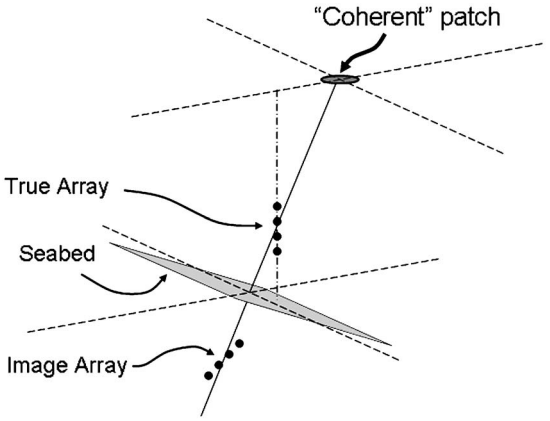


FIG. 3. Geometry equivalent to that in Fig. 2 but for a tilted seabed, showing the tilt of the array image and the horizontal offset of the coherent patch.

Furthermore, in these authors' experience one can obtain depth-accurate subbottom layer profiles from noise correlation without worrying about detailed filtering as long as the lowest frequencies are extracted, but more care needs to be taken to obtain other properties such as reflection coherence.

B. Criterion for detecting a bottom "echo" in uncorrelated background

Away from the correlation peak value there is a decorrelated background, and it is shown in the Appendix that, if the number of independent samples in the cross correlation is N_s, then the background level for the normalized cross-correlation function [Eq. (1)] is $N_s^{-1/2}$. Thus the criterion for detecting a sediment layer or "echo" is that the quantity in Eq. (5) must be greater than $N_s^{-1/2}$, preferably a lot greater. To detect the peak height estimated in Sec. II A (0.016) one needs about 4000 independent samples. Conversely a 10 s random time series sampled at 12 kHz results in a background level of $(120\,000)^{-1/2}=0.0029$, i.e., about one-fifth of the peak value. Of course, in practice to obtain a useful impulse response containing many peaks, one requires the background to be below the weakest peak, so ultimately there is a constraint on the number of samples required and the relative motion of the receiver and target.

C. Additional layers

The addition of a layer is a trivial extension to the theory. It can be seen that the standard deviation of the noise (denominator) responds to the combined reflected power from all the layers whereas the peak value only responds to the layer at the delay of the peak in question. Therefore the cross correlation is a proportional representation of the seabed's impulse response; the layer echoes scale in the same way as with an echo sounder, and they retain their signs.

D. Bottom tilt, tilted facets

In order to see the effects of reflecting facets a tilted plane seabed is considered first. The geometry in Fig. 3 shows that there are still three areas determining the normalized cross correlation. The small coherent area is centered on the point where the projection of the line joining the receiver

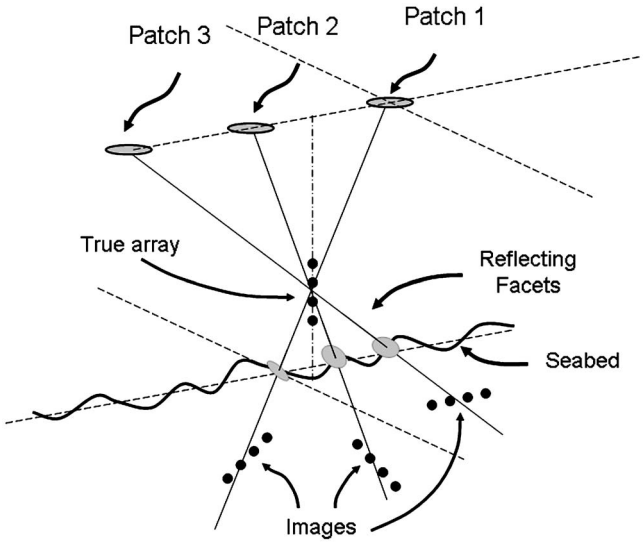


FIG. 4. Images and coherent areas for a rough faceted surface with multiple specular reflections. The three example facets each have their own corresponding tilted image array and coherent patch.

and its image meets the sea surface. However it can only contribute if the upward and downward beams are steered appropriately. For example, with a bottom slope α this coherent patch appears in the center of the upward beam steered to α (from vertical). The image array is tilted at 2α to the vertical, as shown in Fig. 3, so the coherent area appears in the center of the image array's beam (also steered at α from the vertical). It is therefore appropriate to use the same steer angle for the upward and downward beams. The incoherent areas contributing to the denominator of Eq. (1) then depend on the steer angles, but despite quite complicated geometry, the dependence is weak. In the case of a rough surface composed of facets with a distinct Fermat extremum path as in Fig. 4, the above-mentioned reasoning can be applied to each facet. Thus one expects a correlation peak for each correctly (i.e., specularly) oriented facet. One also expects the peak to be resolvable in steer angle given adequate array aperture.

E. Bottom reflection coherence

The analysis in the Appendix assumes that the reflecting layer is perfectly flat. It also assumes that the bottom structure does not change during the collection of the samples to be correlated. In one of the experimental examples of sec. IV it is suspected that neither of these assumptions is true. In principle, given a rough surface, stationary geometry, and unlimited time, one might still expect performance comparable with an echo sounder even though the reflector is not specular. According to the Rayleigh criterion (Brekhovskikh and Lysanov, 1982) if the vertical roughness scale is greater than $\lambda/4\pi$ the surface behaves like many reflecting facets as seen in Fig. 4. By itself this simply spreads the energy of the single peak (expected for a specular reflector) amongst several others. Thus the echo is smeared in time. If the receiver is moving, these nonvertical arrivals shift in travel time and therefore blur. By considering a tilted specular reflector one can see that when the roughness scale is very large the motion-blurring effect is quite distinct from the roughness-smearing since it depends on drift speed. Rather than attempt to explain these effects quantitatively they are demonstrated by simulation in Sec. III.

F. Sound source coherence

So far it has been assumed that the sound sources are completely incoherent. Here the effect of significant source coherence is considered. Imagine a single impulsive source at the sea surface with no other background. The normalized cross correlation will have peak value unity [i.e., much greater than predicted by Eq. (5)], since it is normalized. Therefore it is conceivable that from time to time, or in particular weather states, there can be another more obvious, and nonemergent, correlation mechanism. This is not investigated further but it will be borne in mind in Sec. IV.

III. SIMULATION

A. Data generation

Having established formulas for peak height and background the cross correlation is now investigated by simulation. The approach here differs from earlier simulations (e.g., Siderius *et al.* (2006)) in placing more emphasis on the randomness of the sources than on ducted propagation. The geometry of the array and sources is as in Fig. 1, but the orientation of the seabed will vary from case to case. The sea surface sources are spread uniformly but randomly within a circular area centered on the extrapolated line through the receiver and its image (see Figs. 1 or 3, as appropriate). From each point emanates a unit variance, Gaussianly distributed random sequence of 131 072 ($=2^{17}$) samples. Each sequence is assumed to be sampled at 12 kHz and to propagate from this surface point to each of the hydrophones on the array and their images. The time series are delayed according to geometry by phase shifting in the frequency domain. The 32-element array has hydrophone separation 0.18 m (design frequency 4167 Hz) and is centered at depth 50 m in 80 m of water. It is well known that the sum of power contributions from monopole sources on an infinite flat surface does not converge (Harrison, 1996) unless there is some loss mechanism. In this simulation, partly for this reason and partly for the sake of realism, dipole sources are assumed, and otherwise convergence is ensured by relying on the steered beam directionality. Unless otherwise stated the radius of the circular area of sources is 150 m. The end result is a file containing 32 time series approximately 11 s long, one for each hydrophone. To study the effects of coherently or incoherently adding the correlation functions, the whole process was repeated with new random number seeds to form 81 files.

B. Data processing

Data processing for the simulated time sequences is identical to that already used for experimental data. Generally the time series for each hydrophone is filtered, then it is time-domain beam-formed with hamming shading, and finally cross correlated in the frequency domain, and differenced to form a finite difference time differential. If the time

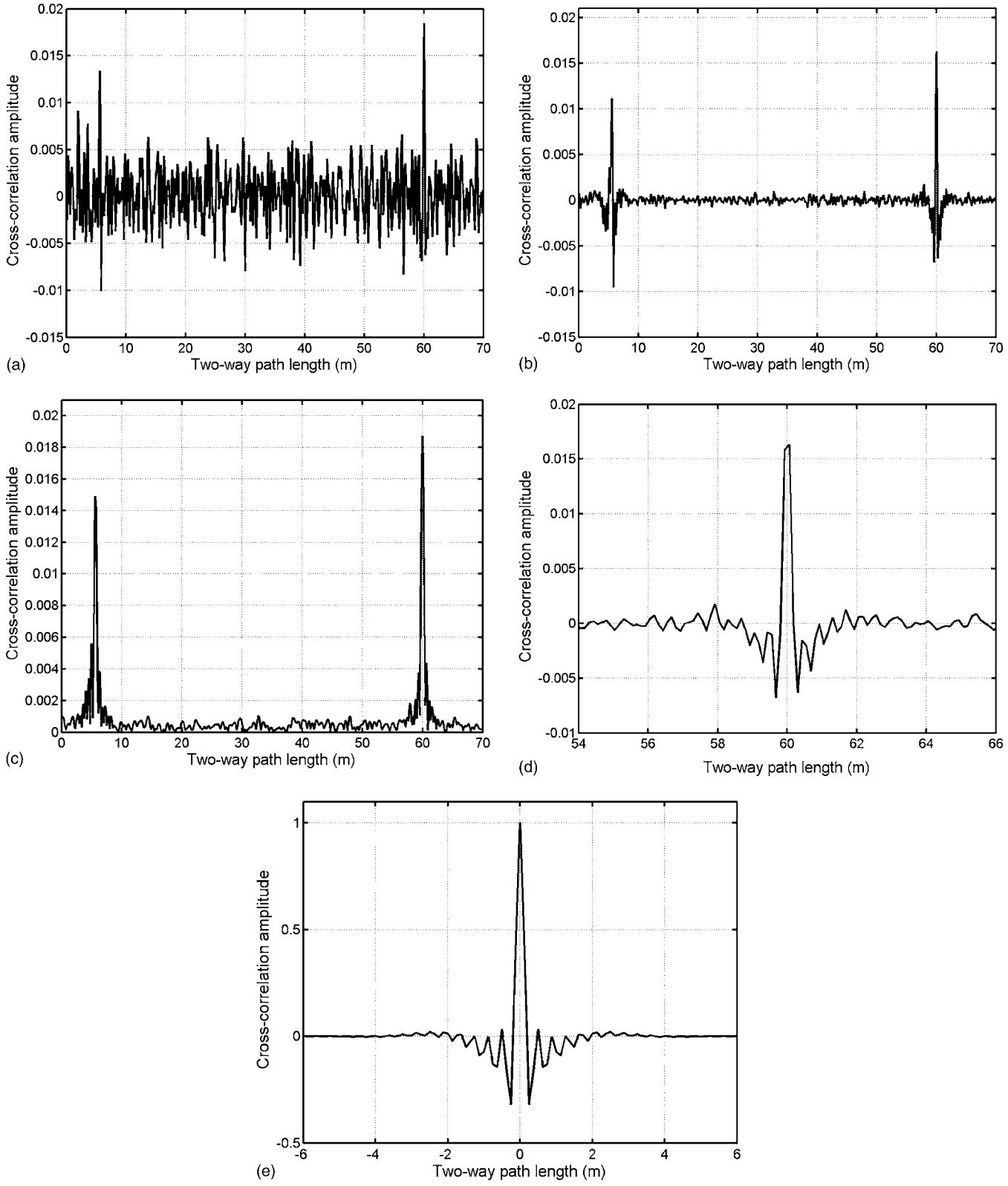


FIG. 5. Simulation of the cross correlation resulting from (a) a single 11 s file showing a peak at two-way travel time corresponding to 60 m, (b) 81 coherently added 11 s files, (c) the Hilbert transform corresponding to 81 files, (d) a blow-up of the peak in (b), and (e) the impulse response of the initial bandpass filter.

series is to be Hilbert transformed then the initial filter is used to exclude the near zero frequencies. Since the spectrum of the simulated sources is already flat there is no need to normalize the spectrum, but it is important to cut out frequencies close to and above the design frequency. For this reason a bandpass filter between 400 and 3900 Hz was used.

C. Test cases

1. Horizontal plane reflector

Figure 5(a) shows the time differential of the cross correlation between the "up" and "down" vertical beams evaluated according to Eqs. (1)–(5) for a single file (lasting 11 s).

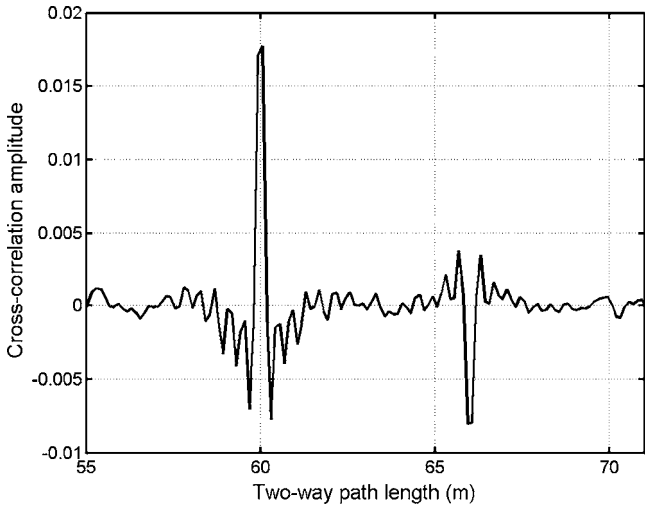


FIG. 6. Simulation of the cross correlation resulting from two layers, one at 30 m below array center with a reflection coefficient of 0.1, the other at depth 33 m with a reflection coefficient of −0.05.

Figure 5(b) shows the reduction in the background on coherently adding 81 files (approximately 15 min), and Fig. 5(c) shows its Hilbert transform. In both cases the peak is at a delay corresponding to 60 m (i.e., z_2-z_1) and its height is approximately the same in both cases ~0.018, which agrees closely with the earlier estimate in Sec. II A and is independent of the assumed reflection coefficient which was 0.1. At much shorter range (5.7 m) there is apparently another peak. This is an artifact that corresponds to the physical length of the array, as can be verified by removing or altering the beam shading.

The standard deviations of the background levels in Figs. 5(a) and 5(b) are, respectively, 0.0029 and 0.000325. Their ratio (=8.9) clearly follows the $N^{1/2}$ prediction of $\sqrt{81}=9$. Low-pass filtering at 3.9 kHz of a signal sampled at 12 kHz broadens the autocorrelation peak to approximately three samples, so the number of independent samples in a single file is 131 072/3 and the absolute background level $N^{-1/2}$ is 0.0048, which agrees well with Fig. 5(a).

A blow-up of the peak arrival shape is shown in Fig. 5(d), and this can be seen to be almost identical to the impulse response of the initial bandpass filter [Fig. 5(e)], as one would expect.

2. Additional plane reflector

A second layer at sediment depth 5 m with reflection coefficient −0.05 is simulated by adding the appropriately delayed source sequences to the existing layer response. No attempt is made to account for multiple reflections in this demonstration. Adding 81 files coherently two impulses can be seen in Fig. 6, the first corresponding to the reflector at depth 80 m (round-trip path length 60 m) with $R=0.1$, the other corresponding to the second reflector at depth 83 m (round-trip path length 66 m) with $R=-0.05$. Clearly the delays and the relative peak amplitudes, including the sign, are correctly reproduced.

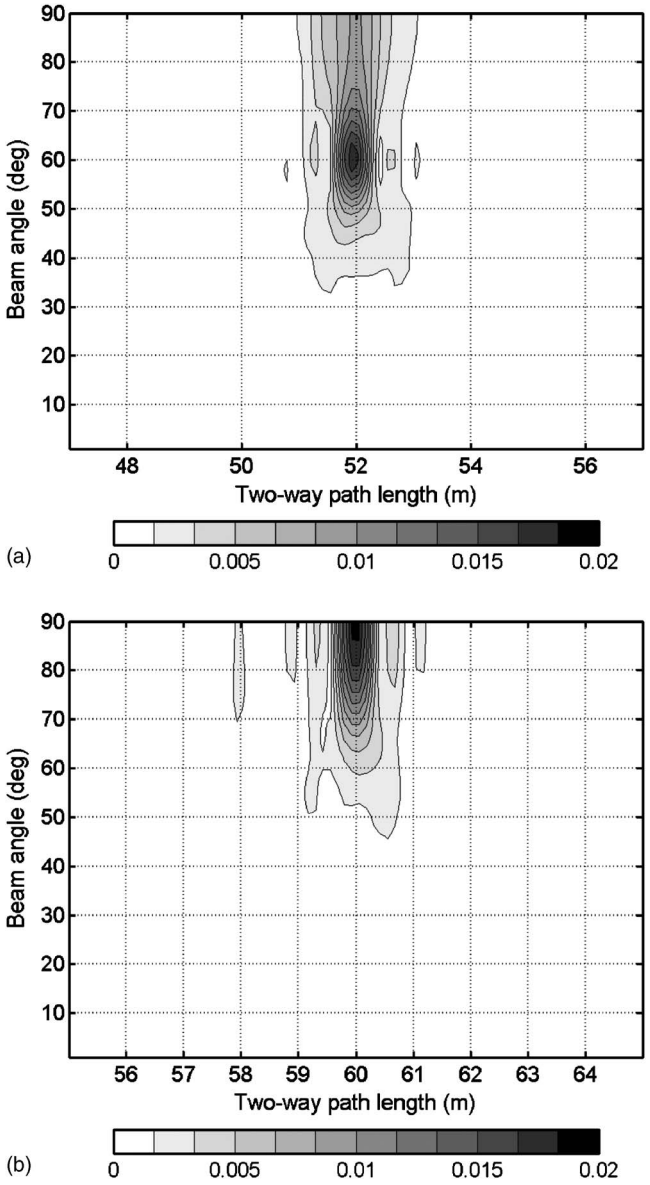


FIG. 7. Correlation amplitude against round-trip path length and steered beam angle for (a) seabed tilted at 30° and (b) horizontal seabed.

3. Tilted plane reflector

To make the point, a large tilt angle of 30° is assumed, so the image array is displaced as shown in Fig. 3, and the area of sources is centered on the small coherent area which is similarly displaced. The result is dependent on steer angle as well as delay and is shown in Fig. 7(a). Since the water depth at the array location is still 80 m the path length to the peak is $2\times30\times\cos(30)=51.9$ m. The peak in Fig. 7(a) agrees with this delay, and also it is centered on a steer angle of 60°. It is instructive to compare this with the corresponding plot for the horizontal seabed Fig. 7(b). This shows a peak at the obvious delay and angle. In these examples angle resolution is relatively poor because the simulation is of an existing realistic system. However there are no restrictions on improving the resolution by increasing the number of hydrophones. According to Eq. (5) (which, of course, already includes array gain effects) this will also increase the peak height.

One might consider extending simulation to a point target. Interestingly, this requires no extra work since the results would differ from the tilted plane reflector only by a time shift and a calculable change in amplitude. Suppose the plane reflector is removed and the image receiver is replaced with a point target. The downbeam path then goes from source to target to receiver, and the upbeam path is unchanged. All that has changed is the addition of a constant delay between the target and the receiver. Thus a point target in this orientation would appear as the peak in Fig. 7(a) but with amplitude according to its target strength. There is, in fact, strong experimental evidence in the second and third examples of Sec. IV that targets can be detected by this method. Note that it is the particular normalization of Eq. (5) that makes the bottom peak height independent of the seabed reflection coefficient. Other reflectors or scatterers will be reduced in proportion.

4. Rough reflector

In principle it would be possible to extend the current numerical simulation to a rough surface by exchanging the downward specularly reflected path for the many paths connecting each sound source with each hydrophone via a large number of scattering facets [for instance, using the Kirchhoff approximation (Brekhovskikh and Lysanov, 1982)]. Because of the large computation time a simpler approach is preferred, since in this context the only interest is in the effect of decorrelation on peak height. For similar reasons horizontal motion of the array is neglected. According to the Rayleigh criterion the coherence is affected only by the vertical scale of the roughness compared with the wavelength. So a crude way to model this is to add a zero-mean, Gaussianly distributed distance with variance σ^2 to the path difference for each sound source. Because the sources are bandpass filtered with the low pass at 3900 Hz the limiting "roughness" is expected to be $c/(4\pi 3900)=0.18/(2\pi)\sim 0.03$ m. A set of roughnesses σ was chosen between 0 and 0.5 m, and 81 files generated for each. Selecting a single file for each roughness it is difficult to see much dependence on σ because the sample length is not long enough for convergence. The effect is clearer after coherent integration over the 81 files. The change in peak amplitude is plotted against σ in Fig. 8; the symbols indicate amplitude with, and without, Hilbert transformation. The main effect of the roughness in the time domain is a time smearing, so one might expect smearing proportional to the roughness and peak height proportional to the peak width or some power of it. Superimposed on the plot is an exponential fit and a power law fit. These have no significance other than to reinforce the fact that the peak height is more or less inversely proportional to the roughness (as modeled here).

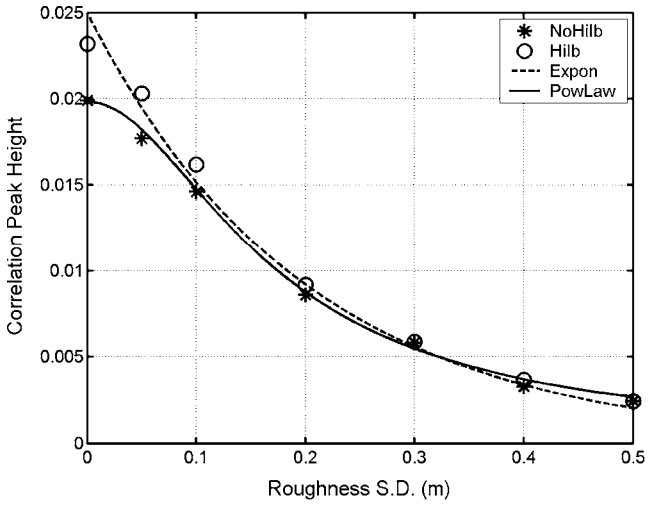


FIG. 8. Plot of correlation peak heights against roughness parameter σ with (○) and without (*) Hilbert transform. Two possible line fits are shown.

IV. EXPERIMENTAL DATA

Three experiments have been carried out in the Mediterranean using a drifting vertical array (see Fig. 9). The first two started from more or less the same place on the Malta Plateau, a smooth layered sediment seabed, south of Sicily (Site 1). In April 2002 32 elements at 0.5 m separation (design frequency 1500 Hz) were taken from the center portion of a 62 m nested vertical array (VLA), and a drift of 11 h resulted in a 9 km track. In July 2003 a medium frequency array (MFA) with 32 elements spaced at 0.18 m (design frequency 4167 Hz) drifted for 13 h resulting in a 6.5 km track. In May 2004 the second array drifted on two occasions (12 and 13 May) over parts of the Ragusa Ridge, a very rough rocky area with two main ridges and many sediment filled pools. The first drift covered 5 km in 10 h; the second covered 14 km in 14 h. As ground truth, seismic boomer layer profiles are available near the 2002 drift, and as accurately as possible, exactly along the 2003 drift track. A chirp sonar is available along the 2004 drifts, however it shows little, if any, bottom penetration. Better detail of the bottom roughness is shown by side-scan sonar. The noise data collected in

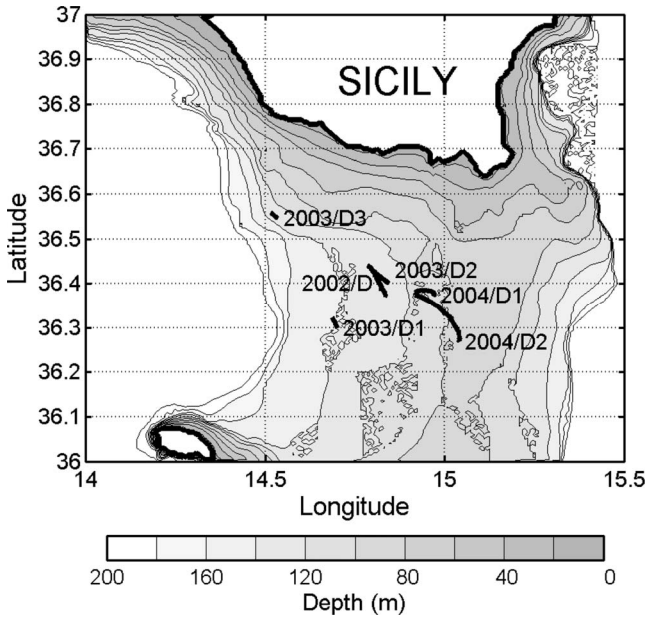


FIG. 9. Map showing the three drift experiments. The 2002 track using a 62 m VLA on the Malta Plateau is labeled 2002/D. The 2003 track using a 6 m MFA on the Malta Plateau is labeled 2003/D2. The two tracks using the MFA on the Ragusa Ridge are labeled 2004/D1 and 2004/D2.

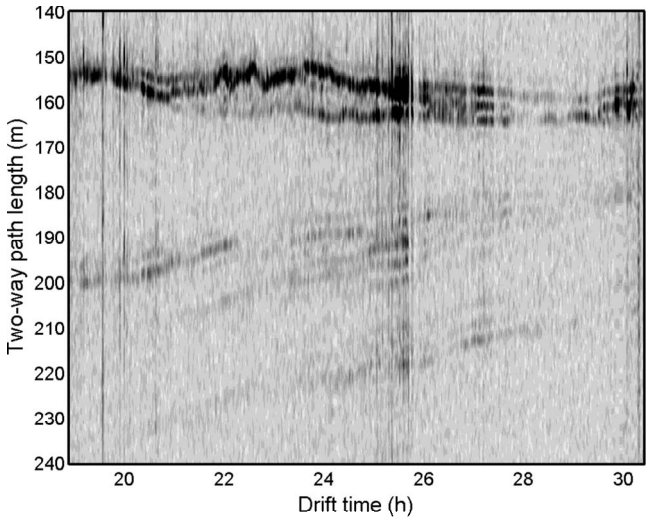


FIG. 10. Subbottom profile from drifting VLA on the Malta Plateau (2002) showing deep echoes.

2002 have been reported in the context of a different processing technique (Harrison, 2004); the 2003 data were also discussed in that context (Harrison, 2005) and in the context of cross-correlation techniques (Siderius *et al.*, 2006). Favorable comparisons have already been made between noise inversion techniques and the various ground truths.

A. Malta Plateau 2002

Each file is approximately 11 s long (65 536 samples at a sampling rate of 6 kHz) so there is the freedom to analyze file by file or to concatenate contiguous files (or equivalently add the processed results coherently) or to smooth the resulting profile. Here it is chosen to analyze file by file and then to process in various ways. Figure 10 shows a profile where a postprocess horizontal smoothing has been applied (incoherent over about 10 files). The seabed is seen at a two-way path length of about 160 m from array center. As well as strong layering in the first 5 m (10 m two-way path as shown) there are clear indications of deep layers at 25 and even 40 m (i.e., 50 or 80 m longer path than the seabed's). Bearing in mind that these calculated depths are simply travel times converted with sound speed in water (assumed 1500 m/s) the actual layer depths are likely to be somewhat greater.

A typical Hilbert transformed correlation amplitude showing a strong, deep second layer echo (at drift time 19:12:00) is shown in Fig. 11(a). A blow-up of the main peak with Hilbert envelope is shown in Fig. 11(b). Another example from 27:00:00 shows a triple echo (see Fig. 10) and its Hilbert envelope. Because the processing used a narrower band (half the design frequency) than the simulated example in Fig. 5 the impulse response is slightly oscillatory, but even so, in Fig. 11(c) it is possible to see differences in phase or sign in the three echoes.

The peak amplitude (averaged over 100 files) is slightly variable throughout the drift (Fig. 12) with a mean between about 0.02 and 0.03. The expected value from Eq. (5) with $\gamma=6000/1500$, $\beta=1.87$ (bandwidth is half the design frequency), $z_2-z_1=150$ m is 0.0285, and this agrees well with

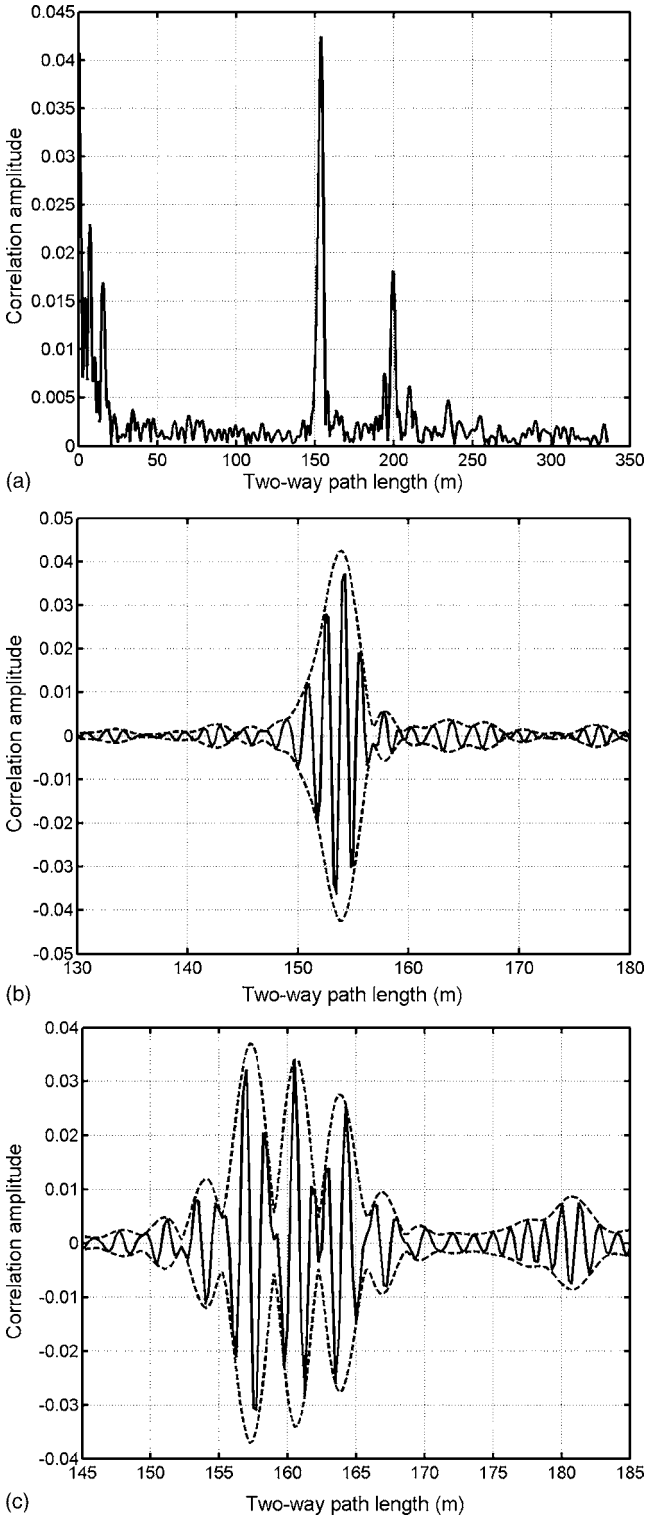


FIG. 11. (a) Hilbert transformed impulse response (average of 100 files) during the 2002 drift (time 19:12:00), (b) blow-up of the main peak showing correlation amplitude with Hilbert envelope, and (c) a triple echo with Hilbert envelope (27:00:00).

the experimental mean. One might expect a slight upward bias of the experimental data in this kind of presentation because the weak peaks are never seen since they are swamped by the background. The variation, in itself, is no surprise and could in principle be averaged out with a stationary array. It is conceivable that from time to time an

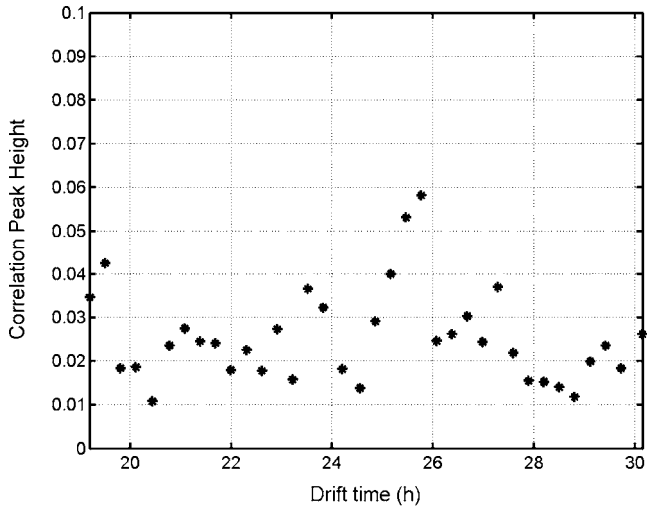


FIG. 12. Correlation peak amplitude (averaged over 100 files) vs drift time during the 2002 drift with the VLA.

exceptionally coherent "clap" from an individual wave might deviate strongly from the normal background "hiss" of wind noise. From the nature of Eq. (1) it is clear that correlation peak heights increasing right up to one are mathematically allowed; they are just extremely unlikely with wind sources.

B. Malta Plateau 2003

Figure 13 shows the profile resulting from the drift of the MFA in 2003. Detailed comparisons have already been made with the profile generated by a seismic boomer subsequently towed along the same track (Siderius *et al.*, 2006). Again two-way paths of 50 m, indicate layer depths of at least 25 m.

A typical Hilbert transformed correlation amplitude (averaged over 100 files around 23:00:00) is shown in Fig. 14(a). Each file is approximately 10 s long (122 880 kHz samples at a sampling rate of 12 kHz). The double peak with Hilbert envelope is blown up in Fig. 14(b). Although the design frequency is now 4167 Hz the relative band is the same as in the VLA case and so the impulse response has the same shape. Again, despite its complexity the phase of the impulse response relative to the envelope can be distinguished.

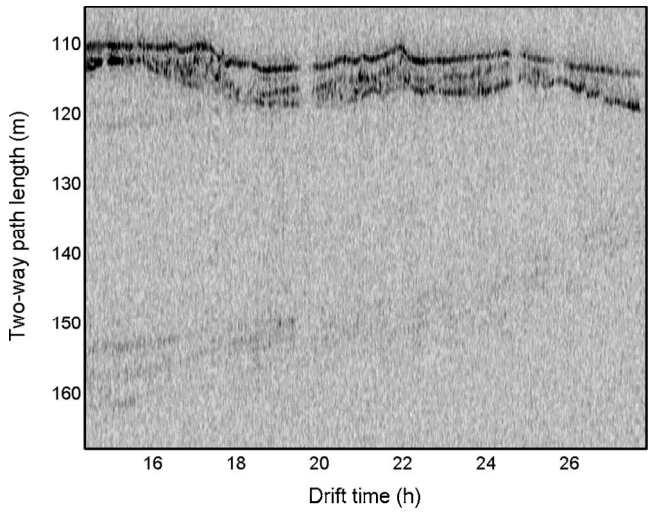


FIG. 13. Subbottom profile from drifting MFA on the Malta Plateau (2003) showing deep echoes.

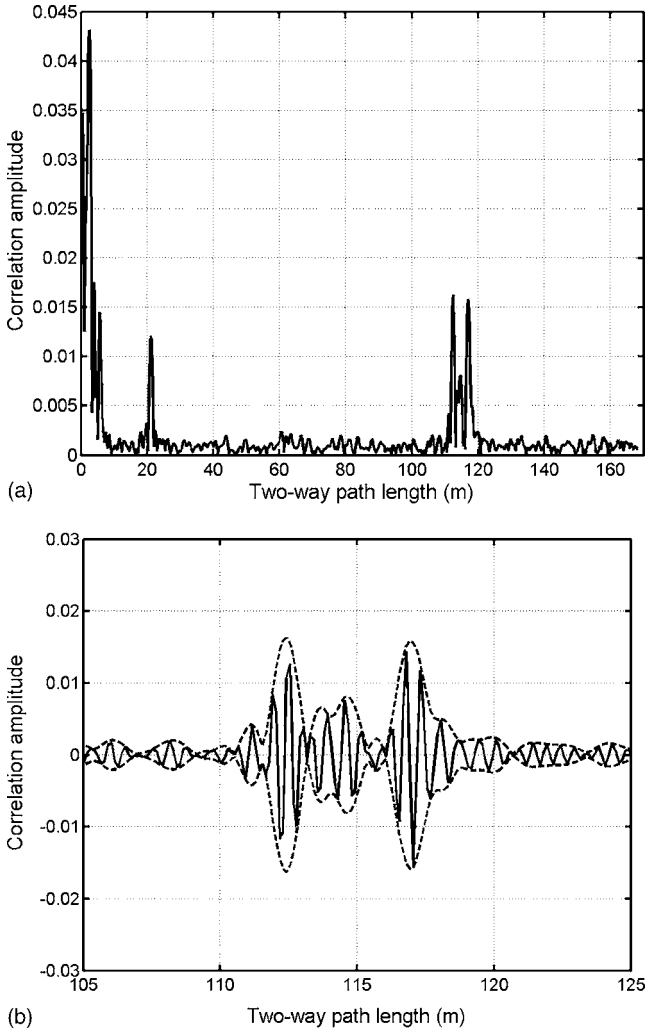


FIG. 14. (a) Hilbert transformed impulse response (average of 100 files) during the 2003 drift (time 23:00:00) and (b) blow-up of the main peak showing correlation amplitude with Hilbert envelope.

In passing, it is interesting to note that the peak at 21 m (two-way path) in Fig. 14(a) is persistent throughout the 13 h drift and is believed to be a reflection from the weight at the bottom of the array. There is an equivalent peak in the 2004 measurements but at 24 m, probably because of minor differences in cable length. The fact that the same equipment and processing was used on the two occasions suggests that this is not a processing artifact but a true target detection using beam–beam cross correlation of noise.

Variation of peak amplitude with drift time is shown in Fig. 15. The mean is between about 0.021 and 0.015. The expected value according to Eq. (5) with $\gamma = 12\,000/4166.7$, $\beta = 1.87$ (bandwidth is half the design frequency), $z_2 - z_1 = 110$ m is 0.0194, and again this agrees well with the experimental mean.

C. Ragusa Ridge 2004

These two drifts were a deliberate attempt to see how various noise inversion techniques would fare with a rough

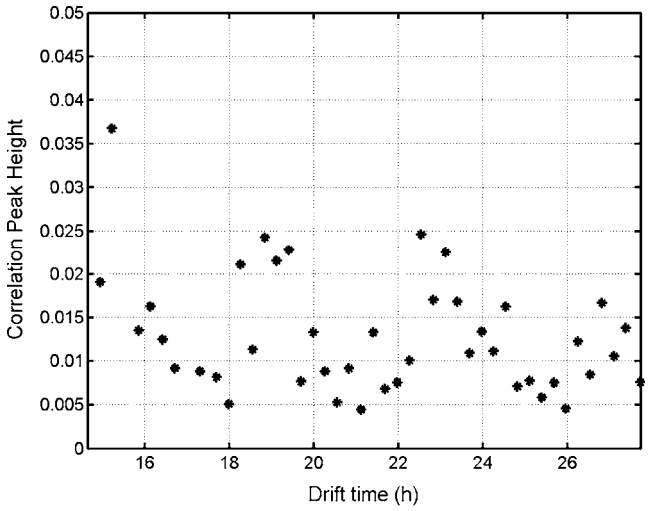


FIG. 15. Correlation peak amplitude (averaged over 100 files) vs drift time during the 2003 drift with the MFA.

seabed. A qualitative indication of the roughness is shown by the sidescan image in Fig. 16. The scales are known to be of order 1–10 m in the vertical and 10–20 m in the horizontal. Thus a coherent average along a drift track of, say 100 m, could be subject to very large vertical roughness excursions compared with those considered in Fig. 8.

From the point of view of cross-correlation techniques there were some additional potentially undesirable problems with acquisition and drop-outs, not to mention occasional nearby ships with singing propellers. On top of this, strong winds, which are usually ideal for generating sound, produced discrete, audible crashing waves. Nevertheless performance was distinctly better on the first day (12 May) than on the second (13 May). The reasons for this will become clear, and fortuitously provide some insight into the conditions under which a moving array can work.

1. Ragusa Ridge 2004 drift1

Figure 17(a) shows the profile obtained over 8 h. The blow-up in Fig. 17(b) emphasizes the variability in strength at the latter end of the drift and resembles an echo sounder record of a rough surface. The flatter sections in Fig. 17(a) between 20:00 and 21:00 and near 16:00 are thought to be small pools of sediment between rock outcrops, and one can see evidence of a weak second reflection a few meters later at both times.

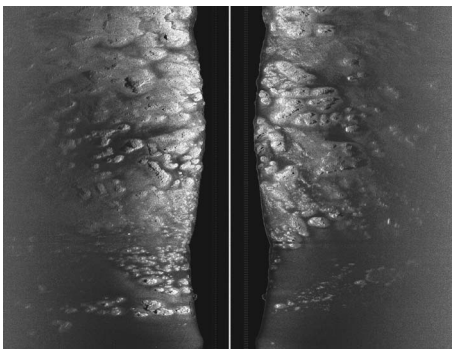


FIG. 16. Side-scan sonar image of part of the Ragusa Ridge showing features of 10–20 m in horizontal extent. Full cross-track range is 430 m.

There is a weakening of the echo in the central part of Fig. 17(a) which, from the considerations of Secs. II and III, cannot be caused by geographical changes in reflection coefficient, although they could be caused by changes in roughness. There were also no changes in instrumentation problems or weather conditions. The probable cause is suggested by the performance during the second drift (a day later).

2. Ragusa Ridge 2004 drift2

The 13 h second drift starting at almost the same location is shown in Fig. 18. The echo is so weak compared with the background that the contrast needed to be adjusted in order to see the bottom echo at all.

If anything, the instrumentation problems and weather conditions were less severe than during the first drift, but a clue as to the most likely cause of this varying performance is the relative lengths of the drift tracks (see Fig. 9) which were, after all, obtained for comparable durations. The average drift speed was just over twice as fast on the second day as on the first, as shown in Fig. 19.

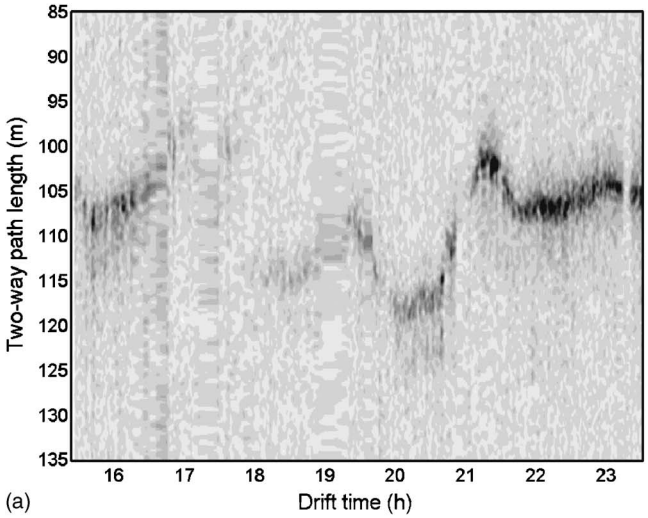
The array's drift is driven by currents rather than wind since there is about 50 m of cable and array hanging vertically. In this area current variations of this magnitude from day to day are common (Lermusiaux and Robinson, 2001). As already discussed (Sec. II) the speed of the drift has little effect when the reflecting surface is flat since the Fermat travel time is independent of position, but when the surface is rough the Fermat path changes rapidly with position (and may be multivalued) so it may not be possible to average for long enough in each position for numerical convergence. The prime suspect for the weak echoes in this case is therefore the drift speed.

In retrospect the fade in the middle of Fig. 17(a) can also be attributed to variation in speed. Although the average speed for the 5 km was about 0.14 m/s there was some variation on 12 May as shown in Fig. 19. The rise in speed between 17:00 and 21:00 is closely matched by the fade. The smoother, flat bottomed section between 20:00 and 21:00 would be expected to survive by being more tolerant to high drift speed. In one sense this fading because of drift speed is a limitation of the technique when the surface is rough. In another sense it is a strength since the fading is unambiguously associated with drift speed and the fluctuations caused by a rough surface. As described here, this association is qualitative, but the effect is, in principle, quantifiable.

Although the reflecting surface is rough, steep angle returns appear not to be steep enough to register outside the rather broad endfire beam of this experimental arrangement. However the considerations of Sec. III suggest that, in principle, angle discrimination is possible given more amenable array designs.

V. CONCLUSIONS

The main point of this paper has been to understand the amplitude of the cross correlation between steered beams of



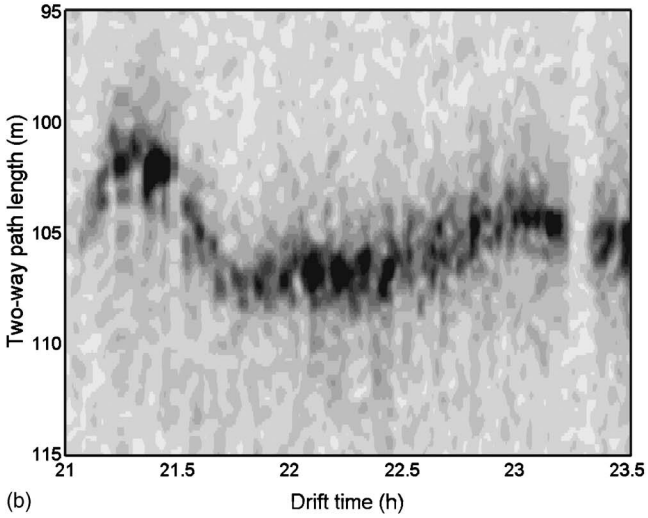


FIG. 17. Subbottom profile from the first MFA drift track on the Ragusa Ridge (2004) (a) showing fading bottom return during the centre portion and (b) a blow-up of the last 2 h.

ambient noise. The proposed normalization [Eq. (1)] results in a formula for peak height given by Eq. (5) which depends on the array size, the distance from the array to the reflector, the ratio of sample frequency to design frequency, but not the reflection coefficient, though relative strength of layers and their signs are retained, and there is sensitivity to the surface roughness through the reflection coherence.

A detailed theory is developed in the Appendix and summarized in Sec. II. This leads to a criterion for detecting a bottom echo in an uncorrelated background and an understanding of the effects of surface roughness, multiple layers, and tilted surfaces.

By representing the sheet of surface sources as many random time sequences emanating from random locations on a plane it was possible to simulate the direct and bottom reflected arrivals at the hydrophones of a vertical array. These were subsequently filtered, beam formed, and cross correlated using exactly the same algorithms as used for experimental data to confirm the theoretical predictions in all the earlier cases.

Finally experiments from three separate sea trials in the Mediterranean using two different arrays over smooth and

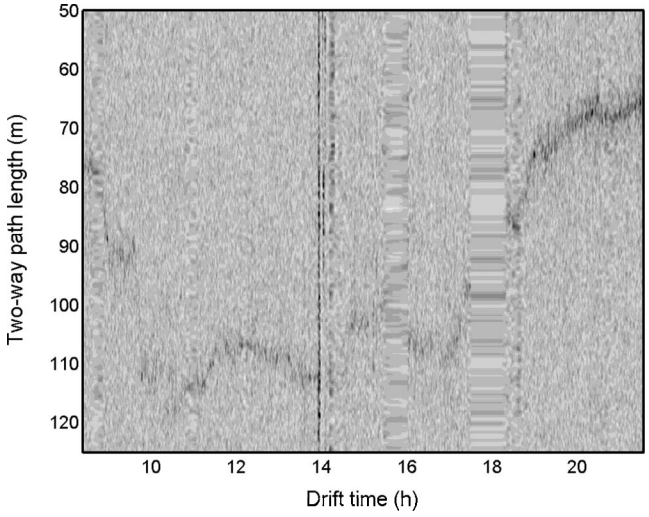


FIG. 18. Subbottom profile from the second MFA drift track on the Ragusa Ridge (2004) showing very weak bottom returns throughout.

rough seabed were processed to show bottom profiles and to investigate the correlation peak amplitudes versus drift time. By comparing the time differential of the beam–beam cross correlation with its Hilbert transform it was possible to distinguish phase changes between the layer reflections despite the rather oscillatory impulse response resulting from the prefiltering. The experimental amplitudes match the theoretical predictions well.

Obviously the correlation results are improved by longer integration times if the array and environment are stationary. When the bottom is smooth, horizontal motion of the array produces minimal effects, and therefore long integration times (coherent integration of many files) are feasible. When the bottom is rough, long integration times only enhance the echo if the array is motionless. The examples over the Ragusa Ridge exhibited large differences in drift speed during the two experiments that were clearly correlated with varia-

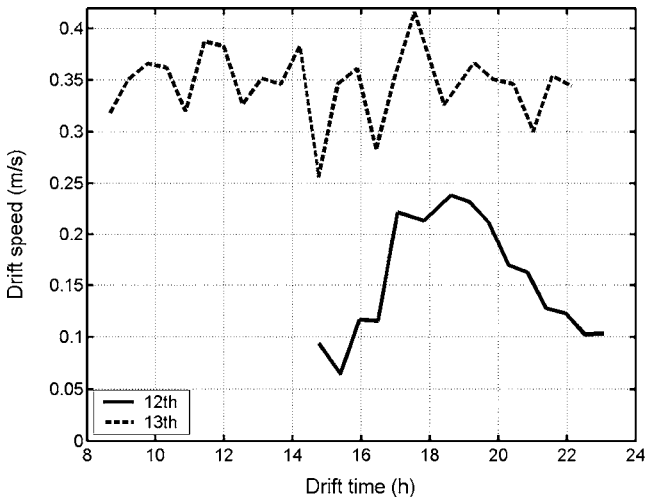


FIG. 19. Drift speed plotted against drift time for the two drifts over the Ragusa Ridge showing the marked difference and also the change during the first drift (12 May).

tion in echo strength. Thus a weak echo is a clear indication of a rough surface, regardless of the detailed motion or roughness dependence.

ACKNOWLEDGMENTS

The authors thank the Captain and crew of the *NRV Alliance*, Luigi Troiano and Enzo Michelozzi, for engineering coordination, Piero Boni for data acquisition, and particularly Peter Nielsen who acted as Scientist-in-charge during the three cited BOUNDARY experiments. They are also indebted to Roberto Rossi and Mark Prior for the side scan image.

APPENDIX: DERIVATION OF CROSS-CORRELATION PEAK AMPLITUDE

In this Appendix the numerator and denominator of the beam–beam cross correlation, Eq. (A1), are derived separately in terms of the source positions and physical delay times,

$$C(\tau) = \frac{\int g_1(t) g_2(t+\tau) dt}{\sqrt{\int g_1^2(t) dt \int g_2^2(t) dt}}. \tag{A1}$$

The physics of the source wave form is deliberately simplified in order to separate out the effects due to delay time and "pure" randomness.

1. Definitions

At each randomly located source point n on the surface a time series $s_n(t)$ is emitted. This is uncorrelated with emissions from any other point m, i.e.,

$$\int s_n(t) s_m(t+\tau) dt = 0 \tag{A2}$$

for $m \neq n$ and all τ greater than some limiting value τ_0. In other words the time series though random may be spectrally "pink." Taking the integral over a time T, $s_n(t)$ is related to its standard deviation through

$$\int s_n(t) s_n(t) dt = \sigma^2 T \tag{A3}$$

and the normalized autocorrelation function of the individual sound sources is therefore given by

$$C_s(\tau) = \int s_n(t) s_n(t+\tau) dt / (\sigma^2 T). \tag{A4}$$

The directional receiver at depth z in water of depth H has an upward beam and a downward beam. The downward beam is represented by its image in the seabed at depth H, i.e., an upward looking beam centered at depth $2H-z$. Thus the received amplitude for the generalized up/down beam $j=1,2$ is

$$g_j(t) = \sum_n s_n(t - r_{jn}/c) b_{jn} R_{jn} / r_{jn}, \tag{A5}$$

where r_{jn}, b_{jn} represent, respectively, the range and combined beam and source directionality factor associated with the jth beam and the nth noise source. The R_{jn} are generalized reflection coefficients. For the upward beam there is, of course, no reflection and so $R_{1n}=1$.

2. Numerator

The numerator of Eq. (A1) is constructed from Eq. (A5) by making use of Eqs. (A2)–(A4) to get rid of the double sum and integral

$$\begin{aligned}
&\int g_1(t) g_2(t+\tau) dt \\
&= \int \sum_m \sum_n s_n(t - r_{1n}/c) s_m(t - r_{2m}/c + \tau) \\
&\quad \times \frac{b_{1n} b_{2m} R_{1n} R_{2m}}{r_{1n} r_{2m}} dt \\
&= \int \sum_n s_n(t') s_n(t' + (\tau - \tau_r)) \frac{b_{1n} b_{2n} R_{1n} R_{2n}}{r_{1n} r_{2n}} dt' \\
&= \sigma^2 T \sum_n C_s(\tau - \tau_r) \frac{b_{1n} b_{2n}}{r_{1n} r_{2n}} R_{2n},
\end{aligned} \tag{A6}$$

where τ_r is the time difference between arrivals from the nth source and the two receivers, $\tau_r = (r_{2n} - r_{1n})/c$. Since these contributions only occur near the vertical the directionality factors can be replaced by the vertical beam power $b_{1n}^2 = b_{2n}^2 = b(0)$, and the vertical path lengths (z_1, z_2) substituted for (r_1, r_2). One can also assume that R_{2n} is the vertical reflection coefficient and drop the subscripts. To evaluate the sum each source point is assumed to occupy an elementary area A such that the sum can be written in terms of an integral over surface area,

$$\int g_1(t) g_2(t+\tau) dt = \frac{\sigma^2 T}{A} \frac{b(0) R(0)}{z_1 z_2} \int_0^\infty C_s(\tau - \tau_r) 2\pi\rho d\rho, \tag{A7}$$

where ρ is a polar coordinate in the surface plane centered on the point above the receiver.

The travel time difference is related to the radius ρ by the Fresnel approximation

$$c\tau_r = r_{2n} - r_{1n} \cong z_2 - z_1 + \frac{\rho^2}{2}\left(\frac{1}{z_2} - \frac{1}{z_1}\right) \tag{A8}$$

so

$$\rho d\rho = -c d\tau_r \frac{z_1 z_2}{z_2 - z_1}. \tag{A9}$$

According to the above-mentioned Fresnel approximation a uniform distribution in area (i.e., $\rho d\rho$) results in a uniform distribution in time. It can be shown that, surprisingly, this is not true with exact Pythagoras path lengths although fortunately this is not important here.

Equation (A7) becomes

$$\int g_1(t) g_2(t+\tau) dt = -\frac{2\pi\sigma^2 T}{A} \frac{b(0) R(0) c}{(z_2 - z_1)} F(\tau) \tag{A10}$$

$$F(\tau) = -\int_{\tau_z}^{\infty} C_s(\tau - \tau_r)d\tau_r = \int_{\tau - \tau_z}^{\infty} C_s(\tau')d\tau', \qquad \text{(A11)}$$

where $\tau_z = (z_2 - z_1)/c$ and the dummy variable is $\tau' = \tau - \tau_r$. If the noise sources had a true uniform spectrum then C_s would be a Kronecker delta function, and so the integral would be unity for $\tau < \tau_z$ and zero for $\tau > \tau_z$, i.e., a step function. To obtain the Green's function as in, for example, Eqs. (1) and (2) of Roux *et al.* (2005), one needs to differentiate Eq. (A10) with respect to τ, in which case F itself becomes a Kronecker delta,

$$\frac{\partial}{\partial \tau}\int g_1(t)g_2(t+\tau)dt = \frac{2\pi\sigma^2 T}{A}\frac{b(0)R(0)c}{(z_2 - z_1)}C_s(\tau). \qquad \text{(A12)}$$

Numerically this can be found from the forward difference (indicated by Δ) of the correlation divided by the sample interval. In the (continuous) frequency domain it is equivalent to multiplication by frequency. The Discrete Fourier Transform (DFT) equivalent of differentiation is convolution by two opposite signed Kronecker delta functions separated by one sample (interval T_s), which is equivalent to multiplication by $(1 - \exp(-i2\pi f T_s))$ in the frequency domain followed by division by T_s. Thus numerically one would find

$$\Delta\left\{\int g_1(t)g_2(t+\tau)dt\right\} = \frac{2\pi\sigma^2 T}{A}\frac{b(0)R(0)c}{f_s(z_2 - z_1)}C_s(\tau) \qquad \text{(A13)}$$

and the peak value would be

$$\max\left[\Delta\left\{\int g_1(t)g_2(t+\tau)dt\right\}\right] = \frac{2\pi\sigma^2 T}{A}\frac{b(0)R(0)c}{f_s(z_2 - z_1)}. \qquad \text{(A14)}$$

Because R retains its sign, the term "max" is used here to mean the maximum of the absolute value multiplied by the sign.

Otherwise if one retains the original spectrum, $\bar{s}(\omega) = \int_{-\infty}^{\infty} s(\omega)\exp(-i\omega t)d\omega$, but sets the near zero frequencies to zero, $|\bar{s}(0)|^2 = 0$, then an identity that follows from the Wiener–Khintchine theorem (p.141, Skudrzyk, 1971) states that

$$\int_{-\infty}^{\infty} C_s(\tau')d\tau' \equiv |\bar{s}(0)|^2 = 0. \qquad \text{(A15)}$$

The function F in Eq. (A10) is therefore still zero for large positive or negative $\tau - \tau_z$. Where $\tau \sim \tau_z$ the function may oscillate, but the absolute value of its Hilbert transform, being the envelope of the oscillation, provides a good representation, though slightly widened, of the Kronecker delta. The penalty is loss of the sign of the impulse response.

Thus there are two processing options, one is to opt for robustness and retain the Hilbert transform (with or without the time differentiation). The other is to perform the differentiation without Hilbert transform and thus retain a signed impulse response.

This time domain derivation has essentially assumed a broad band. If the spectrum of the source term is assumed to be flat (σ^2 is the source variance for the given band), then narrowing its band (by filtering) reduces the height of the correlation peak (through σ^2) and therefore the amplitude response [Eqs. (A10) and (A11)] in proportion to the band, and the power response in proportion to bandwidth squared (as can easily be seen by consideration of a Gaussian spectral shape and its corresponding Gaussian autocorrelation function).

3. Denominator

The amplitude of the numerator of Eq. (A1) is not much use alone since it contains the unknowns σ, T, A. The normalization, i.e., the denominator of Eq. (A1), resolves this because it is proportional to the same unknowns. Each of the two components of the denominator of Eq. (A1) is evaluated as

$$\begin{aligned}\int g_j^2(t)dt &= \int \sum_m \sum_n s_n(t - r_{jn}/c)s_m(t - r_{jm}/c) \\ &\quad \times \frac{b_{jn}b_{jm}R_{jn}R_{jm}}{r_{jn}r_{jm}}dt \\ &= \sigma^2 T \sum_n \frac{b_{jn}^2 R_{jn}^2}{r_{jn}^2}, \end{aligned} \qquad \text{(A16)}$$

where use has been made of Eqs. (A2) and (A3) to reduce the integral and double sum to a single sum. To evaluate the sum each source point is again assumed to occupy an elementary area A such that the sum can be written in terms of an integral over surface area. So now

$$\int g_j^2(t)dt = \frac{2\pi\sigma^2 T}{A}\int_0^{\infty} \frac{b_j^2 R_j^2}{r_j^2}\rho d\rho, \qquad \text{(A17)}$$

where ρ is the polar coordinate in the surface plane centered on the point above the receiver. Notice that in Eq. (A17) the reflection coefficient is squared so that its sign is lost, in contrast with behavior in Eq. (A14). The area integral can be transformed into an angle integral (which does not depend on the distance of the receiver from the surface)

$$\int_0^{\infty} \frac{b_j^2}{r_j^2}\rho d\rho = b(0)\int_0^{\pi/2} B(\theta)N(\theta)\cos\theta d\theta \qquad \text{(A18)}$$

and this is recognized as the integral that appears in the array gain (or noise gain) formula (Urick, 1975), where N is the noise directionality [so that the complete noise source term is $\sigma^2 N(\theta)$] and B is the array's beam pattern, normalized to unity in the steer direction. It can be evaluated straightforwardly by expressing it as the sum of the terms in the noise's normalized cross-spectral density matrix C_{ij} weighted by the array shading w_i and with steering phases ϕ_{ij} (Urick, 1975),

$$\begin{aligned}\int_0^{\pi/2} B(\theta)N(\theta)\cos\theta d\theta &= \frac{\sum_i^M \sum_j^M \{w_i w_j C_{ij}\exp(i\phi_{ij})\}}{\left(\sum_i^M w_i\right)^2} \\ &\quad \times \int_0^{\pi/2} N(\theta)\cos\theta d\theta \end{aligned} \qquad \text{(A19)}$$

If it is assumed that the surface noise sources have a dipole

directionality then, since $N=\sin\theta$, the noise integral is $\int_0^{\pi/2}N(\theta)\cos\theta d\theta=1/2$, and the normalized C_{ij} are (Cron and Sherman, 1962; 1965)

$$C_{ij}=C(kd_{ij})=2\{\exp(ikd_{ij})/(ikd_{ij}) + (\exp(ikd_{ij})-1)/(kd_{ij})^2\}. \quad \text{(A20)}$$

Assuming hamming shading the integral of Eq. (A19) is just a number,

$$\beta \equiv M\frac{\sum_i^M\sum_j^M\{w_iw_jC_{ij}\exp(i\phi_{ij})\}}{2\left(\sum_i^M w_i\right)^2}, \quad \text{(A21)}$$

where M is the number of hydrophones. This result can be inserted directly into Eq. (A17) for $j=1$, but for $j=2$, it is noted that the Rayleigh reflection coefficient is a function of vertical wave number, which is necessarily slowly varying near vertical, so it will be assumed that it remains constant over the endfire beam. It can therefore be taken out of the integral and the same noise gain integral is obtained for both beams.

The final result is

$$\sqrt{\int g_1^2(t)dt\int g_2^2(t)dt}=\frac{2\pi\sigma^2Tb(0)|R(0)|}{A}\frac{\beta}{M}. \quad \text{(A22)}$$

Assuming the spectrum of the source term to be flat, as in the second section of the Appendix (σ^2 is the source variance for the given band), a narrowing of the band (by filtering) reduces the response in proportion, again through σ^2. An additional effect is due to the dependence of beam width on frequency (in fact, inverse proportionality). This controls the numerical value of β, [Eq. (A21)]. For a narrow band at the design frequency $\beta=1.38$; for a band extending from the design frequency down to half the design frequency it is $\beta=1.87$; for a band extending down to almost zero (design frequency/200) it is $\beta=4.08$. This additional effect is therefore merely an averaging over frequency.

4. Complete formula for peak value

Combining Eqs. (A10) and (A22) to form Eq. (A1) gives

$$\max[\Delta\{C(\tau)\}]=\frac{cM\ \text{sign}(R)}{(z_2-z_1)f_s\beta} \quad \text{(A23)}$$

As explained earlier the main bandwidth effects have canceled out leaving the minor effect of frequency averaging the beam width incorporated in β. Since the hydrophone separation a and the design frequency are related by $2f_0a=c$ Eq. (A23) can be written in terms of the array's acoustic length ($L=Ma$) as

$$\max[\Delta\{C(\tau)\}]=\frac{2f_0L\ \text{sign}(R)}{(z_2-z_1)f_s\beta}. \quad \text{(A24)}$$

The ratio of sample frequency to design frequency is also a number $\gamma=f_s/f_0$ so the final peak value is

$$\max[\Delta\{C(\tau)\}]=\frac{2L\ \text{sign}(R)}{(z_2-z_1)\gamma\beta}. \quad \text{(A25)}$$

For the equipment used here there are 32 hydrophones separated by 0.18 m with a design frequency of 4166.7 Hz and a sampling frequency of 12 kHz. A height above the seabed of 30 m leads to $z_2-z_1=60$ m, and assuming the band is half the design frequency ($\beta=1.87$), the final peak height is 0.0357. Alternatively, assuming the band to be the full design frequency ($\beta=4.08$), the final peak height is 0.0164.

5. Complete formula for background

Away from the peak cross correlation with a finite number N of samples the background will not be exactly zero as implied by Eqs. (A6) and (A1). In a loose sense it is related to the number of independent samples; more exactly the background [i.e., the standard deviation of Eq. (A1)] is derived as follows. Each background sample ν_m [i.e., realization of C in Eq. (A1)] is the sum of the product of two (potentially correlated) sequences $b_m=\Sigma_n^N f_{n,m}$, where $f_{n,m}=p_nq_{n+m}$. Although the probability distribution of the product is not Gaussian it can be shown that the variance of the product is the product of the individual variances, say σ^2. The variance of the background is the mean of the squares of these sums, i.e.,

$$m\sigma_b^2=\sum_m\left(\sum_n^N(f_{n,m})\right)^2=\sum_m\sum_n^N\sum_{n'}^N f_{n,m}f_{n',m}$$
$$=\sigma^2\sum_m\sum_n^N\sum_{n'}^N\rho_{(n-n'),m}. \quad \text{(A26)}$$

The last double sum is the sum over the correlation coefficients ρ which for large N leads to $\sigma_b^2=N\sigma^2\Sigma_{j=-N}^N\rho_j$. Since the peak value is $N^2\sigma^2$ the normalized background variance is $\Sigma_{j=-N}^N\rho_j/N$ which can be thought of as the reciprocal of the number of independent samples in f.

Brekhovskikh, L., and Lysanov, Yu., (**1982**). *Fundamentals of Ocean Acoustics* (Springer, New York).

Buckingham, M. J., and Jones, S. A. S., (**1987**). "A new shallow-ocean technique for determining the critical angle of the seabed from the vertical directionality of the ambient noise in the water column," J. Acoust. Soc. Am. **81**, 938–946.

Cato, D. H., (**2000**). "Sea-surface generated noise dependence on wind speed and white cap coverage," J. Acoust. Soc. Am. (A) **107**, 2922.

Cron, B. F., and Sherman, C. H., (**1962**). "Spatial correlation functions for various noise models," J. Acoust. Soc. Am. **34**, 1732–1736.

Cron, B. F., and Sherman, C. H., (**1965**). "Addendum: Spatial correlation functions for various noise models [J. Acoust. Soc. Am. 34, 1732–1736 (1962)]," J. Acoust. Soc. Am. **38**, 885.

Ding, L., and Farmer, D. M., (**1994**). "Observations of breaking surface wave statistics," J. Phys. Oceanogr. **24**, 1368–1387.

Finette, S., and Heitmeyer, R. M., (**1996**). "Angle-time-frequency resolution of the noise field generated by wind-induced breaking waves," J. Acoust. Soc. Am. **99**, 209–222.

Frison, T. W., Abarbanel, H. D. I., Cembrola, J., and Neales, B., (**1996**). "Chaos in ocean ambient 'noise'," J. Acoust. Soc. Am. **99**, 1527–1539.

Harrison, C. H., (**1996**). "Formulas for ambient noise level and coherence," J. Acoust. Soc. Am. **99**, 2055–2066.

Harrison, C. H., (**2004**). "Sub-bottom profiling using ocean ambient noise," J. Acoust. Soc. Am. **115**, 1505–1515.

Harrison, C. H., (**2005**). "Performance and limitations of spectral factorization for ambient noise sub-bottom profiling," J. Acoust. Soc. Am. **118**,

2913–2923.
Harrison, C. H., and Simons, D. G., (**2002**). "Geoacoustic inversion of ambient noise: a simple method," J. Acoust. Soc. Am. **112**, 1377–1389.
Lermusiaux, P. F. J., and Robinson, A. R., (**2001**). "Features of dominant mesoscale variability, circulation patterns and dynamics in the Strait of Sicily," Deep-Sea Research, Part I **48**, 1953–1997.
Melville, W. K., and Matusov, P., (**2002**). "Distribution of breaking waves at the ocean surface," Nature (London) **417**, 58–63.
Roux, P., and Kuperman, W. A., (**2004**). "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003.
Roux, P., Sabra, K. G., Kuperman, W. A., and Roux, A., (**2005**). "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84.
Sabra, K. G., Roux, P., and Kuperman, W. A., (**2005**). "Emergence rate of the time-domain Green's function from the ambient noise cross-correlation function," J. Acoust. Soc. Am. **118**, 3524–3531.
Sabra, K. G., Roux, P., Kuperman, W. A., Hodgkiss, W. S., and D'Spain, G. L., (**2004**). "Using ocean ambient noise for array element localization," J. Acoust. Soc. Am. (A) **115**, 2507.
Siderius, M., Harrison, C. H., and Porter, M. B., (**2006**). "A passive fathometer technique for imaging seabed layering using ambient noise," J. Acoust. Soc. Am. **120**, 1315–1323.
Skudrzyk, E., (**1971**). *The Foundations of Acoustics*, (Springer, New York).
Urick, R. J., (**1975**). *Principles of Underwater Sound*, 2nd ed. (McGraw-Hill New York), Chap. 7.
Weaver, R. L., and Lobkis, O. I., (**2004**). "Diffuse fields in open systems and the emergence of the Green's function," J. Acoust. Soc. Am. **116**, 2731–2734.

Passive fathometer processing

Peter Gerstoft and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Martin Siderius
HLS Research Inc., 3366 North Torrey Pines Court, Suite 310, La Jolla, California 92037

Chen-Fen Huang
Department of Marine Environmental Informatics, National Taiwan Ocean University, Keelung, Taiwan

Chris H. Harrison
NURC, La Spezia, Italy

(Received 21 September 2007; revised 10 December 2007; accepted 10 December 2007)

Ocean acoustic noise can be processed efficiently to extract Green's function information between two receivers. By using noise array-processing techniques, it has been demonstrated that a passive array can be used as a fathometer [Siderius, *et al.*, J. Acoust. Soc. Am. **120**, 1315–1323 (2006)]. Here, this approach is derived in both frequency and time domains and the output corresponds to the reflection sequence. From this reflection sequence, it is possible to extract seabed layering. In the ocean waveguide, most of the energy is horizontally propagating, whereas the bottom information is contained in the vertically propagating noise. Extracting the seabed information requires a dense array, since the resolution of the bottom layer is about half the array spacing. If velocity sensors are used instead of pressure sensors, the array spacing requirement can be relaxed and simulations show that just one vertical velocity sensor is sufficient.
 [DOI: 10.1121/1.2831930]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 1297–1305

I. INTRODUCTION

In recent years it has been demonstrated that the time domain Green's function between a source and a receiver can be extracted from noise, for reviews see Refs. 1 and 2. The theory for this is now relatively well-developed[3–6] and the approach has been demonstrated in ocean acoustics (including ocean waveguides),[7–10] ultrasound,[3] and seismology.[11–18]

The idea is to cross correlate noise wave fields between two receivers and by time averaging over a sufficiently long time only the propagation paths between the two receivers remain. The theory prescribes that the noise should be isotropic. In an ocean acoustic waveguide, the noise is nonisotropic with most noise originating from the surface; this can be modeled as a sheet of sources located close to the surface.[19] The vertically propagating noise is partially reflected from layers of the seabed and eventually dissipated through attenuation losses. For noise propagating at grazing angles lower than critical, this noise is trapped in the ocean waveguide.

Siderius *et al.*[10] introduced a coherent broadband array processing method for a vertical array that limits the effect of the horizontally propagating noise, this is the so-called passive fathometer method. The method is based on relating the down- and up-going signals on the array and can be implemented in the time or frequency domains. In order to develop this fathometer method into a practical processing approach, a number of issues have to be addressed in detail. Several of these are addressed here:

(1) Theory is developed in both time and frequency domains.
(2) It is established that the wavelet is a sinc function.
(3) Adaptive beamforming is introduced for frequency domain processing and shown to give good results.
(4) The important issue of spatial aliasing is discussed.
(5) It is demonstrated that due to strong presence of horizontally propagating noise, the maximum frequency should not be higher than twice the design frequency for a conventional hydrophone array.
(6) Vertical velocity sensors are not sensitive to the horizontally propagating noise and based on simulations, it is shown that just one sensor is sufficient to resolve the seabed layering.

II. SIMPLE EXAMPLE

A simple example motivates the approach and demonstrates that a vertical array can be used as a fathometer, see Fig. 1. Consider a simple 100-m deep ocean with a square pulse (width 0.01 s, amplitude 1) propagating down, reflected, and then propagating up (amplitude 0.5) and recorded on a 50-m long array with 20 receivers as shown in Fig. 1. The processing[10] calls for stacking the down- and up-going signals, Figs. 1(b) and 1(c). This is done by time delaying and summing at a reference depth, in this case (and all examples here) the ocean surface. Both stacked signals show a main peak, corresponding to the wave propagating with that speed, and a spread out waveform, corresponding to the signature obtained from stacking the opposite propagating wave.

Finally, in Fig. 1(d) we cross correlate the down- and up-going signals. This shows a triangular pulse of width

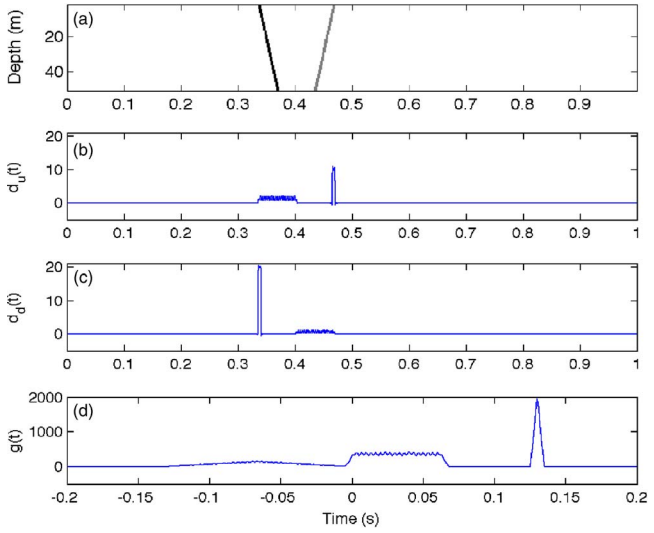


FIG. 1. (Color online) Time domain processing of a simple arrival in a 100-m ocean with 20 receivers at 0–50 m depth. (a) A square pulse 0.1-s wide with a down-going component of amplitude 1 and an up-going component with amplitude 0.5. (b) Stacking the signal at the surface in up-going direction. (c) Stacking the signal at the surface in down-going direction. (d) Cross correlation of the stacked down- and up-going signals.

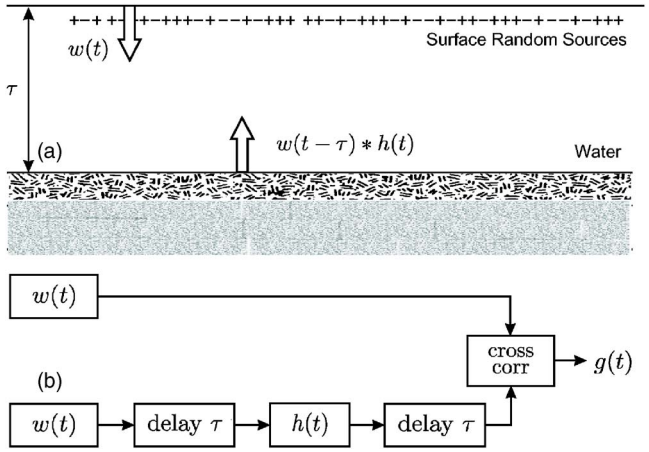


FIG. 2. (Color online) Environment and block diagram. $w(t)$ represents ocean surface noise, $h(t)$ the reflection sequence, and τ the vertical travel time across the ocean.

0.02 s appearing at a travel time corresponding to two water depths ($2\times100/1500=0.13$ s). The triangular shape is due to the convolution of two square pulses. In addition, there is some spurious components visible from about 0 to 0.067 s, due to the convolution of signals that did not propagate in the stacking direction. The time extent of this noise corresponds to twice the travel time across the array length. The corresponding spurious component can be seen in Fig. 1(b) from 0.33 to 0.4 s. Within this interval the level is fairly constant with small variations in amplitude. Though not visible in the plot, the shape at the peaks is a square wave with 19 peaks. The number of small peaks is proportional to the number of hydrophones in the array and depends also on the source wave form.

III. THEORY

First, in Sec. III A a one-dimensional (1D) model is used to demonstrate that the reflection sequence can be extracted by cross correlating the down- and up-going wave fields at a receiver. One way to extract the down- and up-going wave fields is to use beamforming[10] as discussed in the remainder of this section.

The cross-correlation approach can cause spurious components, as demonstrated using interferometric approaches.[9,20,21] A mild example of a spurious component was shown in Fig. 1(d). However, the spurious components can be reduced by cross-correlating down- and up-going wave fields.[22] It is feasible to separate the down- and up-going wavefields by combining the response from a vertical geophone and a hydrophone.[22] Our beamforming approach results in a similar decomposition.

A. 1D model

The environment under consideration is shown in Fig. 2(a). Assuming a simple 1D model, the corresponding block diagram in Fig. 2(b) is obtained, where we cross correlate the down-going signal $d_d(t)$ with the upgoing signal. $d_u(t)$ using a receiver at the surface as a reference. As the noise $w(t)$ is generated near the surface, $d_d(t)=w(t)$. This down-going signal is time delayed by τ corresponding to the travel time to the ocean bottom from the receiver, convolved with the reflection sequence $h(t)$ at the ocean bottom, and time delayed by τ as the signal propagates back to the receiver. This gives ($*$ represents convolution)

$$d_u(t) = w(t-2\tau) * h(t). \tag{1}$$

Cross correlating the down- and up-going signals yields

$$c_{ud}(t) = d_u(t) * d_d(-t) = w(t-2\tau) * h(t) * w(-t)$$
$$= [w(t) * w(-t)] * h(t-2\tau). \tag{2}$$

The first term in square brackets is the autocorrelation function of the surface noise time series $w(t)$ (inverse Fourier transform of the surface noise power spectral density function). The last term is the time-delayed seabed reflection sequence, which is related to the Green's function, indicating that by cross correlating the down- and up-going wave fields the seabed reflection sequence and its layering can be extracted. The 2τ term is twice the travel time from the receiver to the ocean floor and this is the idea behind the ocean bottom fathometer,[10] but as indicated in Eq. (2) the whole reflection sequence is available.

Assuming the noise from the sea surface is spectrally flat ("white") then the autocorrelation of a low-pass white noise process with bandwidth b has a sinc($2bt$) wave form.[23] The width of the main peak in the sinc function is $1/b$. Translating this $1/b$ resolution to the two-way travel time to depth gives a depth resolution of about $1/b(c/2)=c/2b$ (c is the sound speed).

Generally the noise cross correlation approach has been successful in estimating travel times, but less successful in obtaining amplitudes. In the literature there appears to be differences in opinion, as it has been suggested that the time domain Green's function is proportional to the noise cross correlation, its time derivative, time integral, or even a fractional derivative. Most current applications are concerned

with obtaining kinematically correct (arrival time) expressions. This does not depend on the time derivative.

Some of this variation in the use of time derivative is related to the dimension of the physical domain in which the wave is propagating and the effective dimension of the source distribution. For example, using a three-dimensional (3D) physical domain and a 1D source as in Refs. 7, 20, and 21 results in a fractional time derivative (a factor $1/\sqrt{i\omega}$ in the frequency domain).

Consider a uniform 1D medium with speed c and a point source [zero-dimensional (0D)] at $z=0$, then the time domain Green's function $g_1(t)$ at distance z would be

$$g_1(t) = \delta\left(t - \frac{z}{c}\right). \tag{3}$$

The cross correlation in the frequency domain between the downward propagating noise at receivers z_1 and z_2 then would be

$$C_{12}(\omega) = e^{i(\omega/c)(z_2 - z_1)}. \tag{4}$$

In the time domain,

$$c_{12}(t) = \delta(t - (z_2 - z_1)/c) = g_1(t). \tag{5}$$

Thus, for a 1D medium with a single point source the Green's function is proportional to the noise cross correlation. As the fathometer closely resembles a 1D medium, this assumption was used in Ref. 10 and is used here as well.

B. Time domain stacking

Extracting the reflection sequence from noise recordings for a vertical array with N sensors at depth z_i is developed in the time domain $p_i(t)$ in this section and in the frequency domain $p_i(\omega)$ in Sec. III C. The Fourier transform pair is given by

$$p(\omega) = \int p(t) e^{-i\omega t} dt, \tag{6}$$

$$p(t) = \frac{1}{2\pi} \int p(\omega) e^{i\omega t} d\omega. \tag{7}$$

The factor $1/2\pi$ is suppressed in the following. Assuming vertical propagation, the time t is converted to depth by dividing by $2c$, where c is a representative speed.

Stacking or summing over the z-coordinate with a time delay corresponding to a slowness $+s=1/c$ for downgoing waves, gives

$$d_d(t) = \sum_{i=1}^{N} p_i(t + s(z_i - z_0)), \tag{8}$$

where the time series are projected into the time series corresponding to depth z_0. We prefer to use slowness because the processing is linear with slowness. Stacking for upgoing waves using the negative slowness $-s$ gives

$$d_u(t) = \sum_{i=1}^{N} p_i(t - s(z_i - z_0)). \tag{9}$$

In the frequency domain the down- and up-going waves can be expressed as

$$d_d(\omega) = \int \sum_{i=1}^{N} p_i(t + s(z_i - z_0)) e^{-i\omega t} dt$$

$$= \sum_{i=1}^{N} p_i(\omega) e^{+i\omega s(z_i - z_0)}, \tag{10}$$

$$d_u(\omega) = \int \sum_{i=1}^{N} p_i(t - s(z_i - z_0)) e^{-i\omega t} dt$$

$$= \sum_{i=1}^{N} p_i(\omega) e^{-i\omega s(z_i - z_0)}. \tag{11}$$

Cross correlation between down- and up-going waves is computed for the whole observation period T

$$c_{ud}(t) = d_u(t) * d_d(-t) = \int_0^T d_u(\tau) d_d(\tau - t) d\tau, \tag{12}$$

where the $1/T$ normalizing factor is neglected. As we are concerned with comparing signals from the top of the ocean with reflected signals from the bottom, the averaging time T should be much larger than twice the travel time over the ocean depth. In the frequency domain Eq. (12) becomes

$$C_{ud}(\omega) = d_u(\omega) d_d^*(\omega). \tag{13}$$

Inverse Fourier transforming $C_{ud}(\omega)$ to the time domain give, $c_{ud}(t)$, which is the reflection sequence convolved with the noise autocorrelation function, see Eq. (2). From this the seabed layering can be estimated.

C. Frequency domain beamforming

By beamforming we can obtain the down- and up-going beams corresponding to down and up-going wave fields, respectively:

$$d_d(\omega) = \mathbf{w}_d^H \mathbf{p}, \tag{14}$$

$$d_u(\omega) = \mathbf{w}_u^H \mathbf{p}, \tag{15}$$

where $\mathbf{p}(\omega) = [p_1(\omega), p_2(\omega, \ldots]^T$ where superscripts T, H, and the asterisk refer to the transpose, complex conjugate transpose, and complex conjugate operators, respectively. $\mathbf{w}_d$ and $\mathbf{w}_u$ are the steering vectors for down- and up-going wave fields, respectively.

Inserting Eqs. (14) and (15) into Eq. (13) gives the cross correlation in the frequency domain:

$$C_{ud}(\omega) = d_u(\omega) d_d^H(\omega) = \mathbf{w}_u^H \mathbf{p} \mathbf{p}^H \mathbf{w}_d, \tag{16}$$

where both down- and up-going steering vectors are used.

Using conventional processing the downward steering vector is $\mathbf{w}_d(\omega) = \mathbf{w}(\omega) = [e^{-i\omega s z_1}, e^{-i\omega s z_2}, \ldots]^T$ and the upgoing is $\mathbf{w}_u(\omega) = \mathbf{w}^*(\omega)$. Inserting these expressions into Eq. (13) gives the cross correlation in the frequency domain:

$$C_{ud}(\omega) = \mathbf{w}^T \mathbf{p}\mathbf{p}^H \mathbf{w}. \quad (17)$$

This expression is very similar to the conventional beamformer $\mathbf{w}^H \mathbf{p}\mathbf{p}^H \mathbf{w}$ except that $\mathbf{w}^H$ is replaced with $\mathbf{w}^T$.

As with conventional beamforming[24] there is a spatial aliasing issue which requires the frequency to be less than $c/2d$ (d is element spacing). For broadband methods and a single source, this requirement usually can be relaxed, as aliasing from several angles will average out in the results. However, in the present application the noise comes from all directions and this aliasing component can destroy the response, as illustrated in the next section (Fig. 5).

To reduce spatial aliasing, it is beneficial to apply a depth-dependent shading factor pre-multiplied on the observed pressure vector $\mathbf{p}$. A Kaiser–Bessel window ($\alpha=1.5$) is used here.

D. Averaging

Often the time series is segmented into J time series which are processed separately and then the output is averaged. In that case Eq. (12) will become

$$c_{ud}(t) = \frac{1}{J}\sum_j^J d_{uj}(t) * d_{dj}(-t) = \frac{1}{J}\sum_j^J \int_0^T d_{uj}(\tau) d_{dj}(\tau - t) d\tau, \quad (18)$$

where it is assumed that each time series is sufficiently long that truncation effects can be neglected. In the frequency domain the averaging is efficiently expressed via the estimated cross-spectral density matrix $\mathbf{C}=(1/J)\Sigma_j^J \mathbf{p}_j \mathbf{p}_j^H$. Thus the frequency domain expression becomes

$$C_{ud}(\omega) = \frac{1}{J}\sum_j^J d_{uj}(\omega) d_{dj}^H(\omega) = \mathbf{w}^T \mathbf{C}\mathbf{w}. \quad (19)$$

E. High-resolution adaptive beamforming

The frequency domain response can also be obtained by using high-resolution beamforming methods in Eq. (16) with two adaptive steering vectors. These steering vectors can be obtained using any one of several adaptive processing methods. One possible choice is to use the minimum variance distortionless response with white noise gain constraint (MVDR-WNC) beamformer, see e.g., Ref. 24. Although not explored in this article, we have observed that high-resolution beamforming has a tendency toward giving better resolved reflection sequences in the time domain. An example of this processing is shown in the time series at the right of Fig. 3 in the next section.

F. Frequency band

It is noted from Urick[25] that the ambient noise due to wind decays about 6 dB per octave. Thus, if the frequency spectrum is not normalized the response will be dominated by the lower frequencies. Here, the cross spectral density matrix is normalized by its trace to obtain a flat spectrum. Other normalizations are also possible; for a discussion see Ref. 26.

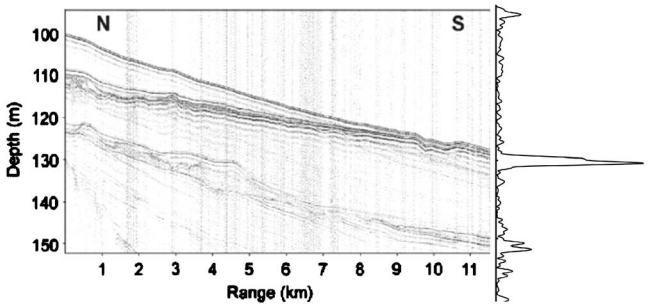


FIG. 3. Seismic section for the Mapex2kbis area (Ref. 29). The location of the vertical array is about 1 km further to the south. To the right the extracted reflection sequence derived from noise, obtained using the MVDR-WNC high resolution adaptive beamformer on the vertical array data.

The minimum and maximum frequency $f_{\min}$ and $f_{\max}$ must be selected for processing. A large bandwidth, i.e., a low $f_{\min}$ and a high $f_{\max}$, is preferable in order to obtain a sharper pulse in the time domain.

The minimum frequency can be selected from the following empirical rules

(1) At low frequencies the noise is dominated by shipping. Urick[25] gives a typical value of 200 Hz as the point where wind noise starts to dominate.
(2) The beam pattern from the end fire beam should be sufficiently narrow that it is not influenced by horizontal propagating noise, which typically propagates with grazing angle of $\pm 30°$ or less. Thus, the lower frequency should be chosen high enough that the width of the end-fire beam is less than $2(90-30)=120°$.

The maximum frequency should be sufficiently low so that the array will not have grating lobes in the horizontal propagating noise direction. Assuming the horizontal propagating noise is within 15° from horizontal then from Eq. (20) a $f_{\max}$ of $2/(1+\sin 15°)f_d = 1.6 f_d$, where $f_d = c/2d$ (array design frequency) corresponds to the frequency where the array elements have spacing $d=\lambda/2$. In practice, it has been found that $f_{\max}=2f_d$ works well.

Finally, in Sec. III A, the depth resolution was found to be $c/2b$ (b is bandwidth and c is sound speed). If b is close to twice the design frequency of the array $b=2f_d=c/d$ then the resolution of the seabed and sediment interface is about $d/2$. Smaller array spacing leads to a higher usable bandwidth and thus better depth resolution.

IV. EXPERIMENT

The data used to demonstrate the approach and its limitations are from the Mapex2kbis experiment on 22 November 2000 (Refs. 27–29), where the vertical array was moored in 130-m deep water with the center of the array at 96-m depth. The site is a sandy bottom in the South Sicily area. The array was nested and two configurations with 32 phones and either 0.5-m spacing or 1-m spacing are used in the analysis.

The experiment took place on the Malta Plateau; for a map of the area see Fig. 1 of Ref. 29. A seismic profile[29] (Fig. 3) was obtained with the array moored 1 km further to

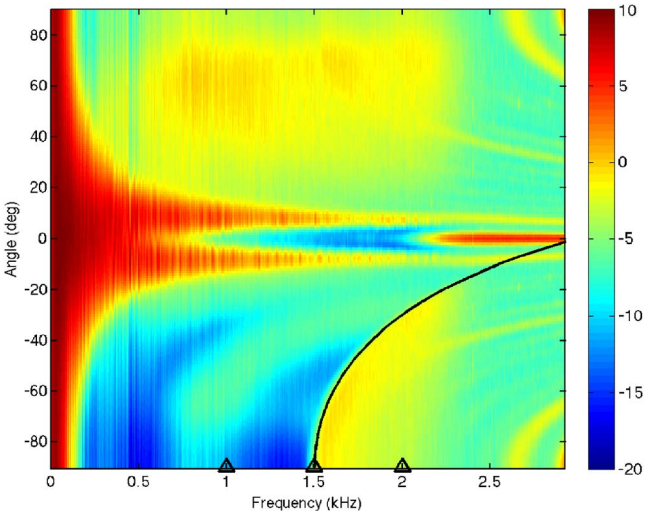


FIG. 4. (Color online) Beamformer output (dB) for the Mapex2kbis data. The grating lobes curve Eq. (20) (solid) is in the lower right corner. This is for a 32-element array with $d=0.5$ m spacing, corresponding to a 1500-Hz design frequency.

the south. For reference, the envelope of the reflection sequence obtained at the array based on the observed ambient noise high-resolution adaptive beamforming (MVDR-WNC) is shown.

For demonstration of the approach, one hundred 10-s samples were used to construct the cross-spectral density matrix. The quality of the estimated time domain Green's functions depends on the observation time. It has been shown[4,6] and observed[14] that the time domain Green's function convergence is proportional to the square root of the observation time. Thus, a much smaller observation time can be used (30 s gives reasonable estimates) at the expense of quality of the estimates. The data were sampled at 6 kHz.[30]

A. Spatial aliasing

In a series of papers Harrison,[27,31,32] demonstrated that bottom properties could be extracted by comparing the down- and up-going energy from a simple beamformer output. The motivation for this approach is that upward propagating ambient noise has one more bottom bounce than downgoing ambient noise at the same angle. Note that this approach is incoherent in the frequency domain whereas the present method is coherent.

An example of the frequency domain conventional beamformer output is shown in Fig. 4 based on the Mapex2kbis data.[27,31] Down-going signals correspond to positive angles and up-going to negative angles. A Kaiser–Bessel window was applied across the phones. The cross-spectral density matrix has been normalized at each frequency by its trace, so that the energy for each frequency is the same (the unnormalized beam response can be seen in Ref. 27). The antialiasing filter begins to roll off at 2 kHz. The nature of the beamformer output changes somewhat around 2.2 kHz, indicating that the compensation for the antialiasing filter also has enhanced data acquisition system self-noise. The nonpropagating electric self-noise can be seen at around 0° at frequencies greater than 2.2 kHz.

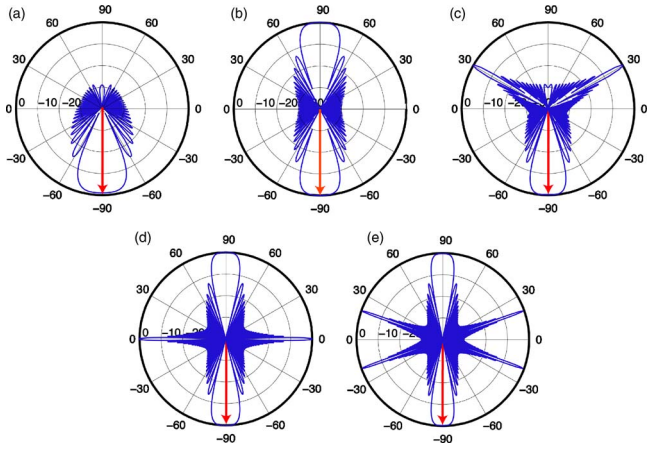


FIG. 5. (Color online) Beam pattern for the downward beam (−90°) at (a) 1 kHz, (b) 1.5 kHz (f_d), (c) 2 kHz ($4/3f_d$), (d) 3 kHz ($2f_d$), and (e) 4.5 kHz ($3f_d$). The first three frequencies used are indicated with △ in Fig. 4. Grating lobes can be seen in (b) at +90°, (c) at +30°, (d) +90°, 0°, and (e) +90°, ±20°.

For frequencies below the design frequency $c/2d$ there is no spatial aliasing, but above this frequency spatial aliasing starts to appear for frequencies and angles larger than the grating angle curve given by[24]

$$f = \frac{c}{d}\frac{1}{1+|\sin\theta|}. \tag{20}$$

This relationship is shown in Fig. 4. Below this curve, the upgoing energy also appears as downgoing energy. This energy from the surface will destroy the ability to extract reflection loss curves,[27] as that processing is based only on magnitude. It may inhibit the ability to extract reflection sequences in the time domain,[10] as this processing is based on both magnitude and phase.

When estimating reflection sequences,[10] one beam is pointing downwards (−90°) and one is pointing upwards (90°). Both of these have similar beam patterns. For the downward beam we explore the variation in beam pattern as the frequency is increased, Fig. 5. Below the design frequency there are no grazing lobes [Fig. 5(a)], but at the design frequency the grazing lobes start to appear [Fig. 5(b)]. As the frequency increases further the grazing lobes appears at lower grazing angles [Figs. 5(c)–5(e)]. The energy that comes in through these grazing lobes can dominate the true response. Thus, the present coherent approach (both phase and magnitude are used) might tolerate minor contributions from the grazing lobes, but the strong energy grazing lobes near horizontal could be a problem.

B. Time response

Reflection sequences were extracted from the noise based on a 0.5-m spacing (Fig. 6) and a 1-m spacing (Fig. 7). The processing is done in the frequency domain. The design frequency f_d is 1500 Hz for the 0.5 m spacing and 750 Hz for 1 m spacing. In the processing the top hydrophone is used as reference depth. The horizontal axis represents depth converted from the two-way traveltime, which is the output of the processing.

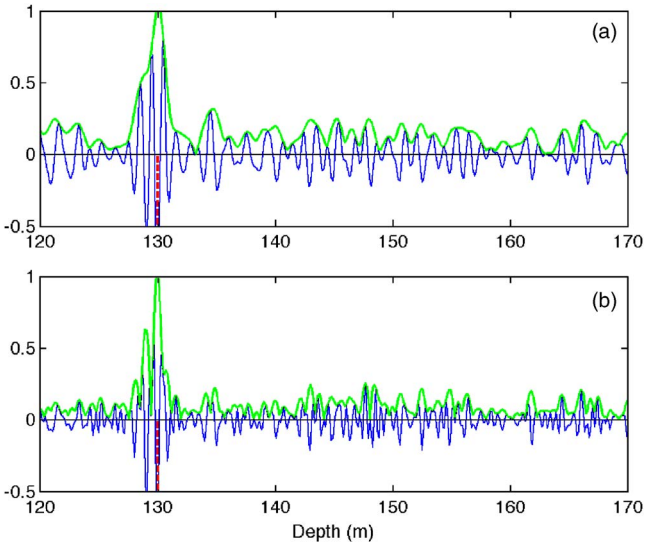


FIG. 6. (Color online) Wave form (thin solid) and envelope (thick solid) impulse response for Mapex2kbis vertical array for 32 phones using 0.5-m spacing and processing in a frequency interval with lower frequency 100 Hz and upper frequency (a) 1500 Hz (f_d) or (b) 3000 Hz ($2f_d$). Bottom is indicated at 130 m depth and a sediment layer can be seen at 145–150 m.

The bottom reflection appears clearly at 130-m depth and the reflection from one or two interfaces at 145–150 m can be seen clearly in Figs. 6 and 7. During the experiment,

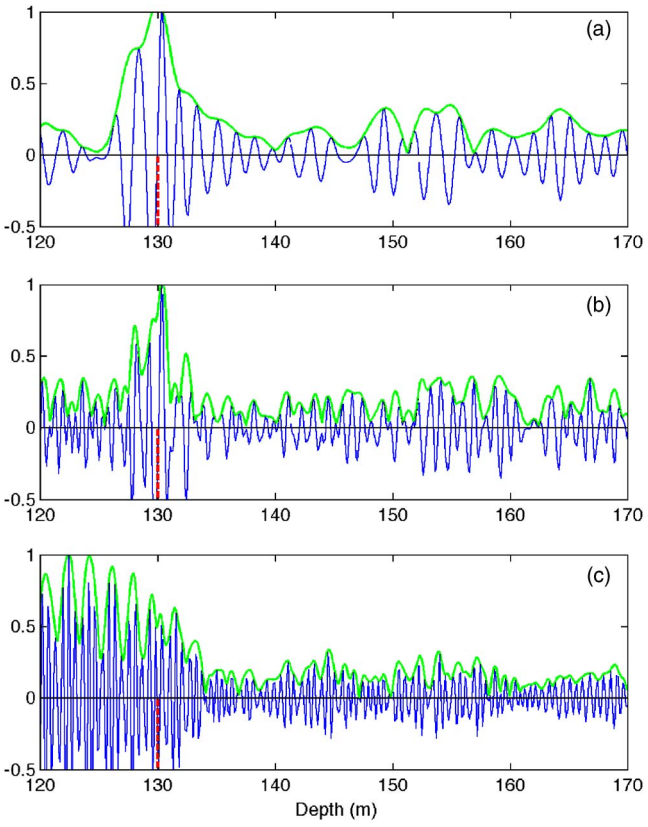


FIG. 7. (Color online) Wave form (thin solid) and envelope impulse (thick solid) response for Mapex2kbis for 32 phones using 1-m spacing and frequency interval (a) 100–750 Hz, (b) 100–1500 Hz, and (c) 100–3000 Hz. With lower frequency 100 Hz and upper frequency (a) 750 Hz (f_d), (b) 1500 Hz ($2f_d$), or (c) 3000 Hz ($3f_d$). Bottom is indicated at 130-m depth and a sediment layer can be seen at 145–150 m.

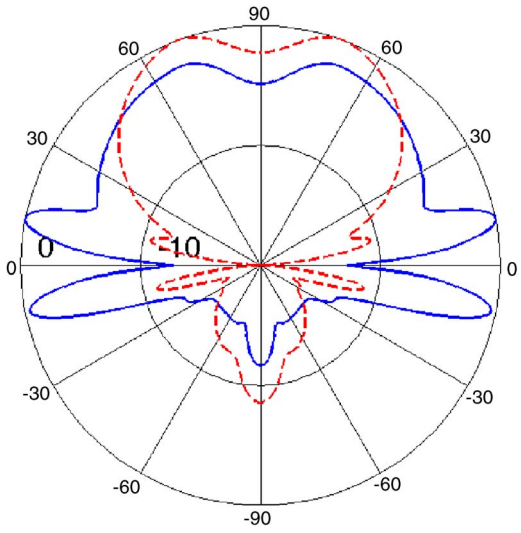


FIG. 8. (Color online) Maximum beampower output (dB) for array of 32 pressure (solid) or vertical velocity sensors (dashed) at frequency 1460 Hz. Each curve is normalized so that the maximum is 0 dB.

a seismic reflection profile was obtained,[29] see Fig. 3. The location of the horizontal array is 1 km south of the southern part of the refection section. The seismic section shows a set of sediment reflectors that are following the ocean bottom at a depth of 15–20 m. This is in clear agreement with the reflectors obtained from the noise processing (Figs. 6 and 7).

The processing was done for several values of the maximum frequency, corresponding to the design frequency f_d [Figs. 6(a) and 7(a)], $2f_d$ [Figs. 6(b) and 7(b)], and $3f_d$ [Fig. 7(c)]. The time series for Figs. 6(b) and 7(b) (corresponding to $2f_d$) seems to give the best resolution indicating that a maximum frequency above the design frequency f_d gives better resolution. The observation for real data is that noise processing gives higher resolution if frequencies above the design frequency are used and about two times the design frequency seems optimal.

V. SIMULATION EXAMPLE

The purpose of the simulations is to further illustrate the processing sensitivity to hydrophone numbers and spacing as well as explore the processing using vertical velocity sensors.[30] Based on the Kuperman–Ingenito model,[19] we simulated the noise field using OASES[33] with the same environment as used in Ref. 10. In the Kuperman–Ingenito model the ocean noise is modeled as a sheet (here at depth 1 m) of spatially and temporally independent sources. This example presents a best case scenario as the simulations assume infinite time averaging.

A geometry similar to the Mapex2kbis experiment is examined: A vertical array with 32 elements with 0.5-m spacing and first element at 84 m. The design frequency is 1500 Hz. The ocean sound speed is constant 1500 m/s. The seabed is at 130 m with layer interfaces at 145 and 150 m depths and bottom sound speed is 1600/1650/1700 m/s and density 1500/2000/2500 kg/m^3 in three bottom layers, respectively.

First, in Fig. 8, we compare the beamformer output at a single frequency (1460 Hz) below the design frequency for

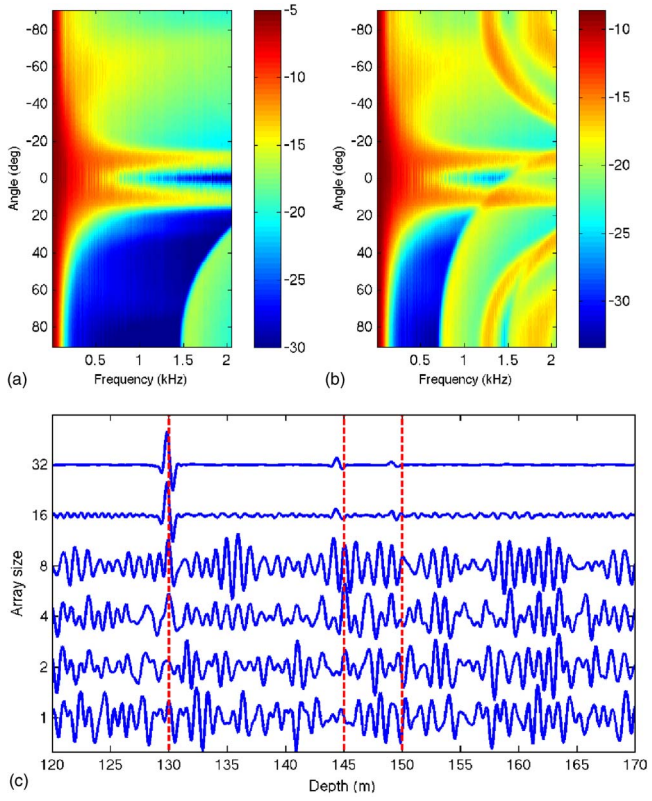


FIG. 9. (Color online) Pressure sensors. Beam response for array with (a) 32 and (b) 16 sensors. (c) Time series for a pressure sensor array. Each trace shows the result of using 32, 16, 8, 4, 2, and 1 sensor, respectively. The array aperture was held constant 15.5 m (except for the 1 sensor). For all signals the time response was computed in the frequency interval 10–1500 Hz and normalized so that the maximum amplitude was one. The interfaces (dashed) are indicated at 130, 145, and 150 m.

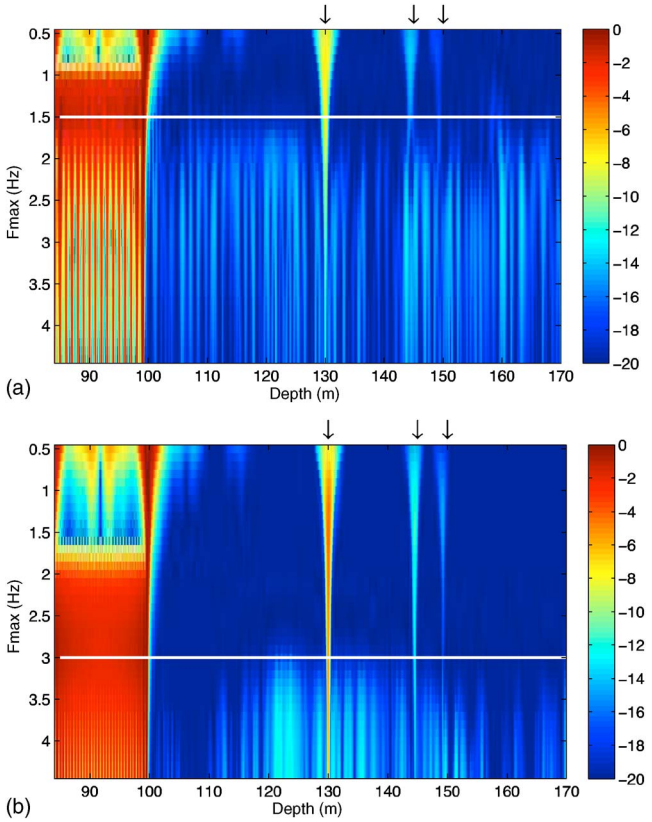


FIG. 10. (Color online) Envelope (dB) of the time response as a function of maximum frequency used for (a) 16 and (b) 32 elements. The minimum frequency was constant 10 Hz and the aperture 15.5 m. Each horizontal line is normalized so that the maximum is 0 dB. The arrows at 130, 145, and 150 m indicates the interfaces. The horizontal white line indicates $2f_d$.

an array consisting of pressure hydrophones (solid) and vertical velocity (dashed) sensors. Due to the waveguide effect there is much more energy in the horizontal direction for the pressure sensors. At higher frequencies, the incoming energy would have a similar pattern, but, due to spatial aliasing (as demonstrated in Fig. 5), this energy appears to come from other directions.

A vertical velocity sensor has a sensor pattern similar to vertical dipole receiver and is thus blind to the horizontal propagating energy. Therefore, the energy is more focused in the vertical direction as shown by the dashed line in Fig. 8. Above the design frequency, the velocity sensor signal would also be aliased, but the effect would be much less, as there is little horizontal signal to be aliased. This reduces the effect of grating lobes away from end-fire.

A. Pressure sensor array

The beam response from the 32 hydrophone array is computed in Fig. 9(a), where the grating lobe is seen starting at the design frequency f_d=1500 Hz. When decreasing the number of phones to 16 (f_d=750 Hz), the aliasing in the beam response, Fig. 9(b), starts appearing at the lower design frequency and it is clear that the beam response is changed below the grating curve that starts at the design frequency. As in the Mapex2kbis data case, Fig. 4, below the grating lobe curve upgoing energy appears as downgoing energy.

The effect on the time domain response of changing the number of phones is investigated in Fig. 9(c). Each of the time series is produced using the same frequency interval 10–1500 Hz (upper band correspond to the design frequency at 1-m spacing). For fixed aperture and fixed upper frequency, when reducing the number of phones the corresponding design frequency reduces by the same factor.

Using all 32 phones (maximum frequency is f_d), the reflections from all three interfaces (130, 145, and 150 m) are seen clearly. (The arrivals for the subbottom bounces [145 and 150 m] appear slightly too early because a constant speed [1500 m/s] instead of the correct sediment speed was used to convert the time axis to depth.) With 16 phones (maximum frequency is $2f_d$), more noise starts appearing on the trace. For fewer phones (maximum frequency is $4f_d$ or more), the subbottom layers cannot be observed and the ocean bottom can barely be observed. The reason for this is that much of the horizontal energy comes in through the sidelobes and this contaminates the calculation of the down- and up-going energy.

The envelope of the time domain response is explored versus the maximum frequency used in the processing for 32 [Fig. 10(a)] and 16 [Fig. 10(b)] elements. First the spurious component can be seen to a depth of 100 m (corresponding to the array length). This tooth-shape response depends on the number of hydrophones, see Sec. II. The reflections from the subbottom start becoming contaminated at a frequency corresponding to $2f_d$ (white horizontal line).

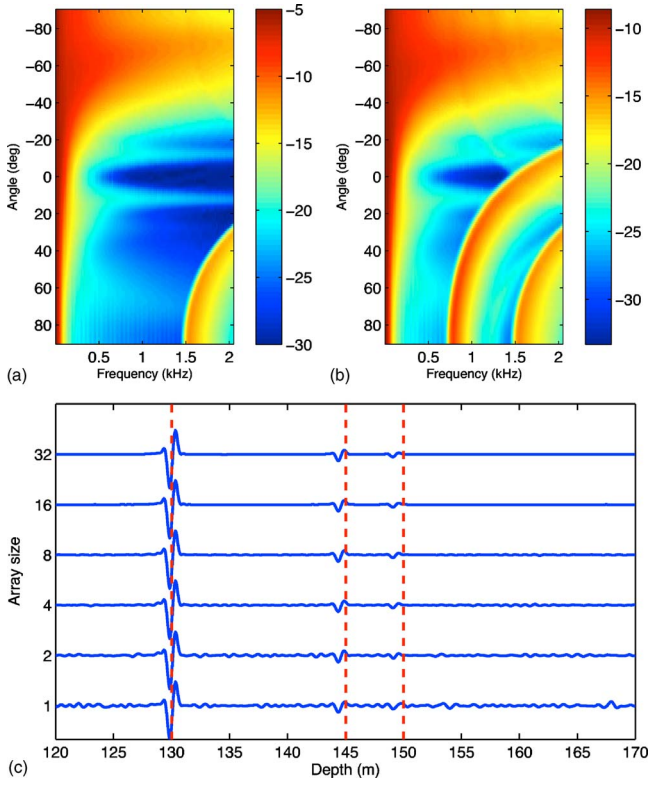


FIG. 11. (Color online) Vertical velocity sensors. Beam response for array with (a) 32 and (b) 16 sensors. (c) Time series for a vertical velocity sensor array. For description see caption in Fig. 9.

B. Vertical velocity sensor array

The same processing as carried out on the hydrophone array (Sec. V A) was carried out with a vertical velocity sensor array, see Fig. 11. As discussed earlier, a vertical velocity sensor is not as sensitive to horizontally propagating energy. This is seen in the beam response in Figs. 11(a) and 11(b).

Figure 11(c) shows the reflection response for the vertical velocity sensors using all 32 phones and then repeating the processing with 16, 8, 4, 2, and 1 phones, similar to the hydrophone array processing (Fig. 9). Note the change in polarity of the wave form relative to the pressure array. It is seen that for a vertical velocity sensor array it is possible to extract the reflection response with just one sensor. The reason being that the vertical velocity sensor is not sensitive to horizontally propagating energy in the waveguide.

VI. CONCLUSION

By array processing of ocean acoustic noise on a vertical array, it is possible to extract the reflection sequence convolved with the noise autocorrelation function. From this, information about the bottom depth and possibly deeper layers can be extracted. A problem in ocean acoustics is that much of the noise is horizontally propagating and this noise dominates the information-carrying vertically propagating noise. Using vertical array processing it is possible to mitigate the impact of the horizontally propagating noise.

When doing noise processing on a vertical array, the reflection response is sensitive to the phone spacing. Generally, the maximum frequency should not be larger that the design frequency of the array. However, higher resolution of the noise reflection sequence often can be obtained with a higher maximum frequency (here about two times the design frequency) and possibly the use of high resolution beamforming techniques.

A way to reduce the spatial sampling requirements is to use a vertical velocity sensor array that rejects horizontally propagating energy. A simulation example indicates that just one vertical velocity sensor is sufficient.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-05-1-0264.

[1]A. Curtis, P. Gerstoft, H. Sato, R. Snieder, and K. Wapenaar, "Seismic interferometry—Turning noise into signal," The Leading Edge **25**, 1082–1092 (2006).
[2]G. D. Bensen, M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Yang, "Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements," Geophys. J. Int. **169**, 1239–1260 (2007).
[3]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[4]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[5]K. Wapenaar, "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301 (2004).
[6]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).
[7]P. Roux and W. A. Kuperman, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[8]K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[9]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[10]M. Siderius, C. H. Harrison, and M. B. Porter, "A passive fathometer technique for imaging seabed layering using ambient noise," J. Acoust. Soc. Am. **120**, 1315–1323 (2006).
[11]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **229**, 547–549 (2003).
[12]E. Larose, A. Derode, M. Campillo, and M. Fink, "Imaging from one-bit correlations of wideband diffuse wavefields," J. Appl. Phys. **95**, 8393–8399 (2004).
[13]N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface wave tomography from ambient seismic noise," Science **307**, 1615–1617 (2005).
[14]K. G. Sabra, P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, "Extracting time-domain Green's function estimates from ambient seismic noise," Geophys. Res. Lett. **32**, L03310 (2005).
[15]P. Roux, K. G. Sabra, P. Gerstoft, W. A. Kuperman, and M. C. Fehler, "P-waves from cross correlation of seismic noise," Geophys. Res. Lett. **32**, L19303 (2005).
[16]P. Gerstoft, K. G. Sabra, P. Roux, W. A. Kuperman, and M. C. Fehler, "Green's functions extraction and surface-wave tomography from microseisms in southern California," Geophysics **71**, S123–S131 (2006).
[17]K. Wapenaar and J. Fokkema, "Green's function representations for seismic interferometry," Geophysics **71**, SI33–SI46 (2006).
[18]D. Draganov, K. Wapenaar, W. Mulder, J. Singer, and A. Verdel, "Retrieval of reflections from seismic background-noise measurements," Geophys. Res. Lett. **34**, L04305 (2007).
[19]W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).

[20]R. Snieder, K. Wapenaar, and K. Larner, "Spurious multiples in seismic interferometry of primaries," Geophysics **71**, SI111–SI124 (2006).
[21]L. A. Brooks and P. Gerstoft, "Ocean acoustic interferometry," J. Acoust. Soc. Am. **121**, 3377–3385 (2007).
[22]K. Mehta, A. Bakulin, J. Sheiman, R. Calvert, and R. Snieder, "Improving the virtual source method by wavefield separation," Geophysics **72**, V79–86 (2007).
[23]A. Papoulis, *Probability, Random Variables and Stochastic Processes* (McGraw-Hill Book Company, New York, 1965).
[24]H. L. Van Trees, *Optimum Array Processing* (Wiley, New York, 2002).
[25]R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1975).
[26]C. H. Harrison and M. T. Siderius, "Bottom profiling by correlating beam-steered noise sequences," J. Acoust. Soc. Am., in press (2008).
[27]C. H. Harrison and D. G. Simons, "Geoacoustic inversion of ambient noise: A simple method," J. Acoust. Soc. Am. **112**, 1377–1389 (2002).
[28]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).
[29]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).
[30]P. Hursky, "Modeling and processing for vector sensors," Office of Naval Research, Ocean Acoustics Annual Report 2006 (also available at http://www.onr.navy.mil/obs/reports/docs/06/oahursky.pdf, last visited 20 September 2007).
[31]C. H. Harrison, "Sub-bottom profiling using ocean ambient noise," J. Acoust. Soc. Am. **115**, 1505–1515 (2004).
[32]C. H. Harrison, "Performance and limitations of spectral factorization for ambient noise sub-bottom profiling," J. Acoust. Soc. Am. **118**, 2913–2923 (2005).
[33]H. Schmidt, *OASES Version 3.1 User Guide and Reference Manual* (MIT, Cambridge, 2004).

Joint time/frequency-domain inversion of reflection data for seabed geoacoustic profiles and uncertainties

Jan Dettmer[a] and Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria, BC V8W 3P6, Canada

Charles W. Holland
The Pennsylvania State University, Applied Research Laboratory, State College, Pennsylvania 16804-0030

(Received 27 August 2007; revised 14 December 2007; accepted 17 December 2007)

This paper develops a joint time/frequency-domain inversion for high-resolution single-bounce reflection data, with the potential to resolve fine-scale profiles of sediment velocity, density, and attenuation over small seafloor footprints (~100 m). The approach utilizes sequential Bayesian inversion of time- and frequency-domain reflection data, employing ray-tracing inversion for reflection travel times and a layer-packet stripping method for spherical-wave reflection-coefficient inversion. Posterior credibility intervals from the travel-time inversion are passed on as prior information to the reflection-coefficient inversion. Within the reflection-coefficient inversion, parameter information is passed from one layer packet inversion to the next in terms of marginal probability distributions rotated into principal components, providing an efficient approach to (partially) account for multi-dimensional parameter correlations with one-dimensional, numerical distributions. Quantitative geoacoustic parameter uncertainties are provided by a nonlinear Gibbs sampling approach employing full data error covariance estimation (including nonstationary effects) and accounting for possible biases in travel-time picks. Posterior examination of data residuals shows the importance of including data covariance estimates in the inversion. The joint inversion is applied to data collected on the Malta Plateau during the SCARAB98 experiment.
 [DOI: 10.1121/1.2832619]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 1306–1317

I. INTRODUCTION

Single-bounce acoustic reflection methods have great potential to resolve local sediment geoacoustic structure. Holland and Osler[1] developed an experimental technique to collect high-resolution seabed reflectivity data in shallow water using a bottom-moored hydrophone and ship-towed impulsive source, with a seafloor experimental footprint of ~100 m. Because of this local scale, theory errors and the effects of spatial and temporal variability in the water column and seabed are greatly reduced compared to long-range acoustic measurements. The reflection experiment yields seismo-acoustic data that can be analyzed in the time domain (travel-time data) and in the frequency domain (reflection-coefficient data). This paper develops a Bayesian joint time/frequency-domain inversion that is applied to measured seismo-acoustic data.

Bayesian inversion combines information from measured data, formulated in terms of a likelihood function, with prior information regarding the geoacoustic model parameters. The goal is to determine the posterior probability density (PPD) to quantify the state of knowledge of the model in terms of parameter estimates, marginal probability distributions, credibility intervals, parameter covariance/correlations, etc. Prior information can constrain parameters within physically realistic values (e.g., with bounded, uniform distributions), or can treat some parameter values as more probable than others. Sequential Bayesian methods use the PPD from one inversion as prior information for subsequent inversions. For nonlinear inverse problems, such as geoacoustic inversion, Bayesian inversion generally requires numerical optimization and integration of the PPD. Markov-chain Monte Carlo methods, such as Gibbs sampling, provide an efficient and effective approach.

Dettmer *et al.*[2] developed a Bayesian ray-tracing inversion to recover sound velocity profiles from reflection travel-time data. The inversion rigorously accounted for data uncertainties employing non-parametric estimation of error covariance matrices from residual analysis, and included nuisance parameters to account for data biases. Quantitative *a posteriori* statistical tests were applied to examine the validity of the underlying assumptions.

Dosso and Holland[3] applied a Bayesian inversion approach to frequency-averaged reflection-coefficient data, recovering the physical parameters of an elastic sediment half space using a plane wave reflection-coefficient forward model, and showed that shear properties did not significantly affect the inversion. Holland *et al.*[4] developed a Bayesian plane-wave inversion for broadband reflectivity data to recover velocity and density gradients in the uppermost sediments (~1 m). Dettmer *et al.*[5] developed a full wave-field (spherical-wave) reflection-coefficient Bayesian inversion to recover geoacoustic profiles, and applied the approach to

[a]Author to whom correspondence should be addressed. Electronic mail: jand@uvic.ca

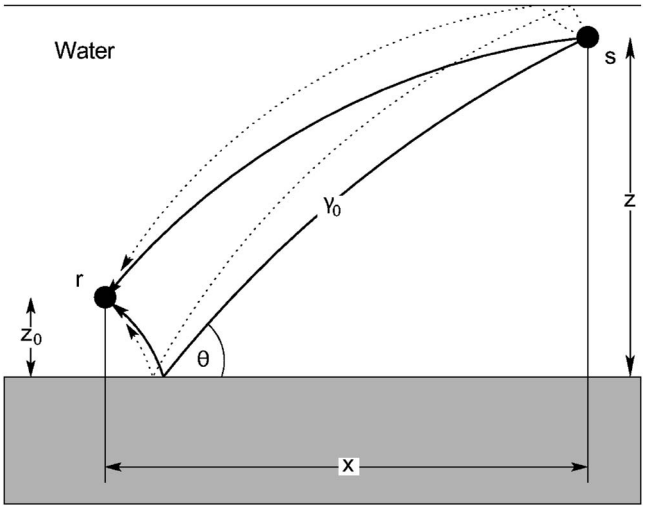


FIG. 1. Experiment geometry to measure reflection coefficients versus grazing angle θ. The direct and the bottom bounce path γ_0 are shown as solid lines, the surface reflected paths as dotted lines. Source **s** moves away from receiver **r** which is mounted at height z_0 above the seabed and transmits at ranges x.

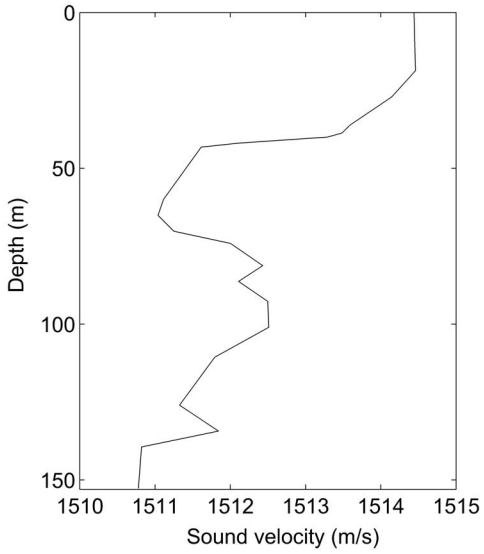


FIG. 2. Malta Plateau (Site 2) sound velocity profile.

synthetic data. Modeled reflection coefficients were obtained by processing full-wave acoustic fields (computed numerically) in the same manner as the measured data. This approach is general and efficient, and allows arbitrary water-column sound velocity profiles and sediment interface roughness. Using simulations, Dettmer *et al.*[5] showed that accounting for full-wave effects was necessary to obtain reliable inversion results at greater depths for typical shallow-water geometries.

This paper combines components of previous work[2–5] to develop a joint time/frequency-domain Bayesian inversion to recover complex fine-scale geoacoustic profiles parametrized in terms of layer thickness (h), seabed sound velocity (c), density (ρ), and attenuation (α) from broadband reflection data. The inversion is based on a multi-stage (sequential Bayesian) strategy. In the time domain, travel-time data are picked for multiple reflectors in the seismo-acoustic traces and ray-tracing inversion[2] is carried out for velocities and thicknesses of a multi-layer model. Posterior credibility intervals from the time-domain inversion are then passed on as prior information to a subsequent frequency-domain inversion.

The frequency-domain inversion is based on full-wave reflectivity modeling[5] and a layer-packet stripping approach. For this inversion, the seismo-acoustic time series are processed to yield reflection coefficients as a function of angle and frequency. Time windowing is applied to process data packets which contain the effects of sets of layers from the seafloor to a sequence of increasing depths (e.g., a given packet contains the layers of the previous packet plus additional deeper layers). In the present case, two packets are used, with the first packet containing two layers and the second packet containing four layers extending to approximately 10 m depth. Each layer is parameterized in terms of layer thickness, velocity, density, and attenuation (the last layer is a half space with infinite thickness). The layer-packet stripping inversion steps down through the packets, using the results (PPD) from one packet inversion as prior information to constrain the upper (common) layers of the subsequent packet inversion. Parameter information is passed from one layer-packet inversion to the next in terms of marginal probability distributions, rotated into a principal-component space. This provides an efficient approach to (partially) account for inter-parameter correlations with one-dimensional, numerical (i.e., sampled) distributions. All inversion stages (time and frequency domain) employ data error covariance estimates, including effects of non-stationarity, and the travel-time inversion accounts for possible picking biases. Finally, rigorous *a posteriori* statistical tests are applied to examine the validity of the covariance estimation and underlying assumptions on the form of the likelihood function.[2,4,6–8]

II. EXPERIMENT AND DATA

This paper applies joint time/frequency-domain inversion to seismo-acoustic reflection data collected during the SCARAB98 experiment on April 22, 1998, on the Malta Plateau (Site 2: 36°30′39″ N 14°32′59″ E), Mediterranean Sea. The experiment geometry is shown in Fig. 1. The source was an electro-mechanical impulsive device (EG&G model 265 Uniboom, commonly referred to as a "boomer") with a short pulse length (<1 ms) and broad bandwidth. The data used here were recorded at a single receiver of a vertical line array of 16 Benthos AQ-4 hydrophones. The source depth was 0.35 m, the hydrophone depth was 112 m and the water depth was 153 m. The water column sound velocity profile is given in Fig. 2, and shows less than 4 m/s variation over the entire column.

Seabed reflection data were collected for two roughly perpendicular ship tracks, both crossing the coordinates of the receiver. This results in four data sets: two with the ship approaching the receiver, and two with the ship departing from the receiver. All four data sets were considered, but the results are very similar and, for brevity, only one set is con-

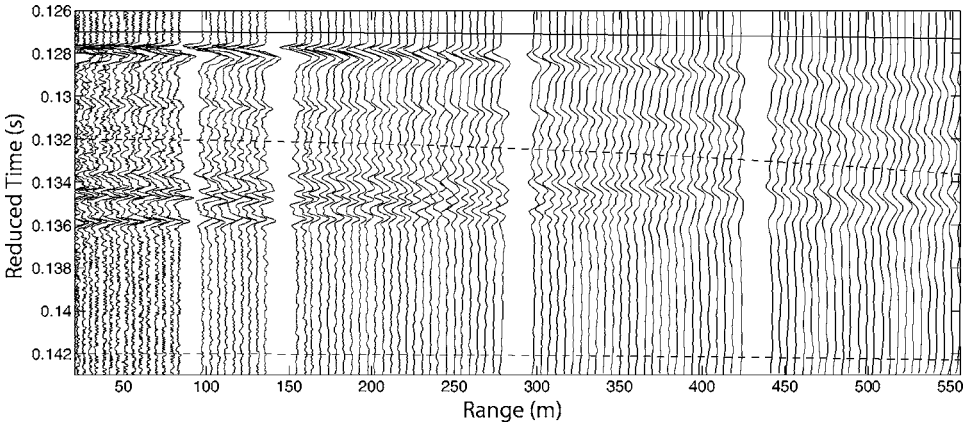


FIG. 3. Malta Plateau seismo-acoustic data in reduced time. The solid line indicates the upper limit of the layer stripping packets, and dashed lines represent the lower limits of the first and second packets.

sidered here. The seismo-acoustic traces are given in reduced travel time in Fig. 3 for a reducing velocity of $c_R = 1511$ m/s according to

$$t_R = \sqrt{t^2 - x^2/c_R^2}, \tag{1}$$

where x is the range. In Fig. 3, the water-sediment interface reflection can be seen at 0.1275 s. The data show low noise levels, and a number of sub-bottom reflections are clearly visible. Gaps due to missing traces in Fig. 3 are caused by data being written to disk, which interrupted the recording.

The seismo-acoustic data are inverted for an environmental model consisting of multiple flat-lying layers. The analysis presented in this paper is applied to the upper four reflectors, representing approximately 10 m depth into the seabed. Although some deeper reflectors are evident in the data, an independent high-resolution seismic survey of the experiment site indicated complicated three-dimensional geological structure at greater depths that violates the assumption of flat-lying layers. The data processing applied to obtain reflection-coefficient data follows Holland.[9]

III. INVERSION TECHNIQUE

A. Bayesian inversion

This section provides a brief overview of the Bayesian formulation used here for the geoacoustic inversion;[10] more general treatments of Bayesian theory can be found elsewhere.[11,12] Bayes' rule can be written

$$P(\mathbf{m}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{P(\mathbf{d})}, \tag{2}$$

where $\mathbf{m} \in \mathbb{R}^M$ and $\mathbf{d} \in \mathbb{R}^N$ are random variables that represent the model parameters and data, respectively. $P(\mathbf{m}|\mathbf{d})$ is the PPD, $\mathcal{L}(\mathbf{d}|\mathbf{m})$ is the likelihood function, $P(\mathbf{m})$ is the model prior distribution, and $P(\mathbf{d})$ is the data prior (a constant factor once the data are measured). The likelihood function can generally be expressed as $\mathcal{L}(\mathbf{m}) \propto \exp(-E(\mathbf{m}))$ where $E(\mathbf{m})$ is an appropriate data error function (considered later). Equation (2) then becomes

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m})]}{\int_{\mathcal{M}} \exp[-\phi(\mathbf{m}')]d\mathbf{m}'}, \tag{3}$$

where the integration is over the model space $\mathcal{M} \subset \mathbb{R}^M$ and

$$\phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m}) \tag{4}$$

is the generalized misfit. The multi-dimensional PPD is generally interpreted in terms of properties such as the maximum *a posteriori* (MAP) model $\hat{\mathbf{m}}$, the model covariance matrix $\mathbf{C}^{(\mathbf{m})}$, and marginal probability distributions $P(m_i|\mathbf{d})$,

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max} P(\mathbf{m}|\mathbf{d}), \tag{5}$$

$$\mathbf{C}^{(m)} = \int_{\mathcal{M}} (\mathbf{m}' - \bar{\mathbf{m}})(\mathbf{m}' - \bar{\mathbf{m}})^T P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{6}$$

$$P(m_i|\mathbf{d}) = \int_{\mathcal{M}} \delta(m_i' - m_i) P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{7}$$

where δ denotes the Dirac delta function and $\bar{\mathbf{m}}$ is the mean model. Higher-dimensional marginal distributions can be defined similar to Eq. (7). Uncertainties of parameter estimates can also be quantified in terms of highest probability density (HPD) credibility intervals, with the β% HPD interval defined as the interval of minimum width that contains β% of the area of the marginal distribution. Interrelations of model parameters can be quantified by the model correlation matrix $R_{ij} = C_{ij}^{(m)} / (C_{ii}^{(m)} C_{jj}^{(m)})^{1/2}$. While analytic solutions to Eqs. (5)–(7) exist for linear inverse problems, nonlinear problems such as geoacoustic inversion must be solved numerically.

MAP estimates [Eq. (5)] can be found by numerical minimization of $\phi(\mathbf{m})$, such as adaptive simplex simulated annealing (ASSA), an efficient hybrid optimization algorithm that combines the local downhill-simplex method within a very fast simulated annealing global search.[13] The integrals of Eqs. (6) and (7) are computed here using the Markov-chain Monte Carlo method of fast Gibbs sampling[3,10] (FGS) to sample $\phi(\mathbf{m})$. FGS applies an adaptive Metropolis Gibbs sampling scheme[14,15] in a principal-component parameter space where the coordinate axes align with the dominant correlation directions.

B. Likelihood function and data uncertainties

Formulating the likelihood function $\mathcal{L}(\mathbf{m})$ requires specifying the data uncertainty distribution, including both measurement errors and theory errors. In general, the Bayesian inversion outlined above can be applied with arbitrary uncertainty distributions. In practice, however, the lack of specific knowledge of the error distribution often suggests a mathematically simple distribution to be assumed. In particular, Gaussian distributions are commonly considered, with their statistical parameters estimated from the data. The validity of this assumption should be checked *a posteriori* using statistical tests.

Consider a data set $\mathbf{d}=\{\mathbf{d}_i, i=1,\ldots,Q\}$ which consists of Q subsets (with N_i data in the ith subset) such that the data error statistics are assumed constant for each subset, but may vary between subsets. In the present work, Q represents the number of reflecting layers in travel-time inversion and the number of frequencies in reflection-coefficient inversion. Assuming unbiased Gaussian-distributed random errors for each data subset, with no error correlations between subsets, the likelihood function for the complete data set is given by

$$\mathcal{L}(\mathbf{m}) \propto \prod_{i=1}^{Q} \exp\left(-\frac{1}{2}[\mathbf{d}_i-\mathbf{d}_i(\mathbf{m})]^{\mathsf{T}}(\mathbf{C}_i)^{-1}[\mathbf{d}_i-\mathbf{d}_i(\mathbf{m})]\right), \tag{8}$$

where $\mathbf{d}_i(\mathbf{m})$ are the replica data computed for model $\mathbf{m}$ and $\mathbf{C}_i$ are the Q $N_i \times N_i$ data covariance matrices. Maximizing the likelihood function is equivalent to minimizing the negative log likelihood, which is termed the data misfit function

$$E(\mathbf{m}) = \sum_{i=1}^{Q} \frac{1}{2}(\mathbf{d}_i-\mathbf{d}_i(\mathbf{m}))^{\mathsf{T}}(\mathbf{C}_i)^{\mathsf{T}}(\mathbf{d}_i-\mathbf{d}_i(\mathbf{m})). \tag{9}$$

The covariance matrices $\mathbf{C}_i$ are usually not known *a priori*. Depending on the data error distribution, different approaches can be applied to estimate covariance matrices. If the data do not have significantly correlated errors, the covariance matrices can be approximated as diagonal, $\mathbf{C}_i=\sigma_i^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. If significant error correlations are present, full data covariance matrices must be estimated. The data covariance matrix can account for nonstationary data errors when estimated in non-Toeplitz form from the data residuals using a nonparametric approach (which does not assume a specific correlation form).[2] Non-Toeplitz covariance matrices can be estimated from the data residuals $\mathbf{r}_i=\mathbf{d}_i-\mathbf{d}_i(\hat{\mathbf{m}})$ (where $\hat{\mathbf{m}}$ is a maximum likelihood model obtained by minimizing a misfit function assuming unknown standard deviations of the data errors[4]) by a three step procedure. First, standard deviations σ_i are estimated from data residuals $\mathbf{r}_i=\mathbf{d}_i-\mathbf{d}_i(\hat{\mathbf{m}})$ as a running root mean square (rms) average and used to standardize the jth residual of the ith data subset according to

$$n_{ij} = \frac{r_{ij}}{\sigma_{ij}}. \tag{10}$$

Second, computing the autocovariance function for the standardized residuals $\mathbf{n}_i$ yields Toeplitz data covariance matrices $\tilde{\mathbf{C}}_i$. Third, these matrices are scaled by the rms standard deviations to yield non-Toeplitz data covariance matrices

$$[\mathbf{C}_i]_{jl} = [\tilde{\mathbf{C}}_i]_{jl}\sigma_{ij}\sigma_{il}. \tag{11}$$

In some cases, the assumption of unbiased data errors may not be valid (e.g., inversion of picked travel times[2]). Biases can be accounted for in the data misfit function by including an unknown offset parameter a_i for each independent data subset according to

$$E(\mathbf{m}) = \sum_{i=1}^{Q} \frac{1}{2}[\mathbf{d}_i+a_i-\mathbf{d}_i(\mathbf{m})]^{\mathsf{T}}(\mathbf{C}_i)^{-1}[\mathbf{d}_i+a_i-\mathbf{d}_i(\mathbf{m})]. \tag{12}$$

Including these offset parameters has the effect of increasing the posterior parameter uncertainties to account for the reduction in data information due to possible biases. The resulting PPD then contains information about the bias parameters and indicates if biases are resolved.

The likelihood function in Eq. (8) assumes Gaussian-distributed data errors with estimated covariance matrices $\mathbf{C}_i$. The validity of these assumptions should be examined *a posteriori* by applying statistical tests to the data residuals (which can be considered a realization of the data errors).[2,4,8] Comparing test results for raw and standardized residuals examines the effect of applying the data covariance matrix estimates in the inversion. Standardized residuals are scaled according to $\tilde{\mathbf{r}}_i=\mathbf{L}_i^{-1}\mathbf{r}_i$, where $\mathbf{L}_i$ is the lower triangular matrix of the Cholesky decomposition (matrix square root)

$$\mathbf{C}_i = \mathbf{L}_i\mathbf{L}_i^{\mathsf{T}}. \tag{13}$$

Raw residuals are scaled only by their standard deviation. The quantitative statistical tests used here examine the form of the residual distribution (e.g., Gaussian) through the Kolmogorov–Smirnov (KS) test,[16,17] and examine the randomness of the residuals through the runs test.[17] For both tests, a level of significance of $p \geqslant 0.05$ is generally accepted as little or no evidence against the initial assumption of randomly distributed Gaussian errors with covariance $\mathbf{C}_i$. Further, the form of the distribution and randomness of the errors can be considered qualitatively by examining histograms and autocorrelations of the data residuals, respectively.

C. Prior information

Prior information can originate from several sources. In an initial inversion, prior parameters can simply be chosen as uniform distributions over physically reasonable bounds so as to let the data predominantly determine the inversion results. Prior information can also be derived from previous inversion results for a different data set. In this case the prior can be given in parametric form (e.g., a Gaussian distribution for linear inversion) or nonparametric form. Numerical inversions, such as Gibbs sampling, provide a nonparametric representation of the PPD in terms of a large sample of models, which can be used to provide prior information in a subsequent inversion. However, for high-dimensional problems, it can be computationally infeasible to compute multidimensional priors directly from the PPD sample. To avoid this problem, one-dimensional marginal probability distribu-

tions can be used to provide prior information at the cost of loosing information about parameter inter-relationships. For instance, the prior probability $P(\mathbf{m})$ for a candidate model $\mathbf{m}$ can be computed from the probabilities p_i of the one-dimensional marginal distributions of all model parameters m_i as

$$P(\mathbf{m}) = \prod_{i=1}^{M} p_i(m_i). \tag{14}$$

Using one-dimensional marginal distributions [Eq. (7)] for the calculation of the prior probability is a simplification that does not account for the full M-dimensional probability structure, including correlations between parameters. To partially include parameter correlation information, a principal-component rotation can be performed. The rotation matrix is obtained from the eigenvector decomposition of the model covariance matrix determined from the PPD

$$\mathbf{C}^{(\mathbf{m})} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathsf{T}}. \tag{15}$$

This matrix $\mathbf{U}$ is then used to rotate the PPD sample into principal components, and marginal distributions are computed for the rotated parameters. The prior information then consists of the principal-component marginals and the prior rotation matrix $\mathbf{U}$. When evaluating the misfit, the prior probability $P(\mathbf{m})$ is computed for the rotated model $\mathbf{m}'$ according to Eq. (14). The transformation pair between rotated and physical space is

$$\mathbf{m}' = \mathbf{U}^{\mathsf{T}}\mathbf{m}$$

$$\mathbf{m} = \mathbf{U}\mathbf{m}'. \tag{16}$$

Approximating the M-dimensional PPD by principal-component marginal distributions is expected to capture the general structure of the inter-parameter correlations. For instance, in the case of a linear inverse problem with Gaussian distributed errors, the PPD is also an M-dimensional Gaussian. In this case, marginal distributions of principal components contain all the information of the M-dimensional PPD. For nonlinear inverse problems, the correlation structure can vary over the parameter space, and may not be fully described in terms of principal components. However, principal-component marginals should provide the best overall representation of the PPD in terms of one-dimensional distributions.

For the joint inversion developed in the present study, an initial inversion of travel-time data is carried out using uniform prior information within wide bounds. The resulting PPD is then used to constrain layer thickness information by using 100% HPD credibility intervals as uniform prior information for the reflection-coefficient inversion. The reflection-coefficient inversion uses a layer stripping approach where the inversion result for one packet of layers constrains the subsequent inversion of the next packet. In particular, the layers that were included in the previous packet inversion are constrained by the principal-component marginal distributions; the new, deeper layers have uniform priors with wide bounds. To ensure a conservative approach to layer stripping, the estimated principal-component prior distributions can be widened to allow a greater range of models to be considered in the inversion and decrease the likelihood of ruling out physically meaningful models at an early stage as a result of the layer stripping.

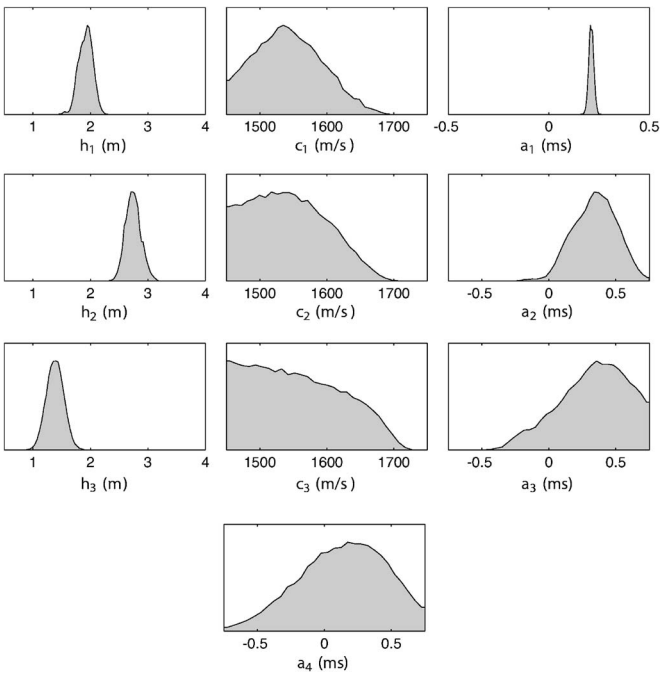


FIG. 4. Marginal probability distributions for the inversion of picked travel-time data for layer thickness h, sound velocity c, and bias parameters a. Prior bounds correspond to plot widths.

IV. TRAVEL-TIME INVERSION

Travel-time inversion was applied previously to the Malta Plateau Site 2 seismo-acoustic data in Dettmer *et al.*[2] This section gives a brief overview of new inversion results for the uppermost four reflectors (deeper reflectors are not considered due to complicated geological structure at greater depth evident in independent seismic data). The results are applied as prior information in the reflection-coefficient inversion in Sec. V.

For the travel-time inversion, four reflectors including the sediment water interface were picked. This resulted in travel-time data consisting of 83 points per reflector for offsets between 30 and 420 m with a range spacing of ~4.7 m. The maximum offset was limited to just over 400 m because travel-time differences (due to relatively thin layers) became too small for reliable picking at larger offsets. Bayesian inversion was applied to the travel-time data addressing biased, correlated, and non-stationary data errors.

The resulting marginal distributions for layer thicknesses, sound velocities, and bias parameters are given in Fig. 4. Layer thicknesses are resolved well within their prior bounds, but sound velocities show limited sensitivity. The bias parameters (a_i for the ith layer) are resolved reasonably well within their prior bounds. Prior bounds for the biases were chosen wide according to the confidence in picking for certain parameters, which depends on the noise level and separation of individual reflectors. The MAP model and 100% HPD credibility intervals are given in Table I for layer

TABLE I. Numerical values of prior information and inversion results (MAP model and 100% HPD credibility intervals) for picked travel-time data.

	h (m)			c (m/s)		
Layer	Prior	MAP	100% HPD	Prior	MAP	100% HPD
1	0.50-4.00	1.77	1.45-2.28	1450-1750	1614	1450-1688
2	0.50-4.00	2.79	2.33-3.19	1450-1750	1452	1450-1705
3	0.50-4.00	1.58	0.81-1.89	1450-1750	1472	1450-1723

thicknesses and sound velocities. The HPDs for layer thicknesses are used as prior information in the reflection-coefficient inversion in the following section. Since the travel-time inversion did not significantly change the information about velocities, velocity priors are not changed.

The statistical assumptions about the data errors were tested with raw and standardized residuals according to Sec. III B. The raw residuals fail the runs test for every reflector with p values $\sim 10^{-4}$. After applying the data covariance matrix estimates, the standardized residuals for all reflectors pass the runs test at the 0.05 level. For the standardized residuals, three out of four reflectors pass the KS test at the 0.05 level, whereas the raw residuals fail the KS test for all reflectors. The tests indicate that the assumptions appear reasonable and that the inversion likely gives useful estimates and uncertainties for the geoacoustic parameters given the model parameterization. Hence, the recovered information is applied as prior knowledge for the following spherical reflection-coefficient inversion in the frequency domain.

V. SPHERICAL WAVE REFLECTION-COEFFICIENT INVERSION

This section develops a layer packet stripping inversion algorithm for reflection-coefficient data. The approach is based on windowing acoustic time series to isolate reflections from sequences of sediment layers (packets) containing information from the sub seafloor to increasing depths. The acoustic data are processed to represent reflected energy across frequency and grazing angle.

The travel-time inversion showed that layer thicknesses for three layers can be resolved in the upper 10 m of the sediment. Hence, the frequency-domain inversion uses a model with four sediment layers (the lowest layer being an underlying half space). The model consists of layers defined by thickness, sound velocity, density, and attenuation. To reduce ambiguity and keep the numerical effort manageable in the inversion, a Bayesian layer stripping approach is applied to recover geoacoustic depth profiles. Figure 3 shows the seismo-acoustic data and the packet boundaries used in the inversion. Due to rapid changes of the reflection coefficient with angle and a relatively large sub-bottom depth ($\sim$10 m) in the Malta Plateau data, a full wave-field (spherical-wave) reflection-coefficient forward model[5] is used. The data misfit function applied in the inversion is given in Eq. (9) with covariance matrices estimated from residual analysis.

The first layer packet is time windowed to about 3 m depth below the seafloor, containing the effects of the uppermost layer. The second packet contains the response from the seafloor to about 10 m depth and contains the effects of the uppermost four layers which were identified in the travel-time inversion. These time windows were then processed according to Holland,[9] to yield the reflection-coefficient data shown in Fig. 5. The data are averaged into five frequency bands from 400 to 1600 Hz using a Gaussian frequency average[5,18] with a fractional bandwidth of 1/20. This bandwidth was found to be appropriate as it was a good compromise between retaining structure in the reflection coefficient and reducing noise. For the computation of replica data, the frequency average is replaced by a range average[18] to save computational time, as discussed by Dettmer *et al.*[5] The data are interpolated onto a uniform spacing in angle with 130 data points at each frequency, and points with a signal to noise ratio of less than 6 dB were excluded. Further, interpolated data that fall into the recording gaps are excluded from computing the data misfit. Figure 5 shows that the reflection coefficient becomes more complicated with greater depth (i.e., the second packet which extends more deeply) and that the level generally increases with depth. This behavior is expected since more energy is reflected upwards for packets extending to increasing depth, and the additional layers complicate the shape of the reflection-coefficient curve due to interference effects. Figure 5 also illustrates the general idea of the layer packet stripping approach. The uppermost packet has lower information content but also less noise. Detail, information content, and noise increase with increasing depth. Hence, extracting the shallow structure should be more straightforward from the shallow packet than from the more complicated deeper packet.

A. First packet inversion

The fit to the data obtained in the inversion for the first packet is shown in Fig. 6. In general, the data are well fit,

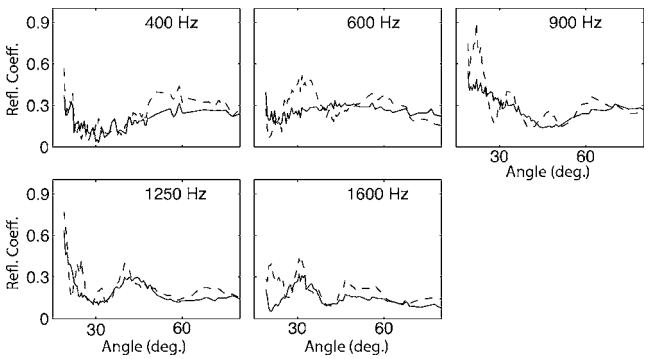


FIG. 5. Reflection coefficient data processed into two packets that extend to increasing depth (packets one and two indicated by solid and dashed lines, respectively).

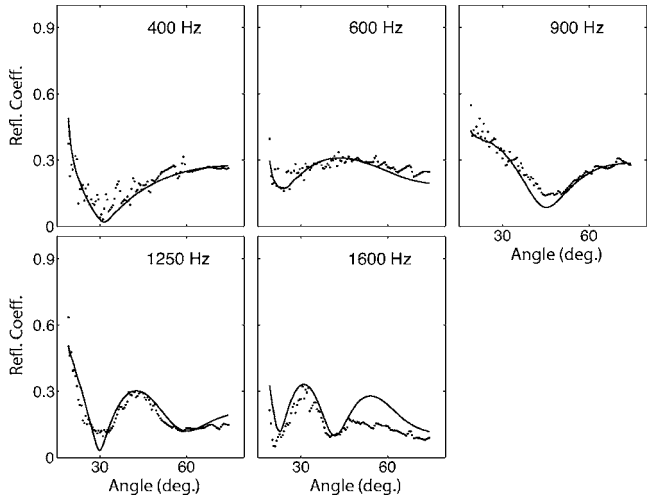


FIG. 6. Measured reflection coefficients and MAP model replica (solid line) for the first layer packet consisting of one layer over a half space.

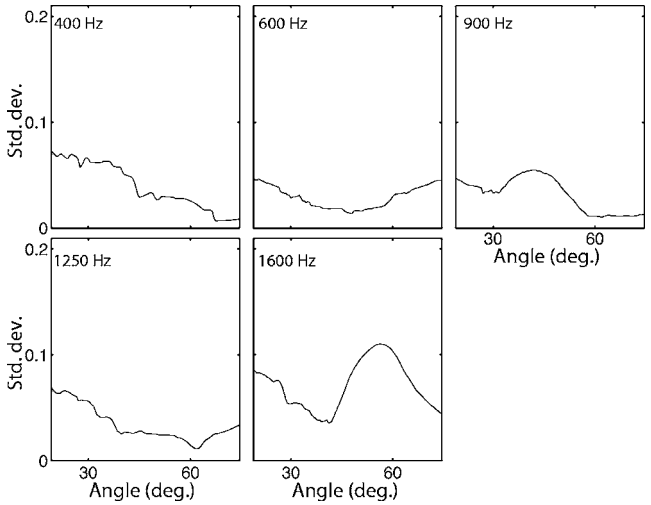


FIG. 8. Estimated running rms average standard deviation (400–1600 Hz) for the first packet. The variation with angle indicates nonstationary residuals.

particularly at low frequencies. However, at 1600 Hz, the replica data are offset towards higher reflection-coefficient values. Several possible reasons for this were considered. One reason could be scattering. To investigate the effect of scattering, a rough interface with a Gaussian roughness spectrum[19,20] was included in the forward model between the first layer and the half space, with a rms roughness length of 0.05 m. ASSA optimization yielded a MAP model which was then used to conduct a sensitivity study. For the sensitivity study, replica data were calculated using the MAP model but changing the rms roughness length. The results of this study (not shown) indicate that roughness cannot account for the offsets since the data misfit was not improved significantly for any value of roughness. In particular, roughness does not seem to be able to account for the offset in level at high angles at 1600 Hz. Another possible explanation for the offset is that gradients in density and sound velocity might be present in the uppermost part of the sediment and might affect the reflection coefficients in a way that is not properly modeled by a single layer. Recovering gradients in multilayer inversions of reflectivity data is an interesting and challenging problem,[4] but is beyond the scope of this study.

The replica and measured data yield the data residuals shown in Fig. 7. These residuals were used to estimate the data error standard deviation as a function of angle (Fig. 8). Figure 8 shows that the residuals are nonstationary in amplitude. The residuals and rms standard deviations were used to estimate the data covariance matrices shown in Fig. 9 using Eqs. (14) and (15). The covariance matrices were estimated over two iterations to ensure a stable result. These data covariance matrices were then applied in sampling the PPD via FGS.

The inversion result for packet one is presented in terms of marginal probability distributions in Fig. 10. The model parameters are resolved well within the prior bounds (given in Table II). *A posteriori* statistical tests were performed for raw and standardized residuals and are similar to those which

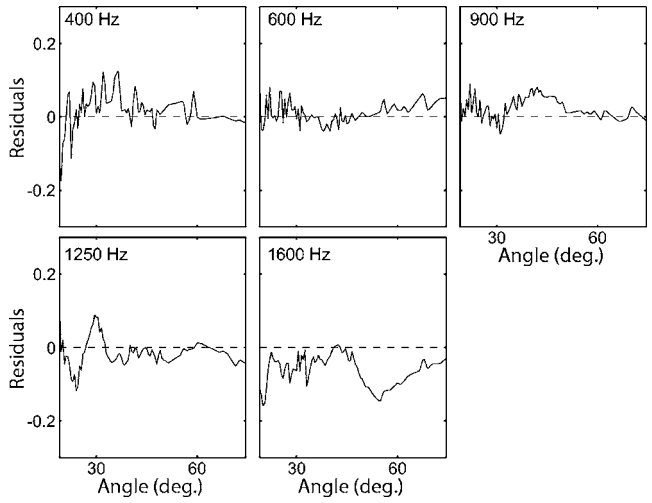


FIG. 7. Data residuals for the first packet. The residuals are strongly correlated.

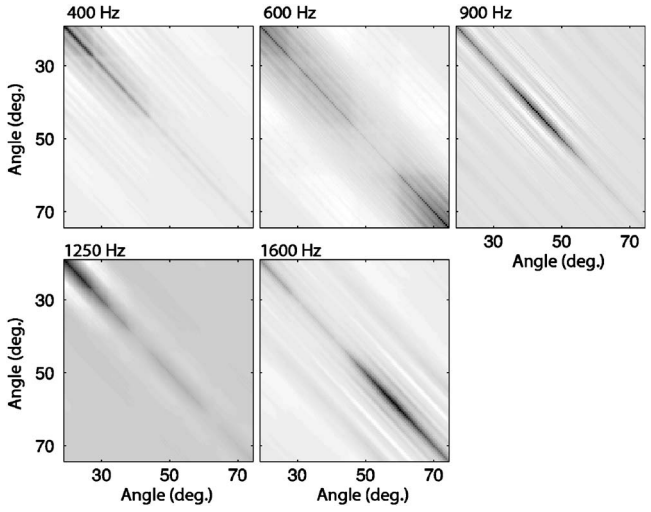


FIG. 9. Estimated non-Toeplitz covariance matrices for all frequencies, packet one.

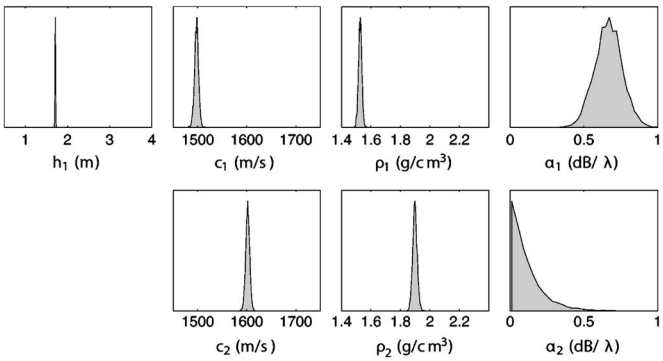


FIG. 10. Marginal probability distributions from reflection-coefficient inversion, packet one.

will be presented in detail for the inversion of packet two in Sec. V B.

Once the final PPD for the first packet is obtained, it is processed to be used as prior information in the inversion for the second packet since subsequent packets include the layers of the previous packet, as well as additional deeper layers (constrained by uniform prior bounds). The information about the half space is generally discarded since the half space is not as well determined as the overlying layers (as there is no interference information for the half space due to the missing lower reflector). To define the prior for the second packet, principal-component marginal distributions for the parameters of the first layer were calculated according to Sec. III C and widened by a factor of 2 around their center values to allow wider searching in the second packet inversion.

B. Second packet inversion

For the inversion of the second packet, in addition to the first layer, two more layers and a half space are included in the model, resulting in a total of 15 parameters in the inversion. The data misfit for the MAP model is shown in Fig. 11. The general structure of the data is well represented by the replica data and the main features agree. However, at low frequencies (400 and 600 Hz) the fit is not as good as at higher frequencies. At these low frequencies and large grazing angles, the level of the reflection coefficient is consistently too low in the modeled data. This could be caused by model mismatch at depth or be a cumulative effect over all layers, since the lower frequencies penetrate most deeply. The structure (i.e., interference pattern) of the reflection-coefficient curves is reasonably well fit even at these low frequencies. These data errors are accounted for by the non-Toeplitz covariance matrices derived from the data residuals as discussed previously.

The marginal probability distributions for the packet two inversion are shown in Fig. 12. Layer thicknesses, sound velocities, and densities for the upper three layers are resolved well within the prior bounds (Table II). In the fourth layer (half space), the sound velocity is well resolved; however, the layer shows a distribution for the density that peaks at high values. Several other inversions were run with higher upper bounds and showed that this density continued to increase as the bounds were opened up. This behavior is likely due to the complicated deep geological structure at the measurement site. It is not considered physically reasonable to allow higher densities than given in Table II and the inversion results were computed using those values. Attenuation in Fig. 12 shows limited sensitivity within the prior bounds. The second layer (with the largest layer thickness) shows the best resolution and peaks at values of approximately 0.15 dB/λ.

Examining correlations between model parameters is useful in terms of understanding an inverse problem. Figure 13 shows the correlation matrix for the packet two inversion result. The strongest correlations are observed between parameters in the same layer or between parameters of adjacent layers. The farther layers are separated, the weaker the correlations. Figure 14 shows joint marginal distributions for the parameters with the four strongest correlations. The shapes of the joint marginals are relatively simple and should be captured well by principal-component prior distributions as discussed in Sec. III C. Some inter-parameter correlations can be understood qualitatively. The positive correlation of 0.77 between c_1 and h_1 is due to conservation of travel time within the first layer [Fig. 14(a)]. The negative correlation of −0.61 between α_3 and c_2 [Fig. 14(b)] maintains the same reflected energy given that c_3 is greater than c_2: Lower attenuation in a layer means an acoustically softer reflector

TABLE II. Numerical values of prior bounds and final inversion results (MAP model and 95% HPD credibility intervals).

		h (m)			c (m/s)			ρ (g/cm^3)	
Layer	Prior	MAP	95%HPD	Prior	MAP	95%HPD	Prior	MAP	95%HPD
1	1.45-2.29	1.70	1.66-1.74	1450-1750	1484	1473-1498	1.4-1.9	1.594	1.525-1.639
2	2.33-3.23	2.96	2.91-3.00	1500-1750	1552	1542-1556	1.5-2.1	1.870	1.798-1.936
3	0.80-2.03	1.35	1.16-1.51	1500-1750	1621	1605-1642	1.5-2.4	2.214	2.116-2.298
4				1550-1750	1594	1551-1625	1.6-2.4	2.399	2.274-2.400

		α (dB/λ)	
	Prior	MAP	95% HPD
1	0.01-1.0	0.253	0.097-0.515
2	0.01-1.0	0.134	0.010-0.241
3	0.01-1.0	0.057	0.010-2.298
4	0.01-1.0	0.319	0.010-0.990

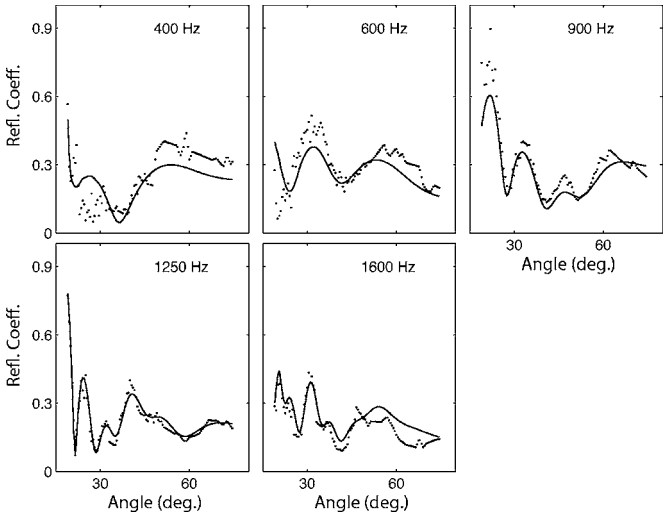


FIG. 11. Measured reflection coefficient and MAP model replica (solid line) for the second layer packet consisting of three layers over a half space.

which can be counteracted by higher sound velocity in the layer above. The negative correlation of −0.64 between c_1 and ρ_1 is due to their coupling as the acoustic impedance of the layer [Fig. 14(c)].

To validate the underlying assumptions of the inversion, the runs test and KS test were performed and plots of residual autocorrelations and histograms were examined at each frequency for raw and standardized residuals. Figure 15 shows the autocorrelation function of the raw and standardized data residuals. It is evident that the center peak is much narrower for the standardized residuals. The standardized residuals passed the runs test at all frequencies at a $p=0.05$ level, whereas the raw residuals failed the test at all frequencies (Fig. 16).

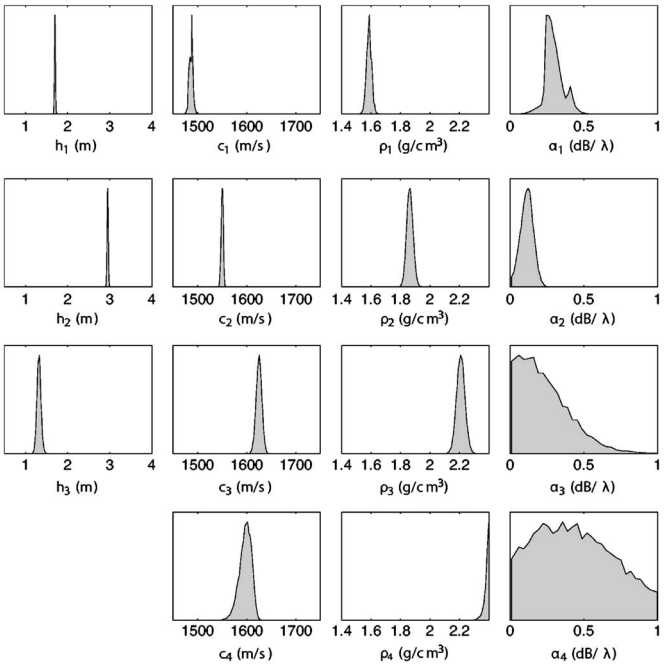


FIG. 12. Marginal probability distributions recovered from reflection-coefficient inversion of layer packet two.

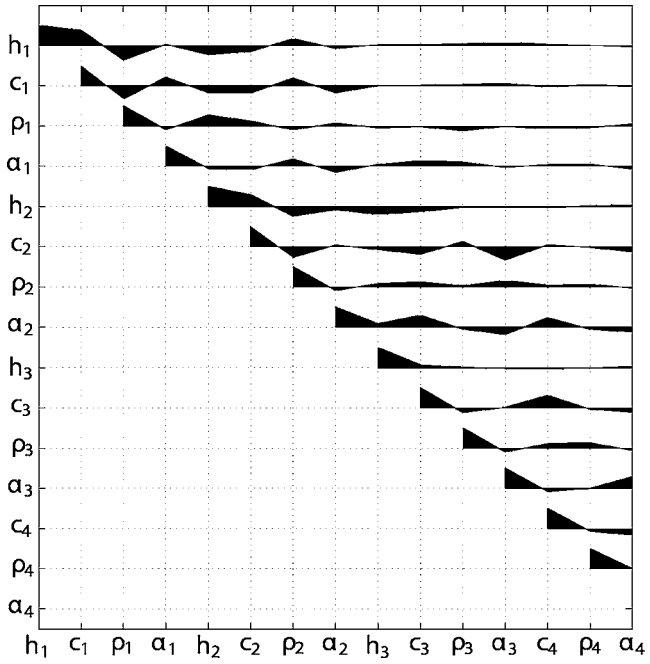


FIG. 13. Upper diagonal of the parameter correlation matrix for packet two reflection-coefficient inversion.

Figure 17 shows histograms of the residuals superimposed with a theoretical Gaussian distribution. The resulting p values for the KS test are given in Fig. 18. The test is passed at three of five frequencies for the standardized residuals, while the raw residuals fail at four frequencies. The residual histograms shown in Fig. 17 do not appear strongly non-Gaussian. Hence, it can be concluded that while the covariance matrix estimates do not fully describe the data error statistics, they significantly improve the inversion. In particular, there is no evidence of large outliers in the residuals, which are problematic in L_2-norm inversions. Hence, the inversion results are expected to be reliable. Further, the re-

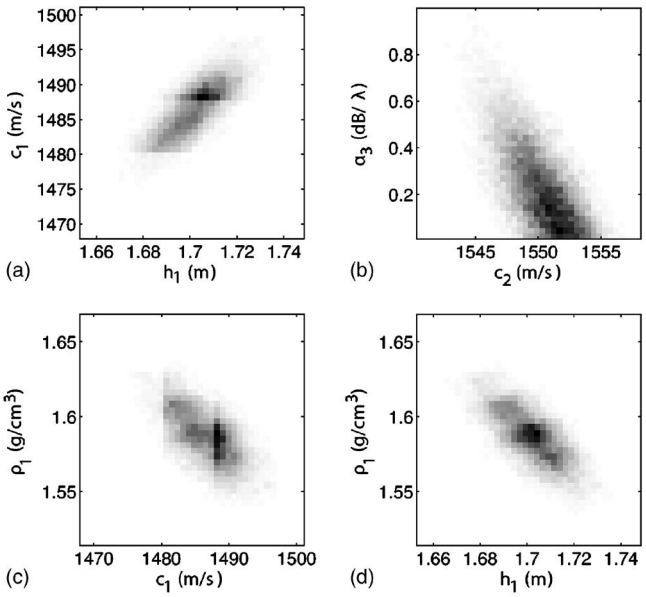


FIG. 14. Joint marginal probability distributions for selected parameters.

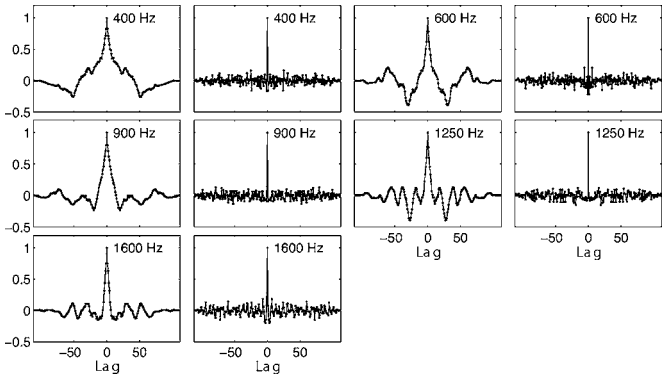


FIG. 15. Autocorrelation function for raw (columns one and three) and standardized (columns two and four) residuals as a qualitative measure of data error covariances.

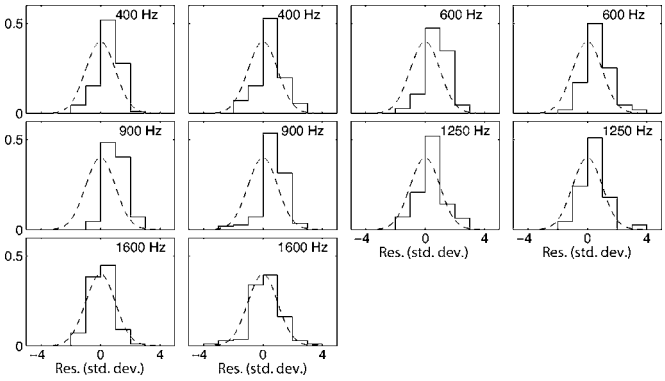


FIG. 17. Histograms of raw (columns one and three) and standardized (columns two and four) residuals as a qualitative measure of Gaussianity of data error.

sidual histograms do not suggest that any other simple distribution (e.g., exponential) would be more appropriate.

C. Final results and cores

Figure 19 shows the final inversion results for four layers with profile marginal distributions given in color (blue corresponds to low and red to high probability). Profile marginals were extracted from the PPD by computing the profile density in area of the plot. The MAP model is shown as a solid black line and is given in Table II. In the fourth layer, the MAP density estimate is found at the upper prior bound (see Fig. 12) and is not regarded to be well determined. However, the other parameters appear to be well determined. In general, the uncertainties of all parameters increase with depth. However, the resolution of sound velocity and attenuation improves in the second, relatively thick (~3 m) layer. The resolution in the fourth layer is limited due to the half space. Since the lower reflector is absent in a half space, less information is available to constrain the parameters of that layer.

Figure 19 also compares the inversion results to three cores taken at the experiment site. Two gravity cores are shown in red and green. Density measurements on the red core were taken with a multisensor core logger. The density for the green core was sampled by splitting the core and extracting ~8 cm^3 samples by hand which were then dried and weighed. The blue core is a 4 m piston core, and density values were also sampled by hand. Error bars of all cores indicate approximate uncertainties of the measurement technique and do not include the result of disruptions to core properties that result from the core collection process. In particular, the consistent shift of sound velocities and densities at the bottom of the blue core towards lower values is likely due to expansion of the core in extraction. The reflectivity inversion results generally agree well with the cores for both sound velocity and density profiles, and the first layer discontinuity at 1.7 m depth clearly appears in the blue core.

VI. SUMMARY

This paper develops a joint time/frequency-domain inversion for seabed reflection data. The approach is based on sequential Bayesian inversion of picked reflection travel times and multi-frequency reflection-coefficient data. The travel-time data are inverted for the thicknesses and velocities of a layered model, with the results subsequently passed on as prior information to the reflection-coefficient inversion in terms of credibility intervals. The reflection-coefficient data are inverted for layer thicknesses, velocities, densities, and attenuations. The inversion employs a layer packet stripping approach, with the results from one packet inversion passed on to constrain the upper layers of the next packet inversion in terms of principal-component marginal probability distributions. In both the time- and frequency-domain inversions, data error covariance matrices are estimated from residual analysis, including nonstationary effects. Potential biases in the travel-time data are accounted for by including nuisance parameters in the inversion. The validity of the estimated covariance matrices and underlying assumption of

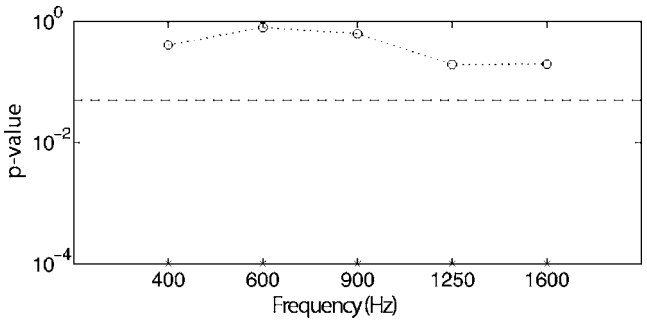


FIG. 16. Runs test results: crosses and circles show p values for the raw and standardized residuals, respectively. Dashed line shows the 0.05 level.

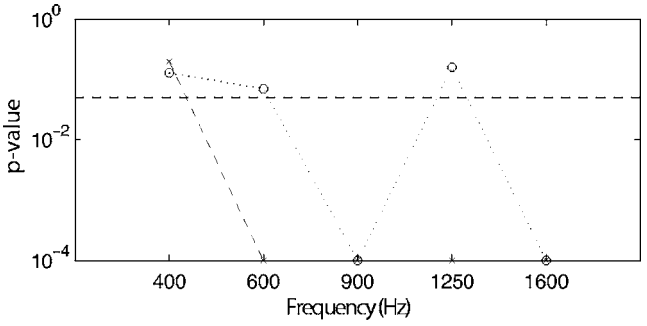


FIG. 18. KS test results: Crosses and circles show p values for raw residuals and standardized residuals, respectively. Dashed line shows the 0.05 level.

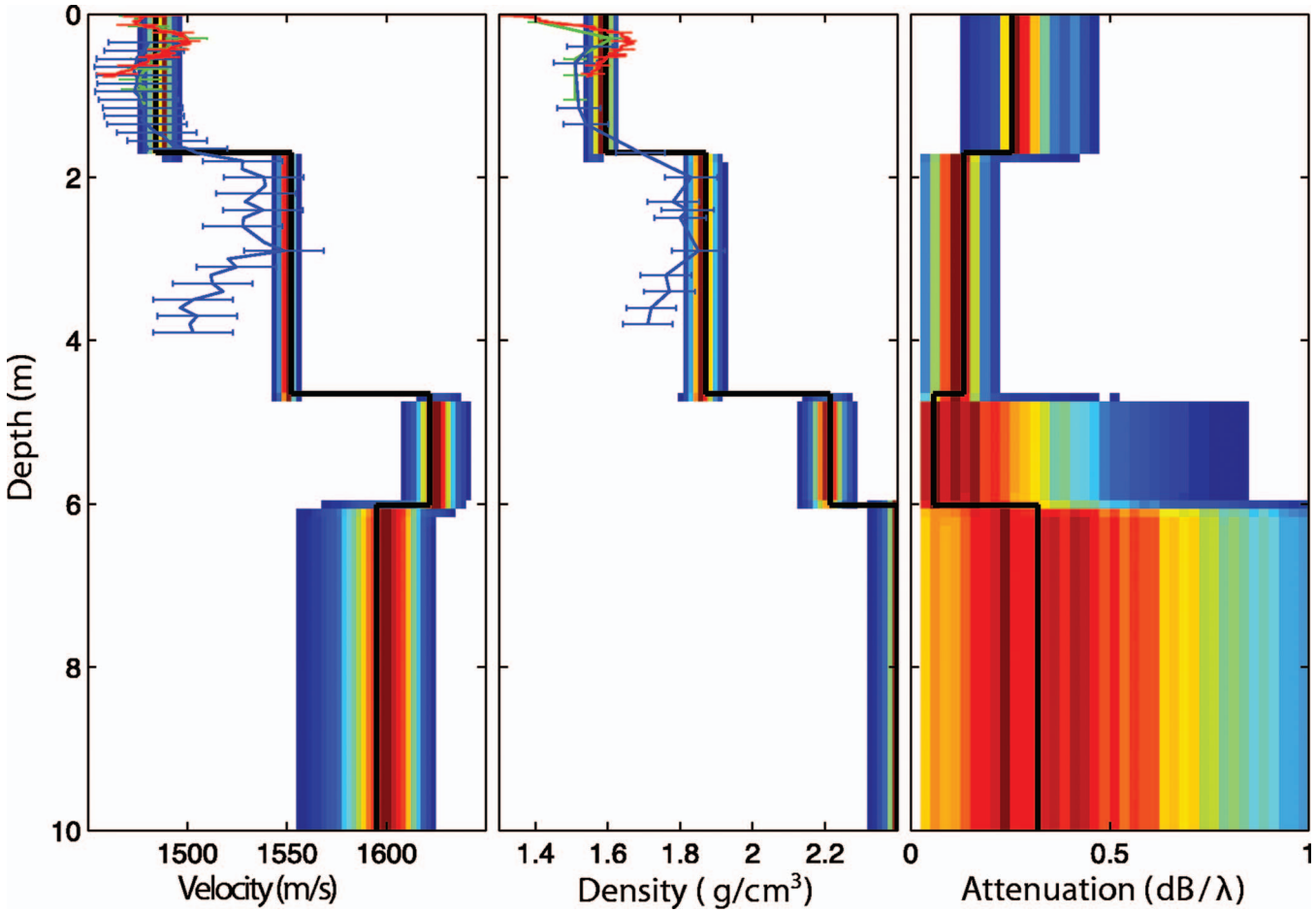


FIG. 19. Inversion results compared to three shallow cores taken at the experiment site (piston core in blue, two gravity cores in red and green). The MAP result is given as the solid black line; profile marginals are given as colored areas with red regions corresponding to high probability. Note that the shifts to low velocity and density over the bottom meter of the blue core are likely extraction artifacts.

Gaussian-distributed errors is examined using qualitative and quantitative statistical tests which showed the importance of including data covariance matrix estimates in the inversion.

The time- and frequency-domain inversions complement each other. The time-domain data primarily constrain the sediment layer thickness structure and provide some information on velocities. In the frequency-domain (reflection-coefficient) data, layer thickness information is often ambiguous due to interference effects. However, the frequency-domain data provide information on all parameters (including density and attenuation) and resolve velocities better than the travel-time data. The sequential Bayesian approach includes the information from both data sets without resorting to more computationally intensive simultaneous inversion. Further, the joint time/frequency-domain approach utilizes travel-time and amplitude information without the very high computational costs associated with full-waveform inversion.

The joint inversion is applied to seismo-acoustic data collected in shallow water of the Malta Plateau in the Mediterranean Sea. The seabed model comprised four layers over approximately 10 m, and two layer packets were employed in the frequency-domain inversion. Due to the geometry of the experiment and the rapidly varying reflection coefficients, a numerically efficient full-wave forward model was used to compute spherical reflection coefficients. The inversion results are interpreted in terms of one- and two-dimensional marginal distributions, MAP parameter estimates, inter-parameter correlations, and profile marginal distributions. In general, layer thicknesses, velocities, and densities are well resolved. The inversion showed limited sensitivity to attenuation, with the best estimates found in the thickest layers. Also, the attenuation results are regarded as an effective measure of the overall loss processes (including scattering and environmental inhomogeneities), not necessarily the intrinsic attenuation of the sediments.[3] The recovered parameter profiles are consistent with the empirical data in Hamilton.[21] Available piston core data for the upper 4 m agree well with the inversion results for velocity, density, and layering structure.

ACKNOWLEDGMENTS

The authors thank the NATO Undersea Research Center for supporting the at-sea data collection. The first author was supported by the Natural Sciences and Engineering Research Council of Canada and the University of Victoria during Ph.D. studies. The authors gratefully acknowledge the support of the Office of Naval Research postdoctoral fellowship.

[1]C. W. Holland and J. Osler, "High-resolution geoacoustic inversion in shallow water: A joint time- and frequency-domain technique," J. Acoust. Soc. Am. **107**, 1263–1279 (2000).
[2]J. Dettmer, S. E. Dosso, and C. W. Holland, "Uncertainty estimation in seismo-acoustic reflection travel-time inversion," J. Acoust. Soc. Am. **122**, 161–176 (2007).
[3]S. E. Dosso and C. W. Holland, "Geoacoustic uncertainties from viscoelastic inversion of seabed reflection data," IEEE J. Ocean. Eng. **31**, 657–671 (2006).
[4]C. W. Holland, J. Dettmer, and S. E. Dosso, "Remote sensing of sediment density and velocity gradients in the transition layer," J. Acoust. Soc. Am. **118**, 163–177 (2005).
[5]J. Dettmer, S. E. Dosso, and C. W. Holland, "Full wave-field reflection coefficient inversion," J. Acoust. Soc. Am. **122**, 3327–3337 (2007).
[6]J. Dettmer, S. E. Dosso, and C. W. Holland, "Geoacoustic inversion with strongly correlated errors," Can. Acoust. **32**, 194–195 (2004).
[7]J. Dettmer, S. E. Dosso, and C. W. Holland, "Bayesian inversion of reflection data for seabed properties of multi-layered systems," in *Proceedings of the Eighth European Conference on Underwater Acoustics*, Carvoeiro, Portugal, 473–478 (2006).
[8]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).
[9]C. W. Holland, "Seabed reflection measurement uncertainty," J. Acoust. Soc. Am. **114**, 1861–1873 (2003).
[10]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[11]A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).
[12]K. Mosegaard and M. Sambridge, "Monte Carlo analysis of inverse problems," Inverse Probl. **18**, R29–R54 (2002).
[13]S. E. Dosso, M. J. Wilmut, and A.-L. S. Lapinski, "An adaptive-hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).
[14]D. H. Rothman, "Automatic estimation of large residual statistics corrections," Geophysics **51**, 337–346 (1986).
[15]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).
[16]F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," J. Am. Stat. Assoc. **46**, 68–78 (1951).
[17]J. E. Freund, *Modern Elementary Statistics* (Prentice-Hall, Englewood Cliffs, NJ, 1967).
[18]C. H. Harrison and J. A. Harrison, "A simple relationship between frequency and range averages for broadband sonar," J. Acoust. Soc. Am. **97**, 1314–1317 (1995).
[19]W. A. Kuperman and H. Schmidt, "Self-consistent perturbation approach to rough surface scattering in stratified elastic media," J. Acoust. Soc. Am. **86**, 1511–1522 (1989).
[20]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, Series in Modern Acoustic and Signal Processing (AIP, New York, 1993).
[21]E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).

Controlled and *in situ* target strengths of the jumbo squid *Dosidicus gigas* and identification of potential acoustic scattering sources

Kelly J. Benoit-Bird[a)]
College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Administration Building, Corvallis, Oregon 97331

William F. Gilly[b)]
Hopkins Marine Station, Department of Biological Sciences, Stanford University, Oceanview Boulevard, Pacific Grove, California 93950

Whitlow W. L. Au[c)]
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

Bruce Mate[d)]
Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Drive, Newport, Oregon 97365

(Received 26 April 2007; revised 3 December 2007; accepted 12 December 2007)

This study presents the first target strength measurements of *Dosidicus gigas*, a large squid that is a key predator, a significant prey, and the target of an important fishery. Target strength of live, tethered squid was related to mantle length with values standardized to the length squared of −62.0, −67.4, −67.9, and −67.6 dB at 38, 70, 120, and 200 kHz, respectively. There were relatively small differences in target strength between dorsal and anterior aspects and none between live and freshly dead squid. Potential scattering mechanisms in squid have been long debated. Here, the reproductive organs had little effect on squid target strength. These data support the hypothesis that the pen may be an important source of squid acoustic scattering. The beak, eyes, and arms, probably via the sucker rings, also play a role in acoustic scattering though their effects were small and frequency specific. An unexpected source of scattering was the cranium of the squid which provided a target strength nearly as high as that of the entire squid though the mechanism remains unclear. Our *in situ* measurements of the target strength of free-swimming squid support the use of the values presented here in *D. gigas* assessment studies.
 [DOI: 10.1121/1.2832327]

PACS number(s): 43.30.Sf, 43.30.Ft, 43.20.Fn [KF] Pages: 1318–1328

I. INTRODUCTION

Dosidicus gigas, the jumbo or Humboldt squid, ranges in adult size from 35 cm to about 2.5 m in total length and can weigh up to 50 kg, making it the largest of the nektonic squids. Individual squid grow extremely rapidly, reaching these sizes in only 1 to 2 years. They are found in the productive and relatively shallow waters of the Eastern Pacific from 125–140°W, 40°N to 45°S. They typically inhabit from surface waters to depths of about 500 m, with pronounced diel vertical migrations bringing them shallower at night (Gilly *et al.*, 2006) but Remotely Operated Vehicle (ROV) sightings have been made deeper than 1000 m (B. Robison, private communication). *D. gigas* is an active migratory predator, feeding primarily on myctophids and other fish, crustaceans, and squid (Markaida and Sosa-Nishizaki, 2003). As adults, they are a food resource for large fish including tuna and swordfish, and marine mammals such as sperm whales and pilot whales. *D. gigas* has been shown to be a keystone species, transferring energy between planktivorous fish and the largest pelagic predators (Shchetinnikov, 1988).

Dosidicus gigas is probably the most abundant nektonic squid in the Eastern Pacific (Nigmatullin *et al.*, 2001), and it currently supports the world's largest cephalopod fishery, with landings of 800 000 tonnes in 2004 (FAO data at ftp://ftp.fao.org/fi/stat/summary/a1e.pdf) Approximately 15% of this catch comes from the Guaymas Basin in the Gulf of California, Mexico (Markaida and Sosa-Nishizaki, 2001), although annual landings can be quite variable (Nevarez-Martinez *et al.*, 2006).

Total standing biomass of *D. gigas* across its large range is unknown, but annual production in the Gulf of California has been estimated to be 210,000 tons/y (Sanchez-Juarez, 1991). This figure undoubtedly is an underestimate of the current situation, given the fact that commercial fishing has reached half of this figure, and consumption by resident sperm whales alone may be comparable (D. Gendron, private communication).

a)Corresponding author. Electronic mail: kbenoit@coas.oregonstate.edu
b)Electronic mail: lignje@stanford.edu
c)Electronic mail: wau@hawaii.edu
d)Electronic mail: bruce.mate@oregonstate.edu

Despite the economic importance of the *Dosidicus* fishery and the inherent difficulties in managing a highly variable population affected by strong environmental fluctuations, extensive migrations, and interactions with other fisheries (Morales-Bojorquez *et al.*, 2001), data on behavior, natural history, and biomass of the species remain limited. These factors are critical components to successful management of any fishery.

Assessment of biomass in squid fisheries has relied on commercial landing data, trawls, and quantitative handfishing (jigging) surveys, each of which has limitations (Boyle and Rodhouse, 2005). Although acoustic methods have been widely used in the assessment of fish stocks (Misund, 1997), application of acoustics to study the biology of squid has received much less attention, probably because of the belief that squid are particularly weak scatterers (Fristrup and Harbison, 2002). Nonetheless, acoustic methods have been successfully employed to observe spawning aggregations (Jefferts *et al.*, 1987; Lipinski and Prowse, 2002), to assess recruitment by quantifying benthic egg beds (Foote *et al.*, 2006), and to characterize acoustic features of different groups of squid for possible future efforts (see, for example, Boyle and Rodhouse, 2005; Goss *et al.*, 2001).

The first step for quantitative acoustic assessment of any marine species is measurement of its target strength distribution, and no such data exist for *D. gigas.* Our goals in this study were: (1) to obtain controlled measurements of the target strength of a wide size range of live *D. gigas* at four frequencies, (2) identify possible sources of scattering in *D. gigas*, and (3) to confirm the validity of target strength measurements *in situ* with free swimming squid.

II. METHODS

Research was conducted between 16 and 24 March 2007 from the 25 m R/V Pacific Storm in Guaymas Basin in the Gulf of California, Mexico. Individual squid were captured with weighted, luminescent jigs generally between dusk and 0100 local time. One squid (the smallest studied) was captured at the surface using a dip net. Squid were typically maintained in individual, covered tanks with a continuous input of fresh seawater overnight before their acoustic properties were measured during daylight hours.

A. Acoustic system

Acoustic characteristics of the squid were measured using a four-frequency, split-beam echosounder system (Simrad EK60s). The 38 kHz echosounder used a 1024 μs pulse, the 70 kHz echosounder used a 512 μs long pulse, while the 120 and 200 kHz echosounders used a 256 μs pulse. The 38 kHz system had a 12° conical split beam while each of the higher frequencies had a 7° conical split beam. The transducers were affixed over the side of the vessel on a rigid pole mount so that they were 1 m beneath the water's surface. The system was calibrated using an indirect procedure incorporating a 38.1-mm-diam tungsten carbide reference sphere as prescribed by Foote *et al.* (1987). In addition, the target strength of the sphere was also measured attached to the mounting line that normally held the squid (described in the following) to rule out any possible interference of the mounting arrangement. All the mounting lines were also placed in the water without the sphere or a squid to measure any potential backscatter from the experimental setup.

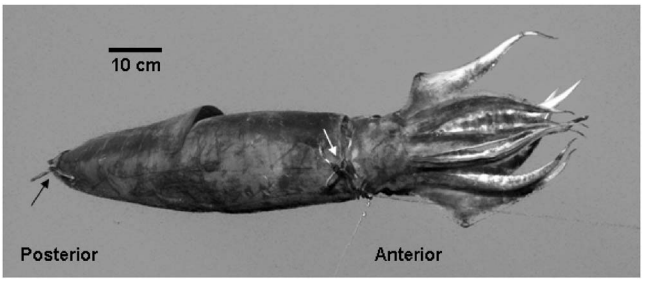


FIG. 1. A mounted squid underwater. A single loop of Tygon® tubing tied to a horizontal length of monofilament is visible at the posterior of the animal at the left of the photo and two loops tied together and then to the monofilament is visible at the anterior of the mantle, near the squid's head in the center of the photo.

B. Mounting of live squid

An individual squid was removed from its holding tank and placed ventral side down for attachment of the mounting hardware. During this time (<5 min total) the gills were continuously irrigated with flowing seawater. The squid did not struggle, and no anesthesia was necessary. Two locations on either side of the midline of the dorsal mantle surface near the anterior edge, and one location through the posterior tip of the gladius, were pierced with a stainless-steel cannula (6 mm o.d.; Floy Tag, Seattle WA), allowing a small length of 6 mm diameter Tygon tubing to be passed through each hole. A plastic cable tie was passed through the tubing and then cinched to form a loop, with the soft tubing acting as a "bushing" to minimize tissue damage (Fig. 1).

For dorsal-aspect measurements, the loops through the squid were then attached to two loops tied approximately the length of squid's mantle apart near the center of a 7.5 m horizontal length of monofilament. The squid, attached to the monofilament mounting line, was then gently lifted onto a cloth stretcher and lowered into the water. Great care was taken to avoid the introduction of air into the squid's mantle during the mounting procedure. The mantle was continuously flushed with gently running seawater while on deck and then held compressed by the stretcher until the squid was released in the water by removing the stretcher. The squid was then observed just below the surface to ensure that it was actively swimming in the proper orientation, that no lines were tangled, and that no air was observed leaving the mantle or around the animal's appendages. The animals were all observed to be vigorously alive and their attempts to swim away from the mounting rig resulted in strong mantle compressions and jetting, likely removing any potential residual air. After observation of the squid near the surface, the horizontal mounting line was lowered to a depth of between 6 and 16 m, depending on currents and other sea conditions. In all cases, the depth was at least twice what was necessary for the entire target to be within the 3 dB beamwidth of the narrowest transducer and in the far field of all transducers

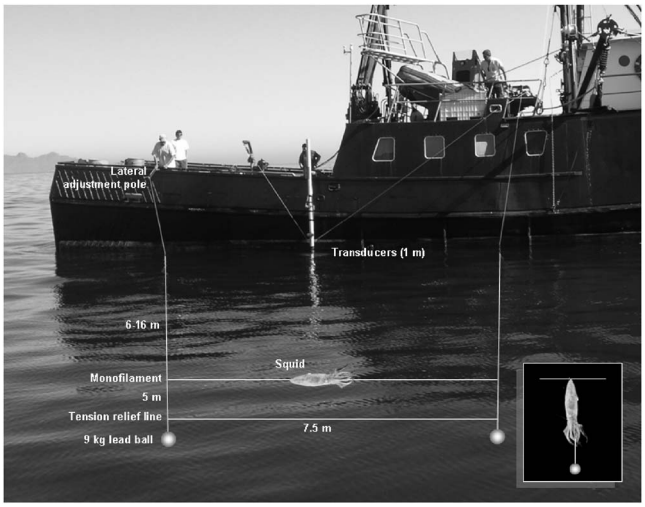


FIG. 2. The squid was mounted to a horizontal piece of monofilament stretched apart by two vertical lines lowered by lead balls. The transducers were mounted 1 m below the water's surface on a rigid pole over the side. Lateral aspect setup is shown in the main figure. Changes for posterior aspect (head down) measurements are shown in the inset on the lower right on the same scale.

when considering the cross-sectional area of the scattering surface perpendicular to the incident signal at the wavelengths used.

As indicated in Fig. 2, the horizontal mounting line was held taut between two vertical nylon lines spaced 10 m apart. Each vertical line was weighted with a 9 kg lead ball 5 m below the attachment point of the mounting line. An additional 7.5 m length of polypropylene line connected the lead weights and provided strain relief. The inboard/outboard position of the lines was adjusted with poles affixed across the rail of the vessel. Dorsal-aspect measurements were taken from a total of 14 squid that were alive both upon entry into the water and after removal from the water postexperiment. Measurements were also made from one animal that had recently expired. Only live animals were used for the acoustic measurements presented in the figures and equations.

After measuring the dorsal aspect of the squid, a subset of the animals were measured from the posterior aspect by suspending the squid in a head-down vertical orientation. In this case, the posterior attachment loop was retained, but the anterior loops were detached from the mounting line, and the corresponding end of that line was disconnected from its vertical line, and the end was directly attached to a 5 m length of monofilament leading to a 5 kg lead ball.

Upon completion of all measurements, we measured dorsal mantle length (DML), length from the tip of the mantle to the end of the outstretched arms (not tentacles), maximum mantle width, and total mass. After these measurements were completed, the squid was euthanized by rapid decapitation, and sex and maturity stage (Lipinski and Underhill, 1995) were visually determined. Effects of maturity and gender on target strength were assessed with a multivariate analysis of variance (ANOVA).

C. Acoustic scattering source assessments

Measurements were made in an attempt to determine the important individual sources of the acoustic scattering from squid. In several cases, following the whole-animal measurements, the dorsal aspects of the head and body were measured separately. Each portion was suspended from the mounting line in a manner similar to that described earlier. Care was taken to avoid the introduction of air into the body parts since it could dramatically impact acoustic measurements. The dense, cartilaginous posterior of the gladius (the conus) was removed and the body was remeasured. The beak, followed by the eyes, arms, and all soft tissue were successively removed from the head with measurements repeated at each step so that ultimately, only the cartilaginous cranium (braincase) remained.

D. Acoustic measurements

Acoustic measurements were either made while the boat was anchored in at least 30 m of water or drifting in calm seas. The position of the squid was adjusted so that it was maintained within 2° of the center of each of the four transducers as identified in real time via the split beam target identification that is part of the Simrad recording software. A minimum of 100 measurements of target strength were made from each squid in each position. Most often, at least 500 measurements were made. Analysis of target strength was only made for echoes that were within 2.5° of the center of the beam along both axes and were at the known depth of the squid. The average target strength and standard deviation (both calculated from linearized data) were determined for each squid. The target strengths of scattering from different parts of the squid were compared using t-tests based on target strengths from individual echoes of the same squid rather than from mean values.

For all measurements, targets were extracted using SonarData's ECHOVIEW program. The single target detection criteria were based on the transmitted pulse length, measurements of target strength of the calibration sphere in the squid mounting setup, and through empirical minimization of the standard deviation of the target strength of individual tethered squid when their depth relative to the transducer's was known and surrounding acoustic clutter was at a minimum. The target strength threshold was set to −55 dB. A "pulse length determination level," the value in decibels below peak value considered when determining the pulse length, or envelope, of a single-target detection, of 12 dB was used. "Normalized pulse lengths," the measured pulse length divided by transmitted pulse length, were required to be between 0.8 and 2.0. The "maximum beam compensation" for correcting for transducer directivity was set to 12 dB. To confirm all sources of scattering within the measured pulse length were from a single target, all samples within this pulse envelope must have had a standard deviation in angular position of less than 3° in both the along and athwart ship directions of the beam. We confirmed that these settings permitted the integrated pulse envelope to enclose the entire body of the squid by observing the standard deviation of the target strength measurements; when the pulse envelope was too short to encompass the entire squid the variation in estimated target strength was extremely high.

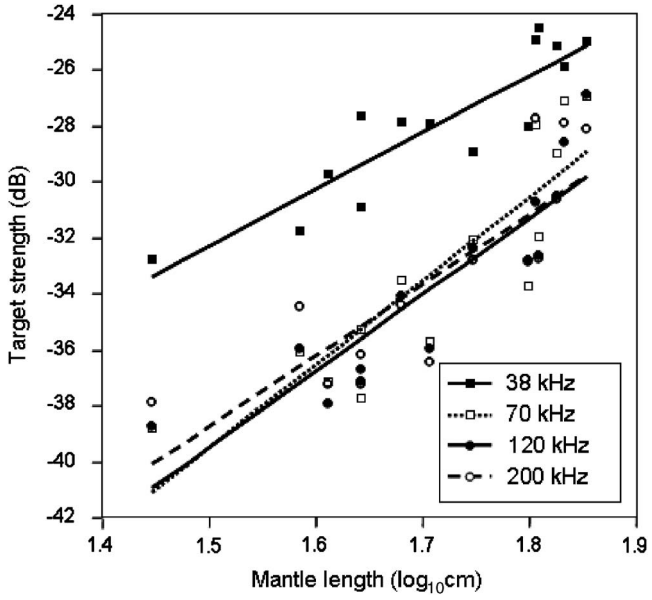


FIG. 3. Dorsal aspect target strength as a function of the log of squid mantle length for each of the four frequencies measured. Standard deviations of the means of all measurements at each frequency for each squid are too small to be shown. Linear regressions through the mean values for each squid are shown. There was no significant difference between the 70, 120, and 200 kHz target strength measurements.

E. *In situ* measurements

In situ measurements of targets that were highly likely to be squid were also made. One night, fishing (jigging) was conducted by two persons while the vessel was at anchor in approximately 30 m of water. Over 15 min, 20 squid were captured and DML measured for each. Targets that met the criteria to be analyzed individually during that time period on the echosounders were extracted and the results compared with those made under controlled circumstances. Results were also compared with *in situ* measurements made while many (15–20) small commercial fishing boats surrounded our vessel and two to three fishermen per boat continuously landed *D. gigas*. Unfortunately, no length measurements were made during this second *in situ* experiment.

III. RESULTS

A. Controlled measurements

The empty squid-holding setup was almost undetectable at all frequencies, and the target strength of the sphere measured in the same arrangement compared well with standard calibration measurements in terms of both the mean and the distribution. This provides convincing evidence that the measurements of squid target strength described in the following are not affected by the hardware setup.

Dorsal-aspect target strength measurements were obtained from 14 live squid ranging in mantle length from 28.0 to 71.5 cm (41.5 to 118.0 cm total length). The average difference in mantle length between successively larger animals was approximately 3.6 cm, and in no case was greater than 10 cm. Mean values of target strength from the dorsal aspect target of each of the 14 live squid are shown in Fig. 3. The standard deviations of the target strength measurements for each individual squid were extremely small (0.005–0.00008 calculated on the linearized data and converted to decibels) and are not shown. An ANOVA with post-hoc tests corrected using the Bonferroni method for multiple comparisons showed that standard deviation was significantly affected by frequency with each frequency having a higher standard deviation in target strength than the frequencies above it. Increasingly larger squid also had significantly higher standard deviations in target strength at a given frequency.

The relationships between mantle length and target strength can be described by the equations:

$$\mathrm{TS}_{38\ \mathrm{kHz}} = 20.4 \log_{10}(\mathrm{DML}) - 62.8, \quad R^2 = 0.81, \tag{1}$$

$$\mathrm{TS}_{70\ \mathrm{kHz}} = 29.9 \log_{10}(\mathrm{DML}) - 84.4, \quad R^2 = 0.77, \tag{2}$$

$$\mathrm{TS}_{120\ \mathrm{kHz}} = 27.3 \log_{10}(\mathrm{DML}) - 80.5, \quad R^2 = 0.80, \tag{3}$$

$$\mathrm{TS}_{200\ \mathrm{kHz}} = 25.4 \log_{10}(\mathrm{DML}) - 76.8, \quad R^2 = 0.69, \tag{4}$$

where TS is the target strength at the frequency noted in the subscript, DML is dorsal mantle length in centimeters, and R^2 is a unit-less descriptor of the goodness of fit of the linear regressions illustrated in Fig. 3 and can be interpreted as the amount of variability in target strength that can be explained by variability in DML. F-tests showed that the slope of each regression was significantly different from zero ($p < 0.05$). If we force the slope of the line to equal 20, expecting the relationship to be related to the cross-sectional area of the squid (McClatchie *et al.*, 2003), the relationships can be described as

$$\mathrm{TS}_{38\ \mathrm{kHz}} = 20 \log_{10}(\mathrm{DML}) - 62.0, \quad \Delta R^2 = -0.01, \tag{5}$$

$$\mathrm{TS}_{70\ \mathrm{kHz}} = 20 \log_{10}(\mathrm{DML}) - 67.4, \quad \Delta R^2 = -0.09, \tag{6}$$

$$\mathrm{TS}_{120\ \mathrm{kHz}} = 20 \log_{10}(\mathrm{DML}) - 67.9, \quad \Delta R^2 = -0.06, \tag{7}$$

$$\mathrm{TS}_{200\ \mathrm{kHz}} = 20 \log_{10}(\mathrm{DML}) - 67.6, \quad \Delta R^2 = -0.03. \tag{8}$$

Although this results in a small but significant ($p < 0.05$) decrease in the goodness of fit (ΔR^2) for all frequencies except 38 kHz, these standardized values can be more readily compared to those from the literature.

Of the squid measured, nine were female and five were male, eight were sexually immature, and six were mature. The effect of gender and maturity were assessed on target strength after the effect of length was removed by subtracting $20 \log_{10}(\mathrm{DML})$ from each target strength value. Individual measurements of target strength, rather than the mean, were used in the statistical analysis. There was no significant effect of gender or sexual maturity on the target strengths at any frequency ($p < 0.05$ for each comparison).

The dorsal-aspect target strength values of a single freshly deceased squid were compared to the measurements of the two live squid closest in length (1.5 cm larger and 3.0 cm smaller) using t-tests. Target strength of the dead squid was not statistically significantly different from that of the live squid at any frequency ($p < 0.05$ for each comparison).

TABLE I. Summary of comparisons between various target strength measurements in decibels. *N* represents the number of measurements. ns is shown for comparisons that did not show a significant difference at the $p<0.05$ level.

	Posterior vs dorsal	Body vs dorsal	Body - Pen vs body	Head vs dorsal	Head vs body	Head vs head - Beak	Head vs head - Beak eyes	Head vs head - Beak - eyes - arms	Head vs cranium	Cranium vs dorsal
N	5	5	4	5	4	4	4	2	4	4
38 kHz	0.9	−1.3	−0.5	ns	1.1	−2.2	−2.1	−6.0	−5.7	−6.0
70 kHz	−3.0	−1.5	−2.6	−1.0	−3.1	ns	ns	ns	−1.5	−2.2
120 kHz	−3.1	ns	−4.9	−1.1	ns	ns	−5.0	−4.9	ns	ns
200 kHz	−1.8	ns	ns	ns	ns	ns	ns	ns	ns	ns

Posterior end-on measurements of the target strengths of five live squid ranging from 41 to 68 cm in mantle length were measured. Target strengths were compared to the dorsal aspect target strength of the same individuals. The results are shown in the first column of Table I. The posterior-aspect target strength measurements were significantly different at all frequencies with slightly higher target strengths from the posterior aspect relative to the dorsal aspect at 38 kHz and lower target strengths from the posterior aspect for all other frequencies measured. Although standard deviations of posterior-aspect target strength at individual frequencies were small, they were significantly larger than those of dorsal aspect measurements.

After the measurements were made on the live squid, some were sacrificed and immediately had various parts of their bodies measured. A summary of the statistical comparisons of various target-strength measurements is shown in Table I. Any measurement type in the header that is preceded by a minus sign means that this body part was removed from the other body part after a measurement had been made, and a comparison measurement was then repeated. In the body of Table I, a negative value for the observed change in target strength means that the target strength corresponding to the upper row of the header was lower than that corresponding to the lower row, and a positive sign means the opposite. Indicated values represent the mean difference (calculated on linearized data) in target strength for all squid. Any values that were not statistically significant at the $p<0.05$ level are indicated "ns."

B. *In situ* measurements

During 15 min of fishing effort by two individuals on 23 March 2007, 20 squid were captured and their dorsal mantle lengths measured. These squid ranged in mantle length from 38.5 to 53.5 cm with a mean mantle length of 46.4 cm. A histogram of the captured squid sizes is shown in Fig. 4. During this same 15 min, a total of 40 718 individual targets fitting the criteria for individual detection [e.g., only one target per acoustic reverberation volume in for each pulse (Sawada *et al.*, 1993)] were detected by the 12°, 38 kHz system and 66 136 were detected by each of the higher frequency, 7° systems. These values correspond to an average detection rate of 45 squid/s at 38 kHz and 74 squid/s at the higher frequencies. The narrower beams had higher detection rates because they had more than one target in each resolvable volume of water less often than the wider beam of the 38 kHz echosounder.

The mean target strength at each frequency for individually identified targets is shown in Fig. 5. The target strength distribution for single targets at each individual frequency was unimodal. While the standard deviations of the *in situ* target strength measures were substantially larger than those from the controlled experiment, they were still too small to be represented in the graph (<0.7 in $\log_{10}$ space for all cases, representing a 95% confidence limit of ± 1.4 dB). Also shown in the graph are the target strengths predicted by the unforced regressions developed from the controlled measurement experiment for the mean mantle length measured from the 20 captured squid. The differences between the predicted measurements and the actual measurements were small, with the largest difference a 1.3 dB greater than expected target strength at 38 kHz.

Earlier the evening of 23 March 2007, our stationary research vessel was surrounded by approximately 15–20 small squid fishing boats, each with two to three fisherman continuously landing jumbo squid. Squid were also visible at

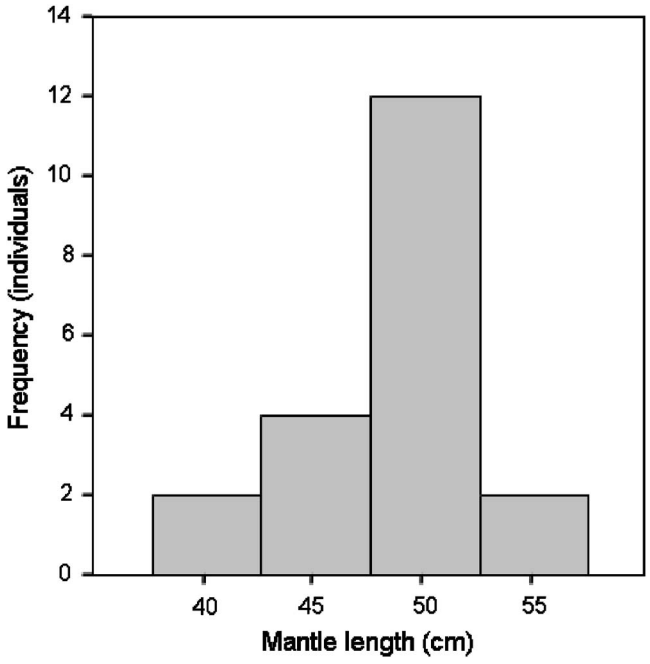


FIG. 4. Histogram of squid mantle lengths for the 20 squid captured during 15 min of *in situ* acoustic measurements.

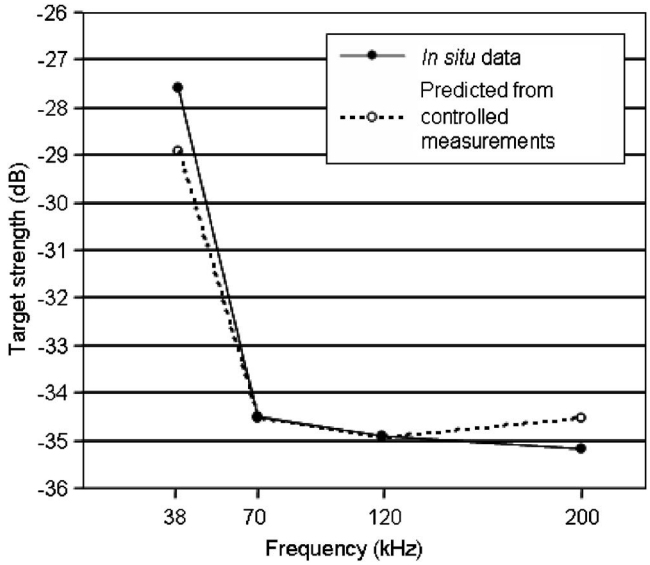


FIG. 5. Mean *in situ* measurements of target strength (closed circles) collected between 23:45 and 00:00 local time on 23 March 2007. Fishing efforts during this same time period resulted in the capture of 20 squid with a mean length of 46.4 cm. Using the unforced regressions for target strength vs length from the controlled measurements, a predicted target strength for each frequency was calculated for this size squid, shown by the open circles.

the surface in high densities. Mean *in situ* measurements of target strength during three, 15 min intervals are shown in Fig. 6. In all measurements, the distribution of target strength was unimodal with a 95% confidence interval of less than ± 2 dB, suggesting a single size class of a uniform target (e.g. one species). Using the unforced regression relationships from the controlled measurement experiment [Eqs. (1)–(4)], the approximate mean squid mantle length that these measurements would equate to are 35, 45, and 55 cm for 2100–2115, 2130–2145, and 2200–2215 h local time, respectively. The distribution of target strength measurements of single targets at each individual frequency was unimodal at all time intervals. However, from 2100 to 2215 h the detection rate of individual targets increased by approximately tenfold from 34 to 361 targets/s on the 7° transducers despite the boat being stationary and no changes in the mean target depth. Echograms of target strength from the beginning (2100 h) of the sampling time when target detection rate was moderate and the end (2213 h) of the sampling time when the detection rate was very high are shown in Fig. 7. It is difficult to determine the number of individual squid detected during these time periods, as opposed to the number of single targets because of our stationary position. Animals likely moved into and out of the beam many times during the observation period. In addition, the conical shape of the transducers' beams means that substantially more area is covered at 75 m (32-m-diam circle for the 38 kHz sensor and 18 m diameter for the higher frequencies) than at 5 m (2.1–1.2 m diameter, respectively) further complicating quantitative assessment from this small sample set.

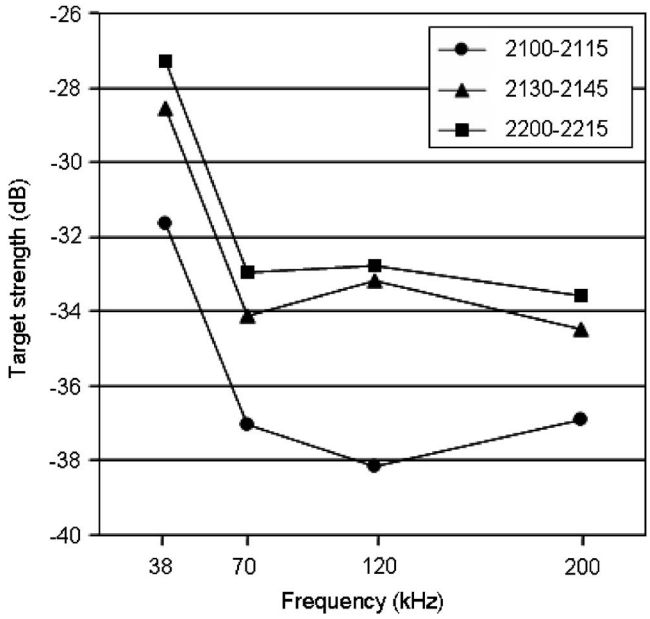


FIG. 6. Mean *in situ* measurements of target strength during three, 15 min intervals on 23 March 2007. During this time, the research vessel was surrounded by approximately 15–20 small squid fishing boats, each with two to three fisherman continuously landing jumbo squid.

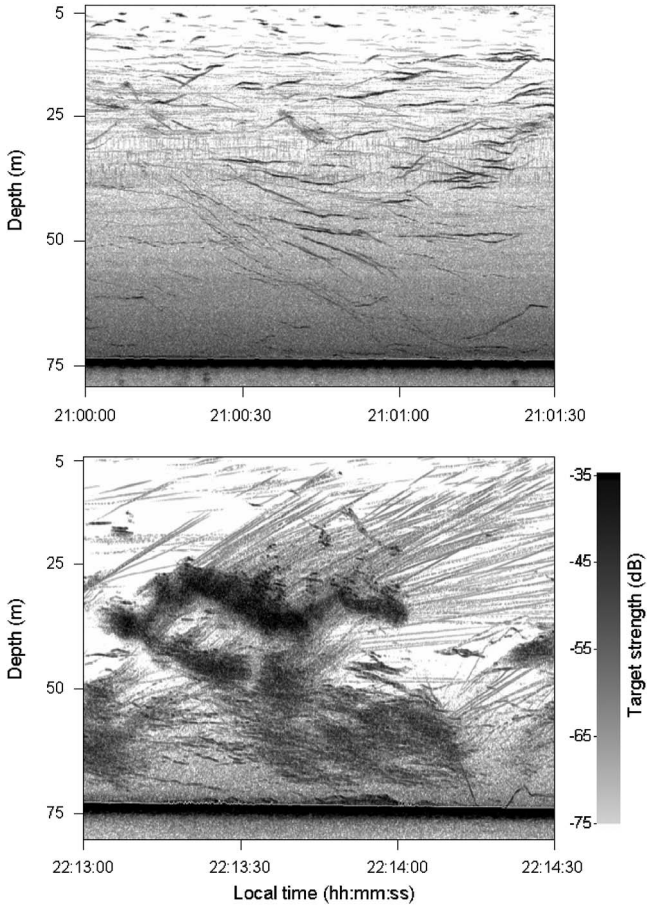


FIG. 7. Two echograms collected from the stationary vessel at 120 kHz for two, 1.5 min intervals on 23 March 2007. The bottom depth was approximately 74 m. Data are expressed with a range compensation function of $40 \log r + 2\alpha r$ to show target strength values for individual targets.

IV. DISCUSSION

We present controlled measurements of individual *Dosidicus gigas* covering nearly the full size range of the species as sampled by commercial fisheries, as well as *in situ*

TABLE II. Published standardized dorsal aspect target strengths as a function of the square of mantle length compared with the results of the standardized dorsal aspect target strengths obtained in this study.

Source	Condition	Frequency (kHz)	Slope	Intercept	Min ML (cm)	Max ML (cm)	Species
This study	Live	38	20	−62.0	28.0	71.5	*Dosidicus gigas*
Benoit-Bird and Au, 2001	Anesthetized	200	20	−62.1	1.2	4.2	Various mesopelagic spp.
Lee *et al.*, 1991	Dead	200	20	−65.7	4.4	16.0	*Loligo edulis*
This study	Live	70	20	−67.4	28.0	71.5	*Dosidicus gigas*
This study	Live	200	20	−67.6	28.0	71.5	*Dosidicus gigas*
This study	Live	120	20	−67.9	28.0	71.5	*Dosidicus gigas*
Kawabata, 2005	Live	38	20	−73.1	18.0	28.4	*Todarodes pacificus*
Kang *et al.*, 2005	Anesthetized	120	20	−73.5	21.0	27.0	*Todarodes pacificus*
Kawabata, 2001	Live	38	20	−73.7	20.0	28.0	*Todarodes pacificus*
Kang *et al.*, 2005	Anesthetized	38	20	−75.4	21.0	27.0	*Todarodes pacificus*
Sawada *et al.*, 2006	*in situ*	70	20	−81.9	18.0	37.0	*Gonatopsis borealis*

measurements of free-swimming individuals along with a simultaneously obtained estimate of size.

A. Controlled measurements

Controlled measurements of *D. gigas* dorsal aspect target strength show a strong, positive linear relationship with the log of squid dorsal mantle length. Target strength increases slightly more rapidly with squid length at 70, 120, and 200 kHz than at 38 kHz. Target strength at 38 kHz was also significantly higher by about 5.5 dB than the target strength at the higher frequencies measured. The reasons for this difference are not apparent. Measurements of a standard sphere do not suggest any instrumental or analysis artifacts and the similarity of the response in free-swimming squid *in situ* eliminates potential artifacts from squid capture, mounting, or air entrapment. The target strengths at 70, 120, and 200 kHz were not significantly different from each other.

Relatively few measurements of squid target strength have been made at multiple frequencies. Arnaya *et al.* (1989a) found that two species of squid that they measured after thawing dead specimens had a 5–10 dB higher target strength at 200 kHz relative to 50 kHz. Similarly, Kang *et al.* (2005) found a 0.7–2.5 dB higher target strength at 120 kHz relative to 38 kHz for live specimens of one species also measured by Arnaya *et al.* (1989a). While the squid measured in these two studies were much smaller than the *D. gigas* measured here, there was some overlap in mantle length. As shown in Fig. 3, the patterns in relative target strength we observed were maintained when considering only these animals of smaller body size suggesting that the frequency response of target strength is species specific, not size specific.

Target strength can be normalized by forcing the slope of the regression to 20, representing a relationship with the square of squid mantle length. The intercepts of these relationships [Eqs. (5)–(8)] can then be compared with normalized target strength measurements from other studies, regardless of animal size. Table II provides such a standardized comparison of dorsal-aspect target strengths from studies in the literature that showed (1) a relationship between target strength and mantle length, and (2) a slope of 20° or enough data to refit the curve with a 20° slope. Because most squid in the present study were substantially larger than those in previous work, only our smaller individuals overlapped with the largest squid measured in other studies, if at all. Despite this, our standardized target strengths were similar to those measured for very small individuals: various mesopelagic species (Benoit-Bird and Au, 2001) and Loligo edulis (Lee *et al.*, 1991). Standardized target strengths for *Todarodes pacificus* were 5–10 dB lower than data reported here (Kang *et al.*, 2005; Kawabata, 2001, 2005), and *Gonatopsis borealis* standardized target strengths were 15–20 dB lower (Sawada *et al.*, 2006).

Comparisons of target strength for species with very different size distributions are more difficult when the slope of the relationship between body size and target strength is not simply related to the square of length. To compare our results with those from published studies of squid dorsal-aspect target strength, the target strength of a 28 cm dorsal mantle length squid was calculated using the equations from each of the studies as well as Eqs. (1)–(4) from this study (Table III). For most of the published studies, 28 cm was at or very slightly above the high end of the size distribution measured. For this study, it was the smallest animal measured. This minimized extrapolation to allow the best comparison of data. The target strengths measured here were relatively low when compared with these studies. The target strength values are, however, very similar to the predicted target strengths from those studies that measured target strength *in situ*. All of the curves generated from animals that had been previously preserved (usually frozen followed by thawing for the measurements) had substantially higher predicted target strengths than those obtained from live animals either tethered or *in situ*.

Many measurements of squid target strength in the literature have been made on dead animals. A squid we measured shortly after its death (with no evidence of rigor) did not have significantly different dorsal-aspect target strengths than similarly sized live animals, suggesting that target strength does not change immediately following death, however it may change as a result of preservation or with greater time delay. This result also suggests that the target strength values we obtained were not a result of the behavior of the squid or any alteration of their angle due to swimming against our mounting setup.

The posterior-aspect target strength of five squid was

TABLE III. Target strengths of a 28 cm mantle length squid calculated from the published equations describing the target strength length relationship from various species as well as predicted and actual measurements from this study.

Source	Condition	Frequency (kHz)	Slope	Intercept	Min ML (cm)	Max ML (cm)	Species	Predicted TS 28 cm ML (dB)	Actual TS 28 cm ML (dB)
Arnaya *et al.*, 1989a	Dead/preserved	200	24.7	−64.8	18.0	30.0	Ommastrephes bartrami	−29.0	
Kaiiwara *et al.*, 1990	Dead/preserved	?	19.0	−56.8	?	?	Ommastrephes bartrami	−29.3	
Arnaya *et al.*, 1989a	Dead/preserved	200	42.0	−91.5	15.0	35.0	Todarodes pacificus	−30.7	
Arnaya *et al.*, 1989a	Dead/preserved	50	23.4	−65.0	18.0	30.0	Ommastrephes bartrami	−31.2	
Arnaya *et al.*, 1989a	Dead/preserved	50	47.5	−101.9	15.0	35.0	Todarodes pacificus	−33.2	
This study	Live	38	20.4	−62.8	28.0	71.5	*Dosidicus gigas*	−33.3	−32.8
Kawabata, 1999	*in situ*	38	40.9	−94.2	18.0	24.0	Todarodes pacificus	−35.0	
Lee *et al.*, 1992	*in situ*	420	33.7	−88.5	3.0	17.0	Loligo edulis	−39.7	
This study	Live	120	25.4	−76.8	28.0	71.5	*Dosidicus gigas*	−40.0	−38.9
This study	Live	200	27.3	−80.5	28.0	71.5	*Dosidicus gigas*	−41.0	−38.7
This study	Live	70	29.9	−84.4	28.0	71.5	*Dosidicus gigas*	−41.1	−37.9

measured. Standard deviations of these measurements were small but considerably larger than those for the corresponding measurements of dorsal-aspect target strength. Animals in the anterior aspect measurements had relatively more freedom of movement than animals being measured from the dorsal aspect. We suspect this led to the more variable target strengths, both because the animals presented more angles to the signal but also because they were harder to keep centered in the beam. A small but significant increase in target strength relative to the dorsal aspect of the same individuals was found at 38 kHz. The scattering mechanism for an increase in target strength at 38 kHz is unclear.

We found a 2–3 dB decrease in posterior-aspect target strength relative to dorsal aspect at 70, 120, and 200 kHz. This compares well with changes of 3–5 dB between dorsal and tail-on measurements at 200 kHz found by Lee *et al.* (1991) in *Loligo edulis*. However, they found that the difference in target strength increased with increasing body size. Their study encompassed a total span in mantle length span of about 12 cm. Despite covering a mantle length range of more than 43 cm, we saw no body-size effect on the differences in target strength as a function of orientation.

B. Potential sources of scattering

Many sources of scattering in squid have been proposed including the flesh itself, chitinous pen (gladius) and beak, thickened suckers on the arms (Goss *et al.*, 2001), internal organs especially the liver and reproductive organs (Iida *et al.*, 1996; Tang and Sato, 2006), outstretched fins (Lee *et al.*, 1991), and even their wake (Selivanovsky and Ezersky, 1996). The fins can be ruled out for the results of the current study because the target strengths of a freshly dead squid were not different from those of live animals of the same size and its fins were not outstretched. The wake can be similarly ruled out. Although the liver and reproductive organs have been shown to be the most important contributors to acoustics scattering in squid at much higher frequencies (Iida *et al.*, 1996; Tang and Sato, 2006), our data do not suggest this. Despite extreme differences in gonadal size between mature and immature individuals of similar body size, and between males and females, we found no significant differences in normalized target strength as a function of gender or sexual maturity, suggesting that the gonads are of limited importance in determining target strength. The effects of the liver could not be investigated in this study.

In order to elucidate the potential contributors to scattering in squid empirically, we measured the target strengths of various parts of several squid immediately after they were studied alive and then sacrificed. Our results suggest that it is indeed the inanimate parts of the squid, not behavior, that causes the backscattering. The results show that both the body and the head strongly contributed to the scattering at all frequencies. At 38 and 70 kHz, the body had a target about 1 dB less than the entire squid, while at 120 and 200 kHz, the target strength of the body was not different from that of the entire squid. Similarly, the head had a target strength 1 dB less than the entire squid at 70 and 120 kHz while the target strengths of the head versus the entire squid at the other frequencies were not different. These results show that the scattering strength of the entire squid is not a simple linear summation of individual parts but does indicate the parts of the animal that could contribute to the backscattering process.

In order to maintain the integrity of the body, only an easily accessible and distinctive part of the pen, the cartilagenous conus, was removed. This had a significant effect on the backscatter at all except the highest frequency suggesting that this dense structure may indeed be an important source of scattering. *Dosidicus* also has a highly muscularized body so the difference in the acoustic impedance of the muscle and seawater may also be an important source of backscatter. The density difference between muscle and seawater is at least 5% and there is a significant difference in sound speed between the muscle and seawater (Arnaya and Sano, 1990).

Many parts of the head showed some effects on scattering. We removed parts of the head successively—beak first, followed by eyes, then arms so we cannot look at the effects of each of these parts on the strength of the reflection but we can examine their potential contribution to the backscattering process. The beak appeared to affect scattering only weakly and only at 38 kHz. The additional removal of the arms had

FIG. 8. Lateral (left) and anterior (right) view of the cranium, or braincase, of the largest squid we measured (71.5 cm mantle length, 118 cm total length).

a stronger effect, also only at 38 kHz. The arms have numerous sucker-cups, each of which has a chitinous ring of teeth, instead of simply being thickened flesh like those of *Loligo* and *Martialia* (Goss *et al.*, 2001). These hardened suckers could thus be an even stronger source of scattering than originally proposed by Goss *et al.* (2001). Removal of the large eyes of the squid had the strongest effect of any removal but only at 120 kHz. Removal of the eyes had no effect on the target strength at any of the other frequencies studied.

Finally, we measured the target strength of only the cranium, or braincase, of *D. gigas*. The braincase is made of very dense cartilage and feels much like stiff plastic (Fig. 8). It contains the statocysts, organs that form the vestibular system. These small dense structures contain statolith crystals made primarily of the aragonite crystal form of calcium carbonate. The target strength of the cranium was 6 dB less than the target strength of the entire squid at 38 kHz. This relative reduction in target strength decreased with frequency to only about 2 dB at 70 kHz and then to 0 dB at higher frequencies. The cranium appears to be the single most important source of acoustic scattering in the head of *D. gigas*. It is remarkable that a single part of the body of this fleshy animal can nearly equal the backscattering strength of the entire animal.

Our results suggest that it should not be surprising that the posterior-aspect target strengths were similar to those from the dorsal aspect. The pen, which seems to be a significant source of backscatter from the dorsal aspect, will not contribute much to the acoustic reflection when the squid is ensonified from the posterior aspect since the cross section perpendicular to the incident signal in this orientation is relatively small. However, the cranium of the head, shown Fig. 8, has a shape that would present a relatively large target to an acoustic signal propagating along the longitudinal axis of the squid from any orientation. The backscatter with the squid from the posterior aspect was probably emanating from the head of the squid.

C. *In situ* measurements

In situ measurements of target strength made during 15 min of concerted fishing effort revealed that the target strengths were remarkably consistent both in amplitude and in frequency response to those predicted by the relationship of length to target strength obtained from animals in controlled conditions. This supports the validity of the tethered measurements, suggesting that there were no artifacts such as air entrapment in the tethered squid or effects of the mounting rig. Arnaya *et al.* (1989b) found that the target strengths of free-swimming squid were lower than those obtained from the same species that were tethered. Lee *et al.* (1992) similarly found a 10 dB lower target strength *in situ* as opposed to tethered. However, in both cases, the animals in tethered measurements were dead, frozen, and thawed before measurement which may cause significant changes in the material properties of the squid. In addition, the species in these studies showed much more significant changes in target strength with orientation than we observed here. Thus, behavioral effects of swimming and orientation adjustments on the *in situ* target strengths in *D. gigas* may be significantly less than in other species. Our *in situ* results combined with the measurements of the length of the squid validate our measurements of tethered animals. These results suggest that it is appropriate to use our target strength curves for the estimation of squid mean sizes *in situ* from mean target strengths of single targets as well as for biomass estimates from volume scattering by squid of known length.

We can begin by applying these measurements to *in situ* measurements of single targets likely to be squid earlier the same night. During this time period of several hours, squid were visible actively feeding at the surface and appeared to be separated by no more than one to three body lengths, at least within the range of illumination provided by the vessel's deck lights. No other large targets were visible near the surface. Fishermen, two to three to a small boat, were each pulling up squid at a remarkable rate, often exceeding 1 squid per minute. With a typical body mass of 5–10 kg, the rate of squid captured per boat would be consistent with the reported figure of 1.0–1.2 ts in 2–4 h of fishing effort (Morales-Bojorquez *et al.*, 2001). This observation suggests that the density of squid in the immediate area of our stationary vessel was extremely high and the rate at which we were able to detect individual targets supports this. We measured the mean target strength of all large, individually detectable targets (see the above-mentioned single target detection) at each frequency in three, 15 min intervals over a period of 2.25 h. During this time, the rate of target detection increased by an order of magnitude, paralleling an apparent increase in fishing success for *D. gigas*, suggesting the majority of large targets were indeed *Dosidicus*. Measured target strengths also changed during this time. The mean target strengths in each 15 min interval showed the same frequency response seen in our controlled experiments (and our *in situ* measurements combined with squid length measurements), but mean target strength at each frequency increased in each successive 15 min interval. Though no independent measurements of squid length were obtained during any of these sampling intervals, the data strongly suggest that the targets were primarily *Dosidicus* so we can extrapolate from both our controlled experiment and verified *in situ* results to infer that squid size also increased during this time. The mean target strengths would equate to mean squid mantle lengths of approximately 25, 35, and 45 cm at 2100–2115, 2130–2145, and 2200–2215, respectively. These differences are not likely due to changes in behavior because of the limited ef-

fects on target strength we observed as a function of orientation. However, these differences were accompanied by changes in aggregation behavior of squid with dense, distinct groups observed only later in the night. Alternative explanations of the limited results include the mixture of other species. The individual target frequency response and the unimodal and narrow (e.g., small confidence interval) distribution of target strength do not suggest a large number of other targets. However, we cannot rule out that the differences in density over time could cause a target strength bias due to single target detection errors with increasing multiple echoes. Over the sampling period, there were also obvious changes in swimming tracks of individual squid (see examples in Fig. 7). This may prove to be the greatest strength of our ability to observe these squid *in situ*—the possibility of observing their behavior in the wild.

V. CONCLUSIONS

This study presents the first target strength measurements of *Dosidicus gigas*, a large squid that is an ecologically significant predator, a key prey resource, and the target of an economically important fishery. Our results show a strong relationship between squid length and target strength over a wide range of sizes encompassing all but the largest individuals of this species. Individual *Dosidicus* showed little variation in target strength at 70 kHz, 120, and 200 kHz but the target strength at 38 kHz was substantially higher, regardless of squid size (Fig. 5). We found little change in target strength when the squid was dorsal versus when it was tilted completely head down. We also found no difference between live and freshly dead squid.

The scattering mechanisms in squid have been long debated with many possibilities suggested. From our results, we can infer that the reproductive organs, or at least variability in their size and state, had little effect on squid target strength. We empirically tested several hypotheses and found that the body and the head are nearly equally important sources of scattering. Our results support the hypothesis that the pen may be an important source of acoustic scattering in squid. We can also support the hypothesis that the beak, eyes, and arms (probably via the sucker rings) play some role in acoustic scattering though their effects were relatively small and affected scattering only at one of four measured frequencies. We found an unexpected source of scattering in the cranium of the squid which provided a target strength nearly as high as that of the entire squid. Measurements of the material properties of this structure are not available so the mechanism of scattering remains unclear. However, the cranium does house the extremely dense statocysts, or balance organs, that should be further investigated.

Our *in situ* measurements of target strength paired with jigged squid samples support the use of the values presented here in squid stock assessment studies. The ease of detection of these squid and their relatively unique frequency response in target strength suggest acoustic studies of *D. gigas* are feasible. With the substantial target strengths measured, depending on density, it should be possible to detect individual squid to the full depth range of most high-frequency scientific echosounders (up to 1000 m for the 38 kHz Simrad EK60) and to detect aggregations of squid to those depths as well. Based on the scattering mechanisms observed in this and other studies for squid, there is no reason to suspect a change in acoustic properties with this increased depth. Acoustics have been used as an effective sampling method for many fish species but only rarely for invertebrates, especially large, commercially exploited species. A fundamental requirement for application of acoustics to quantitative assessment is knowledge of the target strength distribution of the species in question. These results provide that information.

Even more understanding of the biology of these animals could come from *in situ* observations of squid in relatively shallow water where individual animals can be observed and potentially tracked. Individual target tracks are clearly visible in Fig. 7 with many different diving patterns observable. The results provide an enticing glimpse into the behavioral dynamics of individual squid and with further investigation, the possibility of understanding the behavior of populations of *D. gigas*.

ACKNOWLEDGMENTS

We thank Unai Markaida for determinations of sexual maturity stage. Ashley Booth, Abram Fleishman, Ladd Irvine, Dana Shulman, and Chad Waluk provided assistance in the field. Bonnie Anderson-Becktold, Cyndee Pekar, and Cesar Salinas and his students from CIBNOR provided logistical support. Bob Pedro, the captain of the R/V Pacific Storm, and Willi Schlecter, its engineer, provided invaluable assistance both in preparation for, and during the cruise. Funding was provided by donors to the Oregon State University Marine Mammal Institute. The R/V Pacific Storm was donated by Scot and Janet Hockema. This work was also supported by grants from the National Science Foundation (No. OCE 0526640) and the David and Lucile Packard Foundation to W.F.G.

Arnaya, I. N., and Sano, N. (**1990**). "Studies on acoustic target strength of squid V. Effect of swimming on target strength of squid.," Bull. Fac. Fish. Hokkaido Univ. **41**, 18–31.

Arnaya, I. N., Sano, N., and Iida, K. (**1989a**). "Studies on acoustic target strength of squid. II. Effect of behaviour on averaged dorsal aspect target strength," Bull. Fac. Fish. Hokkaido Univ. **40**, 83–99.

Arnaya, I. N., Sano, N., and Iida, K. (**1989b**). "Studies on acoustic target strength of squid. III. Measurement of the mean target strength of small live squid," Bull. Fac. Fish. Hokkaido Univ. **40**, 100–115.

Benoit-Bird, K. J., and Au, W. W. L. (**2001**). "Target strength measurements of animals from the Hawaiian mesopelagic boundary community," J. Acoust. Soc. Am. **110**, 812–819.

Boyle, P., and Rodhouse, P. G. (**2005**). "Fishing methods and scientific sampling," in *Cephalopods: Ecology and fisheries*, edited by P. Boyle, and P. G. Rodhouse (Blackwell, London), pp. 259–276.

Foote, K. G., Hanlon, R. T., Iampietro, P. J., and Kvitek, R. G. (**2006**). "Acoustic detection and quantification of benthic egg beds of the squid Loligo opalescens in Monterey Bay, California," J. Acoust. Soc. Am. **119**, 844–856.

Foote, K. G., Vestnes, G., Maclennan, D. N., and Simmonds, E. J. (**1987**). "Calibration of acoustic instruments for fish density information: A practical guide," in International Council for the Exploration of the Sea Cooperative Research Report No. 144, Copenhagen, Denmark.

Fristrup, K. M., and Harbison, G. R. (**2002**). "How do sperm whales catch squids?," Marine Mammal Sci. **18**, 42–54.

Gilly, W. F., Markaida, U., Baxter, C. H., Block, B. A., Boustany, A.,

Zeidberg, L., Reisenbichler, K., Robison, B., Bazzino, G., and Salinas, C. (**2006**). "Vertical and horizontal migrations by the jumbo squid *Dosidicus gigas* revealed by electronic tagging," Mar. Ecol.: Prog. Ser. **324**, 1–17.
Goss, C., Middleton, D., and Rodhouse, P. G. (**2001**). "Investigations of squid stocks using acoustic survey methods," Fish. Res. **54**, 111–121.
Iida, K., Mukai, T., and Hwang, D. (**1996**). "Relationship between acoustic backscattering strength and density of zooplankton in the sound scattering layer," ICES J. Mar. Sci. **53**, 507–512.
Jefferts, K., Burczynski, J., and Pearcy, W. G. (**1987**). "Acoustical assessment of squid (*Loligo opalecens*) off the central Oregon coast," Can. J. Fish. Aquat. Sci. **44**, 1261–1267.
Kang, D., Mukai, T., Iida, K., Hwang, DJ., and Myoung, JG. (**2005**). "The influence of tilt angle on the acoustic target strength of the Japanese common squid," ICES J. Mar. Sci. **62**, 779–789.
Kawabata, A. (**1999**). "Measurement of the target strength of the Japanese flying squid, *Todarodes pacificus* Steenstrup," Bull. Tohoku Nat. Fish. Res. Inst. **61**, 29–40.
Kawabata, A. (**2001**). "Measurement of the target strength of live squid, *Todarodes pacificus* Steenstrup, in controlled body tilt angle," Bull. Tohoku Natl. Fish. Res. Inst. **64**, 61–67.
Kawabata, A. (**2005**). "Target strength measurements of suspended live ommastrephid squid, *Todarodes pacificus*, and its application in density estimations," Fish. Sci. **71**, 63–72.
Lee, K. T., Liao, C. H., Shih, W. H., and Chyn, S. S. (**1992**). "Application of dual-beam acoustic survey techniques to assess the size distribution of squid, *Loligo edulis*," J. Fish. Soc. Taiwan **19**, 25–34.
Lee, K. T., Shih, W. H., Liao, C. H., and Wang, J. P. (**1991**). "Studies on the hydroacoustic scattering properties of individual squid, *Loligo edulis*," J. Fish. Soc. Taiwan **18**, 215–225.
Lipinski, M. R., and Prowse, M. (**2002**). "Direct hydroacoustic observations of chokka squid *Loligo vulgaris reynaudii* spawning activity in deep water," S. Afr. J. Mar. Sci. **24**, 387–393.
Lipinski, M. R., and Underhill, L. (**1995**). "Sexual maturation in squid: Quantum or continuum?," S. Afr. J. Mar. Sci. **15**, 207–223.
Markaida, U., and Sosa-Nishizaki, O. (**2001**). "Reproductive biology of jumbo squid *Dosidicus gigas* in the Gulf of California, 1995–1997," Fish. Res. **54**, 63–82.
Markaida, U., and Sosa-Nishizaki, O. (**2003**). "Food and feeding habits of jumbo squid *Dosidicus gigas* (Cephalopoda: Ommastrephidae) from the Gulf of California, Mexico," J. Mar. Biol. Assoc. U.K. **83**, 507–522.
McClatchie, S., Macaulay, G., and Coombs, R. F. (**2003**). "A requiem for the use of 20logLength for acoustic target strength with special reference to deep-sea fishes," ICES J. Mar. Sci. **60**, 419–428.
Misund, O. A. (**1997**). "Underwater acoustics in marine fisheries and fisheries research," Rev. Fish Biol. Fish. **7**, 1–34.
Morales-Bojorquez, E., Cisneros-Mata, M. A., Nevarez-Martinez, M. O., and Hernandez-Herrera, A. (**2001**). "Review of the stock assessment and fishery biology of *Dosidicus gigas* in the Gulf of California, Mexico," Fish. Res. **54**, 83–94.
Nevarez-Martinez, M. O., Mendez-Tenoria, F. J., Cervantes-Valle, C., Lopez-Martinez, J., and Anguiano-Carrasco, M. L. (**2006**). "Growth, mortality, recruitment, and yield of the jumbo squid (*Dosidicus gigas*) off Guaymas, Mexico," Fish. Res. **79**, 38–47.
Nigmatullin, C. M., Nesis, K. N., and Arkhipkin, A. I. (**2001**). "A review of the biology of the jumbo squid *Dosidicus gigas* (Cephalopoda: Ommastrephidae)," Fish. Res. **54**, 9–19.
Sanchez-Juarez, E. (**1991**). "Preliminary estimate of the size of the adult population and spawning biomass of jumbo squid *Dosidicus gigas* in the Exclusive Economic zone of the Mexican Pacific," (Spanish Translation "Estimacion preliminar del tamano de la poblacion adulta e biomasa reproductora de calamar gigante *Dosidicus gigas* en la Zona Economica Exclusiva del Pacifico Mexicano, en diciembre de, 1990), Secretaria de Pesca, Instituto Nacional de la Pesca, Informe Tecnico, CRIP, El Sauzal. B.C., Mexico, pp. 1–13.
Sawada, K., Furusawa, M., and Williamson, N. J. (**1993**). "Conditions for the precise measurement of fish target strength *in situ*," Fish. Sci. **20**, 15–21.
Sawada, K., Takahashi, H., Abe, K., and Takao, Y. (**2006**). "In situ measurement of target strength, tilt angle, and swimming speed of Boreopacific gonate squid (*Gonatopsis borealis*)," J. Acoust. Soc. Am. **120**, 3107.
Selivanovsky, D., and Ezersky, A. (**1996**). "Sound scattering by hydrodynamic wakes of sea animals," ICES J. Mar. Sci. **53**, 377–381.
Shchetinnikov, A. S. (**1988**). "Feeding and food relations of abundant squids in southeastern part of the Pacific Ocean," *Autoreferat Dissertatsii na Soiskani Uchenoj Stepeni Kandidata Biologicheskikh* (Nauka, Moscow).
Tang, Y., and Sato, M. (**2006**). "Acoustical imaging of nonbladdered animals using underwater acoustic camera," J. Acoust. Soc. Am. **120**, 3058.

Demonstration of the invariance principle for active sonar

Jorge E. Quijano[a] and Lisa M. Zurk[b]
Northwest Electromagnetics and Acoustics Research Laboratory, Department of Electrical Engineering, Portland State University, Portland, Oregon 97201-0751

Daniel Rouseff[c]
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105-6698

(Received 22 May 2007; revised 21 December 2007; accepted 27 December 2007)

Active sonar systems can provide good target detection potential but are limited in shallow water environments by the high level of reverberation produced by the interaction between the acoustic signal and the ocean bottom. The nature of the reverberation is highly variable and depends critically on the ocean and seabed properties, which are typically poorly known. This has motivated interest in techniques that are invariant to the environment. In passive sonar, a scalar parameter termed the waveguide invariant, has been introduced to describe the slope of striations observed in lofargrams. In this work, an invariant for active sonar is introduced. This active invariant is shown to be present in the time–frequency structure observed in sonar data from the Malta Plateau, and the structure agrees with results produced from normal mode simulations. The application of this feature in active tracking algorithms is discussed. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2836763]

PACS number(s): 43.30.Vh, 43.30.Bp, 43.30.Zk, 43.60.Ac [RCG] Pages: 1329–1337

I. INTRODUCTION

In shallow water environments, the performance of active sonar systems is strongly affected by the presence of reverberation. This is the result of acoustic energy traveling through multiple paths and interacting with the sea surface and seabed. In active sonar, an acoustic source is utilized to insonify a target and the time of arrival of the reflected wave can be used to estimate its location. At the receiver, the reverberation noise can obscure the desired return from the target.

Although there is considerable structure in sonar returns (both from the target and bottom reverberation) the high variability and lack of knowledge about exact parameters of the ocean environment typically makes model-based processing unfeasible. To improve the robustness of active sonar, it is desirable to identify the properties of acoustic propagation that are insensitive to specific knowledge of the ocean environment. In the past, the theory of waveguide invariants has been used to improve the processing of passive sonar. In this paper, the waveguide invariant concept is extended to active sonar and tested using experimental data.

The waveguide invariant concept was developed for passive sonar. It is a quantity introduced by Chuprov[1] and described by Brekhovskikh[2] that describes the properties of the waveguide in a single scalar (β) that relates the intensity of the propagating wave, range from source to receiver, and frequency band of the source. The equation that defines β can be used to estimate the horizontal range between source and receiver, and this relation will remain unaffected by small perturbations in the sound speed profile or the geoacoustic properties of the channel. The waveguide invariance can be observed as constant intensity lines (striations) that are a function of range and frequency. The value of β is measured as the slope of the striations, and for many shallow water environments $\beta \approx 1$.

The waveguide invariant has been recently applied to passive sonar. For example, Rouseff and Leigh[3] use it for coherent processing with beamforming of a horizontal array of receivers. Yang[4] presents a comparison between the spectrogram obtained from a single hydrophone and the spectrogram obtained from the output of a beamformer and shows that the invariant striation pattern is enhanced in the latter case due to the gain of the array and as expected, the slope of the striations does not change. Yang[4] also utilized the invariant to detect the presence of two sources located at different ranges based on the different time–frequency pattern associated to each source. Another application of the passive sonar invariant is presented by Thode[5] for range independent bathymetries, in which the invariant is utilized along with the concept of a virtual receiver to create a virtual aperture for estimation of the range of an unknown source. The virtual aperture algorithm was successfully applied to experimental data, and simulations suggest that the slope of the striation pattern at the output of the virtual aperture does not change substantially as a result of different sound speed profiles. A similar approach is presented by Zurk and Tracey,[6] where the invariance relation is utilized to shift in range the transfer function between a guide source and a receiver. This is combined with a depth shifting technique and allows the measurement of the water channel response in several positions other than the guide source location. For range dependent bathymetries, D'Spain and Kuperman[7] show that the value of the invariant is affected, but this can be modeled using an adiabatic approach for the pressure field from the source to the receiver.

[a]Electronic mail: jorgeq@pdx.edu.
[b]Electronic mail: zurkl@pdx.edu.
[c]Electronic mail: rouseff@apl.washington.edu.

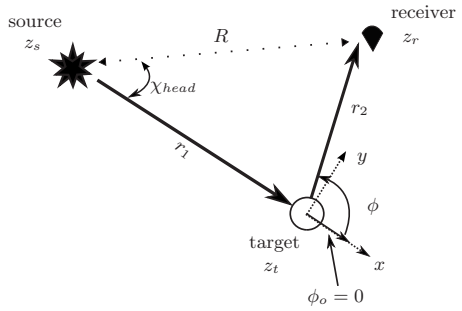


FIG. 1. Top-down view of a bistatic active sonar network. The bistatic range r_b is defined as $r_b = r_1 + r_2$. The x/y axis is used as a reference for the source and target azimuth angle in the next section.

Extending the waveguide invariant concept to active sonar involves considering more complicated geometries than are encountered in passive sonar. Sound must propagate both from the source to the target and from the target to the receiver. Figure 1 shows a bistatic active sonar geometry where the source is separated by range R from a receiver in the presence of a target. The invariant structure is developed for the full bistatic range $r_b = r_1 + r_2$, which is the sum of the path lengths from the source to the target and from the target to the receiver. The question for active sonar is whether the bistatic range and frequency content are described by an invariant as in the passive case.

To test the active invariant concept, this paper presents spectrograms produced from the Shallow Water Active Classification (SWAC) experiment and shows that these exhibit intensity striations that are similar to the ones found in passive sonar lofargrams. This paper is organized as follows: In Sec. II, the concept of the waveguide invariant is reviewed. A far-field approximation for the bistatic pressure is also introduced and the procedure to obtain simulated active sonar spectrograms is outlined. In Sec. III the SWAC experiment is described and in Sec. IV the experimental results are compared to simulated spectrograms. Section V discusses the potential use of the bistatic time–frequency invariant to improve the performance of active sonar, followed by conclusions.

II. THE INVARIANT AND ITS APPLICATION TO BISTATIC GEOMETRIES

In the following, the invariant as presented by Brekhovskikh and Lysanov[2] is first reviewed. The concept is then extended to a general bistatic geometry and is utilized to simulate bistatic spectrograms.

A. Single path invariant

For passive sonar, the pressure at range r due to a point source can be written using normal modes. Assuming an $e^{-i\omega t}$ time convention,

$$p_0(r,z;z_s;\omega) = C \sum_{m=1,M} \Psi_m(z_s)\Psi_m(z)\frac{e^{ik_m r}}{\sqrt{k_m r}},$$

$$\equiv \sum_{m=1,M} B_m e^{ik_m r}, \tag{1}$$

where $z = z_s$ is the depth of the source, k_m is the horizontal wave number for the mth propagating mode, $\Psi_m(z)$ is the corresponding depth dependent mode function, and $C = ie^{-i\pi/4}/\rho\sqrt{8\pi}$ is a normalizing constant. The more compact form using B_m will prove useful. The acoustic intensity $I_0(r,z;z_s) = |p_0(r,z;z_s;\omega)p_0^*(r,z;z_s;\omega)|$ can be expressed as

$$I_0(r,z;z_s;\omega) = \sum_{m=1}^{M} B_m B_m^* + 2\sum_{m,n} B_m B_n^* \cos((k_m - k_n)r). \tag{2}$$

This relation explains the striation patterns observed in passive lofargrams which depict intensity as a function of time (equivalent to range for a moving source) and frequency. The interference between any pair of modes is represented by the cosine term in Eq. (2) which predicts constructive interference when

$$r = \frac{2\pi q}{(k_m - k_n)}, \tag{3}$$

where q is an integer. Recognizing that $v_m = \omega/k_m$ is the phase velocity of the mth mode, Eq. (3) provides a simultaneous constraint on the frequency.

The invariant is obtained by considering the conditions where all the propagating modes combine constructively as a function of both frequency and range, and it was developed by defining an equation to describe constant intensity striations in a lofargram as

$$\frac{d\omega}{dr} = \frac{\partial I/\partial r}{\partial I/\partial \omega}. \tag{4}$$

The modes are weak functions of frequency, but for limited bandwidths a reasonable approximation is to treat Ψ_m as frequency independent. Substituting Eq. (2) into Eq. (4) and introducing a Taylor series approximation to relate the phase velocity to the group velocity yields, after some manipulation,[2] the invariant β,

$$\beta = -\frac{d(1/v)}{d(1/u)} = \frac{r}{\omega}\frac{\Delta\omega}{\Delta r}, \tag{5}$$

where Δr and $\Delta\omega$ represent small perturbations in range and frequency and u and v are the average value of the group and phase velocity, respectively.

The interpretation of Eq. (5) is that given the intensity $I(r,\omega)$, any change $\Delta\omega$ in the frequency of the source must be compensated by Δr in range to keep a constant intensity. As mentioned before, one of the key features of the invariant β is that its value is not strongly dependent on environmental parameters and for many shallow water environments, $\beta \approx 1$.

B. Active sonar simulations

The complexity of the waveguide invariant expression increases for active sonar because the acoustic wave travels over two paths (source to target, target to receiver) instead of

one, and the modes interact when they scatter at the target. There is a scattering factor that couples the two sets of traveling modes in a target-dependent fashion, and therefore it is not straightforward that a robust interference pattern will emerge for active geometries.

A simulation of the frequency-range dependency at the receiver in Fig. 1 requires the computation of the bistatic pressure. The pressure measured at the position of the target can be written in terms of normal modes as in Eq. (1), and a similar expression exists for propagation between target and receiver.

Here, the notation for the scattered field from a rigid sphere used by Ingenito[8] is adopted, where $S(\alpha_m,\phi;\alpha_n,\phi_0)$ represents the scattering matrix, α_n is the grazing angle corresponding to the incoming nth mode, ϕ_0 is the azimuth angle of the incoming wave (Fig. 1), and α_m and ϕ are the corresponding grazing and azimuth angles for the scattered mode. An approximation for the scattered pressure for small grazing angles ($\alpha_m \approx 0$ and $\alpha_n \approx 0$) can be written as

$$p(r_1,z_t;r_2,z_r;z_s;\omega) = C^2 \sum_m \sum_n \Psi_m(z_s)\Psi_m(z_t)\frac{e^{ik_m r_1}}{\sqrt{k_m r_1}} \times S(\alpha_m,\phi;\alpha_n,\phi_0)\Psi_n(z_t)\Psi_n(z_r)\frac{e^{ik_n r_2}}{\sqrt{k_n r_2}}. \quad (6)$$

Note that Ingenito[8] defined the angles α_n with respect to the vertical and therefore his approximation is for $S(\pi/2,\phi;\pi/2,\phi_0)$ rather than $S(0,\phi;0,\phi_0)$ as in this paper. From Eq. (6) it is evident that the scattered pressure is equivalent to the one generated by a point source located at the position of the target with the strength of the incident field and weighted by the scattering matrix.

In active sonar, the pressure recorded at the receiver is the final two-path acoustic propagation over paths of length r_1 and r_2. Hypothetically, if the energy arriving to the target was recorded before scattering, a spectrogram for the propagation in the first leg (range r_1) could be computed. Similarly, by considering the energy scattered from the target as a source, the pressure at the receiver can be utilized to compute a spectrogram for the second leg (range r_2). Simulated spectrograms from both "single-leg" and the full bistatic path can be constructed for the geometry shown in Fig. 1, where it is assumed that the target moves with constant speed of 3.3 m/s in the direction $\chi_{\text{head}}=80°$. Therefore, ϕ_0 remains constant and $S(\alpha_m \approx 0,\phi;\alpha_n \approx 0,\phi_0=0)$, since the grazing angle α_m is small for shallow water. For this simulation a spherical target is used with $R=10$ km, $z_t=50$ m, $z_s=20$ m, and $z_r=100$ m. A source with a flat spectrum between 130 and 170 Hz is assumed and the scattering strength was computed using Eq. (27) from Ingenito.[8] The environment consists of a flat bathymetry of 200 m depth with constant sound speed of 1500 m/s, with a pressure release at the surface and a sediment half space at the bottom. The geoacoustic properties of the bottom half space were taken from Siderius *et al.*[12] and are summarized in Table I.

Figure 2 shows the single path striation patterns for the r_1 and r_2 legs along with the range as a function of time. In both cases, the slope of the striations can be predicted by Eq. (5). For Fig. 2(a), the invariant $\beta_1=0.95$ was calculated from measuring the actual slope in the images for $r_1=7080$ m, $\Delta r_1=1330$ m, $f=140$ Hz, and $\Delta f=25$ Hz and for Fig. 2(b), $\beta_2=1.1$ for $r_2=16.11$ km, $\Delta r_2=2.18$ km, $f=142$ Hz, and $\Delta f=21$ Hz.

TABLE I. Geoacoustic parameters for the Malta Plateau used to compute simulated spectrograms with KRAKEN (Ref. 9) normal modes.

Variable	Value
Sediment sound speed	1550 m/s
Sediment attenuation	0.2 dB/λ
Bottom density	1.5 g/cm^3

Figure 3(a) shows the spectrogram for the source–target–receiver bistatic geometry estimated from Eq. (6). The resulting bistatic spectrogram preserves characteristics of both single path spectrograms for the r_1 and r_2 legs. In Sec. IV, the procedure described here will be used to simulate spectrograms and compare to experimental spectrograms from the SWAC data. It will be shown that the simulations can reproduce the main features of the experimental striations with very little knowledge of the environment and scattering properties of the target.

III. EXPERIMENTAL DATA AND PROCESSING

The experimental data presented in this paper were provided by the Naval Undersea Warfare Center (NUWC, Newport, RI).[10] The data were collected in October of 1994 near the Malta Plateau as part of the SWAC project. The goal was to characterize the clutter due to bottom reverberation in shallow water and to reduce the number of false alarms for low- and midfrequency systems. In this paper, these data have been used to identify striations related to the invariance principle.

A. Experiment setup

Figure 4(a) shows a top-down view of the experimental setting. For this experiment, the relative position between a

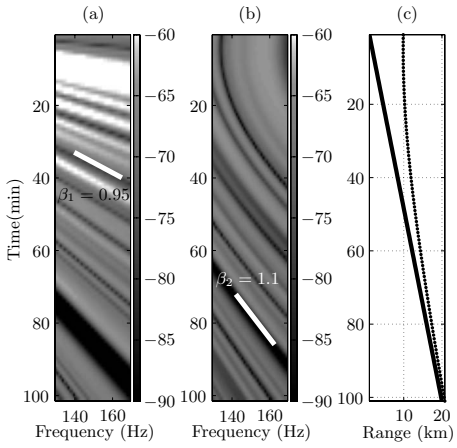


FIG. 2. Single path invariant for the configuration shown in Fig. 1. (a) Time–frequency dependency for the r_1 leg; (b) time–frequency dependency for the r_2 leg; (c) r_1 (solid line) and r_2 (dotted line) as a function of time.

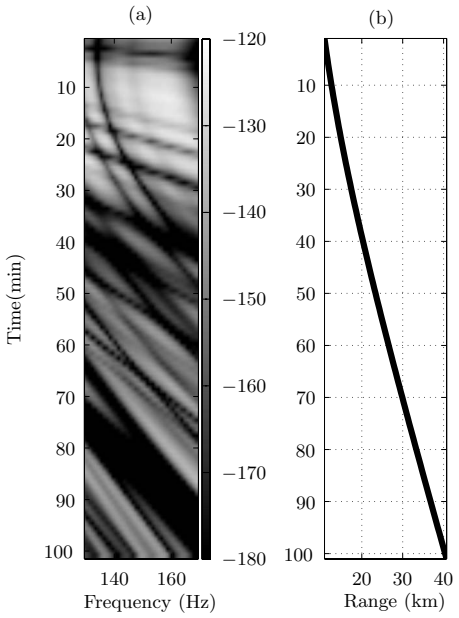


FIG. 3. Simulation of a bistatic spectrogram for the configuration shown in Fig. 1: (a) Bistatic spectrogram for the full path source–target–receiver. The corresponding single leg spectrograms are shown in Fig. 2; (b) $r_b = r_1 + r_2$ as a function of time. The scattering strength $S(\alpha_m, \phi; \alpha_n, \phi_0)$ was computed using Eq. (27) from Ingenito (Ref. 8).

towed active monostatic sonar and several stationary targets was tracked as a function of time. The data set provided by NUWC consists of match-filtered acoustic pressure corresponding to 20 different stationary targets. In total, 661 trials along different towing paths are contained in the data set. In this paper, we present results for three of the targets: two wrecks and an oil rig moored tanker (although all cases were processed and exhibited similar features where appropriate). The sonar system was towed with an average speed of 2.6 m/s following the tracks detailed in Fig. 4(a). Note that this configuration mimics having a moving target and a stationary sonar system as might be the case in real applications. Figure 4(b) shows three measurements of the sound speed profile in this region. In all cases a mixed layer of warm water was present and located at a depth of 45 m (solid line), 36 m (dotted line), and 31 m (dashed line) with the thermocline extending to 74, 60, and 59 m, respectively.

In addition to the acoustic data, the position of the ship for each transmitted ping was measured using GPS. This provides the horizontal range between the sonar and the target. The range information was used along with the experimental sound speed profile and the parameters in Table II to simulate bistatic spectrograms utilizing the method developed in Sec. II.

Figure 5(a) shows an example of the matched filter output for one of the target returns from the moored tanker with a signal-to-noise ratio around 14 dB. Two different chirp signals were utilized in the SWAC experiment: low frequency (LF) with a center frequency between 495 and 600 Hz and a bandwidth between 70 and 400 Hz, and midfrequency (MF) with a center frequency around 3000 Hz and a bandwidth between 400 and 700 Hz.

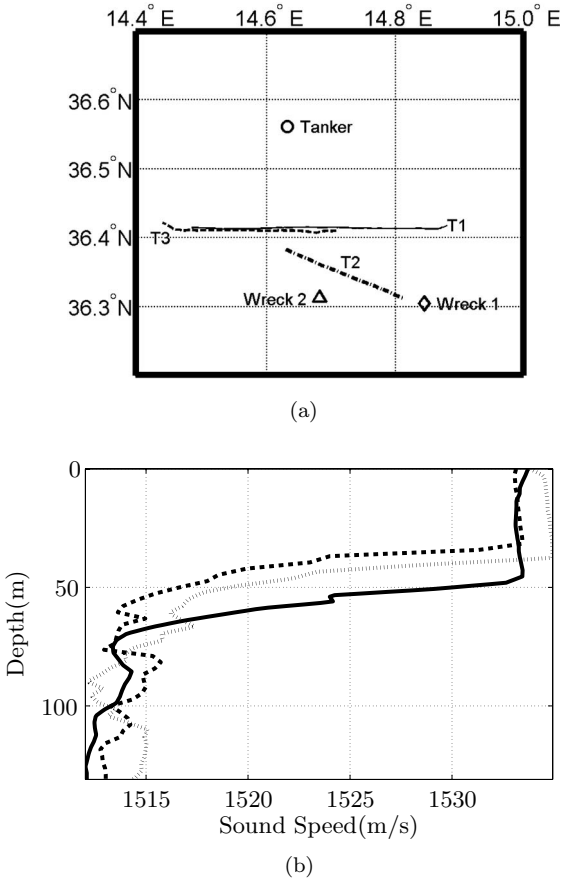


FIG. 4. (a) Map showing the SWAC experiment (Ref. 10). Three different targets are highlighted: An oil rig moored tanker (○), Wreck 1 (◇), and Wreck 2 (△). T1 (solid line), T2 (dash-dot line), and T3 (dashed line) are the corresponding GPS tracks followed by the towed sonar; (b) observed sound speed profile for the SWAC experiment. The three measurements correspond to the days 25 October 1994 at 13:53 Greenwich Mean Time (GMT) (dashed line), 26 October 1994 at 12:01 GMT (solid line), and 26 October 1994 at 18:34 GMT (dotted line). Water depth is 135 m.

The experimental results presented here correspond to the LF band. Striations were not observed at MF in this particular data set for a number of reasons. First, only 159 MF tracks are greater than 20 pings; tracks with less than 20 pings result in spectrograms with coarse granularity in which the striations are not obvious. Second, the signal-to-noise ratio in the MF data is smaller and usually the target arrival is not as clear as the one shown in Fig. 5(b) for the LF data,

TABLE II. Main experimental parameters for the Malta Plateau experiment.

Variable	Value
Sampling rate	222–666 S/s
Pulse (chirp) length	1.2–2 s
Bandwidth (LF)	70–400 Hz
Center frequency (LF)	495–600 Hz
Source, receiver depth	66.5 m
Water depth	135 m
Ping frequency	1 ping/min
GPS tracking	1 sample/min
Moored tanker lon/lat	14.6327E/36.5609N
Wreck 1 lon/lat	14.8454E/36.3041N
Wreck 2 lon/lat	14.6827E/36.3124N

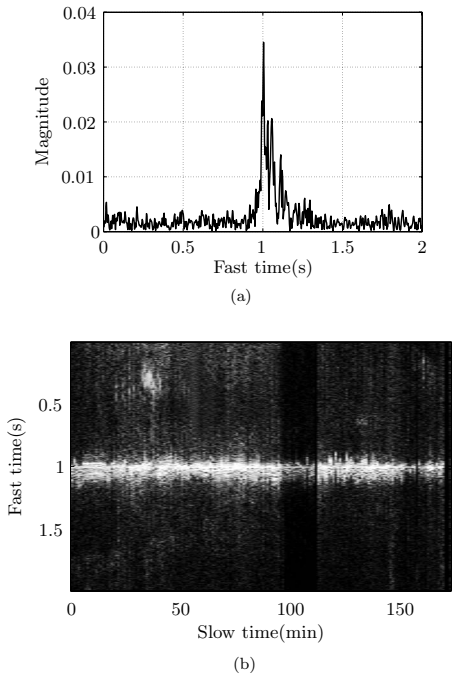


FIG. 5. (a) Output of the matched filter for the first ping in (b); (b) example of the windowed match-filtered output for a track of 174 pings (1 ping/min), associated with the moored tanker target in Fig. 4(a).

which results in a blurry spectrogram. A third reason is that a small bandwidth can result in scarcity of data, which means that not enough target detections are available to construct the line pattern. Perhaps this is the biggest factor influencing the MF data since the predicted slope is governed by $\Delta\omega/\omega$, so the required ping rate at higher frequencies needs to increase to maintain the same image granularity. This proportionality relation is illustrated with a simulation in Sec. IV. While striations were not observed at MF, it is believed they would be observable with appropriate sonar parameters.

In this paper, the term "slow time" refers to the ping rate along a track, while "fast time" refers to the duration of the snapshot centered around each detection as observed in Fig. 5(b), where 174 pulses corresponding to the moored tanker [track T1 in Fig. 4(a)] were stacked horizontally.

B. Spectrogram processing

The energy of the received pings was not uniform along each track, as is evident in Fig. 5(b) between $t=97$ and $t=113$ min. For this reason, a ping to ping normalization was performed before analysis of the frequency spectrum of the arrivals.

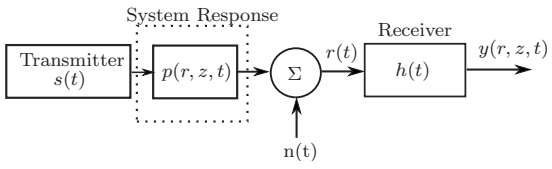


FIG. 6. The system response corresponds to the ocean channel as given by Eq. (6), and includes the two-way wave propagation and the target scattering effects. The impulse response of the receiver is $h(t)=s(\tau-t)$.

Figure 6 shows a block diagram of the system, where the output $s(t)$ of the transmitter is a chirp. The system can be modeled as the total impulse response, which convolves propagation effects with the scattering of the target. At the receiver, the scattered pulse is match filtered using a delayed replica of the transmitted chirp $s(t)$ yielding $y(t)$.

The matched filtered signal at the output of the receiver can be written as the convolution

$$y(t,r,z) = h(t) * [s(t) * p(t,r,z) + n(t)], \tag{7}$$

in which $n(t)$ is the (white) background noise and $p(t,r,z)$ is the combined response of the target scatter and the water column. Note that the noise due to reverberation is included in $p(t,r,z)$. Since $h(t)=s(\tau-t)$, then

$$y(t,r,z) = s(\tau-t) * [s(t) * p(t,r,z) + n(t)], \tag{8}$$

and taking the Fourier transform of Eq. (8)

$$Y(\omega,r,z) = [S^*(\omega)S(\omega)P(\omega,r,z) + S^*(\omega)N(\omega)]e^{-i\omega\tau}. \tag{9}$$

Since $s(t)$ is a linear chirp (with energy evenly distributed over the frequency band), it has a flat band-limited spectrum. Therefore, the spectrum of the chirp will not alter the shape of the time–frequency structure contained in the frequency response of the channel, $P(w,r,z)$. For active sonar, a high signal-to-noise ratio is expected $[n(t) \ll s(t)]$ and then the magnitude of Eq. (9) has the same shape as the system response

$$|Y(\omega,r,z)| \approx |P(\omega,r,z)|, \tag{10}$$

where $|Y(\omega,r,z)|$ is an approximation of the bistatic spectrogram that can be obtained from the experimental data by estimating the power spectral density of each individual snapshot. The frequency analysis was performed using the Welch–Bartlett method with a Hamming window (length $=25\%$ of the data; overlapping$=20\%$ of the data),[11] and the results are shown in Sec. IV.

IV. EXPERIMENTAL AND SIMULATED RESULTS

Following the procedure described in Sec. III, spectrograms were obtained from the SWAC data set. Striation patterns were found in the majority of the processed tracks. For visualization purposes, some features like the aspect ratio or the gray scale of the spectrograms were adjusted manually.

Figure 7 shows the spectrogram for the experimental data corresponding to the moored tanker target. A copy of the spectrogram with guidelines set manually is presented to the right to emphasize striations of constant intensity similar to those observed in passive sonar lofargrams.

Figure 7(c) shows the GPS measurement of the horizontal range, and it matches the shape of the striations in the experimental spectrogram. This suggests that there is a potential of using the spectrogram to predict or confirm the position and the heading of the target, as discussed in Sec. V. The striation pattern has a concave shape due to the change in range as a function of time, where the source was moving toward the target initially, and then it moved away from the target. The resulting spectrogram is symmetric with respect to the turning point at $t=74$ min. If Eq. (4) holds for bistatic

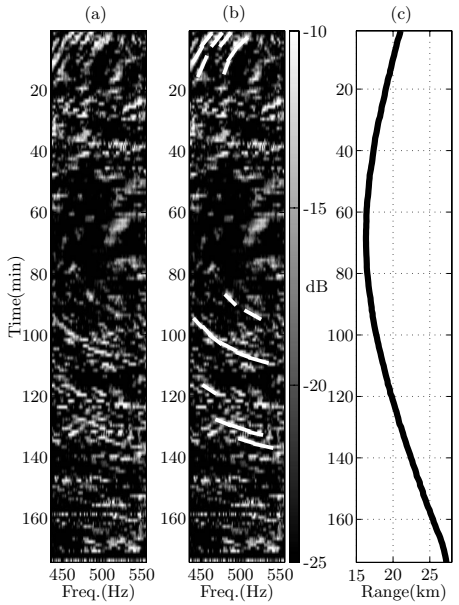


FIG. 7. (a) Spectrogram for a moored tanker (Target N.3, track N.56 from NUWC database). The data were collected on 26 October 1994 at 9:48 GMT; (b) superimposed guidelines highlight the main striations; and (c) GPS measurement of range as a function of time.

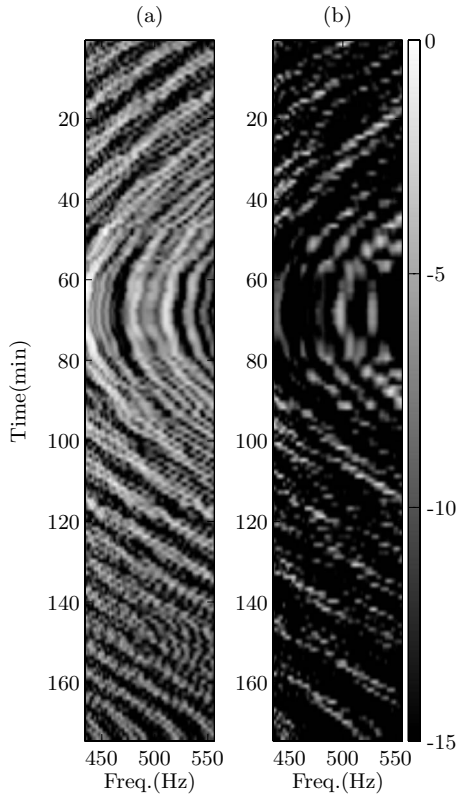


FIG. 8. Spectrogram simulations for the moored tanker track shown in Fig. 7. Two different scattering matrices were considered: (a) No intermode coupling and (b) Lambertian model.

configurations, then the change in range Δr as a function of time can be extracted directly from the spectrogram as in the case of passive sonar.

To test the bistatic invariant concept, a simulated spectrogram was generated that could be compared to the experimental results in Fig. 7. The simulation used the ship track shown in Fig. 7(c) and the sound speed profile shown in Fig. 4(b) (solid line) as input. The GPS measurements of the position of the ship were used as r_1 (for this simulation, $r_1 = r_2$ because the transmitter and receiver were co-located).

The seabed parameters for the Malta Plateau were taken from Siderius *et al.*[12] and are summarized in Table I. Mode calculations were performed using KRAKEN. The remaining task to simulate the spectrogram was to use a specific model for the scattering matrix S in Eq. (6). Due to the lack of knowledge on the scattering properties of the targets, two different scattering models for S_{mn} were evaluated: First, $S_{mn}=\delta(m-n)$ (where δ is the Kronecker operator) can model a flat surface in which no intermode coupling takes place; second, a Lambertian scattering matrix which is more appropriate to model a rough surface and is defined as

$$S_{mn} = \sqrt{\sin(\alpha_m)\sin(\alpha_n)}, \tag{11}$$

where α_m and α_n are the grazing angle of the incoming mth mode and the scattered nth mode.

Figure 8 shows the simulated spectrograms obtained through the procedure detailed in Sec. II. Both simulations exhibit the same striation patterns as observed in the experimental data and the main difference is on the number of striations. Note that the simulation results were computed with minimum knowledge of environmental parameters and no knowledge of the scattering properties of the target utilizing simplified forms for S_{mn}, and the effects of reverberation or the bathymetry were not included. This suggests a potential use to increase the robustness of the time–frequency structure with respect to perturbations in environmental parameters.

Figure 9 is another spectrogram comparison between data and simulation. It shows the case when the range between source and target is monotonically decreasing as the ship is approaching Wreck 1 in Fig. 4(a). In this spectrogram the striations are oriented along straight lines, and the similarity of this spectrogram with the simulation for the Lambertian scattering matrix is apparent. The value of the active "invariant" was calculated for the striation marked with an arrow in Fig. 9(a), for $r=16\,030$ m, $\Delta r=-2290$ m, $\omega=642.5$ Hz, and $\Delta\omega=-132.5$ Hz, resulting in $\beta_{\text{active}}=1.4$.

Figure 10(a) is another example of a curved trajectory for Wreck 2. In this track, the GPS tracking was incomplete so instead of using GPS as the input for the simulation of the active spectrogram, the range estimated by the sonar was utilized. The data drops at $t=27$, 39, 72, and 102 were ignored and correspond to the black horizontal lines in the simulation of Fig. 10(b).

One challenge for the analysis of the active spectrograms is the automatic detection of the striations. A quantitative method could be the use of the Radon transform to identify straight patterns in a picture. Bright spots in the Radon domain correspond to straight lines in the picture under analysis. Figure 11(a) shows the section of the spectrogram in Fig. 9 from $t=10$ to $t=50$ min. A Canny edge detector[13] with a standard deviation of 105 and the Radon transform were applied to this picture. In the Radon trans-

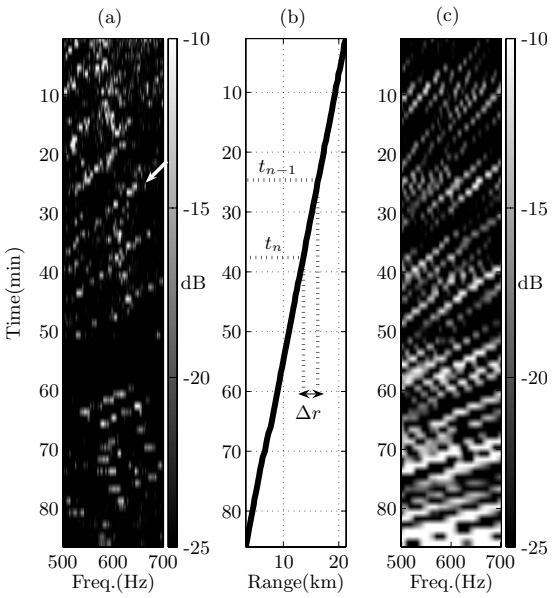


FIG. 9. (a) Bistatic spectrogram corresponding to the tracking of the Wreck 1 (Target N.9, track N.3 from NUWC database). The data were collected on 26 October 1994 at 20:03 GMT; (b) a monotonically decreasing range as a function of time results in striations with the same orientation; and (c) simulated spectrogram using a Lambertian scattering matrix and the sound speed profile in Fig. 4(b) (dotted line).

form domain of Fig. 11(b), the brightest spot corresponds to the straight striation indicated with an arrow in Fig. 11(a). There are other spots of high intensity, and this is due to the high level of noise in the experimental spectrogram. The extraction of features using the Radon transform is limited when the spectrogram is noisier, as in Fig. 7, and in those cases image denoising techniques should also be applied.

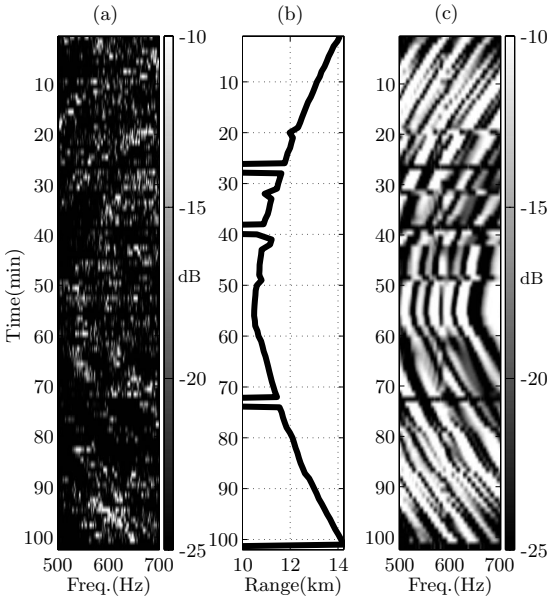


FIG. 10. (a) Bistatic spectrogram corresponding to the tracking of Wreck 2 (Target N.8, track N.23 from NUWC database). The data were collected on 26 October 1994 at 14:12 GMT; (b) the range was estimated by the sonar system instead of GPS; and (c) simulated bistatic spectrogram using a Lambertian scattering matrix and the sound speed profile in Fig. 4(b) (dashed line).

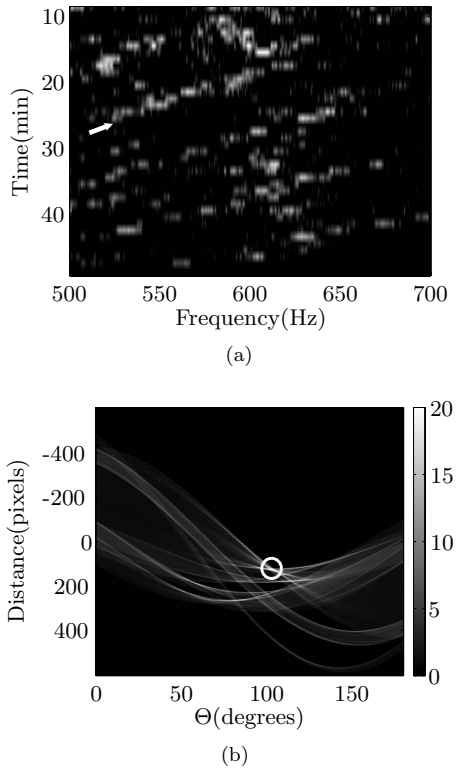


FIG. 11. (a) Straight striation in the spectrogram of Fig. 9, detected with the Radon transform and (b) Radon domain corresponding to the picture shown in Fig. 11(a).

For curved striations as in Figs. 7 and 10, methods like the Hough transform can be utilized to extract the value of the invariant.[14] An approach based on this method is to compute a family of theoretical striations corresponding to different values of β using the GPS measurement of range. Then, the task is to find the theoretical striation that matches the experimental spectrogram more closely. For example, a line integral along the theoretical striation can be performed to find the value of β that yields the largest integral. Another option is to pick the value of β that minimizes the variance along the theoretical striation, which corresponds to constant intensity striations.

An analytical expression for the active sonar invariant was derived by Zurk,[15] and the final expression defines the *bistatic invariant* γ as a quantity analogous to the passive invariant β, that depends on the scattering matrix of the target, in addition to the average value of the phase and group velocities. For some geometries, the bistatic invariant can be reduced to an expression similar to Eq. (5) for passive sonar. For example, for the monostatic case where $r_1=r_2$, the analytical expression reduces to $\gamma=\beta$. A question for future research is on the stability of γ for a truly bistatic geometry with respect to the scattering properties of the target. This will have to be addressed with a different experimental geometry and perhaps initially considering targets of simple geometries (spheres, cylinders).

It is interesting to note that the value of $\beta_{\text{active}}=1.4$ for the waveguide invariant found in the spectrogram of Fig. 9 yields the best match between the model and the data. That optimal performance occurs for something other than the ca-

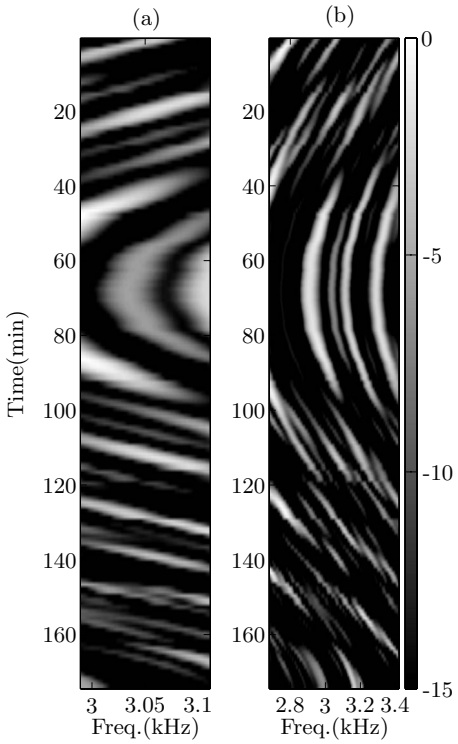


FIG. 12. Simulation of the spectrogram in Fig. 7 for a midfrequency of 3.05 kHz with two different bandwidths: 120 and 740 Hz. In both cases, Lambertian scattering matrix was used.

nonical shallow water value of 1.0 can be explained by the local oceanographic conditions. The canonical value is derived assuming an isovelocity water column. The sound speed profile shown in Fig. 4(b) is depth dependent; a mixed layer of warm, acoustically fast water overlies cooler, slower water. For the present case, with both the acoustic source and receiver below the mixed layer, a value of $\beta > 1$ should be expected as has been previously measured[16] and predicted.[3]

As mentioned before, the striations observed in the low frequency SWAC experimental data were not observed in the midfrequency data, which is pertinent to active sonar systems with frequencies from 3 to 10 kHz. Even if the scattering matrix of the target is independent of frequency, the frequency band can still affect the visualization of the spectrogram because the aspect ratio of the image is proportional to the center frequency. By defining the center frequency for low and midfrequencies ($f_{\text{low}}, f_{\text{mid}}$) and the corresponding bandwidth Δf_{low} and Δf_{mid}, it is observed from Eq. (5) that if β, r, and Δr do not change, then the ratio of the center frequency to the bandwidth must satisfy

$$\frac{f_{\text{low}}}{\Delta f_{\text{low}}} = \frac{f_{\text{mid}}}{\Delta f_{\text{mid}}}. \tag{12}$$

Figure 12 is a simulation of the spectrogram shown in Fig. 7 with a center frequency $f_{\text{mid}} = 3.05$ kHz (instead of $f_{\text{low}} = 495$ Hz) and the same bandwidth $\Delta f_{\text{low}} = 120$ Hz. From Fig. 12 it can be understood that for smaller bandwidths the striations become more horizontal and measurement of Δr becomes difficult. By increasing the bandwidth to $\Delta f_{\text{mid}} = 740$ Hz according to Eq. (12) the slope of the striations is preserved.

There is also a compromise between the ping rate and the maximum speed of the target. If the horizontal range between source and receiver is changing too fast or the ping rate is slow, the resulting spectrogram is a "subsampled" version that is hard to visualize because the points connecting a single striation would be sparse.

V. APPLICATION OF THE BISTATIC "INVARIANT"

The observations provided here motivate examination of the bistatic invariant as a relationship to improve tracking algorithms. It has been suggested by LePage[17] and Goldhahn *et al.*[18] that the bottom reverberation exhibits a time–frequency structure, and this has been applied to simulated data for the detection of targets embedded in strong reverberation.[18] In this paper, the experimental active sonar data show that there is a time–frequency structure corresponding to the target and if this relation can be described by an invariant quantity, it can be used as a constraint for target tracking algorithms. The bistatic "invariant" γ describes the slope of the striations in the bistatic spectrograms

$$\gamma = \frac{r(t_{n-1})\Delta\omega}{\omega(t_{n-1})\Delta r}, \tag{13}$$

where $\Delta\omega = \omega(t_n) - \omega(t_{n-1})$ is the difference between the frequency of two points along a chosen striation and $\Delta r = r(t_n) - r(t_{n-1})$ is the change in the bistatic range, as depicted in Fig. 9. For a tracking formulation, $\omega(t_{n-1})$ and $r(t_{n-1})$ are known at any time t_n except for $n=1$. Thus, the invariant provides a relation between instantaneous range and frequency content of a target return. As described by Quijano and Zurk,[19] this can be used with an extended Kalman filter to further constrain allowable target kinematics and to improve the tracker performance.

VI. CONCLUSION

A time–frequency structure similar to the invariant for passive sonar has been demonstrated for active sonar and tested in a monostatic configuration. The agreement of experimental data to simulations of active sonar spectrograms using limited information about the environment suggests that the slope of striations may be robust to details of the propagation environment or the scattering of the target. The spectrograms are easily obtained by analyzing the spectrum of the returns corresponding to the intended target.

The results presented in this paper correspond to low frequency data and monostatic geometry. In the LF SWAC data, multiple cases were found to clearly show the invariant structure, and this structure agreed with the trajectory obtained from GPS measurements. This provides strong evidence that this relationship is present for active sonar. Furthermore, normal mode simulations accurately produced the observed structure indicating that the phenomena is captured by the invariance of the modal structure. The MF data suffered from an inadequate ping rate to provide sufficient resolution to observe the invariant structure.

A topic of future research is the use of the time–frequency structure as a constraint equation in the state vector of a Kalman filter for active sonar trackers.

ACKNOWLEDGMENT

This project has been sponsored by the Office of Naval Research (ONR), grant No. N000140510886.

[1]S. D. Chuprov, "Interference structure of a sound field in a layered ocean," in *Ocean Acoustics. Current State*, edited by L. M. Brekhovskikh and I. B. Andreeva (Nauka, Moscow, 1982), pp. 71–91.
[2]L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics*, 3rd ed. (Springer, New York, 2002).
[3]D. Rouseff and C. V. Leigh, "Using the waveguide invariant to analyze lofargrams," in Proceedings of Oceans '02, Biloxi, MS, October 2002, pp. IV.2239–IV.2243.
[4]T. C. Yang, "Beam intensity striations and applications," J. Acoust. Soc. Am. **113**, 1342–1352 (2003).
[5]A. Thode, "Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory," J. Acoust. Soc. Am. **108**, 1582–1594 (2000).
[6]L. M. Zurk and B. H. Tracey, "Depth-shifting of shallow water waveguide source observations," J. Acoust. Soc. Am. **118**, 2224–2233 (2005).
[7]G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[8]F. Ingenito, "Scattering from an object in a stratified medium," J. Acoust. Soc. Am. **82**, 2051–2059 (1987).
[9]M. Porter, "The KRAKEN normal mode program," www.hlsresearch.com/oalib/ (version as read on 21 May, 2007).
[10]W. J. Comeau and W. V. Petersen, *False Clutter Database for the Characterization and Reduction of Active False Tracks (CRAFT)* (Naval Undersea Warfare Center Division, Newport, RI, 2004).
[11]D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*, 1st ed. (Artech House, Boston, 2005).
[12]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).
[13]R. Gonzalez and R. Woods, *Digital Image Processing*, 1st ed. (Prentice Hall, Englewood Cliffs, NJ, 2002).
[14]A. Turgut, "Geoacoustic inversion by using broadband ship noise recorded on the New Jersey Shelf (A)," J. Acoust. Soc. Am. **118**, 1857 (2005).
[15]L. M. Zurk, D. Rouseff, J. E. Quijano, and G. Greenwood, "Bistatic invariance principle for active sonar geometries," in Proceedings of the Eighth European Conference on Underwater Acoustics '06, Carvoeiro, Portugal, June 2006, pp. 787–791.
[16]W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, "A long-range and variable focus phase-conjugation experiment in a shallow water," J. Acoust. Soc. Am. **105**, 1597–1604 (1999).
[17]K. D. LePage, "Statistics of broad-band bottom reverberation predictions in shallow-water waveguides," IEEE J. Ocean. Eng. **29**, 330–346 (2004).
[18]R. Goldhahn, G. Hickman, and J. L. Krolik, "Waveguide invariant reverberation mitigation for active sonar," in IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, HI, April 2007, pp. II-941–II-944.
[19]J. E. Quijano and L. M. Zurk, "Use of the invariance principle for target tracking in active sonar geometries," in Proceedings of Oceans '06, Boston, MA, September 2006, pp. 1–5.

Waveguide invariant focusing for broadband beamforming in an oceanic waveguide

Hailiang Tao[a] and Jeffrey L. Krolik[b]
Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708

(Received 2 April 2007; revised 13 December 2007; accepted 14 December 2007)

The performance of broadband sonar array processing can degrade significantly in shallow-water environments when interference becomes angularly spread due to multipath propagation. Particularly for towed line arrays near endfire, elevation angle spreading of multipath interference often results in masking of weaker sources of interest. While adaptive beamforming in a series of narrow frequency bands can suppress coherent multipath interference, this approach requires long observation times to estimate the required narrowband covariance matrices. To form wideband covariance matrices which can be estimated with less observation time, plane-wave focusing methods have been used to avoid interference covariance matrix rank inflation. This paper extends wideband focusing to the case of coherent multipath interference. The approach presented here, called waveguide invariant focusing (WIF), exploits a robust relationship for the frequency dependence of horizontal wave number differences. Unlike matched-field methods, WIF does not model multipath wave fronts but rather makes the interference appear to occupy the same rank-one subspace across frequency. This permits formation of wideband covariance matrices without interference rank inflation. Simulation experiments in a realistic ocean environment indicate that adaptive beamforming using WIF covariance matrices can provide a significant array gain improvement over conventional adaptive methods with limited observation time. [DOI: 10.1121/1.2832335]

PACS number(s): 43.30.Wi, 43.60.Fg, 43.60.Gk, 43.30.Bp [RCG] Pages: 1338–1346

I. INTRODUCTION

This paper addresses the problem of suppressing moving endfire interference incident on a horizontal towed array in a shallow-water acoustic channel. Masking of targets by endfire interference frequently occurs when multipath spread in elevation angle appears as spread in "slant" azimuth on a horizontal array. Mitigation of endfire interference is especially important in passive sonar applications since often the first detection of a target with a towed passive array occurs toward forward endfire despite better bearing resolution at broadside. Although adaptive beamforming (ABF) can mitigate endfire interference, it requires a long observation time in which relative stationarity is maintained. This is a requirement that may be precluded in dynamic environments.[1] Alternatively, matched-field processing (MFP) has been proposed for tow-ship noise suppression, wherein a null is steered in the direction of the interference wave front as predicted using a full-field acoustic propagation model.[2] As with other MFP methods, however, such an approach can be quite sensitive to mismatch in the assumed geoacoustic environmental parameters used in the acoustic model.

For broadband interference, processing using a larger observation time-bandwidth product could, in theory, improve adaptive beamformer convergence rates. However, such a reduction in observation time is not achieved by simply incoherently averaging the output of a set of narrowband adaptive beamformers. The reason is that since narrowband ABF weights are computed at each frequency, the number of adaptive degrees of freedom grows just as fast as the interference bandwidth. Alternatively, forming a broadband data covariance matrix by averaging narrowband snapshot outer products and then computing a single set of adaptive weights achieves faster convergence but is often seriously limited by covariance matrix rank inflation which reduces the expected interference suppression. Broadband covariance matrix rank inflation is especially problematic in the presence of endfire interference which is spread due to multipath. An approach for broadband ABF which avoids rank inflation in the case of plane-wave sources is so-called wideband "focusing."[3] The idea here is to preprocess the sensor outputs so that the interference occupies the same signal subspace at each frequency. Forming a covariance matrix by averaging outer products of such transformed snapshots ideally maintains the interference in a rank-one subspace across the band. The result is that a much smaller number of adaptive degrees of freedom are needed to null the broadband interference with weights which can be computed using shorter observation times. Among wideband focusing methods, the most popular are the coherent signal-subspace (CSS) method[4] and the steered covariance method (STCM).[5] A limitation of the CSS method is that it requires a preliminary estimate of target directions, which could introduce source location bias. The STCM method avoids the need to make an initial estimation of target directions but is more computationally intensive. STCM achieves focusing by inserting delays in the sensor channels which correspond to each steering direction. A broadband target on the steering direction will have signal subspaces aligned to spatial dc for each temporal frequency.

[a]Electronic mail: hailiang.tao@duke.edu
[b]Electronic mail: jk@ee.duke.edu

For plane-wave sources, both CSS and STCM methods, as well as other broadband focusing techniques, have been shown to achieve a reduction of the threshold observation time required for adaptive interference suppression. Extension of focusing techniques to nonplanar coherent multipath arrivals, however, has not previously been achieved and is thus the subject of this paper.

The difficulty in finding a focusing transformation for shallow-water propagation arises from the interference among constituent multipath arrivals from the same source. In this paper, these "crinkly" wave fronts are modeled as plane waves with an unknown complex frequency-dependent spatial modulation. This spatial modulation depends upon the differences of modal horizontal wave numbers which characterize the propagation channel. These wave number differences, which are critical to matched field techniques, are notoriously difficult to predict. However, a robust theory for the *frequency dependence* of these wave number differences does exist and exploits a property described by the so-called waveguide invariant.[6]

The waveguide invariant is a channel parameter describing the dispersive characteristics of a stratified oceanic waveguide in terms of the ratio of phase-slowness differences to group-slowness differences. Remarkably, as noted by several authors,[6,7] the waveguide invariant is relatively constant with respect to variations within a given class of underwater acoustic waveguides. For example, in a Pekeris waveguide, the waveguide invariant, β, is unity but has been shown to vary only slightly in more realistic downward refracting channels.[7] In previous work, waveguide invariant theory has been used to explain the time-frequency spectrogram intensity surface of a moving broadband acoustic source, where constant intensity lines, known as striations, exhibit nearly constant slope.[8–13] In these papers, the waveguide invariant parameter β is defined to obtain a simple relationship between the frequency and range of these striations. In addition, many authors have used this relationship to estimate source parameters from time-frequency spectrograms. For example, Thode[9] achieved source ranging at different times using a vertical array and signal samples from a second broadband source at known location. Thode *et al.*[10] found that the range of Bartlett matched field processing sidelobes could be predicted in the range-frequency plane using waveguide invariant theory. This in turn was exploited to improve the source location performance of Bartlett MFP beamforming. Kim *et al.*[11] used the waveguide invariant to achieve robust time reversal focusing. Kuperman *et al.*[14] used waveguide invariant to mitigate fluctuation effects through adaptive processing. Zurk[15,16] and Yang[12] used the waveguide invariant to combat source motion in adaptive horizontal line array processing. Yang[13] further analyzed the beam outputs of a horizontal line array obtained by conventional beamforming and characterized the waveguide invariant striations in the beam output. Methods operating in the time-frequency spectrogram intensity domain were then developed to suppress interference based on the difference between target and interference striation slopes.

In contrast to previous applications of the waveguide invariant which operate in the intensity domain, in this paper, we use the waveguide invariant to achieve focusing transformations which operate on *complex* narrowband array snapshots containing broadband multipath endfire interference impinging on a long horizontal array. In particular, we exploit the relationship between the waveguide invariant and the frequency dependence of horizontal wave number differences to obtain waveguide invariant focusing (WIF), which achieves broadband focusing in the same sense as previously described plane-wave focusing techniques. A unique feature of WIF that distinguishes it from MFP is that it does not attempt to model the multipath wave front at any frequency. Rather the goal is to make endfire interference appear to occupy the same rank-one subspace across frequency without specifying the focused interference wave front. This is consistent with what ABF methods require, i.e., as long as the interference remains low rank, its details can be learned by the adaptive beamformer and subsequently suppressed. We apply WIF to a broadband plane-wave beamforming scenario where a strong multipath endfire interference masks a significant portion of bearing space. We show via simulation experiments that WIF can provide significant array gain improvements over conventional methods with limited observation time, thus increasing the usable bearing space fraction.

The remainder of the paper is organized as follows. Section II introduces the signal model and some conventional methods. In Sec. III, waveguide invariant focusing is proposed and its application in endfire interference is discussed. Section IV describes some simulation results. Finally, the conclusion is provided in Sec. V.

II. MODELING AND CONVENTIONAL METHODS

We consider a long horizontal array with p equally spaced sensors in a multipath shallow water ocean channel. The long array has an aperture length of more than 30 wavelengths, which has enough resolution capability to resolve targets near endfire directions, and also begins to resolve wave number spreading.[1] Assume a loud broadband interference in the endfire direction is moving along the array axis at constant speed v, as shown in Fig. 1. The objective is to use adaptive beamforming methods to detect targets at azimuths near the endfire direction. As will be explained later, because of the multipath spreading and motion, the energy of the endfire interference will appear in off-endfire directions (up to $\pm 40°$). When the interference is loud, with a received signal-to-interference ratio (SIR) below -20 dB, targets within such azimuth range are commonly masked by the interference.

The received signal is modeled as a superposition of the multipath interference and the weaker target signal, as well as additive white noise. Consider a limited observation time consisting of a total of N snapshots. The nth $p \times 1$ sensor snapshot at frequency ω_j can be expressed as

$$\mathbf{x}(\omega_j,n) = \mathbf{x}_I(\omega_j,n) + \mathbf{x}_T(\omega_j,n) + \mathbf{n}(\omega_j,n), \tag{1}$$

where $\mathbf{x}(\omega_j,n)=[x_1(\omega_j,n)x_2(\omega_j,n)\cdots x_p(\omega_j,n)]^T$ and $x_i(\omega_j,n)$ is the ith sensor output at discrete Fourier transform fre-

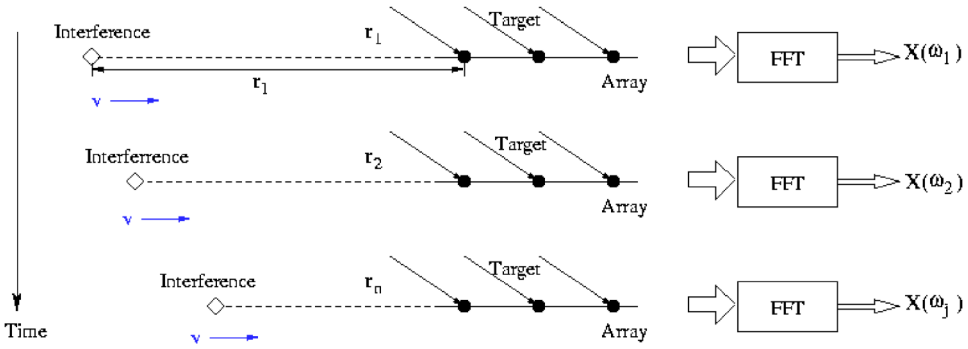


FIG. 1. (Color online) Broadband beamforming with a loud moving end-fire interference.

quency bin ω_j for snapshot n. The received data consist of components, $\mathbf{x}_I(\omega_j,n)$, the multipath interference, the target signal $\mathbf{x}_T(\omega_j,n)=s_T(\omega_j,n)\mathbf{v}_j$, and random noise $\mathbf{n}(\omega_j,n)$. The signal wave front $\mathbf{v}_j$ at frequency ω_j is assumed known, and the temporal Fourier coefficient $s_T(\omega_j,n)$ is treated as a complex Gaussian random process. The noise vector $\mathbf{n}(\omega_j,n)$ is assumed spatially white and uncorrelated across frequencies. The interference, target, and noise are assumed mutually uncorrelated.

In a shallow water waveguide, the multipath interference $\mathbf{x}_I(\omega_j,n)$ can be expressed by a normal mode expansion:[17]

$$\mathbf{x}_I(\omega_j,n) = s_I(\omega_j,n)\sum_m A_m(z_I,z_r)\exp(-jk_m(\omega_j)\mathbf{r}(n))$$

$$= s_I(\omega_j,n)\exp(-jk_1(\omega_j)\mathbf{r}(n))$$

$$\odot \sum_m A_m(z_1,z_r)\exp(-j(k_m-k_1)\mathbf{r}(n)), \qquad (2)$$

where $s_I(\omega_j,n)$ is the interference amplitude at frequency ω_j, $A_m(z_I,z_r)$ is the mth modal depth eigenfunction evaluated at interference depth z_I and receiver depth z_r, and k_m is horizontal wave number of the mth mode. The symbol $\odot$ denotes the element-by-element Hadamard product, and $\mathbf{r}(n)$ is the range vector such that $r_g(n)$=the horizontal distance from the interference to the gth sensor when obtaining snapshot n. The form in Eq. (2) expresses the interference wave front as a product of a plane-wave term and a spatial modulation function due to a coherent sum of multipath arrivals from different elevation angles. This expression will facilitate our later derivation of waveguide invariant focusing. Note this wave front is a function of time n, indicating that at given frequency ω_j it changes for every snapshot due to interference motion.

Two adaptive beamforming methods, one narrowband and one broadband, are compared with the proposed waveguide invariant focusing method under the model of Eq. (1). These methods are compared to WIF because: (1) the narrowband method has the same number of effective snapshots as WIF at its focusing frequency, and (2) the broadband method utilizes the same bandwidth as WIF. The details of this comparison will be discussed in Sec. III.

The narrowband method is the well-known minimum variance distortionless response (MVDR) beamformer. We process each frequency bin individually by forming outer products of $\mathbf{x}(\omega_j,n)$ to obtain the narrowband sample covariance matrix:

$$\hat{\mathbf{R}}_j = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(\omega_j,n)\mathbf{x}^H(\omega_j,n). \qquad (3)$$

MVDR weight vector is designed as

$$\mathbf{w}_{\mathrm{MVDR},j}(\theta) = \frac{\hat{\mathbf{R}}_j^{-1}\mathbf{v}_j(\theta)}{\mathbf{v}_j^H(\theta)\hat{\mathbf{R}}_j^{-1}\mathbf{v}_j(\theta)},$$

where $\mathbf{v}_j(\theta)$ is the target signal steering vector at frequency ω_j. The spatial spectrum output as a function of bearing θ can be expressed as

$$\hat{P}_{\mathrm{MVDR},j}(\theta) = \frac{1}{\mathbf{v}_j^H(\theta)\hat{\mathbf{R}}_j^{-1}\mathbf{v}_j(\theta)}. \qquad (4)$$

When there are an insufficient number of snapshots available to invert $\hat{\mathbf{R}}_j(N<p)$, an effective technique is to apply diagonal loading to $\hat{\mathbf{R}}_j$ to obtain $\hat{\mathbf{R}}_{j,\mathrm{DL}}$:

$$\hat{\mathbf{R}}_{j,\mathrm{DL}} = \hat{\mathbf{R}}_j + \delta^2\mathbf{I}$$

and use $\hat{\mathbf{R}}_{j,\mathrm{DL}}$ in place of $\hat{\mathbf{R}}_j$. The coefficient δ^2 is the diagonal loading factor, and its amount can be chosen using several techniques.[18,19]

It is natural to consider an "incoherent" broadband method based on the narrowband MVDR, i.e., simply averaging the outputs of each narrowband MVDR output across the bandwidth. In the problem we consider this approach does not offer an improvement over each narrowband output, and has less statistical stability compared with the following broadband method.[5]

A broadband covariance matrix combining all snapshots across the whole bandwidth has the potential to improve the statistical stability of Eq. (4). However, before doing so one needs to address the rank inflation problem caused by frequency variation of signal spatial signature. The STCM is designed to mitigate rank-inflation for plane-wave signals.[5] By inserting delays on the sensor outputs corresponding to the signal at each steering direction, STCM tries to align the plane-wave of each frequency to spatial dc, so that energy coming from one direction can be coherently summed up across the bandwidth. The sample covariance matrix of STCM as a function of bearing θ can be obtained by

$$\hat{\mathbf{R}}_{\mathrm{STCM}}(\theta)=\frac{1}{J}\sum_{j=1}^{J}\mathbf{T}_j^H(\theta)\hat{\mathbf{R}}_j\mathbf{T}_j(\theta). \tag{5}$$

The diagonal steering matrix $\mathbf{T}_j(\theta)$ is

$$\mathbf{T}_j(\theta)=\mathrm{diag}([e^{-j\omega_j\tau_1(\theta)}e^{-j\omega_j\tau_2(\theta)}\cdots e^{-j\omega_j\tau_p(\theta)}]), \tag{6}$$

where $\tau_1(\theta),\tau_2(\theta),\ldots,\tau_p(\theta)$ are wave propagation delays in each sensor channel. Because the poststeering signal wave front is dc in the STCM covariance matrix, the spatial spectrum output of STCM with MVDR beamforming is

$$\hat{P}_{\mathrm{STCM}}(\theta)=[\mathbf{1}^H\hat{\mathbf{R}}_{\mathrm{STCM}}(\theta)^{-1}\mathbf{1}]^{-1} \tag{7}$$

where $\mathbf{1}$ is a vector with each element being 1. The STCM method has been shown to reduce the threshold observation time required to suppress plane-wave interference, but its success depends on the knowledge of how wave fronts change over frequencies. Although such knowledge is readily applicable for plane-wave sources, it is usually hard to obtain in coherent multipath propagation environments. In Sec. III, however, we show that there exists a frequency dependence relationship applicable to multipath environments which does not require modeling the actual interference wave front.

III. WAVEGUIDE INVARIANT FOCUSING FOR ENDFIRE INTERFERENCE SUPPRESSION

Grachev[6] found that the waveguide invariant parameter β describes the frequency dependence of horizontal wave number differences $\Delta k_{mn}=k_m-k_n$ according to

$$\Delta k_{mn}=\zeta_{mn}\omega^{-1/\beta}, \tag{8}$$

in which ζ_{mn} is a constant indexed by the mth and nth mode number pair. Equation (8) suggests that the horizontal wave number difference depends on frequency only through the factor $\omega^{-1/\beta}$ for any combination of mode numbers. Moreover, the value of β is found to be approximately constant or invariant for a group of closely spaced modes with respect to frequency and mode wave number differences.[7]

To use Eq. (8) for wideband focusing, express the range vector in Eq. (2) from the array to the endfire interference as

$$\mathbf{r}(n)=r(1)+(n-1)v\Delta T+\Delta\mathbf{d}, \tag{9}$$

where $r(1)$ is the horizontal distance from the interference to a reference sensor when obtaining the first snapshot, ΔT is the time delay between adjacent snapshots, and $\Delta\mathbf{d}$ is the array element spacing vector along array axis measuring from the reference sensor, $\Delta\mathbf{d}=[\Delta d_1\Delta d_2\cdots\Delta d_p]^T$. In this paper, $r(1)$ is generally assumed known *a priori*, although the cases of ill-estimated or completely unknown $r(1)$ are discussed later. Substituting Eqs. (9) and (8) into Eq. (2) gives

$$\begin{aligned}\mathbf{x}_I(\omega_j,n)&=s_I(\omega_j,n)\exp(-jk_1(\omega_j)(r(1)+(n-1)v\Delta T\\&\quad+\Delta\mathbf{d}))\odot\sum_m A_m\exp(-j\zeta_{m1}\omega_j^{-1/\beta}\mathbf{r}(n))\\&=\tilde{s}_I(\omega_j,n)\exp(-jk_1(\omega_j)\Delta\mathbf{d})\odot\mathbf{b}(\omega_j,n),\end{aligned} \tag{10}$$

where

$$\mathbf{b}(\omega_j,n)=\sum_m A_m\exp(-j\zeta_{m1}\omega_j^{-1/\beta}\mathbf{r}(n)) \tag{11}$$

is now the interference direction vector "spatial modulation" at wave number k_1.

If we can align the frequency-dependent and snapshot-dependent vector $\mathbf{b}(\omega_j,n)$ into a common subspace, we can make the spatial modulation term the same, up to a complex scalar, at all temporal frequencies in the band. The problem then is reduced to dealing with the frequency variation of $\exp(-jk_1(\omega_j)\Delta\mathbf{d})$, which is very similar to the plane-wave case. From inspection of Eq. (11), we can accomplish this by first specifying a reference frequency ω_1 within the band, referred to as the focusing frequency, and then picking a frequency ω_n for subsequent snapshot n $(n=2,3,\ldots,N)$, to satisfy

$$\omega_n^{-1/\beta}\mathbf{r}(n)=\omega_1^{-1/\beta}\mathbf{r}(1), \tag{12}$$

so that when substituted into Eq. (11), $\mathbf{b}(\omega_j,n)=\mathbf{b}(\omega_1,1)$. This is a set of overdetermined linear equations, which generally cannot be satisfied exactly. To approximate Eq. (12), choose ω such that

$$\omega_n=\min_{\omega}\|\omega^{-1/\beta}\mathbf{r}(n)-\omega_1^{-1/\beta}\mathbf{r}(1)\|. \tag{13}$$

This optimization problem can be easily solved:

$$\omega_n=\left(\frac{\mathbf{r}^T(n)\mathbf{r}(n)}{\mathbf{r}^T(n)\mathbf{r}(1)}\right)^{\beta}\omega_1. \tag{14}$$

By substituting Eq. (12) into Eq. (11), clearly $\mathbf{b}(\omega_n,n)\cong\mathbf{b}(\omega_1,1)=\mathbf{b}$, so now we can rewrite the interference signal $\mathbf{x}_I(\omega_n,n)$ in Eq. (10) as

$$\mathbf{x}_I(\omega_n,n)\cong\tilde{s}_I(\omega_n,n)\exp(-jk_1(\omega_n)\Delta\mathbf{d})\odot\mathbf{b}. \tag{15}$$

Equation (15) shows that a judicious frequency offset removes the frequency dependence of the modulation term, but the resulting snapshots are still frequency dependent. To further reduce the dimensionality of the interference subspace, a presteering step similar to STCM can be employed. As with plane-wave beamforming, assuming knowledge of the first wave number k_1 at each frequency, we can perform the usual presteering to eliminate this frequency dependence, i.e.,

$$\hat{\mathbf{x}}_I(\omega_n,n)=\mathbf{T}_n\mathbf{x}_I(\omega_n,n)=\tilde{s}(\omega_n,n)\mathbf{b}, \tag{16}$$

where

$$\begin{aligned}\mathbf{T}_n&=\mathrm{diag}[\exp(jk_1(\omega_n)\Delta\mathbf{d})]\\&=\begin{bmatrix}e^{jk_1(\omega_n)\Delta d_1} & & \cdots & 0\\ 0 & e^{jk_1(\omega_n)\Delta d_2} & & \vdots\\ \vdots & & & 0\\ 0 & \cdots & & e^{jk_1(\omega_n)\Delta d_p}\end{bmatrix}.\end{aligned} \tag{17}$$

After frequency offset and presteering, the direction vector $\tilde{\mathbf{x}}_I(\omega_n,n)$ will span the same subspace at each frequency. Thus the broadband interference covariance matrix obtained by performing WIF on the interference wave front, i.e.,

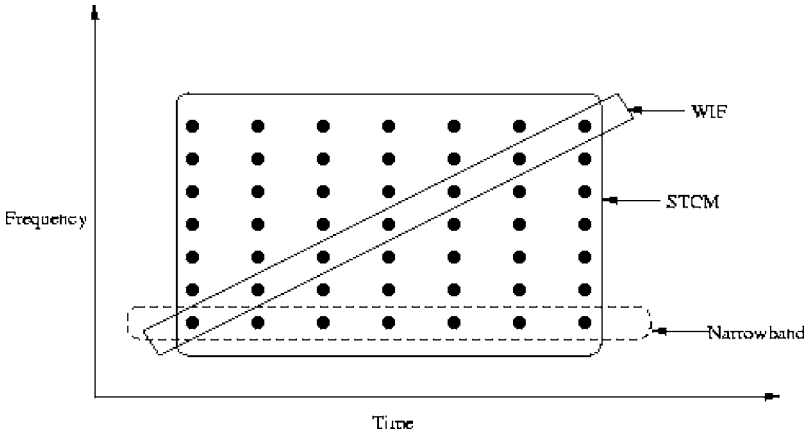


FIG. 2. Illustration of different snapshot subsets used by algorithms.

$$\hat{\mathbf{R}}_I = \frac{1}{N}\sum_{n=1}^{N} \tilde{\mathbf{x}}_I(\omega_n,n)\tilde{\mathbf{x}}_I^H(\omega_n,n) = \frac{1}{N}\sum_{n=1}^{N} |s_I(\omega_n,n)|^2 \mathbf{b}\mathbf{b}^H, \quad (18)$$

will be approximately rank one.

The snapshot subsets of WIF, STCM, and narrowband MVDR are compared in Fig. 2. In Fig. 2, each black dot represents a snapshot obtained at a particular time and frequency. It can be seen that STCM uses snapshots across all time and all frequencies. Narrowband methods use only snapshots from one frequency. Contrary to both of these methods, WIF picks the frequency of each snapshot as a function of time according to Eq. (14). It is worthwhile pointing out that both incoherent narrowband methods and WIF can utilize all snapshots available to STCM. For narrowband methods, we can incoherently average their outputs, but as mentioned earlier this does not improve on suppressing the moving multipath interference. For WIF, we can also incoherently average multiple WIF outputs with different focusing frequencies.

To use WIF for beamforming, the impact of frequency offset and presteering on the target signal must also be accounted for. Assuming the target wave front is known, without loss of generality, consider the case of a plane-wave signal. After the focusing transformation, its direction vectors across frequency becomes

$$\tilde{\mathbf{x}}_T(\omega_n,n) = \tilde{s}_T(\omega_n,n)\exp(-j(k_p(\omega_n)\sin\theta_T - k_1(\omega_n))\Delta\mathbf{d}), \quad (19)$$

where $\tilde{s}_T(\omega_n,n)$ is target signal amplitude at frequency ω_n, $k_p(\omega_n)$ is the target wave number, and θ_T is target bearing. Thus after focusing, the received signal (1) becomes

$$\tilde{\mathbf{x}}(\omega_n,n) = \tilde{s}_T(\omega_n,n)\exp(-j(k_p(\omega_n)\sin\theta_T - k_1(\omega_n))\Delta\mathbf{d}) + \tilde{s}_I(\omega_n,n)\mathbf{b} + \hat{\mathbf{n}}(\omega_n,n). \quad (20)$$

With the above-presented transformation, a broadband covariance matrix, to be subsequently used for computing adaptive beamformer weights, may be formed. The focused broadband covariance matrix is defined as

$$\hat{\mathbf{R}}_f = \frac{1}{N}\sum_{n=1}^{N} \tilde{\mathbf{x}}(\omega_n,n)\tilde{\mathbf{x}}(\omega_n,n)^H = \hat{\mathbf{R}}_I + \hat{\mathbf{R}}_T + \sigma_n^2\mathbf{I}, \quad (21)$$

where $\hat{\mathbf{R}}_I$ is the interference covariance matrix expressed in Eq. (18), which is made rank-one by WIF. The noise covariance matrix is $\sigma_n^2\mathbf{I}$ based on the fact that noise is uncorrelated across frequency, and the presteering step does not alter the noise covariance either, as can be easily shown. $\hat{\mathbf{R}}_T$ is the target covariance matrix:

$$\hat{\mathbf{R}}_T = \frac{1}{N}\sum_{n=1}^{N} |s_T(\omega_n,n)|^2 \tilde{\mathbf{v}}(\omega_n,\theta_T)\tilde{\mathbf{v}}(\omega_n,\theta_T)^H, \quad (22)$$

where

$$\tilde{\mathbf{v}}(\omega_n,\theta) = \exp(-j(k_p(\omega_n)\sin\theta - k_1(\omega_n))\Delta\mathbf{d}) \quad (23)$$

is the postfocusing target signal direction vector. In Eq. (22) only one plane-wave target is considered. If there are Q plane-wave target sources, $\hat{\mathbf{R}}_T$ can be written as

$$\hat{\mathbf{R}}_T = \frac{1}{N}\sum_{n=1}^{N}\sum_{q=1}^{Q} |s_T(\omega_n,q)|^2 \tilde{\mathbf{v}}(\omega_n,\theta_T(q))\tilde{\mathbf{v}}(\omega_n,\theta_T(q))^H \quad (24)$$

and the same direction vector of Eq. (23) will be used to scan across bearing as with conventional beamforming. Note that although WIF induces some rank inflation of the plane-wave target signal, this is typically small for the fractional bandwidth considered [see, e.g., Eq. (27)].

Now that we have a focused broadband covariance matrix, better interference suppression can be obtained. As the target signal has different direction vectors for different frequencies, MVDR weights are designed for each frequency (and for each snapshot):

$$\mathbf{w}_n(\theta) = \frac{\hat{\mathbf{R}}_f^{-1}\tilde{\mathbf{v}}(\omega_n,\theta)}{\tilde{\mathbf{v}}(\omega_n,\theta)^H\hat{\mathbf{R}}_f^{-1}\tilde{\mathbf{v}}(\omega_n,\theta)}. \quad (25)$$

If there is an insufficient number of snapshots to ensure $\hat{\mathbf{R}}_f$ nonsingular, we apply diagonal loading as in the narrowband case. The spatial spectrum output of WIF-MVDR is then:

$$P_{\mathrm{WIF}}(\theta) = \frac{1}{N}\sum_{n=1}^{N} |\mathbf{w}_n^H(\theta)\tilde{\mathbf{x}}(\omega_n,n)|^2. \quad (26)$$

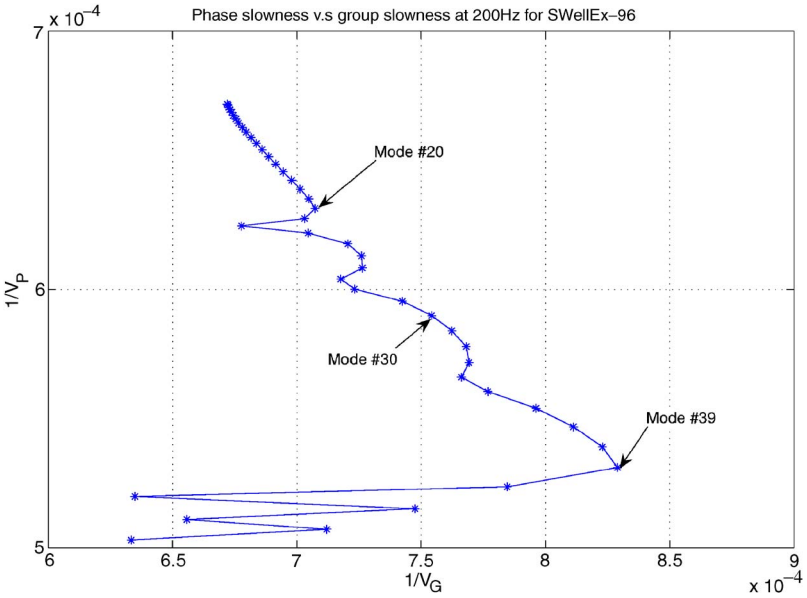


FIG. 3. (Color online) Group slowness vs phase slowness for SWellEx-96 environment at 200 Hz. The waveguide invariant is defined as the slope of this curve.

IV. SIMULATIONS

A. The shallow water channel

To compare the performance of WIF versus conventional methods, simulations were performed assuming the SWELLEX-96 shallow-water environment off the San Diego coast.[20,21] The reader is referred to the references for the details of this shallow water channel. Figure 3 shows the phase slowness $(1/v_p)$ versus group slowness $(1/v_g)$ at 200 Hz for all the propagating modes as computed by the KRAKEN normal mode model.[22] This environment consists of a downward refracting sound-speed profile in a 200-m-deep approximately range-invariant ocean. In Fig. 3 the slope of the curve corresponds to the value of waveguide invariant. For the first 20 modes, β remains almost at a constant value of 1.1. Modes 21–39 appear to have three different propagation groups, each of which has a different β value. The last six modes, however, exhibit irregular slope values, probably due to strong bottom interactions. In the following simulations we assume $\beta=1.1$, because higher order modes exhibit strong attenuation ensuring the low order modes are most dominant.

Another issue in applying waveguide invariants is the bandwidth over which β is approximately constant. The bandwidth cannot be so large that the relationship between $1/v_p$ and $1/v_g$ changes dramatically. Moreover, as bandwidth becomes large, the mode amplitude A_m will no longer be relatively constant across the entire band, which is an underlying assumption of waveguide invariant theory. Following a rule of thumb,[9] we form focused covariance matrices over a band such that

$$\alpha = \frac{\Delta\omega}{\omega_c} \leqslant 0.1, \tag{27}$$

where ω_c is the central frequency.

B. Results of endfire interference suppression using WIF

In all the following simulations, we construct acoustic fields using the environmental profile of Sec. IV A. A horizontal line array is towed at depth 40 m, with 150 elements equally spaced 1 m apart. A plane wave target signal impinges on the array at 18° with signal-to-noise ratio 0 dB. The endfire interference has an INR 10 dB. Its starting distance $r(1)$ is 3 km. The interference is moving away from the array at speed 5 m/s, at depth 21 m. The focusing frequency we chose is 300 Hz. Suppose we have 20 snapshots, one per 1 s, corresponding to total 20 s observation time. The calculation results of Eq. (14) show the picked frequency for snapshot 20 is 310.21 Hz, so the bandwidth satisfies Eq. (27). We assume a flat spectrum for both the target and the interference. All outputs are results averaged over 100 Monte Carlo simulations of the complex Gaussian FFT coefficients at all frequencies and all time instants.

Figure 4 compares the eigenspectrum of WIF broadband covariance matrix (21) with that of narrowband covariance matrix (3) at frequency 300 Hz. The eigenspectra plot their eigenvalues in descending order. Note that the WIF eigenspectrum decays much more rapidly than that of narrowband covariance matrix, and that the value of the dominant eigenvalue in the WIF spectrum is much bigger than other eigenvalues, making the WIF broadband covariance matrix approximately rank one. This is indicative of the focused interference wave fronts being in the same subspace.

Figure 5 shows the spatial spectrum output of WIF-MVDR compared with the other two methods. The narrowband output is obtained using Eq. (4) at frequency 300 Hz. The STCM output is computed according to Eq. (7). It is evident that only WIF-MVDR can suppress the endfire interference to reveal the weaker target at 18°. In the outputs of both narrowband MVDR and STCM, the power of the interference spreads across a wide azimuth space from 0° to 40°, completely masking the target. The multipath interference is not adequately suppressed because their covariance matrices

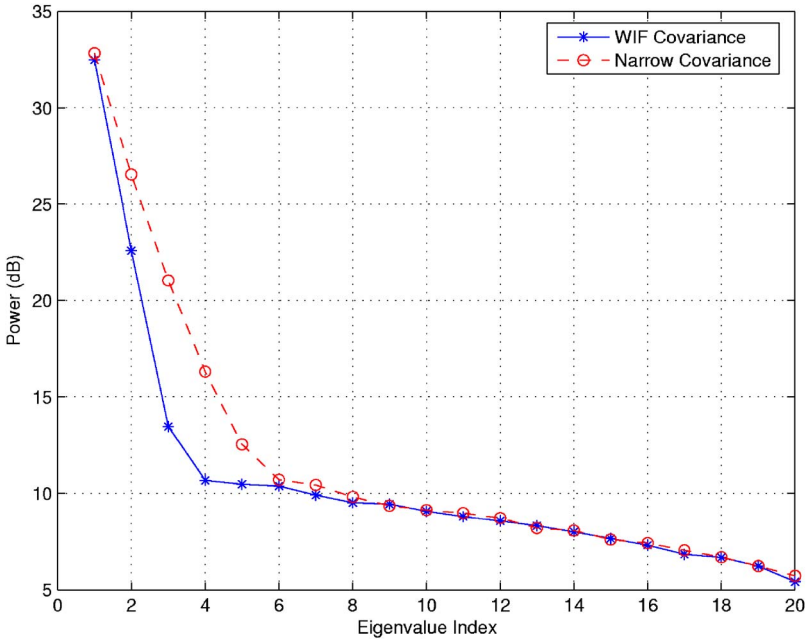


FIG. 4. (Color online) Comparison of the eigenspectra between WIF covariance matrix and narrow-band covariance matrix.

do not contain stable estimation of the time varying wave front of the moving interference, so that their inverse cannot place nulls to suppress the interference. The WIF broadband covariance matrix, in contrast, contains a rank-one interference subspace whose energy is coherently averaged over snapshots. This ensures beamforming weights place a null on the interference.

Array gain is an important performance measure of beamformer performance. The array gain is defined as the ratio of output signal to interference plus noise ratio (SINR) to input SINR at the steering direction. By the design of distortionless response adaptive beamformers, the output signal power is the same as the input signal power. The output interference plus noise power at frequency ω_j is

$$P^o_{\mathrm{IN}}(\omega_j) = \mathbf{w}_j^H \mathbf{R}_{\mathrm{IN}}(\omega_j)\mathbf{w}_j, \tag{28}$$

where $\mathbf{R}_{\mathrm{IN}}(\omega_j)$ is the interference plus noise covariance matrix at frequency ω_j. The array gain is thus calculated by

$$\mathrm{AG} = \frac{\sum_j P^i_{\mathrm{IN}}(\omega_j)}{\sum_j P^o_{\mathrm{IN}}(\omega_j)}. \tag{29}$$

Figure 6 shows the array gain for WIF-MVDR, narrowband MVDR, and STCM-MVDR for the same settings as Fig. 5. Figure 6 shows that WIF outperforms the narrowband method and STCM, especially in the bearing space near endfire. It is noted that the array gain of WIF-MVDR is a function of SNR as with most MV beamformers.[23]

Thus far we have assumed that the range of the interference is known. Now we investigate what happens when that knowledge is incorrect. Figure 7 shows mainlobe-sidelobe differences in the WIF spectrum output as a function of the interference range mismatch:

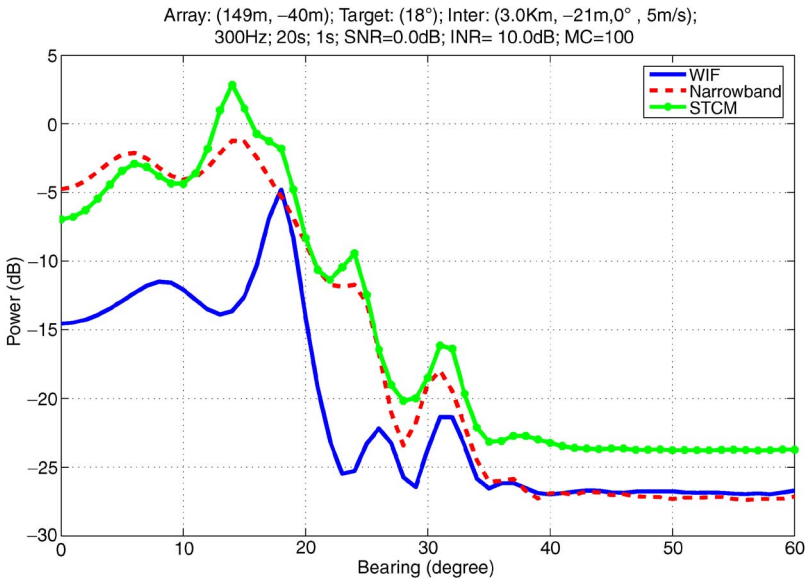


FIG. 5. (Color online) Spatial spectrum output of WIF-MVDR compared with other methods.

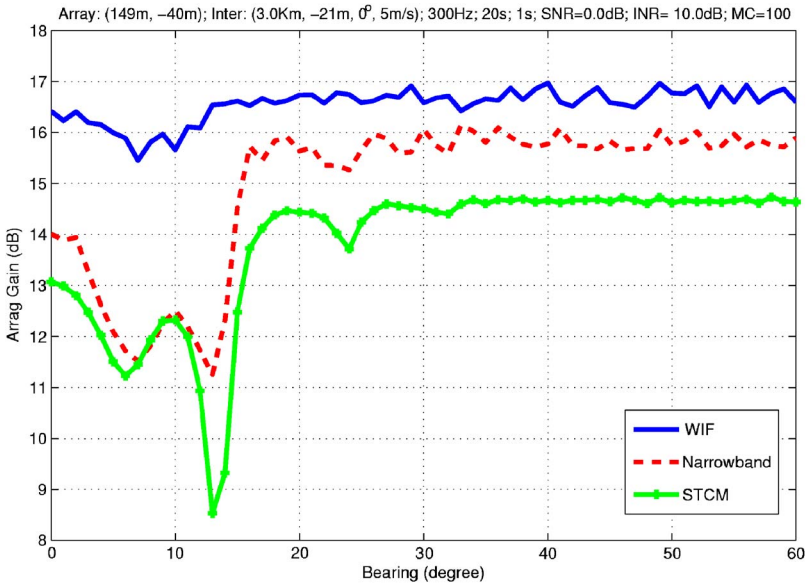


FIG. 6. (Color online) Array gain of WIF-MVDR compared with other methods.

$$r_o = r_a - r(1),$$

where r_a is the assumed range of the interference and $r(1)$ is the true range. It can be seen that WIF performs best at the correct range, and degrades at both smaller ranges and larger ranges. First, we notice that the performance within ± 500 m is relatively stable. Second, the degradation of negative r_o is much more severe than positive r_o. This is because the frequency offset is a decreasing function of the assumed range r_a. Smaller r_a leads to larger frequency offset, thus destroying the focusing operation more rapidly. The plot suggests having a positive r_o is better than a negative r_o. This does not mean we can choose arbitrarily large r_a: Larger r_a dictates a smaller frequency offset, thus requiring longer FFT time to obtain a better frequency domain interpolation.

In practice, a measurement of the interference range may not be available, and even the value of β may be unknown. However, the frequency offset can still be performed using the following method. Equation (15) represents the interference signal for chosen frequency ω_n. If we define a vector $\mathbf{x}_{I,m}(\omega_n, n)$ whose elements are the magnitude of corresponding elements of $\mathbf{x}_I(\omega_n, n)$, we find that $\mathbf{x}_{I,m}(\omega_n, n)$, $n = 1, 2, \ldots, N$, are in the same subspace. This suggests we can find ω_n by matching the subspace of $\mathbf{x}_{I,m}(\omega_n, n)$ and $\mathbf{x}_{I,m}(\omega_1, 1)$ without knowing the starting interference range $r(1)$ and the value of β. Although not pursued further in this paper, this approach merits future investigation.

V. CONCLUSION

This paper has proposed a new approach to achieve broadband focusing for coherent multipath signals frequently encountered in oceanic waveguide. The method, waveguide invariant focusing, is used to reduce the rank of an interference component in broadband covariance matrix estimation without the need to precisely model the channel. It can be used with a towed horizontal line array and essentially amounts to very judicious frequency-dependent averaging of array snapshots over time. In this paper, WIF was applied to focus multipath interference which has large angular spread

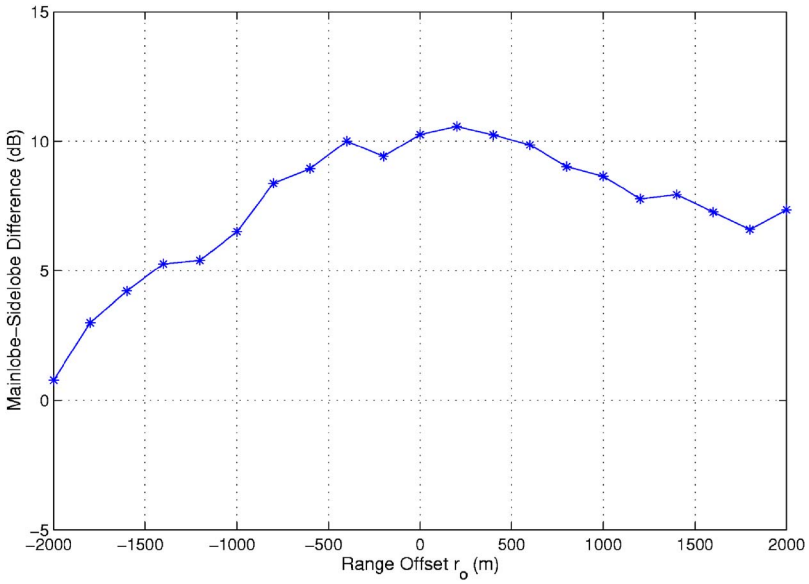


FIG. 7. (Color online) Range mismatch of WIF-MVDR.

at near endfire directions. It was shown to significantly improve interference suppression performance in a typical shallow water environment.

In future work, a beamspaee WIF-MVDR will be developed, which will focus on beams where the angular spread of endfire interference most possibly exists. Because the beamspace WIF-MVDR operates with reduced number of dimensions—the number of relevant beams is often a fraction of the number of array elements, a beamspace WIF approach should offer both further computational and statistical advantages.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research, Code 321US, under Grant No. N000140510023.

[1]A. B. Baggeroer and H. Cox, "Passive sonar limits upon nulling multiple moving ships with large aperture arrays," in *Proceedings of the 33rd IEEE Asilomar Conference on Signals, Systems, and Computers*, 1999, Vol. **1**, Nice, France, pp. 103–108.

[2]I. P. Kirsteins, J. M. F. Moura, J. W. Fay, and S. K. Mehta, "Ownship interference cancellation and source localization," in *Proceedings of IEEE OCEANS'98*, 1998, Nice, France, pp. 153–157.

[3]J. Krolik, "Focused wide-band array processing for spatial spectral estimation," in *Advances in Spectrum Analysis and Array Processing*, edited by S. Haykin (Prentice Hall, Englewood Cliffs, NJ, 1991), Vol. **II**, Chap. 6.

[4]H. Hung and M. Kaveh, "Focussing matrices for coherent signal-subspace processing," IEEE Trans. Acoust., Speech, Signal Process. **36**, 1272–1281 (1988).

[5]J. Krolik and D. Swingler, "Multiple broadband source location using steered covariance matrices," IEEE Trans. Acoust., Speech, Signal Process. **37**, 1481–1494 (1989).

[6]G. A. Grachev, "Theory of acoustic field invariants in layered waveguides," Acoust. Phys. **39**, 33–35 (1993).

[7]L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics* (Spinger, Berlin, 1991), Chap. 6, pp. 140–145.

[8]G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).

[9]A. M. Thode, "Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory," J. Acoust. Soc. Am. **108**, 1582–1594 (2000).

[10]A. M. Thode, W. A. Kuperman, G. L. D'Spain, and W. S. Hodgkiss, "Localization using Bartlett matched-field processor sidelobes," J. Acoust. Soc. Am. **107**, 278–286 (2000).

[11]S. Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, and T. Akal, "Robust time reversal focusing in the ocean," J. Acoust. Soc. Am. **114**, 145–157 (2003).

[12]T. C. Yang, "Motion compensation for adaptive horizontal line array processing," J. Acoust. Soc. Am. **113**, 245–260 (2003).

[13]T. C. Yang, "Beam intensity striations and applications," J. Acoust. Soc. Am. **113**, 1342–1352 (2003).

[14]W. A. Kuperman, S. Kim, G. F. Edelmann, W. S. Hodgkiss, H. C. Song, and T. Akal, "Group and phase speed analysis for predicting and mitigating the effects of fluctations," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance*, edited by N. Pace and F. Jensen (2002), Kluwer Academic Publisher, Dordredht, The Netherlands, pp. 279–286.

[15]L. M. Zurk, "Source motion compensation using waveguide invariant theory," J. Acoust. Soc. Am. **110**, 2717 (2001).

[16]L. M. Zurk, N. Lee, and J. Ward, "Source motion mitigation for adaptive matched field processing," J. Acoust. Soc. Am. **113**, 2719–2731 (2003).

[17]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, Woodbury, NY, 1994).

[18]Y. Kim, S. Pillai, and J. Guerci, "Optimal loading factor for minimal sample support space-time adaptive radar," in *Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, Seattle*, WA, 1998, pp. 2505–2508.

[19]N. Ma and J. Goh, "Efficient method to determine diagonal loading value," in *Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing*, 2003, Hong Kong, pp. 341–344.

[20]The SWellEx-96 Experiment website, URL http://www.mpl.ucsd.edu/swellex96/, 3 December 2007.

[21]G. L. D'Spain, W. S. Hodgkiss, J. J. Murray, P. W. Schey, and N. O. Booth, "Mirages in shallow water matched-field processing," J. Acoust. Soc. Am. **105**, 3245–3266 (1999).

[22]M. B. Porter, "The KRAKEN normal mode program," Memorandum No. SM-245, SACLANT Undersea Research Centre, La Spezia, Italy, 1991.

[23]H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," J. Acoust. Soc. Am. **54**, 771–785 (1973).

Efficient absorbing boundary conditions for Biot's equations in time-harmonic finite element applications

Reiner Wahl
Institut für Werkstoffkunde II, Universität, c/o Forschungszentrum Karlsruhe, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

Martin Spies[a)]
Fraunhofer Institut für Techno- und Wirtschaftsmathematik ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany

Stefan Diebels
Lehrstuhl für Technische Mechanik, Universität des Saarlandes, Geb. E6.3, 66123 Saarbrücken, Germany

(Received 1 March 2007; revised 5 November 2007; accepted 12 December 2007)

Absorbing boundary conditions for two phase media previously presented by Zerfa and Loret [Earthquake Eng. Struct. Dyn. **33**, 89–110 (2004)] have been improved by considering additionally absorbing waves with auxiliary angles of incidence. These angles are defined at each point on the boundaries, so one can easily implement tensor impedances as analogous to those defined by Krenk and Kirkegaard for isotropic, nonporous media [J. Sound Vib. **247**, 875–896 (2001)]. The boundary conditions have been tested and validated in two-dimensional frequency domain simulations.
 [DOI: 10.1121/1.2832325]

PACS number(s): 43.35.Ei, 43.20.Bi, 43.20.Gp, 43.35.Cg [LLT] Pages: 1347–1351

I. INTRODUCTION

The transient or harmonic analysis of wave propagation in porous media is of interest in many fields of research, e.g., in acoustical applications concerning soils, glacial ice, or biological materials such as bones (see, e.g., Refs. 1–6). An efficient description of these phenomena is possible by applying Biot's theory for poro-elastic and poro-viscoelastic media.[7–10] Although empirically found, its applicability has been verified in theory and experiment and therefore it is frequently used for these problems. As being available in specific commercial software, the finite element method (FEM) is often the computational method of choice. To limit the computational domain it is necessary to establish artificial boundaries whereas the waves normally would propagate to infinity in the real problem. These special boundary conditions have to guarantee that the waves leave the domain without any reflections. In using commercial software, local boundary conditions should be implemented on the basis of natural boundary conditions. Lysmer and Kuhlemeyer[11] were the first to develop such boundary conditions which worked as dampers at the border to absorb perpendicularly impinging waves. Krenk and Kirkegaard[12] improved those absorbing boundary conditions through their tensor radiation boundary conditions by additionally absorbing waves of oblique incidence. They developed "tensor impedances" to have their conditions working as viscous dampers. There are many other effective methods for this matter like the use of boundary elements,[13] but they are no longer local. A review for absorbing boundary conditions is given in Ref. 14.

The conditions already mentioned are applicable for solids. For simulations on the basis of Biot's theory we need an absorbing boundary condition for two phase media. Three different approaches have been made by Akiyoshi *et al.*,[15] Degrande and de Roeck,[16] and Zerfa and Loret,[17] which all resulted in a viscous traction-like formulation. Akiyoshi *et al.* fulfilled a paraxial approximation of order zero like Clayton and Engquist[18] did for homogeneous solids. Degrande and de Roeck divided the wave field into an incident and a reflected one and calculated the tractions for a zero reflection amplitude. Both neglected the second (slow) pressure wave. Zerfa and Loret evaluated the viscous tractions of an elastic wave field including this wave without the use of explicit equations of motion. They used the constitutive equations following Bowen's formulation[19,20] and ignored waves of oblique incidence.

In this article, these absorbing boundary conditions are improved by taking into account all angles of incidence. Knowing the source's coordinates allows one to determine the traveling direction of the waves and thus the angle of incidence at each point on the borders.

II. DERIVATION OF THE TENSOR RADIATION BOUNDARY CONDITION FOR SOILS (TRFS)

We start from Biot's equations of motion in the form

$$N\nabla^2\mathbf{u} + (A+N)\nabla\nabla\cdot\mathbf{u} + Q\nabla\nabla\cdot\mathbf{U} = \rho_{11}\ddot{\mathbf{u}} + \rho_{12}\ddot{\mathbf{U}} + b(\dot{\mathbf{u}} - \dot{\mathbf{U}}),$$

$$Q\nabla\nabla\cdot\mathbf{u} + R\nabla\nabla\cdot\mathbf{U} = \rho_{12}\ddot{\mathbf{u}} + \rho_{22}\ddot{\mathbf{U}} - b(\dot{\mathbf{u}} - \dot{\mathbf{U}}), \quad (1)$$

where A and N are the Lamé coefficients of the porous material, Q and R are Biot's additional constants for porous media (their meaning has been analyzed in Ref. 9), and $\mathbf{u}$ and $\mathbf{U}$ are the particle displacement vectors for the solid and the

a)Electronic mail: martin.spies@itwm.fraunhofer.de

fluid, respectively. Furthermore, b is a material-dependent constant and the coefficients ρ_{11}, ρ_{12}, ρ_{22} are related to the fluid and solid densities according to

$$(1-n)\rho_{\text{Solid}} = \rho_{11} + \rho_{12} = \rho_1,$$

$$n\rho_{\text{Fluid}} = \rho_{12} + \rho_{22} = \rho_2,$$

$$\rho_1 + \rho_2 = \rho, \tag{2}$$

where ρ designates the density of the composite.

In the following, we neglect Biot's inertia coupling term. Using the Helmholtz' potential decomposition, equation system (1) is split into two separate systems for the dilatational and the rotational part of the displacement field. Thus—as described in detail in Ref. 7—the dispersion relations can be solved separately in order to get the velocities of the three propagating waves which are the fast pressure wave $P1$, the slow pressure wave $P2$, and the shear wave S.

Decomposing the displacement field $(\mathbf{uU})^T$ and the stresses $(\boldsymbol{\sigma}^{\text{Solid}}\boldsymbol{\sigma}^{\text{Fluid}})^T$ into the different modes, and further considering the solid displacement causing stresses both in the solid and the fluid phase and vice versa, the following expressions are obtained:

$$\mathbf{u}(\mathbf{x},t) = \mathbf{u}^{P1}(\mathbf{x},t) + \mathbf{u}^{P2}(\mathbf{x},t) + \mathbf{u}^{S}(\mathbf{x},t),$$

$$\mathbf{U}(\mathbf{x},t) = \mathbf{U}^{P1}(\mathbf{x},t) + \mathbf{U}^{P2}(\mathbf{x},t),$$

$$\boldsymbol{\sigma}_{\text{Solid}} = (\boldsymbol{\sigma}^{P1}_{\text{Solid–Solid}} + \boldsymbol{\sigma}^{P1}_{\text{Fluid–Solid}}) + (\boldsymbol{\sigma}^{P2}_{\text{Solid–Solid}} + \boldsymbol{\sigma}^{P2}_{\text{Fluid–Solid}}) + \boldsymbol{\sigma}^{S}_{\text{Solid}},$$

$$\boldsymbol{\sigma}_{\text{Fluid}} = (\boldsymbol{\sigma}^{P1}_{\text{Fluid–Fluid}} + \boldsymbol{\sigma}^{P1}_{\text{Solid–Fluid}}) + (\boldsymbol{\sigma}^{P2}_{\text{Fluid–Fluid}} + \boldsymbol{\sigma}^{P2}_{\text{Solid–Fluid}}). \tag{3}$$

It was accounted for the fluid to be free of rotations; additionally, plane wave behavior is assumed for each displacement according to

$$\mathbf{u}^{P1/P2} = \hat{\mathbf{r}}u^{P1/P2}(r - v_{P1/P2}t),$$

$$\mathbf{U}^{P1/P2} = \hat{\mathbf{r}}U^{P1/P2}(r - v_{P1/P2}t),$$

$$\mathbf{u}^{S} = \hat{\mathbf{t}}u^{S}(r - v_S t). \tag{4}$$

Here $\hat{\mathbf{r}}$ is the propagation direction of the waves and $\hat{\mathbf{t}}$ is the direction perpendicular to $\hat{\mathbf{r}}$. Applying the plane wave formalism the corresponding strains can be derived according to

$$\boldsymbol{\epsilon}^{P1/P2}_{Solid} = \hat{\mathbf{r}}\hat{\mathbf{r}}\frac{\partial u^{P1/P2}}{\partial r} = -\frac{1}{v_{P1/P2}}\left(\frac{1}{3}\mathbf{I} + \left(\hat{\mathbf{r}}\hat{\mathbf{r}} - \frac{1}{3}\mathbf{I}\right)\right)\dot{u}^{P1/P2},$$

$$\boldsymbol{\epsilon}^{P1/P2}_{\text{Fluid}} = \hat{\mathbf{r}}\hat{\mathbf{r}}\frac{\partial U^{P1/P2}}{\partial r} = -\frac{1}{v_{P1/P2}}\left(\frac{1}{3}\mathbf{I} + \left(\hat{\mathbf{r}}\hat{\mathbf{r}} - \frac{1}{3}\mathbf{I}\right)\right)\dot{U}^{P1/P2},$$

$$\boldsymbol{\epsilon}^{S}_{\text{Solid}} = \frac{1}{2}(\hat{\mathbf{r}}\hat{\mathbf{t}} + \hat{\mathbf{t}}\hat{\mathbf{r}})\frac{\partial u^S}{\partial r} = -\frac{1}{2v_S}(\hat{\mathbf{r}}\hat{\mathbf{t}} + \hat{\mathbf{t}}\hat{\mathbf{r}})\dot{u}^S. \tag{5}$$

Here the strains were split into an isotropic and a deviatoric part, in order to directly relate them to the corresponding elastic constants; we thus obtain the stresses caused by the different waves as

$$\boldsymbol{\sigma}^{P1/P2}_{\text{Solid–Solid}} = -\frac{1}{v_{P1/P2}}(A\mathbf{I} + 2N\hat{\mathbf{r}}\hat{\mathbf{r}})\dot{u}^{P1/P2},$$

$$\boldsymbol{\sigma}^{P1/P2}_{\text{Solid–Fluid}} = -\frac{Q\mathbf{I}}{v_{P1/P2}}\dot{u}^{P1/P2},$$

$$\boldsymbol{\sigma}^{P1/P2}_{\text{Fluid–Solid}} = -\frac{Q\mathbf{I}}{v_{P1/P2}}\dot{U}^{P1/P2},$$

$$\boldsymbol{\sigma}^{P1/P2}_{\text{Fluid–Fluid}} = -\frac{R\mathbf{I}}{v_{P1/P2}}\dot{U}^{P1/P2},$$

$$\boldsymbol{\sigma}^{S}_{\text{Solid}} = -\frac{N}{v_S}(\hat{\mathbf{r}}\hat{\mathbf{t}} + \hat{\mathbf{t}}\hat{\mathbf{r}})\dot{u}^S, \tag{6}$$

where the Sommerfeld relation $(\partial/\partial t - \partial/\partial r)u = 0$ has been used. Now, it is possible to replace the stresses in the general assumption (3) by the expressions in (6) to formulate

$$\boldsymbol{\sigma}_{\text{Solid}} = -\frac{1}{v_{P1}}((A\mathbf{I} + 2N\hat{\mathbf{r}}\hat{\mathbf{r}})\dot{u}^{P1} + Q\hat{\mathbf{n}}\dot{U}^{P1}) - \frac{1}{v_{P2}}((A\mathbf{I} + 2N\hat{\mathbf{r}}\hat{\mathbf{r}})\dot{u}^{P2} + Q\mathbf{I}\dot{U}^{P2}) - \frac{N}{v_S}(\hat{\mathbf{r}}\hat{\mathbf{t}} + \hat{\mathbf{t}}\hat{\mathbf{r}})\dot{u}^S,$$

$$\boldsymbol{\sigma}_{\text{Fluid}} = -\frac{1}{v_{P1}}(R\mathbf{I}\dot{U}^{P1} + Q\mathbf{I}\dot{u}^{P1}) - \frac{1}{v_{P2}}(R\mathbf{I}\dot{U}^{P2} + Q\mathbf{I}\dot{u}^{P2}). \tag{7}$$

Since the numerical calculation depends on the general displacement field $(\mathbf{uU})^T$, we have to find a relation between the displacement velocities of the two pressure waves and the general displacement velocity. This relation is found through the eigenvectors of the two dilatational modes. In Ref. 17 the pressure wave was considered to be a linear combination of these modes and the splitting of the amplitudes was found to be

$$\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{Q}{\Delta} - \left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{Q}{\Delta} = \dot{u}^{P1} + \dot{u}^{P2} = \dot{u}^P$$

$$-\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{\Gamma_1}{\Delta} + \left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{\Gamma_2}{\Delta} = \dot{U}^{P1} + \dot{U}^{P2}$$

$$= \dot{U}^P, \tag{8}$$

with $\Gamma_{1/2} = (A+2N) - \rho_1 v^2_{P1/P2}$ and $\Delta = \Gamma_2 - \Gamma_1$. Replacing the displacement velocities of the first and second pressure wave by the above-presented relation leads to a connection to the general displacement velocities according to

$$\boldsymbol{\sigma}_{\text{Solid}} = -\frac{1}{v_{P1}}\left((A\mathbf{I} + 2N\hat{\mathbf{r}}\hat{\mathbf{r}})\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{Q}{\Delta} - Q\mathbf{I}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{\Gamma_1}{\Delta}\right) - \frac{1}{v_{P2}}\left(-(A\mathbf{I} + 2N\hat{\mathbf{r}}\hat{\mathbf{r}})\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{Q}{\Delta} + Q\mathbf{I}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{\Gamma_2}{\Delta}\right) - \frac{\mu}{v_S}(\hat{\mathbf{r}}\hat{\mathbf{t}}$$

$$+\hat{\mathbf{t}}\hat{\mathbf{r}})\dot{u}^S,$$

$$\boldsymbol{\sigma}_{\text{Fluid}} = -\frac{1}{v_{P1}}\left(-R\mathbf{I}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{\Gamma_1}{\Delta} + Q\mathbf{I}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{Q}{\Delta}\right) - \frac{1}{v_{P2}}\left(R\mathbf{I}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{\Gamma_2}{\Delta} - Q\mathbf{I}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{Q}{\Delta}\right). \quad (9)$$

Projecting the stresses on an auxiliary surface with normal $\hat{\mathbf{n}}$ results in the traction vector (for all angles of incidence) according to

$$\mathbf{t}_{\text{Solid}} = -\frac{1}{v_{P1}}\left((A\hat{\mathbf{n}} + 2N(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}})\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{Q}{\Delta} - Q\hat{\mathbf{n}}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{\Gamma_1}{\Delta}\right) - \frac{1}{v_{P2}}\left((A\hat{\mathbf{n}} + 2N(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}) \times\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{Q}{\Delta} - Q\hat{\mathbf{n}}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{\Gamma_2}{\Delta}\right) - \frac{\mu}{v_S}\left((\hat{\mathbf{n}}\cdot\hat{\mathbf{r}})\hat{\mathbf{t}} + (\hat{\mathbf{t}}\cdot\hat{\mathbf{n}})\hat{\mathbf{r}}\right)\dot{u}^S$$

$$\mathbf{t}_{\text{Fluid}} = -\frac{1}{v_{P1}}\left(-R\hat{\mathbf{n}}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{\Gamma_1}{\Delta} + Q\hat{\mathbf{n}}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_2}{Q}\right)\frac{Q}{\Delta}\right) - \frac{1}{v_{P2}}\left(R\hat{\mathbf{n}}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{\Gamma_2}{\Delta} - Q\hat{\mathbf{n}}\left(\dot{U}^P + \dot{u}^P\frac{\Gamma_1}{Q}\right)\frac{Q}{\Delta}\right). \quad (10)$$

Now we relate the scalar displacement velocity of the pressure wave to the total displacement velocity in the two phases via $\dot{u}^P = \hat{\mathbf{r}}\cdot\dot{\mathbf{u}}$ and $\dot{U}^P = \hat{\mathbf{r}}\cdot\dot{\mathbf{U}}$. The displacement velocity of the rotational wave arises from the total displacement velocity through $\hat{\mathbf{t}}\dot{u}^S = (\mathbf{I} - \hat{\mathbf{r}}\hat{\mathbf{r}})\cdot\dot{\mathbf{u}}$. Thus we get the absorbing boundary condition for soils in the formulation of viscous tractions through "tensor impedances" (analogous to Ref. 12).

$$\begin{bmatrix}\mathbf{t}^{\text{Solid}}\\ \mathbf{t}^{\text{Fluid}}\end{bmatrix} = \begin{bmatrix}\mathbf{Z}^{SS} & \mathbf{Z}^{SF}\\ \mathbf{Z}^{FS} & \mathbf{Z}^{FF}\end{bmatrix}\cdot\begin{bmatrix}\dot{\mathbf{u}}\\ \dot{\mathbf{U}}\end{bmatrix} \quad (11)$$

with

$$\mathbf{Z}^{SS} = \frac{\Gamma_1\Gamma_2}{\Delta}\left[\left(\frac{1}{\Gamma_2 v_{P2}} - \frac{1}{\Gamma_1 v_{P1}}\right)(A\hat{\mathbf{n}}\hat{\mathbf{r}} + 2N(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}\hat{\mathbf{r}}) + \left(\frac{1}{v_{P1}} - \frac{1}{v_{P2}}\right)\hat{\mathbf{n}}\hat{\mathbf{r}}\right] - \frac{N}{v_S}(\hat{\mathbf{r}}\hat{\mathbf{n}} + (\hat{\mathbf{n}}\cdot\hat{\mathbf{r}})(\mathbf{I} - 2\hat{\mathbf{r}}\hat{\mathbf{r}})),$$

$$\mathbf{Z}^{SF} = \frac{Q}{\Delta}\left[\left(\frac{1}{v_{P2}} - \frac{1}{v_{P1}}\right)(A\hat{\mathbf{n}}\hat{\mathbf{r}} + 2N(\hat{\mathbf{n}}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}\hat{\mathbf{r}}) + \left(\frac{\Gamma_1}{v_{P1}} - \frac{\Gamma_2}{v_{P2}}\right)\hat{\mathbf{n}}\hat{\mathbf{r}}\right],$$

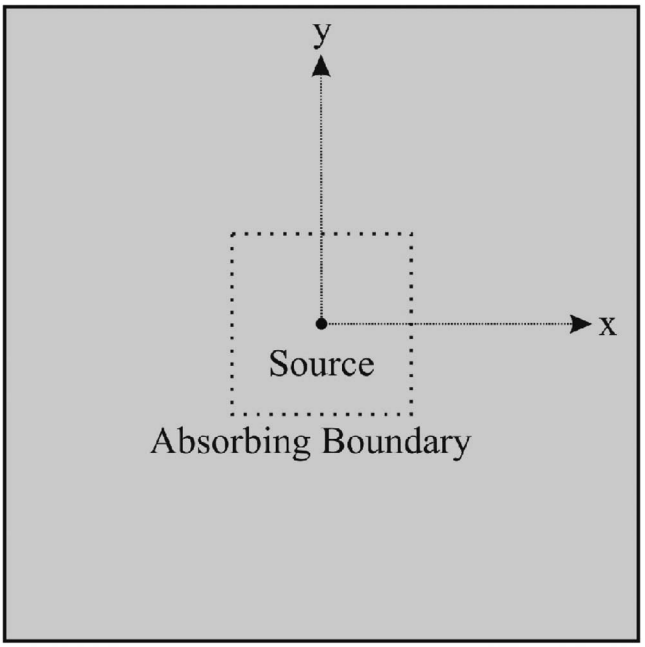


FIG. 1. The computational domain is built by two squares of 6 m and 2 m side lengths, respectively. The source acts on the solid phase in positive x direction.

$$\mathbf{Z}^{FS} = \frac{\Gamma_1\Gamma_2}{\Delta Q}\left[\left(\frac{1}{\Gamma_2 v_{P2}} - \frac{1}{\Gamma_1 v_{P1}}\right)Q^2\hat{\mathbf{n}}\hat{\mathbf{r}} + \left(\frac{1}{v_{P1}} - \frac{1}{v_{P2}}\right)R\hat{\mathbf{n}}\hat{\mathbf{r}}\right],$$

$$\mathbf{Z}^{FF} = \frac{1}{\Delta}\left[\left(\frac{1}{v_{P2}} - \frac{1}{v_{P1}}\right)Q^2\hat{\mathbf{n}}\hat{\mathbf{r}} + \left(\frac{\Gamma_1}{v_{P1}} - \frac{\Gamma_2}{v_{P2}}\right)R\hat{\mathbf{n}}\hat{\mathbf{r}}\right], \quad (12)$$

where in the time harmonic case the derivation with respect to time can be replaced by a factor $j\omega$.

III. TESTING THE TRFS BOUNDARY CONDITIONS

In order to test the new boundary conditions we refer to Fig. 1: here, we produce a quasianalytical solution in the small square (dotted line) by increasing the computational domain to the size of the big square. Additionally, Kelvin–Voigt damping has been implemented and the damping value has been adjusted such that the amplitude is almost zero at the outer border. This result is compared to the small square solution featuring different absorbing boundary conditions: those obtained by Zerfa and Loret, the TRFS derived earlier, and—as a "worst case"—the Dirichlet condition (Fig. 2). The elastic constants have been supplied with the Kelvin–Voigt term $(1+i\omega\alpha_d)$ as shown in Table I, where the damping is controlled by the value of α_d. The frequency used is 1 kHz; the porosity n and the specific weights are also given in Table I.

The finite element simulation was formulated in the frequency domain with a minimum element size of a fifth of the smallest wavelength. The quasianalytical solution resulted in a linear equation system of about 2×10^6 degrees of freedom (DOF) and the small region had about 2.5×10^5 DOF. Figure

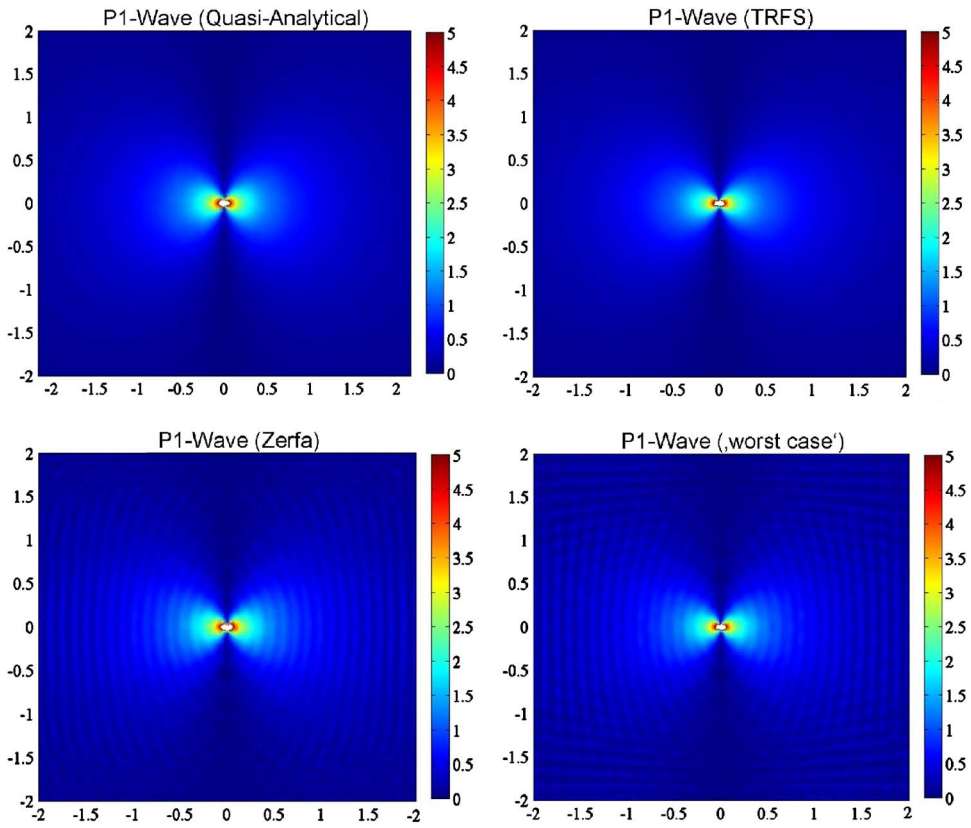


FIG. 2. (Color online) The spatial amplitude distribution of the first pressure wave as calculated using various boundary conditions.

2 shows the spatial amplitude distribution of the different solutions. As expected we see a smooth decay of amplitude in the first plot, while in the last one we see interference patterns due to using hard walls instead of absorbing boundaries.

The result obtained by using the TRFS boundary conditions compares well with the quasianalytical solution, while the conditions from Zerfa and Loret[17] show only little difference from the worst case. Thus it is evident that the effect of additionally absorbing obliquely impinging waves improves the quality of the simulation tremendously. Figure 3 confirms this fact, since there are no recognizable reflections in the TRFS result.

IV. CONCLUSION

In this work the absorbing boundary conditions for two phase media given previously by Zerfa and Loret[17] have been improved by considering additional absorbing waves with auxiliary angles of incidence. This was done by projecting the stress tensor of the medium on an auxiliary plane. The new conditions can be implemented by characterizing the traveling direction of the waves, provided that the source's position is known. Then by defining one additional angle—or two angles if three-dimensional evaluation is considered—on the boundaries the impinging plane can be specified. By using global coordinates (x,y,z) these angles are available for each point on the boundary and thus the additional computational cost is negligible. In the case of extended sources and long distances between source and boundary the source can be considered as a point source. Applying this solution to the case of short distances between boundaries and source can of course lead to failures in the absorbing performance.

In general, the performance of boundary conditions strongly depends on the used parameters, so this also applies to the presented TRFS conditions. These have been tested for a special set of parameters and have proven to work well. Other sets of data have to be checked individually. Although only tested in two dimensions the improved absorbing boundary conditions should work in three dimensions as well, since these have been developed in analogy to the work of Krenk and Kirkegaard.[12] While the TFRS conditions have been established in view of time-harmonic applications, the frequency-dependent speeds of sound used in the derivation may cause problems in time domain.

TABLE I. Material constants with Kelvin–Voigt damping used in the simulation; the constants are similar to those in Ref. 17.

$A^*(1+i\omega\alpha_d)$ (Pa)	$Q^*(1+i\omega\alpha_d)$ (Pa)	$R^*(1+i\omega\alpha_d)$ (Pa)	$N^*(1+i\omega\alpha_d)$ (Pa)	n	ρ_{Solid} (kg/m^3)	ρ_{Fluid} (kg/m^3)	α_d	b
10^9	10^9	10^9	10^8	0.333	3101	2977	9×10^{-6}	0

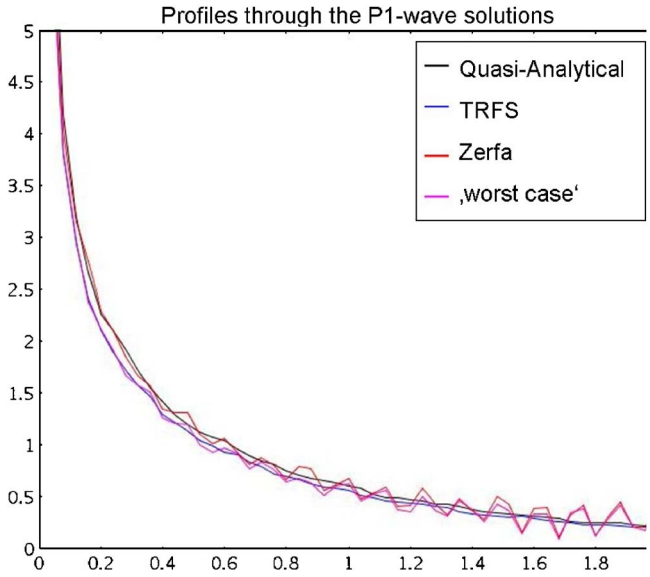


FIG. 3. (Color online) Amplitude profiles along the positive x axis for the different solutions.

ACKNOWLEDGMENTS

This work has been performed at the Fraunhofer Institute for Nondestructive Testing IZFP in Saarbrücken, Germany. The authors would like to thank Professor Walter Arnold, IZFP, for his support of and his interest in this study.

[1]K. L. Williams and R. E. Francois, "Sea ice elastic moduli: Determination of Biot parameters using in-field velocity measurements," J. Acoust. Soc. Am. **91**, 2627–2636 (1992).
[2]Ph. Leclaire and F. Cohen-Tènoudji, "Extension of Biot's theory of wave propagation to frozen porous media," J. Acoust. Soc. Am. **96**, 3753–3768 (1994).
[3]K. I. Lee, H. Roh, and S. W. Yoon, "Acoustic wave propagation in bovine cancellous bone: Application of the Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293 (2003).
[4]Z. E. A. Fellah, J. Y. Chapelon, S. Berger, W. Lauriks, and C. Depollier, "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73 (2004).
[5]K. A. Wear and A. Laib, "Comparison of measurements of phase velocity in human calcaneus to Biot theory," J. Acoust. Soc. Am. **117**, 3319–3324 (2005).
[6]S. Derible, "Debye-series analysis of the transmission coefficient of a water-saturated porous plate obeying Biot's theory," J. Acoust. Soc. Am. **118**, 3430–3435 (2005).
[7]M. A. Biot, "General theory of three-dimensional consolidation," J. Appl. Phys. **12**, 155–164 (1941).
[8]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[9]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[10]M. A. Biot and D. G. Willis, "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601 (1957).
[11]J. Lysmer and R. Kuhlemeyer, "Finite dynamic model for infinite media," J. Engrg. Mech. Div. **95**, 859–877 (1969).
[12]S. Krenk and P. H. Kirkegaard, "Local tensor radiation conditions for elastic waves," J. Sound Vib. **247**, 875–896 (2001).
[13]J. P. Wolf and C. Song, *Finite-Element Modelling of Unbounded Media* (Wiley, New York, 1996).
[14]D. Givoli, "Non-reflecting boundary conditions," J. Comp. Physiol. **94**, 1–29 (1991).
[15]T. Akiyoshi, K. Fuchida, and H. L. Fang, "Absorbing boundary conditions for dynamic analysis of fluid-saturated porous media," Soil Dyn. Earthquake Eng. **13**, 387–397 (1994).
[16]G. Degrande and G. De Roeck, "An absorbing boundary condition for wave propagation in saturated poroelastic media. I. Formulation and efficiency evaluation," Soil Dyn. Earthquake Eng. **12**, 411–421 (1993).
[17]Z. Zerfa and B. Loret, "A viscous boundary for transient analyses of saturated porous media," Earthquake Eng. Struct. Dyn. **33**, 89–110 (2004).
[18]R. Clayton and B. Engquist, "Absorbing boundary conditions for acoustic and elastic wave equations," Bull. Seismol. Soc. Am. **67**, 1529–1540 (1977).
[19]R. M. Bowen, "Incompressible porous media models by use of the theory of mixtures," Int. J. Eng. Sci. **18**, 1129–1148 (1980).
[20]R. M. Bowen, "Compressible porous media models by use of the theory of mixtures," Int. J. Eng. Sci. **20**, 697–735 (1982).

Predicted attenuation of sound in a rigid-porous ground from an airborne source

Kai Ming Li
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 Martin Jischke Drive, West Lafayette, Indiana 47907-2031

(Received 6 March 2007; revised 19 December 2007; accepted 21 December 2007)

An approximate analytical formula has been derived for the prediction of sound fields in a semi-infinite rigid-porous ground due to an airborne source. The method starts by expressing the sound fields in an integral form, which can subsequently be evaluated by the method of steepest descents. The concept of effective impedance has been introduced by using a physically plausible assumption. The integral can then be simplified and evaluated analytically. The analytical solution can be expressed in a closed form analogous to the classical Weyl–Van der Pol formula that has been used for predicting sound fields above a rigid-porous ground. Extensive comparisons with the wave-based numerical solutions according to the fast field formulation and the direct evaluation of the integral have been conducted. It has been demonstrated that the analytical formula is sufficiently accurate to predict the penetration of sound into a wide range of outdoor ground surfaces.
 [DOI: 10.1121/1.2835437]

PACS number(s): 43.35.Mr, 43.35.Ty, 43.28.Js [VEO] Pages: 1352–1363

I. INTRODUCTION

Although there has been sustained interest in sound propagation involving two semi-infinite media separated by a plane interface, many studies have been focused on the propagation of sound in the upper medium over different types of ground surface. This has been pertinent to those interested in the acoustical properties of the soil[1] and other sound absorption materials.[2] Relatively few studies have considered the prediction of sound field in the lower medium, which is significant for those involved with the detection of buried objects such as nonmetallic landmines[3,4] and plastic water pipes.[5]

If the ground surface is modeled as a poroelastic solid, supporting a shear wave and fast and slow dilatational waves,[6] the asymptotic solution becomes intractable analyticaly even for such simple boundary conditions as a half-space[7] and a hard-back layer.[8] On the other hand, the problem can be solved numerically by use of the fast field formulation.[9] To simplify the problem sufficiently to obtain an analytical result, we take advantage of the fact that the stiffness and density of the solid frame are usually much greater than those of air, so that the solid frame may be regarded as rigid and motionless and only supports the fast dilatational wave. This is known as the modified fluid assumption.

The modified fluid assumption cannot be used to predict the seismic motion of solid particles induced by acoustic waves, but it has been widely used for modeling outdoor ground surfaces and other sound absorption materials for indoor acoustics. The modified fluid assumption is reasonably good for many practical situations involving outdoor surfaces and sound absorbents. Moreover, in a recent study, Velea *et al.*[10] adapted the modified fluid model to estimate the coupling effect between acoustic to seismic motions between a landmine and its surrounding soil. Their model gives a good quantitative account of the low-frequency resonance at the surface induced by the acoustic disturbances.

As a first step toward a versatile analytical solution for the two-media problem involving a poroelastic ground, we see the need to develop a simple yet accurate Green function to predict the sound field in the lower medium, which is assumed to be a modified fluid. This paper will be focused on the derivation and validation of such an analytical formula. Other issues, such as the penetration of sound in a layered poroelastic ground, will be a subject of future publication.

II. THEORY FOR SOUND REFRACTION IN AN ABSORPTIVE MEDIUM

A. Formulation

In an earlier studies, Li *et al.*[11] considered the propagation of sound above a rigid-porous ground in the air, the upper medium where the air density, sound speed, and wave number are denoted, respectively, by ρ_0, c_0, and k_0. The lower medium, which may be modeled as a semi-infinite rigid-porous ground, has the complex density, sound speed, and propagation constant (wave number) of ρ_1, c_1, and k_1, respectively. Here, we consider the penetration of airborne sound into the lower effective fluid medium, i.e., the ground, due to an airborne source with the angular frequency of ω_0. The time-dependent factor, $e^{-i\omega_0 t}$, is omitted in the following analysis. Unless stated otherwise, the subscripts 0 and 1 are used to denote the corresponding parameters for the upper and lower media, respectively

Figure 1 shows a schematic diagram of the problem where z_s is the source height above the air/ground interface and r is the horizontal separation between the source and receiver. In Fig. 1, we also display a graphical illustration of all distances and angles used in the subsequent derivation. The receiver can be located above the interface measured upward at a height of z, but it can also be situated in the

 0001-4966/2008/123(3)/1352/12/$23.00

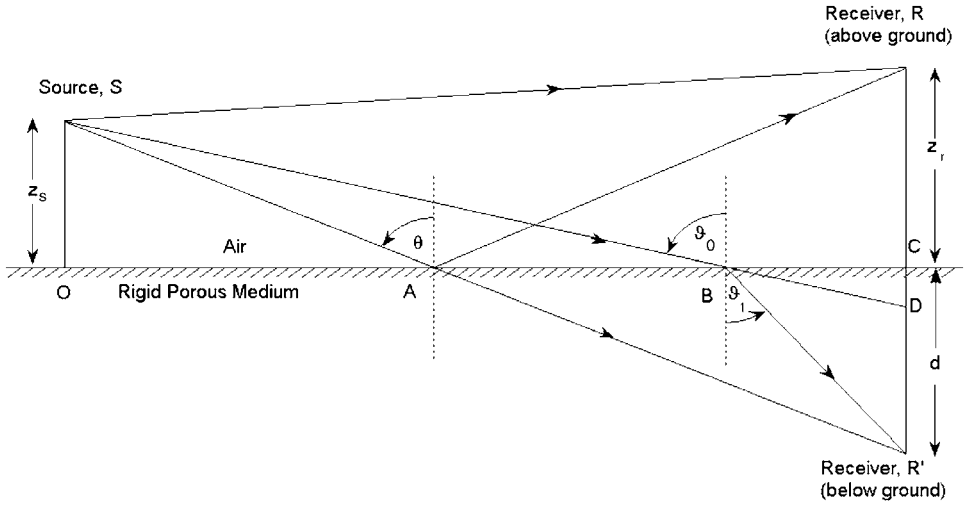


FIG. 1. A schematic diagram showing a source locating in the air at the upper medium, a receiver is situated at R with height z_r above an interface and another receiver at R' is located with depth d below the interface at the lower medium. In the main text, the following symbols are linked to the equivalent distances as shown in the diagram: $\overline{SR}=R_1$; $\overline{SA}+\overline{AR}=R_2$; $\overline{SD}=\bar{R}_2$; $\overline{SB}=R_s$; $\overline{BR'}=R_d$; $\overline{OC}=r$; $\overline{OB}=r_s$; $\overline{BC}=r_d$; $\overline{OS}=z_s$; $\overline{CR}=z_r$; and $\overline{CR'}=d$. The angle θ is the incident angle of the reflected wave that links S to R. For the refracted ray linking S to R', ϑ_0 is the incident angle and ϑ_1 is the refracted angle. Note that a special situation with $z_r=d$ is presented in the diagram for the purpose of illustration.

ground measured downward at a depth of D (i.e., $z=-D$) from the interface. The sound field satisfies the respective Helmholtz equations for the upper and lower media as follows:

$$\nabla^2 p + k_0^2 p = -\delta(z - z_s) \quad \text{for } z > 0,$$

$$\nabla^2 p + k_1^2 p = 0 \quad \text{for } z < 0, \tag{1}$$

where p is the acoustic pressure. At the air/ground interface of $z=0$, the continuity of particle velocity and pressure are imposed such that

$$p|_{z=0^+} = p|_{z=0^-},$$

$$\frac{1}{\rho_0}\frac{\partial p}{\partial z}\bigg|_{z=0^+} = \frac{1}{\rho_1}\frac{\partial p}{\partial z}\bigg|_{z=0^-}. \tag{2}$$

Use of the Sommerfeld radiation condition and the boundary conditions given earlier, we can represent the total sound field in the lower medium as

$$p(r,z_s,D|K_0,K_1) = p_1(r,z_s,D) = \frac{\mathrm{i}}{2\pi}\int_{-\infty}^{\infty} \frac{\kappa \exp[i(K_0 z_s + K_1 D)]}{K_0 + \varsigma_1 K_1} H_0^{(1)} \times (\kappa r) d\kappa, \tag{3}$$

where $H_0^{(1)}$ is the Hankel function of the first kind[12] and ς is defined as the ratio of the air density to the density of the medium, i.e.,

$$\varsigma_0 = \frac{\rho_0}{\rho_0} = 1, \quad \varsigma_1 = \frac{\rho_0}{\rho_1}. \tag{4}$$

The variable K_0 and K_1 are defined as

$$K_0 = +\sqrt{k_0^2 - \kappa^2} \tag{5a}$$

and

$$K_1 = +\sqrt{k_1^2 - \kappa^2}, \tag{5b}$$

where positive roots are taken for K_0 and K_1. The variable κ used in the integral of Eq. (3) can be usefully interpreted as the horizontal component of the respective wave vectors for the upper and lower media, $\mathbf{k}_0$ and $\mathbf{k}_1$. It follows immediately from Eqs. (5a) and (5b) that K_0 and K_1 can be regarded as the respective vertical component of the wave vectors in the upper and lower media, i.e., $\mathbf{k}_0=(\kappa,K_0)$ and $\mathbf{k}_1=(\kappa,K_1)$.

At the same horizontal range r, the sound field for a receiver located at z_r above the interface in the air can be written in the following form:[13–15]

$$p = \frac{e^{ik_0 R_1}}{4\pi R_1} - \frac{e^{ik_0 R_2}}{4\pi R_2} + p_0(r,z_s,z_r), \tag{6}$$

where R_1 is the path length of the direct wave and R_2 is that of the reflected wave. The term $p_0(r,z_s,z_r)$ is a correction term to account for the effect of a porous ground. It is given by

$$p_0(r,z_s,z_r) = p(r,z_s,z_r|K_0,K_0) = \frac{\mathrm{i}}{2\pi}\int_{-\infty}^{\infty} \frac{\kappa \exp[iK_0(z_s + z_r)]}{K_0 + \varsigma_1 K_1} H_0^{(1)}(\kappa r) d\kappa. \tag{7}$$

We note that the factor K_0 is multiplied separately with z_s and z_r in the exponential term of the previous equation. This signifies that the transmission of sound is confined to the upper medium. However, the exponential factor in Eq. (3) contains the summation of $K_0 z_s$ and $K_1 D$, which implies that the sound is refracted from the upper to the lower medium.

The exact solutions for Eqs. (3) and (7) are generally not possible as the integral involves a Hankel function, a complex exponential function and an irrational term, $1/(K_0 + \varsigma_1 K_1)$. However, when $k_0 r$ is large, an asymptotic approximation can be applied to evaluate the integral analytically. In the forthcoming sections, we shall give an overview of two

asymptotic solutions for predicting the sound field in a porous medium in earlier studies. A more general approach for deriving an asymptotic solution will then be given in Secs. III B and III C.

B. Review of analytical solutions (Ref. 16)

Brekhovskikh[17] discussed the use of a wave approach to solve a two-media problem and he employed a ray method to determine the refraction of sound in the lower medium. According to Snell's law, the angle of incidence, ϑ_0, is related to the angle of refraction, ϑ_1 (see Fig. 1 for these two angles) as follows:

$$\sin \vartheta_0 = n_1 \sin \vartheta_1, \tag{8}$$

where n_1 is the index of refraction for sound transmitted from the upper to the lower medium given by

$$n_1 = \frac{k_1}{k_0} = \frac{c_0}{c_1}. \tag{9}$$

By considering the balance of energy fluxes at the interface, Brekhovskikh showed that the transmitted wave can be approximated by

$$p_1(r,z_s,D) = \frac{\exp[ik_0L]}{2\pi\sqrt{\bar{R}_2 R_r}}\left[\frac{\cos\vartheta_0}{(\cos\vartheta_0+\bar{\beta}_e)} + O(1/k_0r) + \cdots\right], \tag{10}$$

where

$$\bar{\beta}_e = \varsigma_1\sqrt{n_1^2-\sin^2\vartheta_0} = \varsigma_1 n_1\cos\vartheta_1, \tag{11a}$$

$$L = z_s/\cos\vartheta_0 + n_1 D/\cos\vartheta_1, \tag{11b}$$

$$\bar{R}_2 = r/\sin\vartheta_0, \tag{11c}$$

$$R_r = \frac{z_s}{\cos\vartheta_0} + \frac{D\cos^2\vartheta_0}{n_1\cos^3\vartheta_1}, \tag{11d}$$

$O(1/k_0r)$ is the term with its magnitude proportional to $(k_0r)^{-1}$, and other higher order terms are not shown in Eq. (10). The physical interpretation of $\bar{\beta}_e$, L, $\bar{R}_2$, and R_r will be given in the next section when we derive an improved asymptotic formula. He further suggested the use of the saddle path method for providing a correction term to the ray method for the refracted waves. In the limit of near-grazing incidence and a large index of refraction, i.e., highly absorptive ground, Eq. (10) can be approximated by[17]

$$p_1(r,z_s,D) = \exp[ik_0D\sqrt{n_1^2-\sin^2\vartheta_0}]p_0(r,z_s,0). \tag{12}$$

Extending the analysis of Paul,[13] Attenborough and co-workers[14,18] investigated the propagation of sound due to an airborne source placed above a two-media boundary. The former publication was concerned with the penetration of sound into the porous ground from a point source in the air. The later dealt with the propagation of sound above an extended reaction ground at near-grazing angles. Attenborough and co-workers[14,18] used essentially the same scheme to approximate the integrals [Eqs. (3) and (7)]. Closed form analytical expressions for the sound fields below and above the interface were obtained, respectively. However, the analytical expression for the asymptotic solution of the integral given in Eq. (3) was so complicated that it did not lend itself for convenient numerical implementation. Following the lead of Brekhovskikh and using the same assumption (a near-grazing incidence and a large index of refraction of the ground), Richards *et al.*[14] suggested the same result as given in Eq. (12) except the term $\sin\vartheta_0$ was approximated by $r/\sqrt{r^2+z_s^2}$. Using this simplified formula, Richards *et al.*[14] compared their asymptotic formulation with indoor experimental measurements of the sound field within a porous material. This approximation is only valid when the receiver is located very close the air/ground interface.

III. ASYMPTOTIC APPROXIMATION

A. Analysis of the saddle path

We endeavor to develop an improved asymptotic formula in this section by applying the method of steepest descent.[17] To approximate the sound field in the lower medium, we recast the integral of Eq. (3) as follows:

$$p_1(r,z_s,D) = \frac{\mathrm{i}}{2\pi}\int_{-\infty}^{\infty}\frac{\kappa G(\kappa)\exp[if(\kappa)]}{K_0+\varsigma_1 K_1}d\kappa \tag{13}$$

where

$$f(\kappa) = \kappa r + K_0 z_s + K_1 D \tag{14a}$$

and

$$G(\kappa) = H_0^{(1)}(\kappa r)e^{-i\kappa r}. \tag{14b}$$

In the integral, the term involving the Hankel function is conveniently grouped with an additional exponential term [see Eq. (14b)] because $G(\kappa)$ can be approximated asymptotically as

$$G(\kappa) = H_0^{(1)}(\kappa r)e^{-i\kappa r} \approx (2/i\pi\kappa r)^{1/2}[1+1/8i\kappa r + O(\kappa r)^{-2}+\cdots]. \tag{15}$$

The saddle point occurs at $f'(\kappa)=0$. Hence, differentiating Eq. (14a) with respect to κ. and setting it to zero, we obtain the following equation:

$$r = \frac{\kappa z_s}{K_0} + \frac{\kappa D}{K_1}. \tag{16}$$

By solving Eq. (16) for the variable κ, we can determine the stationary point along the steepest descent path, i.e., the so-called saddle path. To obtain a physically meaningful solution, it is instructive to consider Snell's law given in the form of Eq. (9) as the variable, κ, used in Eq. (13) can be interpreted as the horizontal component of the wave vectors and Eq. (9) can be rewritten as

$$\kappa = k_0\sin\mu_0 = k_1\sin\mu_1 = \text{constant}, \tag{17}$$

where μ_0 indicates the direction of propagation of each of the incident plane waves and μ_1 is that for the refracted plane waves. Both angles are measured from the vertical axis perpendicular to the interface, i.e., the z-axis.

We can now transform the integral of Eq. (13) from κ to μ_0 by using the relation given in Eq. (17) to yield

$$p_1(r,z_s,D) = \frac{ik_0}{2\pi}\int_{-\pi/2+i\infty}^{\pi/2-i\infty} \frac{\cos\mu_0 \sin\mu_0 G(\mu_0)\exp[if(\mu_0)]}{\cos\mu_0 + \varsigma_1 n_1 \cos\mu_1} d\mu_0, \tag{18}$$

where $f(\kappa)$ and $G(\kappa)$ are transformed to

$$f(\mu_0) = k_0(r\sin\mu_0 + z_s\cos\mu_0 + n_1 D\cos\mu_1) \tag{19a}$$

and

$$G(\mu_0) = H_0^{(1)}(k_0 r\sin\mu_0)e^{-ik_0 r\sin\mu_0}. \tag{19b}$$

For a given location of source height, z_s, the receiver depth, D, and the horizontal separation, r, the saddle point, $\mu_0 = \vartheta_0$, can be determined by transforming Eq. (16) from κ-space to μ_0-space to furnish

$$r = z_s\tan\vartheta_0 + D\tan\vartheta_1, \tag{20}$$

where ϑ_1 is the angle of refraction for the ray entering in the porous medium. These two angles, ϑ_0 and ϑ_1, are related according to Eq. (8). Equation (20) is the same as the expression given by Brekhovskikh[17] who has determined the angle of incidence of a refracted ray by a geometrical consideration of the problem. It is enlightening to show that the ray path determined by the saddle point analysis coincides with that given by the previous study using a ray approach.

Suppose that the imaginary part of n_1 is zero, which means that both the upper and lower layers are lossless media. Then the incident and refracted angles are real. In other words, we can take the horizontal distance, r as the sum of the horizontal distance traveled by the incident ray, r_s, and the horizontal distance traveled by the refracted ray, r_d where

$$r_s = R_s\sin\vartheta_0, \quad r_d = R_d\sin\vartheta_1, \tag{21}$$

where R_s and R_d are the respective distance traveled by the incident ray and the refracted ray, respectively. The physical representation of these distances and angles is shown in Fig. 1. With this interpretation, $f(\mu_0)$ and $G(\mu_0)$ can be expressed as

$$f(\mu_0) = k_0[R_s\cos(\mu_0 - \vartheta_0) + n_1 R_d\cos(\mu_1 - \vartheta_1)] \tag{22}$$

and

$$G(\mu_0) = H_0^{(1)}(k_0 r_s\sin\mu_0 + k_1 r_d\sin\mu_1) \times e^{-i(k_0 r_s\sin\mu_0 + k_1 r_d\sin\mu_1)}. \tag{23}$$

We see that the mathematical analysis used previously is equally valid for the case of porous media in which the propagation constant and the index of refraction are complex. This fact leads to the following conclusion: ϑ_0, ϑ_1, r_s, r_d, R_s, and R_d are all complex because of the geometrical requirements stipulated in Eqs. (20) and (21). However, $\mathrm{Im}(z_s\tan\vartheta_0)$ must be equal to $-\mathrm{Im}(D\tan\vartheta_1)$ because the total horizontal distance traveled by the ray, r, remains a real number.

It may now possible to follow the elaborate scheme suggested by Attenborough *et al.*[18] to evaluate the integral of Eq. (7) asymptotically. However, their scheme does not necessarily lead to a more accurate numerical solution.[11,19] On the other hand, Li *et al.*[11] have demonstrated that the term $\varsigma_1 n_1\cos\mu_1$ in the integral can be approximated heuristically by an effective admittance,

$$\beta_e = \varsigma_1 n_1\cos\theta_1, \tag{24}$$

for the determination of the sound field above the rigid-porous ground. In Eq. (24), θ_1 is the refracted angle due to an incoming wave with an angle of incidence of θ. Use of β_e, which is independent of μ_0, can lead to a close form analytical expression for the sound field above a rigid-porous ground.

Although the concept of the effective admittance has not been explicitly identified by Brekhovskikh,[17] we can see that the effective admittance term $\bar{\beta}_e$ as defined in Eq. (11a) appears in his expression for the sound field below a rigid-porous ground, cf. Eq. (10). Consequently, we make use of the same heuristic approximation by replacing $\varsigma_1 n_1\cos\mu_1$ with $\bar{\beta}_e$ for evaluating the integral in Eq. (18). It is remarkable the respective effective admittances (β_e and $\bar{\beta}_e$) for the sound field above and below the rigid-porous ground are necessarily different because of different angles of incidence for these two situations. The validity of this heuristic approximation will be examined numerically by comparing the solutions obtained from the resulting analytical approximation with the more accurate wave-based numerical solutions.

B. Evaluation of the integral for the sound field in a porous medium

A close examination of the integral expression for the sound field in the lower medium [see Eqs. (3) and (13)] reveals that it has a rather similar form to that in Eq. (7) for the reflection of sound in the upper medium. Therefore, we expect that the solution for the sound field in the lower medium resembles closely with the respective solution for the correction term in the upper medium. An asymptotic analysis for sound scattered by an impedance ground was studied in details by Chien and Soroka.[20] We shall use their approach and the salient steps will only be presented below.

To evaluate the integral of Eq. (18) asymptotically, we express it in the following form:

$$p_1(r,z_s,D) = I_1 - I_2 \tag{25a}$$

where

$$I_1 = \frac{ik_0}{2\pi}\int_{-\pi/2+i\infty}^{\pi/2-i\infty} \phi(\mu_0)\exp[if(\mu_0)]d\mu_0, \tag{25b}$$

$$I_2 = \frac{ik_0\bar{\beta}_e}{2\pi}\int_{-\pi/2+i\infty}^{\pi/2-i\infty} \frac{\phi(\mu_0)\exp[if(\mu_0)]}{\cos\mu_0 + \bar{\beta}_e} d\mu_0, \tag{25c}$$

and $\phi(\mu_0)$ in Eqs. (25b) and (25c) is given by

$$\phi(\mu_0) = \sin\mu_0 G(\mu_0). \tag{25d}$$

The first integral, I_1 can then be evaluated straightforwardly by the standard saddle path analysis because there are no singularities in the function, $\phi(\mu_0)$. As discussed in the previous section, the saddle point can be found by setting $f'(\mu_0)=0$. According to Eq. (19a), this occurs at $\mu_0=\vartheta_0$. We

introduce a dummy variable, X, in I_1 such that

$$\tfrac{1}{2}X^2 = i[f(\vartheta_0) - f(\mu_0)] = ik_0L - if(\mu_0), \tag{26}$$

where L is given by Eq. (11b). The factor, k_0L, in the exponential term of Eq. (11b) represents the total phase shift for the sound waves emanating from the source to the receiver. Therefore, we can interpret L as the acoustical path length linking the source in the upper medium to the receiver in the lower medium.

By changing the variable from μ_0 to X, I_1 can be rewritten as

$$I_1 = \frac{ik_0}{2\pi}e^{ik_0L}\int_{-\infty}^{\infty}\Phi(X)e^{-X^2/2}dX, \tag{27}$$

where the transformed integrand, $\Phi(X)$, is given by

$$\Phi(X) = \phi(\mu_0)d\mu_0/dX = \sin\mu_0 G(\mu_0)d\mu_0/dX. \tag{28}$$

Differentiating Eq. (26) with respect to X, we obtain an expression for $d\mu_0/dX$. Substituting it into Eq. (28), we get

$$\Phi(X) = iX\phi(\mu_0)/f'(\mu_0). \tag{29}$$

Next, we expand $f(\mu_0)$, $f'(\mu_0)$, and $\phi(X)$ in the form of series in powers of $\Delta \equiv \mu_0 - \vartheta_0$ by means of Taylor's expansion:

$$f(\mu_0) = f(\vartheta_0) + \Delta f'(\vartheta_0) + \frac{\Delta^2}{2!}f''(\vartheta_0) + \frac{\Delta^3}{3!}f'''(\vartheta_0) + O(\Delta^4) = k_0L - k_0[R_r\Delta^2/2] + O(\Delta^3), \tag{30a}$$

$$f'(\mu_0) = f'(\vartheta_0) + \Delta f''(\vartheta_0) + \frac{\Delta^2}{2!}f'''(\vartheta_0) + O(\Delta^3), = -k_0R_r\Delta + O(\Delta^2), \tag{30b}$$

and

$$\phi(X) = -iX\frac{\{\sin\vartheta_0 G(\vartheta_0) + [\cos\vartheta_0 G(\vartheta_0) - \sin\vartheta_0 G'(\vartheta_0)]\Delta + O(\Delta^2)\}}{k_0R_r\Delta + O(\Delta^2)}, \tag{30c}$$

where the primes in Eqs. (30a)–(30c) denote the derivatives with respect to their arguments. The parameter, R_r, has already been introduced in Eq. (11c). Based on the ray model, Brekhovskikh[17] showed that this path length was associated with the balance of energy fluxes across the interface for the incident, reflected, and refracted rays. We shall refer R_r as the energy path length. In the saddle path analysis, the energy path length can be found by setting

$$R_r = |f''(\vartheta_0)| = R_s + n_1R_d\left(\frac{d\vartheta_0}{d\vartheta_1}\right)^2, \tag{31}$$

where R_s and R_d are the respective path lengths of the incident and refracted ray, cf. Eq. (21). Figure 1 shows the geometrical interpretation of various distances. It is straightforward to show that Eq. (31) is the same expression as Eq. (11d) by noting the following identity:

$$d\mu_0/d\mu_1|_{\mu_0=\vartheta_0} \equiv d\vartheta_0/d\vartheta_1 = \cos\vartheta_0/(n_1\cos\vartheta_1). \tag{32}$$

Substitution of Eq. (30a) into Eq. (26) and rearrangement of the resulting equation lead to an approximate expression for Δ as follows:

$$\Delta = \frac{X}{\sqrt{ik_0R_r}} + O(X^2). \tag{33}$$

Hence, $\Phi(X)$ in Eq. (29) can be simplified by using Eqs. (30b) and (33) to yield

$$\Phi(X) = \frac{-ie^{i\pi/4}G(\vartheta_0)\sin\vartheta_0}{\sqrt{k_0R_r}} + O(X). \tag{34}$$

Changing the variable from μ_0 to X in the integral I_1, we also need to deform the path of integration from C' to C as shown in Fig. 2, The saddle path, C, can be found by setting $\mathrm{Im}[if(\vartheta_0)] = \mathrm{Im}[if(\mu_0)]$, i.e.,

$$\mathrm{Im}[ik_0L] = \mathrm{Im}\{ik_0[R_s\cos(\mu_0 - \vartheta_0) + n_1R_d\cos(\mu_1 - \vartheta_1)]\}. \tag{35a}$$

which can be expressed in terms of μ_0 as follows:

$$\mathrm{Im}\{ik_0[(R_s - z_s\cos\mu_0 - r_s\sin\mu_0) + (n_1R_d - D\sqrt{n_1^2 - \sin^2\mu_0} - r_d\sin\mu_0)]\} = 0. \tag{35b}$$

Solving Eq. (35b) for μ_0, we can identify the saddle path C.

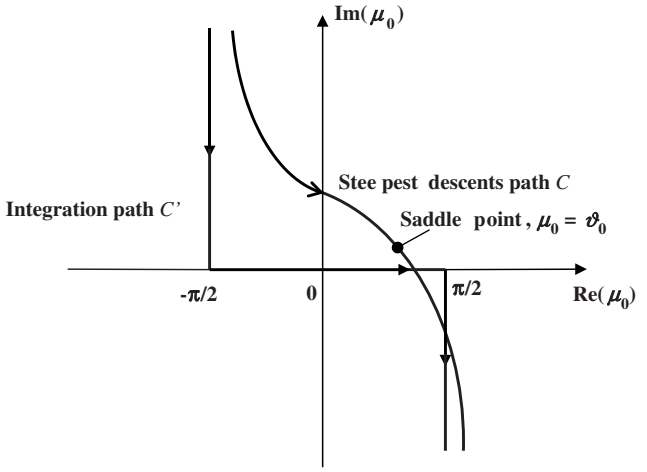


FIG. 2. Schematic diagram shown the path of integration C and C'.

The deformation of the integration path from C' to C is allowable provided that the contributions from the arcs joining C and C' are negligible and the contribution from any pole crossed is included appropriately. There is no pole in the integrand of Eq. (27) because there are no singularities in $\Phi(X)$.

Substituting Eq. (34) into (27), making use of Eq. (15) for $G(\vartheta_0)$ and integrating the expression, we can show that I_1 can be simplified to a rather compact form as

$$I_1 = \frac{e^{ik_0L}}{2\pi\sqrt{\bar{R}_2 R_r}}[1 + O(1/k_0 r) + \cdots]. \tag{36}$$

Here, $\bar{R}_2$ can be treated as the apparent path length which is a function of r and ϑ_0 only. It is equivalent to the apparent distance traveled by the incident ray in the absence of the lower medium.

We are now in a better position to consider the asymptotical expression for I_2 given in Eq. (25c). In light of the above-mentioned analysis for I_1, we can apply the transformation and expand the integrand in the power series of the dummy variable x. However, the presence of the term $(\cos\mu_0 + \bar{\beta}_e)$ in the denominator means that a pole can be identified in the integrand. The pole μ_P, say, is found simply by setting the term to zero to yield

$$\cos\mu_P = -\bar{\beta}_e \tag{37}$$

This pole is usually referred to as the surface wave pole because a similar mathematical treatment is also applied for determining the sound field above an impedance ground[20] or above a porous medium.[11] The presence of the surface wave pole has a significant effect on the asymptotical solution for I_2. When the integration path is deformed from C' to C, an analogous condition for crossing the surface wave pole [cf. Eq. (13) of Ref. 20] can be established by using Eqs. (35b) and (37) to give

$$\mathrm{Im}[X_p^2] < 0, \tag{38}$$

where the location of the pole is given by

$$X_p^2 = ik_0[(R_s + z_s\bar{\beta}_e - r_s\sqrt{1-\bar{\beta}_e^2}) + (n_1 R_D - iD\sqrt{1-n_1^2-\bar{\beta}_e^2} - r_d\sqrt{1-\bar{\beta}_e^2})]. \tag{39a}$$

However, in the case for near grazing propagation and penetration of sound below the ground, i.e., both ϑ_0 and ϑ_1 are close to $\pi/2$, we can see that $\mu_0 - \vartheta_0 \approx \mu_1 - \vartheta_1$. This condition can be met if $r \gg z_s$ and $r \gg D$. Equation (39a) can then be simplified to:

$$X_p^2 = ik_0 R_r[1 + \bar{\beta}_e\cos\vartheta_0 - \sqrt{1-\bar{\beta}_e^2}\sin\vartheta_0]. \tag{39b}$$

Deforming the integration path from C' to C, we have to include the contribution from the pole if the condition given in Eq. (38) is satisified. Hence, the integral I_2 can be rewritten as

$$I_2 = H\{-\mathrm{Im}[X_p^2]\}P + \frac{ik_0\bar{\beta}_e}{2\pi}e^{ik_0L}\int_{-\infty}^{\infty}\frac{\Phi(X)e^{-X^2/2}}{\cos\mu_0 + \bar{\beta}_e}dX, \tag{40}$$

where H is the Heaviside step function and P is the contribution from the surface wave pole which can be calculated readily by using the residue theory to yield

$$P = k_0\bar{\beta}_e H_o^{(1)}(k_0 r\sqrt{1-\bar{\beta}_e^2})\exp\{ik_0[-\bar{\beta}_e z_s + iD\sqrt{1-n_1^2-\bar{\beta}_e^2}]\}. \tag{41}$$

We may expand $1/(\cos\mu_0 + \bar{\beta}_e)$ in a power series of x with the leading terms shown as follows

$$\frac{1}{\cos\mu_0 + \bar{\beta}_e} = \frac{1}{\cos\vartheta_0 + \bar{\beta}_e}[1 + O(1/k_0 r) + \cdots], \tag{42}$$

and the resulting expression can be evaluated in a rather similar way as for I_1. It is straightforward to show that Eq. (40) can be expressed as

$$I_2 = H\{-\mathrm{Im}[X_p^2]\}P + \frac{e^{ik_0L}}{2\pi\sqrt{\bar{R}_2 R_r}}\left[\frac{\bar{\beta}_e}{\cos\vartheta_0 + \bar{\beta}_e} + O(1/k_0 r) + \cdots\right] \tag{43}$$

Hence, an asymptotic expression for the sound field below the rigid-porous ground due to an airborne source can be obtained by substituting Eqs. (36), (43), and (25a) to give:

$$p_1(r, z_s, D) = H\{-\mathrm{Im}[X_p^2]\}P + \frac{e^{ik_0L}}{2\pi\sqrt{\bar{R}_2 R_r}}\left[\frac{\cos\vartheta_0}{\cos\vartheta_0 + \bar{\beta}_e} + O(1/k_0 r) + \cdots\right] \tag{44}$$

For the near grazing propagation where $\vartheta_0 \to \pi/2$, Brekhovskikh[17] has shown that a first higher order term can approximated by:

$$O(1/k_0 r) \approx \frac{i\sqrt{\bar{R}_2 R_r}}{k_0 \varsigma_1 (n_1^2 - 1) r^2}. \tag{45}$$

Under the same condition, we expect that the leading term ignored in Eq. (44) is of the same order as that given in Eq. (45).

C. A uniform asymptotic expression for sound fields below the ground

The asymptotic expression given in Eq. (43) becomes increasingly inadequate when the surface wave pole, X_p, tends to zero. Chien and Soroka[20] used the method of pole subtraction to derive a uniform asymptotic expression for the sound field above an impedance ground. The same mathematical technique can be applied in the present situation. The result can be summarized below. The integral of Eq. (40) can be expressed after the pole subtraction to yield:

$$\frac{ik_0\bar{\beta}_e}{2\pi}e^{ik_0L}\int_{-\infty}^{\infty}\frac{\Phi(X)e^{-X^2/2}}{\cos\mu_0+\bar{\beta}_e}dX$$

$$\approx\frac{P}{2\pi i}\int_{-\infty}^{\infty}\frac{e^{-X^2/2}}{X-X_p}dX+O(1/k_0r). \quad (46)$$

Although Chien and Soroka have also shown derivations of other higher order terms in their analysis and it is tedious, but straightforward, to derive comparable expressions for the present study, these details are omitted here because the accuracy of most asymptotic expressions is usually captured in the first term of the expansion.[21] Any higher order terms play only a minor role in the asymptotic expression. The versatility and accuracy of our formula will be demonstrated by comparing with the computational intensive numerical methods. These numerical comparisons will be shown in Sec. III.

We can evaluate the integral of the right-hand side of Eq. (46) by using an identity for the complementary error function:

$$\int_{-\infty}^{\infty}\frac{e^{-X^2/2}}{X-X_p}dx=-\frac{\pi}{i}e^{-X_p^2/2}\{\mathrm{erfc}(-iX_p/\sqrt{2})$$

$$-2H(-\mathrm{Im}\,X_p)\}. \quad (47)$$

For a relatively hard ground with $\bar{\beta}_e<<1$ and both the source and receiver are close to the interface, i.e., $r\gg z_s$ and $r\gg D$,[20] the asymptotic expression of I_2 can be obtained by combining Eqs. (40), (41), (46), and (47) to give

$$I_2=-2i\bar{\beta}_e\sqrt{\pi}\left(\frac{1}{2}ik_0R_r\right)^{1/2}e^{-\bar{w}^2}\,\mathrm{erfc}(-i\bar{w})\frac{e^{ik_0L}}{4\pi\sqrt{\bar{R}_2R_r}}$$

$$+O(1/k_0r), \quad (48)$$

where the term $\bar{w}$ is known as the numerical distance determined by

$$\bar{w}=\sqrt{\frac{1}{2}ik_0R_r}(\cos\vartheta_0+\bar{\beta}_e). \quad (49)$$

It is noted that the following approximations have been used to obtain Eqs. (48) and (49) in their present forms:

$$1+\bar{\beta}_e\cos\vartheta_0-\sqrt{1-\bar{\beta}_e^2}\sin\vartheta_0\approx\frac{1}{2}(\cos\vartheta_0+\bar{\beta}_e)^2 \quad (50a)$$

and

$$H_0^{(1)}(k_0r\sqrt{1-\bar{\beta}_e^2})e^{-ik_0r\sqrt{1-\bar{\beta}_e^2}}\approx\sqrt{2/i\pi k_0\bar{R}_2}. \quad (50b)$$

Substitution of Eqs. (36) and (48) into Eq. (25a) leads to an expression for the sound field in the lower medium as

$$p_1(r,z_s,D)=2\left[1+i\sqrt{\pi}\left(\frac{1}{2}ik_0R_r\right)^{1/2}\bar{\beta}_e e^{-\bar{w}^2}\,\mathrm{erfc}(-i\bar{w})\right]$$

$$\times\frac{e^{ik_0L}}{4\pi\sqrt{\bar{R}_2R_r}}, \quad (51)$$

where higher order terms are not shown in the previous equation. As mentioned earlier, there is no surface wave pole term in Eq. (10).

Ignoring the contribution due to the surface wave pole, the correction term $[p_0(r,z_s,z_r)$ in Eq. (6)] for the sound field of a receiver located at z_r *above* the interface can be given as follows:[17]

$$p_0(r,z_s,z_r)=\frac{2\beta_e}{\cos\theta+\beta_e}\frac{e^{ik_0R_2}}{4\pi R_2}, \quad (52)$$

where R_2 and θ are, respectively, the distance and the incident angle of the reflected ray that link the source S to the receiver R, see Fig. 1. In Eq. (52), the effective admittance β_e is calculated according to Eq. (24). Including the contribution of the surface wave pole, Li *et al.*[11] suggested a uniform asymptotic formula for the sound field above a rigid-porous ground, i.e., a modified fluid model. Their result can be used to show that a more accurate expression for $p_0(r,z_s,z_r)$ is given by

$$p_0(r,z_s,z_r)=2\left[1+i\sqrt{\pi}\left(\frac{1}{2}ik_0R_2\right)^{1/2}\beta_e e^{-w^2}\,\mathrm{erfc}(-iw)\right]$$

$$\times\frac{e^{ik_0R_2}}{4\pi R_2}, \quad (53)$$

where the numerical distance w for the upper medium is defined somewhat differently as

$$w=\sqrt{\tfrac{1}{2}ik_0R_2}(\cos\theta+\beta_e). \quad (54)$$

Without a correction term due to the surface wave pole, it is well known that the resulting asymptotic formula will not give accurate predictions of sound fields especially for the near-grazing propagation.[22] Comparing Eqs. (53), (52), and (51) with Eq. (10), we therefore expect that the asymptotic formula given in Eq. (10) will not be adequate for predicting the sound fields especially when the source and receiver are located near the air/ground interface.

Equation (51) is the main result of the present study. It represents a uniform asymptotic expression that can be employed to compute the sound field below a ground due to an airborne sound source. The possible contribution due to lateral waves[17] is ignored in the present analysis. This assumption is justifiable if the index of refraction has a large imaginary component. In this case, the lateral wave will decay exponentially at large distances.[14] In the next section, we shall validate our asymptotic formula by comparing numerical results obtained by the more accurate wave-based numerical solutions.

IV. NUMERICAL COMPUTATIONS

A. The fast field program

To establish the validity of the asymptotic solutions derived earlier, it is important to assess its accuracy. Chandler-Wilde and Hothersall[23] described a mathematical rigorous approach of giving an error bound to an asymptotic solution for predicting the propagation of sound above a homogeneous impedance ground. However, its use is somewhat limited because the error bound is a formal solution which is dependent on the acoustical properties of the ground and the source/receiver geometry. A statement of the error bound for a wide range of situations is generally not possible. We shall

therefore not seek a formal solution for the error bound but be contented by comparing the numerical results obtained from the asymptotic solutions with an established computational scheme.

The parabolic equation method and the fast field formulation (FFP) are the two popular wave-based numerical schemes[24] for predicting sound propagation outdoors. As our asymptotic solutions are cast in an integral form, the fast field formulation will be employed more conveniently in the numerical comparisons for the far field. Details for the implementation of the fast field formulation have been discussed elsewhere.[25] The fast field formulation is known to be inaccurate at near fields. With some prior numerical simulations, we find that FFP is inaccurate at the near field when r is less than about 0.4λ, where λ is the wavelength of the sound source. In this case, a straightforward quadrature, such as a trapezoidal scheme, will be employed to compute the integral of Eqs. (25b) and (25c) directly. Just like FFP, the path of integration for the quadrature is indented to avoid the numerical instabilities introduced by the pole of the integrand. We remark that the quadrature is, however, not suitable at longer ranges because of the highly oscillatory nature of the integral. Many more points for the integrand are needed to ensure the convergence of the quadrature.

B. Comparison of numerical results

The asymptotic expression derived in the Sec. II can be used to calculate the sound field below the ground. However, we need to determine the incident and refracted angles (ϑ_0 and ϑ_1) that link an airborne source with a subsurface receiver. Equation (20) supported by Snell's law [Eq. (17)] allows one to determine these angles. It is generally not possible to obtain a simple form of solution for ϑ_0 but it is relatively straightforward to obtain these angles numerically. Details of a stable iterative scheme for determining ϑ_0 are given in the Appendix.

Unless stated otherwise, the Delany and Bazley model[26] will be used to calculate the real and imaginary parts of the index of refraction and the complex density ratio of the porous ground:

$$n_1 = 1 + 10.8(\sigma/f)^{0.7} + 10.3i(\sigma/f)^{059}, \tag{55}$$

$$s_1 = \frac{1}{n_1[1 + 9.08(\sigma/f)^{0.75} + 11.9i(\sigma/f)^{073}]}, \tag{56}$$

where σ is the effective flow resistivity of the porous ground in cgs units (1 cgs unit $\equiv 1$ kPa s m^{-2}) and f is the frequency of the source. We wish to point out that the choice of this semiempirical model is far from ideal to represent a realistic outdoor ground surface because it is not designed to predict the acoustic properties of low-porosity soils. The model can lead to a poor and nonanalytical prediction of sound fields in the low frequency range. It is also likely to predict erroneous value of the refractive index that leads to an incorrect prediction of the amplitude of the transmitted waves into the soil. Nevertheless, the Delany and Bazley model[26] has been chosen for the present study because of its simplicity—only a single parameter is needed for modeling the acoustical properties of the interface boundary. Our primary aims are to examine the accuracy and robustness of the uniform asymptotic formula presented in Eq. (51) for different types of air/ground interfaces with widely varying acoustical properties.

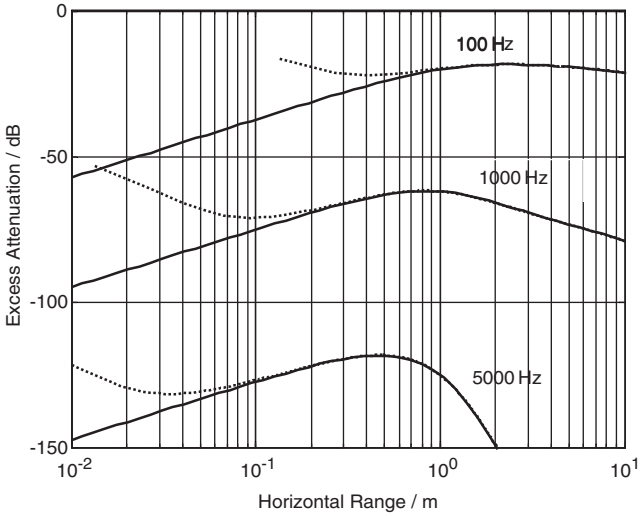


FIG. 3. Predictions of the excess attenuation versus the horizontal range between the source and receiver for frequencies at 100, 1000, and 5000 Hz, respectively. The source is located at 1 m above the porous ground and the receiver is placed at 0.5 below the air/ground interface. The solid lines show the numerical predictions according to the direct numerical integration and dotted lines show the results according to the fast field program. The effective flow resistivity of the ground is 10 kPa s m^{-2}.

With the information given in Eqs. (55) and (56), the sound fields due to an airborne source can be computed by the method of direct numerical integration, FFP, and the analytical formulation developed in Sec. II. The comparisons provide a useful validation of the accuracy and robustness of the derived asymptotic formulas. They can subsequent be used for predicting the penetration of sound by using other more suitable impedance models (see, e.g., Chap. 3 of Ref. 22) to realistically represent outdoor ground surfaces or rigid-porous materials.

We show in Fig. 3 various plots with the source located at 1 m above the rigid-porous ground and the receiver is buried at 0.5 below the air/ground interface. An effective flow resistivity of 10 kPa s m^{-2} at three representative frequencies of 100 Hz (low), 1 kHz (medium), and 5 kHz (high) are used in the predictions. The effective flow resistivity value is chosen to be relatively low deliberately so that sound penetration is significant. Such a value would be typical of snow or forest floors. The curves show comparisons of the numerical results calculated by FFP (dotted lines) and by the direct numerical integration (solid lines). In Fig. 3, the excess attenuation (EA) is plotted against r where EA is defined as

$$\mathrm{EA} = 20\log_{10}\left[\frac{p_1(r,z_s,D)}{e^{ikr}/4\pi r}\right]. \tag{57}$$

As shown in the plots for longer ranges, the two "exact" numerical schemes agree to within the thickness of the line but the FFP results start to deviate when the receiver is lo-

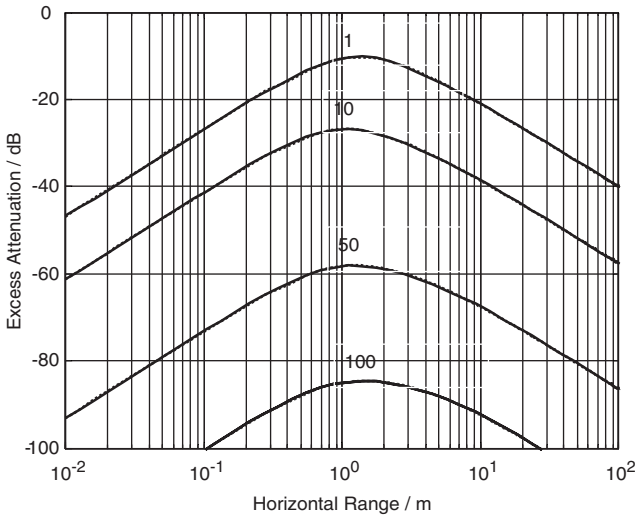


FIG. 4. Predictions of the excess attenuation versus the horizontal range between the source and receiver for the effective flow resistivities of 1, 10, 50, and 100 kPa s m^{-2}, respectively. The source is located at 1 m above the porous ground, the receiver is located at 0.2 m below the air/ground interface and the frequency of the source is 1000 Hz. The predictions according to the integrated exact solution (solid lines) agree to within the thickness of the lines for the predictions according to the asymptotic solution (dotted lines) given in Eq. (51).

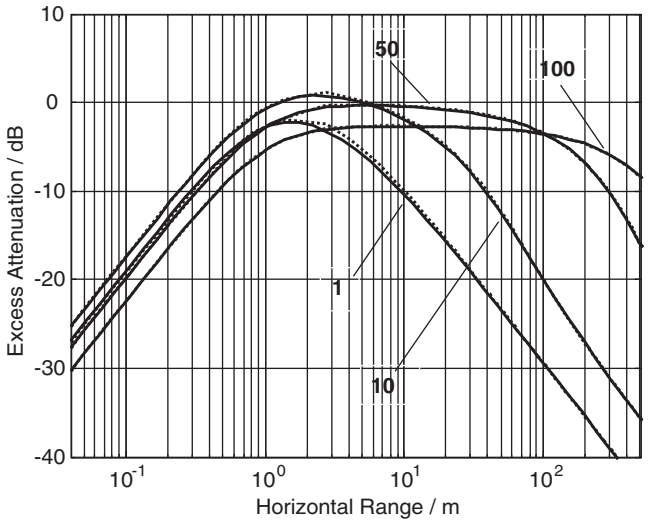


FIG. 5. Same as Fig. 3 except that the receiver is located at 0.05 m below the air/ground interface and the frequency of the source is 100 Hz. In the predictions, the effective flow resistivities of 1, 10, 50, and 100 kPa s m^{-2} are used.

cated at a horizontal range within a fraction of the wavelength from the source. We also wish to point out that the choice of parameters for z_s, D, σ, and f may not represent realistic situations because of the extremely low predicted EA. These parameters are chosen for demonstrating the accuracy of the exact numerical schemes even at such a low sound level of −50 dB and below. Different combinations of source/receiver geometry, the effective flow resistivities and frequencies have been used for computations of EA. All these numerical comparisons lead to the same conclusion described earlier.

Integrated results of FFP and the direct numerical integration will be used to benchmark the asymptotic solution in the following computations. Numerical results from the direct numerical integration will be used when r is less than 0.4 λ and the FFP results will be applied otherwise. This choice ensures a smooth curve for a broad range of r with the use of two different numerical schemes applying at the near and far fields. Hereinafter, we shall refer this hybrid scheme as the integrated exact solution (IES)

With the IES, we can assess the validity of the asymptotic solution derived in Sec. II. Figure 4 shows the predicted EA at a source frequency of 1 kHz with z_s = 1.0 m, d=0.2 m for different effective flow resistivities. We see that the agreement between the asymptotic solution (dotted lines) and IES (solid lines) is remarkable. The predictions due to these two schemes almost overlap over a broad range of r. This excellent agreement supports the neglect of the lateral wave contribution in the theoretical analysis detailed in Sec. II.

It is of interest to point out that although the predicted EA decreases with increase of the effective flow resistivity, this is not always the case for lower frequencies as shown in Fig. 5 when f=100 Hz, z_s=1.0 m, and d=0.05 m. The predicted EA are within a few decibels for various values of σ at a range of less than 1 m. As the range increases the predicted EA shows greater discrepancies. Again, the agreement in this set of predictions is excellent, except when the effective flow resitivity is small and the receiver is located at a distance within about a wavelength from the source. However, the discrepancies are no more than about 0.5 dB in this case with source frequency of 100 Hz. These discrepancies become more pronounced at lower frequencies.

We illustrate in Fig. 6 a numerical simulation of the source frequency of 1 Hz, the effective flow resistivity of 1 kPa s m^{-2}, the source height of 1.0 m and the receiver depth of 0.5 m. Again, we wish to caution the suitability of using Delany and Bazley model[26] for frequencies below 100 Hz. Indeed, the accuracy of this model has not validated experimentally at such a low frequency. However, we reiter-

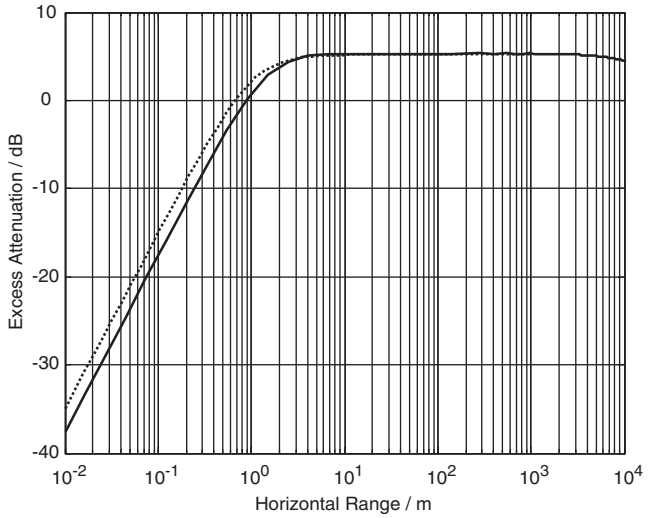


FIG. 6. Same as Fig. 3 except that the receiver is located at 0.5 m below the air/ground interface and the frequency of the source is 1 Hz. In the predictions, the effective flow resistivity of 1 kPa s m^{-2} is used.

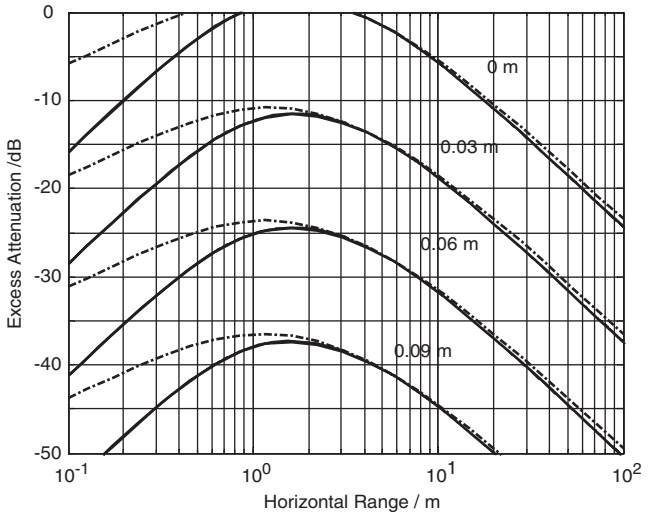


FIG. 7. Comparison of the uniform asymptotic formula with Brekhovskikh's solution, see Eq. (10). The source with frequency of 1 kHz is located at 1 m above ground with receiver located at 0, 0.03, 0.06, and 0.09 m below the porous ground with an effective flow resistivity of 100 kPa s m^{-2}. The predictions according to the asymptotic solution given in Eq. (51) (solid lined) agree to within the thickness of the lines for the prediction according to the integrated exact solution. The predictions according to Brekhovskikh's asymptotic formula are shown in dash-dotted lined.

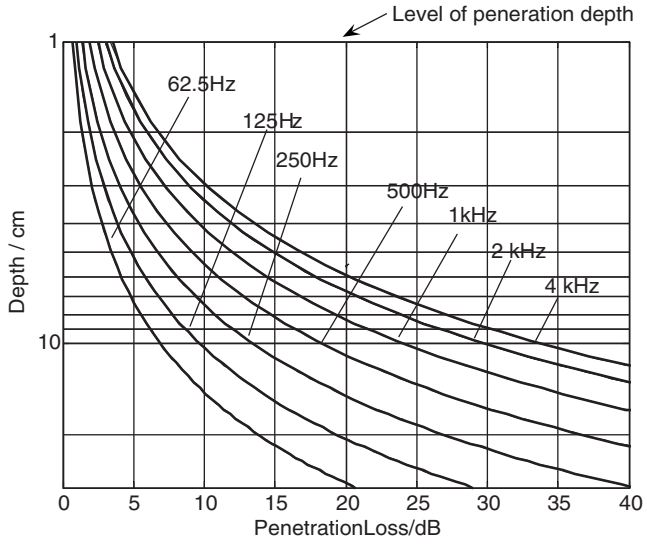


FIG. 8. Penetration loss of sound levels versus the depth of receiver below the ground for various frequencies. The source is located at 1 m above the ground and 5 m horizontal distance from the receiver. For the porous ground, a four-parameter model is used compute its acoustical properties with an effective flow resistivity of 80 kPa s m^{-2}, porosity of 0.37 and tortuosity of 1.8.

ate that our purpose is mainly to test the robustness and accuracy (as compared with other computational intensive schemes) of our uniform asymptotic formula under this extreme situation. From Fig. 6, we see both schemes deviate when the horizontal range is less than about 2 m ($\sim 0.06\lambda$). This is understandable because the asymptotic solution is not valid in this condition.

Figure 7 presents numerical predictions of the sound fields at various depths (0, 3, 6, and 9 cm) in a porous ground along a horizontal range varying from 0.1 to 100 m. The Delany and Bazley model[26] with an effective flow resistivity of 100 kPa s m^{-2} is used in the plots. The source has a frequency of 1 kHz and is located at 1 m above the ground. In this set of plots, we aim to highlight the inadequacy of ignoring the contribution from the surface wave pole in predicting the penetration of sound in the porous ground. Using Eq. (10), we compare the predicted sound fields with the IES. Generally, a reasonably good agreement can be found for the horizontal range varies from 2 to 20 m. For this set of source/receiver geometry, noticeable discrepancies are apparent at longer ranges and the error becomes even more significant below about 1 m for all depths. Indeed, there are many situations with different source/receiver geometries and impedance parameters that Eq. (10) fails to give accurate numerical solutions. These details are not shown for brevity. On the other hand, the current uniform asymptotic formula [see Eq. (51)] agrees with IES to within the thickness of the line in this set of predictions. We wish to make a remark that the numerical predictions according to the uniform asymptotic formula are in excellent agreement with the IES in most situations even though its derivation is based on the following assumptions:

(i) $k_0 r \to \infty$,

(ii) the use of an heuristic approximation by replacing $\mathrm{s}_1 n_1 \cos \mu_1$ with the effective admittance $\bar{\beta}_e$ given in Eq. (11a), and

(iii) the contribution from the lateral wave pole is ignored.

Finally, we end this section by considering a practical situation to predict the penetration of sound in washed sand. The high effective flow resistivity/low-frequency approximation may be used to simplify a four-parameter scheme.[22] In this case, the index of refraction and the ratio of complex densities can be approximated by

$$n_1 = \frac{k_1}{k_0} = \sqrt{\gamma}(a_k q^2 + i\sigma_e/(\omega_0 \rho_0))^{1/2} \tag{58a}$$

and

$$\mathrm{s}_1 = \frac{\rho_0}{\rho_1} = \frac{\Omega\sqrt{\gamma}}{n_1}(q^2 a_z + i\sigma_e/(\omega_0 \rho_0))^{-1/2}, \tag{58b}$$

where γ is the ratio of specific heats, ω is the angular frequency, a_k and a_z are dependent on the shape of pores, σ_e is the effective flow resistivity, Ω is the porosity, and q^2 is the tortuosity. For the ground composed of typical washed sand, the following parameters are chosen for a_k, a_z, σ_e, Ω, and q^2: 1.106, 1.535, 90 kPa m s^{-2}, 0.37 and 1.8, respectively. These parameters are comparable to those used in a study by Frederickson *et al.*[27] for the acoustic characterization of a rigid-porous medium. To present the results at different frequencies, it is convenient to introduce a term known as the penetration loss in sound levels (PL) which is defined as

$$\mathrm{PL} = 20 \log_{10}[p_0(r, z_s, 0)/p_1(r, z_s, D)], \tag{59}$$

where $p_0(r, z_s, 0)$ and $p_1(r, z_s, D)$ can be computed by Eqs. (53) and (51), respectively. In Fig. 8, we display PL versus D with the horizontal range, r set at 5 m for the source frequen-

cies varying at 62.5, 125, 250, 500, 1000, 2000, and 4000 Hz, respectively. We also use the IES to compute $p_0(r,z_s,0)$, $p_1(r,z_s,D)$, and PL as a further step of quality assurance. Numerical results obtained by both methods are in excellent agreement—the lines of plots obtained from both results coincide in Fig. 8. The penetration loss in the sound level increases almost linearly with the depth of the receiver. The gradient of PL reduces with the increasing frequency of the source.

Suppose a reference microphone is placed on the interface boundary. A term known as an effective penetration depth may then be defined as a vertical distance below the ground where the penetration loss in sound levels is 20 dB at that position. We can see from Fig. 8 that the effective penetration depth at 62.5 Hz is about 29 cm. It is decreased to about 6 cm for the source operated at a frequency of 4 kHz. The effective penetration depth of sound may therefore provide an estimate of a physical limit for detecting buried objects in a rigid-porous ground using a microphone. The limitation is not relevant however to techniques of detection using the surface (solid particle) velocity, since the wave induced in the solid frame of the ground is much less attenuated with depth.

V. CONCLUDING REMARKS

An accurate asymptotic formula has been derived to compute the sound field below an extended reaction ground from an airborne source. It is shown by comparing with the fast field program that the asymptotic formula is valid for a wide range of source/receiver geometries even though the condition of $k_0 r \to \infty$ (where k_0 is the wave number and r is the horizontal separation between the source and receiver) is not satisfied. The formula provides a first step for a better understanding of the sound penetration into a porous ground which can be treated as the physical phenomenon of the sound refraction from the air into a rigid-porous ground.

ACKNOWLEDGMENTS

The author is grateful to Keith Attenborough for his comments on the manuscript. He also thanks the two anonymous reviewers and the Associate Editor, Vladimir Ostashev, and for many useful comments.

APPENDIX

For the horizontal separation, r, is large, we can reformulate Eq. (20) in an iterative form as follows:

$$S_{j+1} = z_s\sqrt{1-S_j^2}\left[r - D\sqrt{\frac{1-S_j^2}{S_j^2+n_1^2-1}}\right]^{-1}, \tag{A1}$$

where $S=\cos\vartheta_0$ and the subscript j denotes the value of S at the jth iteration. For different values of z_s, r, D, and n_1, the unknown parameter $\cos\vartheta_0$ can be determined by iterating Eq. (A1) successively until the required accuracy is reached. A close examination of Eq. (A1) suggests that when $r\to\infty$, the term $\sqrt{1-S_j^2}$ tends to 1 and $\cos\vartheta_0$ can be approximated by $z_s[r-D/\sqrt{n_1^2-1}]^{-1}$. This gives an initial value for $j=0$ which will then be used to start the iteration process described earlier.

The previous iteration scheme works rather well for large r but it fails to converge to the solution when r becomes smaller. It is necessary to develop an alternative method for finding the solution for Eq. (20). To obtain the appropriate root, we define an auxiliary function, f, in terms of $S(\equiv\cos\vartheta_0)$ as

$$f(S,D) = z_s\sqrt{(1-S^2)(S^2+n_1^2-1)} + DS\sqrt{1-S^2} - rS\sqrt{(S^2+n_1^2-1)}. \tag{A2}$$

The appropriate solution for Eq. (20) can then be determined by solving the auxiliary function, $f(S)=0$. It can be solved by the standard Newton–Raphson iteration method.

An improved solution for $f(S,D)=0$ is found by using the following iteration until the expected accuracy is obtained:

$$S_{j+1} = S_j - f_j/f_j' \tag{A3}$$

where $f_j \equiv f(S_j,D)$, $f_j' \equiv f'(S_j,D)$ and the prime denotes the derivative of the function with respect to its argument.

The Newton–Raphson method starts with an initial 'guess' value close to the required root in order to get the 'correct' solution. It is remarkable that the incident angle is $\cos^{-1}(z_s/\sqrt{z_s^2+r^2})$ when $D=0$. Hence, a good starting value will be chosen with $S=z_s/\sqrt{z_s^2+r^2}$. For some combination of geometrical configurations, source frequencies and the index of refraction, it is found necessary to replace D with $l\delta D$ where $l=1,2,3,\ldots$, and δD is the step length. We start the iteration with $l=1$ and obtain the correct solution at δD. This correct solution will then be used as the starting value for the next value of l, and so on, until the required depth is reached. Smaller step lengths enhance the chance of obtaining the correct solution at a given D at the expense of higher computational time. A step length between 5 mm and 5 cm is found necessary to obtain a converged solution for most practical situations. We also find that the Newton–Raphson method works well for the receiver located for small r but the solution may become unstable when r is large.

The two iteration schemes described earlier are complementary to each other: Eq. (A1) is employed for longer ranges and Eq. (A3) is used for shorter ranges. There is an overlapped region where either one of these two iteration schemes may be used to give correct solutions. In the present paper, there is no attempt to optimize r for switching from one iteration scheme to the other. Nevertheless, the angle of refraction can be determined for a variety of source/receiver configurations and different types of ground surfaces. This in turn, will enable us to compute the sound fields below the ground through the use of Eq. (51).

[1] K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).
[2] C. J. Hickey, D. Leary, J. F. Allard, and M. Henry, "Impedance and Brewster angle measurement for thick porous layers," J. Acoust. Soc. Am. **118**, 1503–1509 (2005).
[3] D. L. Berry, S. N. Chandler Wilde, and K. Attenborough, "Acsutic scattering by a near surface obstacle in a rigid poeous medium," J. Sound Vib. **170**, 161–179 (1994).
[4] G. Kristensson and S. Ström, "Scattering from buried inhomogeneities—a

general three-dimensional formalism," J. Acoust. Soc. Am. **64**, 917–936 (1978).
[5]C. G. Don and A. J. Rogers, "Using acoustic impulses to identify a buried nonmetallic object," J. Acoust. Soc. Am. **95**, 2837–2838 (1994).
[6]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. High frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[7]K. Attenborough, "Solid particle motion induced by a point source above a poroelastic half-space," J. Acoust. Soc. Am. **86**, 1085–1092 (1987).
[8]J. M. Sabatier, H. E. Bass, L. N. Bolen, K. Attenborough, and V. V. S. S. Sastry, "The interaction of airborne sound with the porous ground: Theoretical formulation," J. Acoust. Soc. Am. **79**, 1345–1352 (1986).
[9]S. Tooms, S. Taherzadeh, and K. Attenborough, "Sound propagation in a refracting fluid above a layered fluid-saturated porous elastic material," J. Acoust. Soc. Am. **93**, 173–181 (1993).
[10]D. Velea, R. Waxler, and J. M Sabatier, "An effective fluid model for landmine detection using acoustic to seismic coupling," J. Acoust. Soc. Am. **115**, 1993–2000 (2004).
[11]K. M. Li, T. Waters-Fuller, and K. Attenborough, "Sound propagation from a point source over extended-reaction ground," J. Acoust. Soc. Am. **104**, 679–685 (1998).
[12]F. W. J. Olver, "Bessel functions of integer order," in *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970), Chap. 9.
[13]D. I. Paul, "Acoustic radiation from a point source in the presence of two media," J. Acoust. Soc. Am. **29**, 1102–1109 (1957).
[14]T. L. Richards, K. Attenborough, N. W. Heap, and A. P. Watson, "Penetration of sound from a point source into a rigid-porous medium," J. Acoust. Soc. Am. **78**, 956–963 (1985).
[15]There are apparent typographical errors in Eqs. (1) and (2) of Ref. 14. The constant term should be $1/4\pi$ instead of $M/4\pi$ in Eq. (1). The constant for the third term at the right-hand side of Eq. (2) should be 2 instead of $2/M$. The horizontal wave number, κ (their K), is missing in the Hankel function of Eq. (1).
[16]In this section, formulas given from other publications have been adapted so that the symbols used are the same as in the present paper.
[17]L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980), pp. 277–286.
[18]K. Attenborough, S. I. Hayek, and J. M. Lawther, "Propagation of sound above a porous half-space," J. Acoust. Soc. Am. **85**, 1493–1501 (1985).
[19]J. Nicolas, J. L. Berry, and G. A. Daigle, "Propagation of sound above a finite layer of snow," J. Acoust. Soc. Am. **77**, 67–73 (1985).
[20]C. F. Chien and W. W. Soroka, "Sound propagation along an impedance ground," J. Sound Vib. **43**, 9–20 (1975).
[21]R. Wong, *Asymptotic Approximations of Integrals* (Academic, London, 1989).
[22]K. Attenborough, K. M. Li, and K. Horoshenkov, *Predicting Outdoor Sound* (Taylor & Francis, London, 2006), Chap. 2.
[23]S. N. Chandler-Wilde and D. C. Hothersall, "A uniformly valid far field asymptotic expansion of the Green function for two-dimensional propagation above a homogeneous impedance plane," J. Sound Vib. **182**, 665–675 (1995).
[24]L. C. Sutherland and G. A. Daigle, "Atmospheric sound propagation" *Handbook of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1998), Chap. 28.
[25]T. L. Richards and K. Attenborough, "Accurate FFT-based Hankel transform for predictions of outdoor sound propagation," J. Sound Vib. **109**, 157–167 (1986).
[26]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[27]C. K. Federickson, J. M. Sabatier, and R. Raspet, "Acoustic chacterization of rigid-frame air-filled porous media using both reflection and transmission measurements," J. Acoust. Soc. Am. **99**, 1326–1332 (1996).

Thermal boundary layer effects on the acoustical impedance of enclosures and consequences for acoustical sensing devices[a)]

Stephen C. Thompson[b)]
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Janice L. LoPresti
Knowles Electronics, LLC, 1151 Maplewood Drive, Itasca, Illinois 60143

(Received 13 August 2007; revised 9 December 2007; accepted 11 December 2007)

Expressions are derived for the acoustical impedance of a rectangular enclosure and of a finite annular cylindrical enclosure. The derivation is valid throughout the frequency range in which all dimensions of the enclosure are much less than the wavelength. The results are valid throughout the range from adiabatic to isothermal conditions in the enclosure. The effect is equivalent to placing an additional, frequency-dependent complex impedance in parallel with the adiabatic compliance. As the thermal boundary layer grows to fill the cavity, the reactive part of the impedance varies smoothly from the adiabatic value to the isothermal value. In some microphones, this change in cavity stiffness is sufficient to modify the sensitivity. The resistive part of the additional cavity impedance varies as the inverse square root of frequency at high frequencies where the boundary layer has not grown to fill the enclosure. The thermal modification gives rise to a thermal noise whose spectral density varies asymptotically as $1/f^{3/2}$ above the isothermal transition frequency.
 [DOI: 10.1121/1.2832314]

PACS number(s): 43.38.Ar, 43.38.Kb, 43.35.Ud [AJZ] Pages: 1364–1370

I. INTRODUCTION

The acoustical impedance of enclosed air volumes is important in determining the performance of many acoustic sensing devices. It is well understood, for example, that the stiffness of the air volume behind the diaphragm of a condenser microphone must be considered to understand its sensitivity and frequency response. In many cases, the adiabatic compliance of the enclosure is sufficient to describe the effect. In small enclosures and at low frequencies, the thermal boundary layer can become a significant fraction of the total volume of the enclosure, causing interesting modifications to the cavity impedance. This paper will discuss the calculation of the acoustical impedance of enclosures with special attention to the frequency dependence of the acoustic resistance. Expressions for the rectangular enclosure and the annular cylindrical enclosure are provided. These shapes find practical application as enclosures in acoustical sensing devices. An example is shown in which the use of these results provides an improved match to the measured sensitivity of a miniature microphone.

Daniels[1] has provided a solution for the impedance of a spherical enclosure, and a method for calculation of the impedance of enclosures of other geometries. This work will follow the methods of Daniels. We are concerned with a harmonic acoustic pressure excitation given by $pe^{j\omega t}$, superimposed on the static pressure p_0. The pressure gives rise to an oscillating temperature variation $Te^{j\omega t}$ in the volume of the enclosure superimposed on the ambient temperature T_0. Combining Daniels' Eqs. (4), (6), and (8), the impedance of the enclosure can be written as

$$Z = \frac{\gamma p_0}{j\omega V\left[\gamma - (\gamma - 1)\dfrac{\bar{T}}{T_a}\right]} = \frac{Z_a}{\left[\gamma - (\gamma - 1)\dfrac{\bar{T}}{T_a}\right]}, \tag{1}$$

where V is the volume of the enclosure; $\gamma = C_p/C_v$ is the ratio of specific heats for the gas in the enclosure; C_p is the specific heat at constant pressure; C_v is the specific heat at constant volume; $\bar{T}$ is the value of the oscillatory temperature amplitude in the enclosure averaged over the volume of the enclosure; $T_a = pT_0(\gamma - 1)/p_0\gamma$ is the oscillatory temperature variation assuming purely adiabatic compression; $C_0 = V/\gamma p_0$ is the adiabatic compliance of the enclosure (when $\bar{T} = T_a$); and $Z_a = 1/(j\omega C_0)$ is the adiabatic impedance of the enclosure.

In the limit of large enclosures where $\bar{T} = T_a$, Eq. (1) reduces to the adiabatic impedance. For smaller enclosures, thermal transfer at the walls causes a significant spatial variation in T in the enclosure that modifies the impedance according to Eq. (1). If we set

$$Y = 1 - \frac{\bar{T}}{T_a} \tag{2}$$

then

[a)]Portions of this work were presented in "Acoustic noise in microphones from the thermal boundary layer resistance in enclosures," J. Acoust. Soc. Am. **119**, 3377 (2006).

[b)]Electronic mail: steve.thompson@psu.edu

$$Z = \frac{1}{j\omega C_0 + j\omega C_0(\gamma - 1)Y} = \frac{1}{\frac{1}{Z_a} + \frac{1}{Z_t}}, \tag{3}$$

which shows that the impedance can be represented as the parallel combination of the adiabatic impedance Z_a, and a thermal correction impedance Z_t given by

$$Z_t = \frac{1}{j\omega C_0(\gamma - 1)Y}, \tag{4}$$

which is caused by thermal transfer to the walls.

This separation of the total impedance into the adiabatic impedance and the thermal correction impedance is only a mathematical convenience, and is not meant to imply that these two impedances can be associated with separate regions of the enclosure volume. Equation (3) is convenient because it allows the impedances to be discussed as if they were to actual impedances connected in parallel. In particular it means that, in a frequency region where Z_a and Z_t differ in magnitude by a large factor, the larger value has a negligible effect on the magnitude of the total impedance.

Equation (1) shows that the cavity impedance can be calculated from the distribution of the oscillatory temperature amplitude in the enclosure. Daniels[1] shows that this temperature distribution can be found as a solution of

$$\nabla^2 T = \beta^2 (T - T_a) \tag{5}$$

where $\beta^2 = j\omega\rho_0 C_p/\kappa$; ρ_0 is the static density; and κ is the thermal conductivity of the gas in the cavity.

The general character of solutions to Eq. (5) can be illustrated in the simplest case of a semi-infinite space with a single isothermal boundary wall in the $z=0$ plane. It is straightforward to show that the solution is

$$T = T_a(1 - e^{-\beta z}) = T_a(1 - e^{-(1+j)z/\delta_\kappa}), \tag{6}$$

where

$$\delta_\kappa = \sqrt{\frac{2\kappa}{\omega\rho_0 C_p}} \tag{7}$$

is the thickness of the thermal boundary layer. Far from the wall, where $z \gg \delta_k$, the temperature variation is in phase with the acoustic pressure, and has its adiabatic value. Within the boundary layer the temperature variation decreases to zero at the wall, and there is a phase difference between the temperature and the pressure which approaches 45° at the wall. Figure 1 shows the magnitude and phase of the oscillatory temperature variation near an infinite wall for a frequency of 1 Hz. At other frequencies, Eq. (7) states that the thickness of the boundary layer varies inversely as the square root of frequency. The top axis in Fig. 1 shows T as a function of z/δ_κ.

Daniels[1] offered solutions for the temperature distribution and the enclosure impedance for three enclosure geometries in which T can be expressed as a function of a single spatial variable. These cases are a sphere, an infinite cylinder, and infinite parallel plates. Of these, only the sphere is truly an enclosure. The two idealized infinite cases can be seen as limiting cases for a cylinder much longer than its diameter

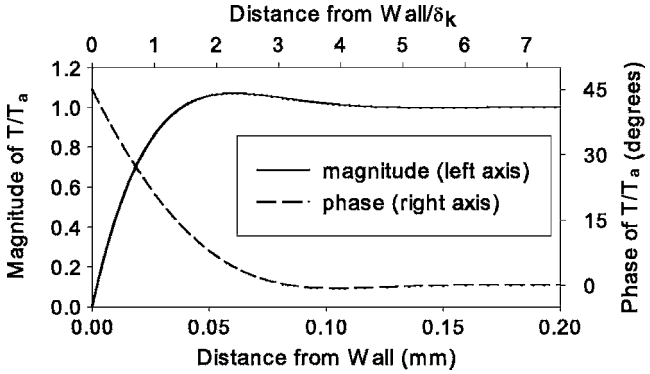


FIG. 1. Magnitude and phase of the oscillatory temperature variation near a wall. The frequency is 1 Hz.

and a thin flat enclosure whose cross-sectional dimensions are much larger than its thickness. A solution for the impedance of a finite cylinder was given by Biagi and Cook.[2] It can be shown that their expression for the impedance of a finite cylinder is limited properly to that for the infinite cylinder when the length is much greater than the radius, and to that for the infinite plates when the radius is much greater than the length. This paper gives solutions for the rectangular box and finite annular cylindrical enclosures.

II. THE SPHERICAL ENCLOSURE

To understand the nature of the enclosure impedance and its potential effect on sensor noise, one may use the example of the spherical enclosure. The temperature distribution in a spherical cavity is given by

$$\frac{T}{T_a} = 1 - \frac{\beta a}{\beta r}\frac{\sinh \beta r}{\sinh \beta a}, \tag{8}$$

where a is the radius of the sphere. Figure 2 shows the magnitude of the thermal distribution at several frequencies in a sphere with radius of 2 mm. Figure 3 shows the phase of this temperature distribution relative to the phase of the driving pressure. At higher frequencies where $\beta a \gg 1$, the boundary layer is confined to the region near the walls and the temperature approaches the adiabatic value through much of the volume. At low frequencies, however, the cavity becomes nearly isothermal at the wall temperature.

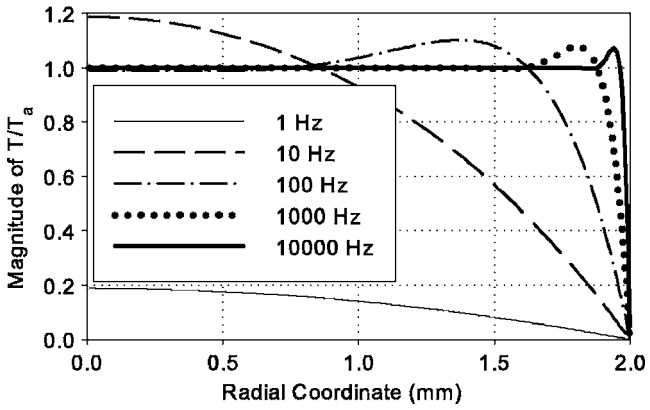


FIG. 2. Magnitude of the thermal distribution in a spherical cavity with radius 2 mm.

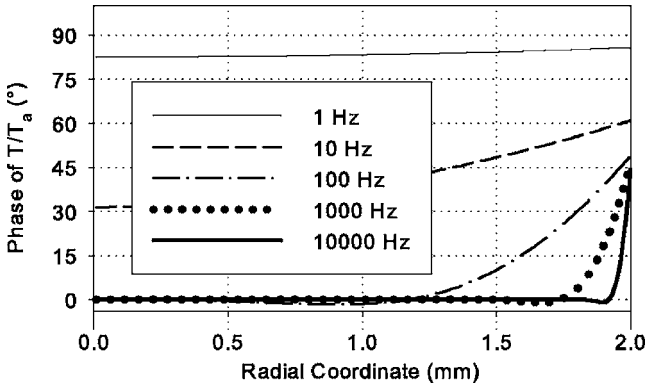


FIG. 3. Phase of the thermal distribution in a spherical cavity with radius 2 mm.

From Eqs. (2), (4), and (8), the thermal correction impedance is

$$Z_t = \frac{\beta^2 a^2}{3j\omega C_0(\gamma-1)(\beta a \coth \beta a - 1)}. \tag{9}$$

The real and imaginary parts of this expression are the resistance and reactance of the thermal correction impedance. The reactive impedance is shown in Fig. 4, and the resistance is shown in Fig. 5. The asymptotic behavior of the thermal correction impedance at low and high frequencies can be calculated from Eq. (9). At low frequencies,

$$Z_t \underset{\omega\to 0}{\to} \frac{1}{j\omega C_0(\gamma-1)} + \frac{a^2\rho_0 C_p}{15C_0(\gamma-1)\kappa}. \tag{10}$$

The resistive impedance is much smaller than the reactance, and can often be neglected. However, the resistance may be important in considering the noise generated by the enclosure impedance as will be discussed in Sec. V. The low frequency thermal reactance is a constant compliance in parallel with the adiabatic compliance according to Eq. (3). The two compliances in parallel create the isothermal compliance at low frequencies.

At high frequencies, Eq. (9) becomes

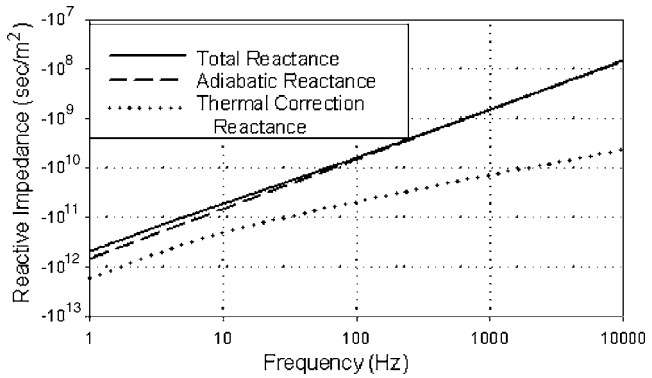


FIG. 4. The adiabatic impedance and the thermal correction impedance combine to yield the total impedance. Values are shown for a sphere of 2 mm radius.

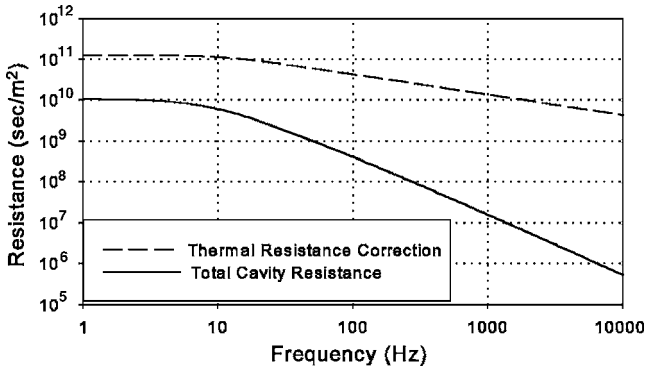


FIG. 5. The thermal correction resistance is constant at low frequencies and falls as $f^{-1/2}$ at high frequencies. The total cavity resistance falls as $f^{-3/2}$ at high frequencies. Values are shown for a sphere of 2 mm radius.

$$Z_t \underset{\omega\to\infty}{\to} \frac{V}{j\omega C_0(\gamma-1)S\delta_\kappa} + \frac{\gamma p_0}{\omega(\gamma-1)S\delta_\kappa}, \tag{11}$$

where S is the surface area of the enclosure. In Eq. (11), the magnitudes of the real and imaginary parts are equal, although they are written differently to facilitate the following explanation. Note that the boundary layer thickness δ_κ varies inversely as the square root of frequency so that at high frequencies, $Z_t \gg Z_a$. From Eq. (3), the much larger value of Z_t has a negligible effect on the total impedance. The thermal resistance is inversely proportional to the surface area of the enclosure and decreases as the square root of frequency. It has little effect on the magnitude of the total impedance, but the resistive loss may be important in determining the Q of a cavity at resonance, or in calculating the thermal noise in the enclosure. The dependence of the thermal resistance on the surface area is reasonable, because significant energy loss occurs only within the boundary layer. At high frequencies the boundary layer is confined to the surface, and the volume of the boundary layer is simply the surface area times the boundary layer thickness.

Note that in Eq. (11), the only reference to the geometry is the total volume and the total surface area. The authors believe that the equation in this form holds for any enclosure shape. This has been proven analytically for a sphere, for an infinite cylinder, and for infinite parallel plates. The numerical calculations in Sec. III show it to be true for a rectangular box of any aspect ratio. Figure 5 shows the thermal correction impedance and the total impedance for a sphere with radius 2 mm. At high frequencies the thermal correction resistance in Eq. (11) varies as $f^{-1/2}$ because of the implicit frequency dependence of δ_κ. By combining Eqs. (11) and (3) it can be shown that the total resistance at high frequencies varies as $f^{-3/2}$.

III. THE RECTANGULAR ENCLOSURE

Stinson[3] has provided a solution for the propagation of waves in a tube of rectangular cross section, including the effects of thermal transfer at the walls. The present derivation has much in common with that work. To find the temperature distribution in a rectangular enclosure with sides having lengths a, b, and c, we seek a solution to Eq. (5) with the boundary conditions

$$T = 0 \text{ when} \begin{cases} x = 0, & x = a \\ y = 0, & y = b \\ z = 0, & z = c. \end{cases} \tag{12}$$

The solution has the form

$$T = \sum_{\ell,m,n=1}^{\infty} A_{\ell mn} \sin \alpha_\ell x \sin \alpha_m y \sin \alpha_n z \tag{13}$$

where the boundary conditions require that

$$\alpha_\ell = \begin{cases} \dfrac{\ell \pi}{2a} & \text{when } \ell \text{ is odd} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

with similar expressions for α_m and α_n. To solve for the $A_{\ell mn}$, one should substitute Eqs. (13) and (14) into Eq. (5), multiply by the solution eigenfunction, and integrate over the volume of the enclosure. The full solution for the temperature distribution in the enclosure is then

$$\frac{T}{T_a} = \frac{64\beta^2}{\pi^3} \sum_{\ell,m,n=1,3,\ldots}^{\infty} \frac{\sin \alpha_\ell x \sin \alpha_m y \sin \alpha_n z}{\ell mn[\beta^2 + (\alpha_\ell^2 + \alpha_m^2 + \alpha_n^2)]} \tag{15}$$

and the average value of the temperature in the enclosure is

$$\frac{\bar{T}}{T_a} = \frac{512\beta^2}{\pi^6} \sum_{\ell,m,n=1,3,\ldots}^{\infty} \frac{1}{\ell^2 m^2 n^2[\beta^2 + (\alpha_\ell^2 + \alpha_m^2 + \alpha_n^2)]}. \tag{16}$$

The value of Y in Eq. (2) is

$$Y = 1 - \frac{\bar{T}}{T_a} = \frac{512}{\pi^6} \sum_{\ell,m,n=1,3,\ldots}^{\infty} \frac{(\alpha_\ell^2 + \alpha_m^2 + \alpha_n^2)}{\ell^2 m^2 n^2[\beta^2 + (\alpha_\ell^2 + \alpha_m^2 + \alpha_n^2)]}. \tag{17}$$

The thermal impedance of the rectangular enclosure is calculated by substituting this expression into Eq. (4). The total impedance of the enclosure is than calculated from Eq. (3). These calculations have been done numerically for five rectangular enclosures that have the same surface area. The first is a 4 mm cube. The others have thicknesses of 2, 1, 0.5, and 0.25 mm. The areas of the two square surfaces of each enclosure are adjusted to provide the same surface area, 96 mm^2. Figure 6 shows the total reactive impedances of the enclosures with the largest volume (the cube) and the smallest volume. These graphs show the same general behavior as the sphere, approaching the adiabatic value at high frequencies and approaching the isothermal value for low frequencies. Figure 7 shows the total resistance and the thermal correction resistance for the five rectangular enclosures. The thermal correction resistances for all of these enclosures have the same high frequency limiting behavior. Not only do the impedances fall with frequency at the same rate, they also have the same impedance value. This supports the statement in Sec. II that Eq. (11) holds for enclosures of any shape.

As the frequency is decreased, the resistance increases until the boundary layer has grown to fill the enclosure. When the cavity has become approximately isothermal, the resistance is constant with further decrease in frequency. The high frequency asymptotic behavior of the rectangular enclosure has the same frequency dependence as the spherical enclosure; namely, that the thermal correction resistance varies as $f^{-1/2}$ and the total resistance varies as $f^{-3/2}$.

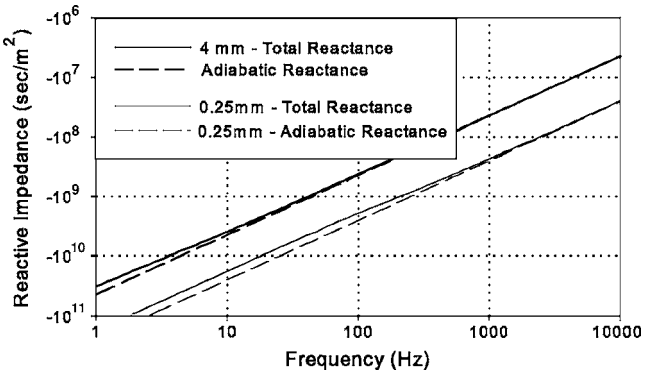


FIG. 6. The total reactive impedance of two different rectangular enclosures that have the same surface area.

IV. THE ANNULAR CYLINDRICAL ENCLOSURE

The right annular cylinder is a shape that is used as the internal volume of some commercial microphones. To calculate the acoustical impedance of this enclosure, we seek a solution to Eq. (5) in cylindrical coordinates with boundary conditions

$$T(r, \phi, z) = 0 \text{ when} \begin{cases} r = r_i, & r = r_o \\ z = 0, & r = 2b \end{cases}. \tag{18}$$

The complete formal method of solution can be found in standard mathematical references such as Hayek.[4] The solution has the form

$$T = \sum_{m,n=1}^{\infty} \left(A_{\text{mn}} J_0\left(\frac{\alpha_m r}{r_o}\right) + B_{mn} Y_0\left(\frac{\alpha_m r}{r_o}\right) \right) \times (\sin k_n z + D_n \cos k_n z). \tag{19}$$

The boundary conditions in z require that $D_n = 0$ and $k_n = n\pi z/2b$. The radial boundary conditions require

$$A_{mn} J_0(\alpha_m) + B_{mn} Y_0(\alpha_m) = 0,$$

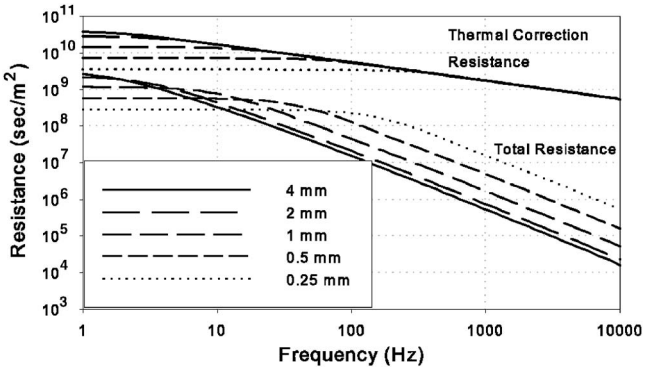


FIG. 7. The thermal resistance of a family of rectangular enclosures with 96 mm surface area. Two sides of the box are squares. The legend gives the thickness of the box.

$$A_{mn}J_0(\sigma\alpha_m) + B_{mn}Y_0(\sigma\alpha_m) = 0, \tag{20}$$

where

$$\sigma = \frac{r_i}{r_o}. \tag{21}$$

In order for Eq. (15) to have a nontrivial solution, it is necessary that

$$J_0(\alpha_m)Y_0(\sigma\alpha_m) - J_0(\sigma\alpha_m)Y_0(\alpha_m) = 0. \tag{22}$$

The radial eigenvalues α_m are the roots of this equation. These roots are tabulated for some values of σ, but must be calculated numerically in the general case. From Eq. (22) we may write

$$\frac{J_0(\alpha_m)}{Y_0(\alpha_m)} = \frac{J_0(\sigma\alpha_m)}{Y_0(\sigma\alpha_m)} = C_m. \tag{23}$$

The boundary conditions then allow the calculation of B_{mn} as

$$B_{mn} = -A_{mn}C_m \tag{24}$$

and T becomes

$$T = \sum_{m,n=1}^{\infty} A_{mn}\left[J_0\left(\frac{\alpha_m r}{r_o}\right) - C_m Y_0\left(\frac{\alpha_m r}{r_o}\right)\right]\sin\frac{n\pi z}{2b}. \tag{25}$$

The procedure to determine the A_{mn} is

(1) substitute this expression for T into Eq. (5);
(2) multiply both sides of the resulting equation by the eigenfunction with indices $k\ell$;
(3) integrate over the volume of the annular enclosure, and;
(4) use the orthogonality of the eigenfunctions to solve for A_{mn}.

This process is straightforward, if tedious, and results in

$$A_{mn} = \frac{8\beta^2 T_a}{n\pi\alpha_m}\frac{1}{\left[\beta^2 + \left(\frac{\alpha_m}{r_o}\right)^2 + \frac{n^2\pi^2}{4b^2}\right][G_m(\alpha_m) + \sigma G_m(\sigma\alpha_m)]}, \tag{26}$$

where

$$G_m(\alpha_m) = J_1(\alpha_m) - C_m Y_1(\alpha_m). \tag{27}$$

The average of the temperature over the volume of the annulus is

$$\bar{T} = \frac{1}{\pi(r_o^2 - r_i^2)2b}\int_0^{2b}\int_0^a 2\pi r T dr dz = \frac{32\beta^2 T_a}{\pi^2(1-\sigma^2)} \times \sum_{m,n=1}^{\infty}\frac{1}{n^2\alpha_m^2}\frac{[G_m(\alpha_m) - \sigma G_m(\sigma\alpha_m)]}{\left[\beta^2 + \frac{\alpha_m^2}{r_o^2} + \frac{n^2\pi^2}{4b^2}\right][G_m(\alpha_m) + \sigma G_m(\sigma\alpha_m)]} \tag{28}$$

from which the impedance of the enclosure can be calculated using Eq. (1).

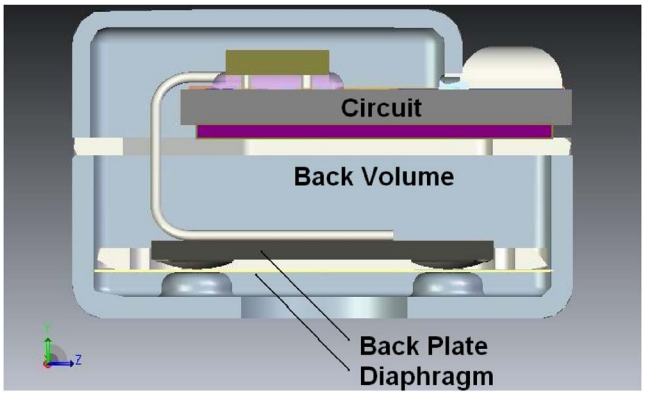


FIG. 8. (Color online) The microphone is a Knowles EM modified to have a single large hole in the case to admit sound to the diaphragm.

V. EFFECT ON SENSING DEVICES

The acoustical impedance of enclosed air volumes is important in determining the performance of many acoustic sensing devices. It is well understood, for example, that the stiffness of the air volume behind the diaphragm of a condenser microphone must be considered to understand its sensitivity and frequency response. Most often, however, the models used to calculate the performance of sensors use only the adiabatic compliance of the enclosure.[5–8] There are at least some cases in which the neglect of thermal effects leads to significant errors.

An example of this behavior is an experimental miniature microphone made by Knowles Electronics shown in Fig. 8. It is very similar to the "salt shaker" microphone previously described by Thompson *et al.*,[6] except that it has a single large hole in the case to admit sound pressure to the diaphragm. The internal construction of the microphone is that of the Knowles EM family. A cross section of this microphone is shown in Fig. 9. Sound enters the microphone through the large hole in the case at the bottom of Fig. 9. The diaphragm is suspended in front of an electret coated backplate, also sometime called the charge plate. The diaphragm is very thin and difficult to see in Fig. 8, but it is present at the position indicated. The thin air layer between the diaphragm and backplate is connected to the air in the backvolume by small channels that are not shown in Fig. 8. The

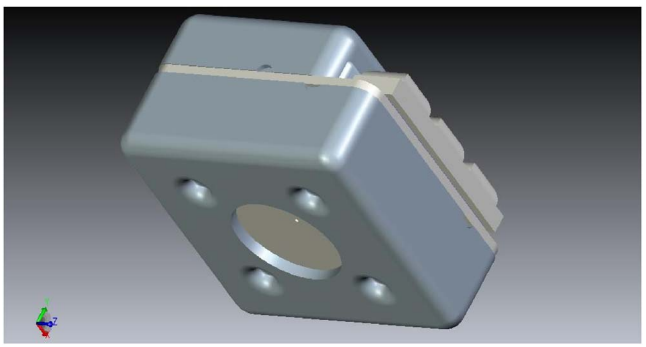


FIG. 9. (Color online) Cross section of the special microphone. The backvolume consists of two connected sections, one immediately above the backplate and the other near the top of the figure surrounding the amplifier circuit.

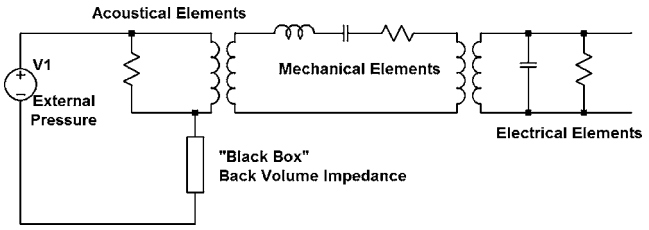


FIG. 10. The analog circuit model for the microphone has elements in the acoustical, mechanical, and electrical domains. The backvolume impedance, shown as a black box, is calculated by the methods of this paper.

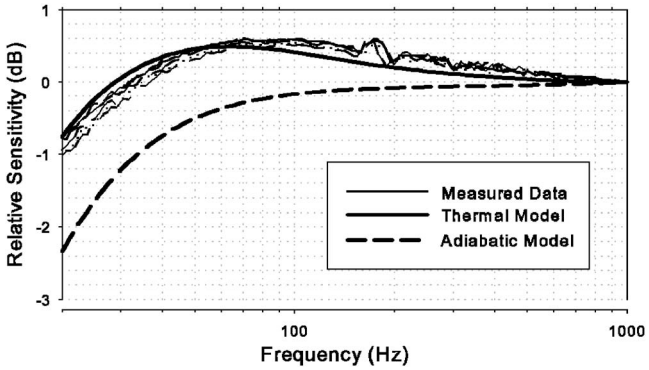


FIG. 11. Measured sensitivity data for five samples of the EM microphone compared with two theoretical calculations. The curve labeled "Adiabatic Model" uses only the adiabatic impedance for the backvolume. The curve labeled "Thermal Model" uses a numerical approximation to Eq. (17).

backvolume consists of two connected spaces above and below the circuit mounting plate. Each of these spaces is approximately rectangular. A wire connected to the backplate makes electrical connection to the amplifier chip mounted inside the backvolume. The diaphragm has a thin metallic coating to make it conductive, and this coating is electrically connected to ground potential. The diaphragm and backplate thus form the two plates of a capacitor whose output voltage drives the amplifier to produce the output signal.

The external acoustic pressure drives a small motion of the diaphragm. The diaphragm moves near the charged electret on the backplate and creates the signal voltage that is amplified to become the microphone output signal. The resistance of the air film between the diaphragm and backplate provides some damping to the motion of the diaphragm. The air film connects through holes at the edges and center of the diaphragm to the air contained in the microphone backvolume. There is also a small hole pierced in the diaphragm to equalize static pressure in the backvolume to the ambient air pressure.

The analog circuit that models the behavior of the microphone is shown in Fig. 10. The impedance analogy is used throughout the model. In the acoustical domain, the variables are acoustic pressure and volume velocity. As external pressure enters the microphone, a part of the volume flow goes through the diaphragm vent hole into the backvolume. The other part of the volume flow moves with the mechanical motion of the diaphragm. The acoustic flow through the vent and the diaphragm both enter the backvolume. The impedance of the backvolume is shown as a "black box" because, in general, it cannot be represented as a single simple component. Instead, the impedance of this backvolume enclosure should include the effects of thermal transfer as discussed in this paper. The acoustomechanical transformer in Fig. 10 performs the area transformation to the mechanical variables of force and velocity of the diaphragm. The mechanical elements represent the mass, compliance, and mechanical resistance of the diaphragm. The electromechanical transformer converts the diaphragm force and velocity into the electrical voltage and current that are the output of the microphone. The electrical terminals in Fig. 10 would normally be connected to an internal buffer amplifier which is not included here.

The thickness dimension of the open space in the backvolume is approximately 0.5 mm, although the actual thickness varies somewhat due to components on the circuit board and other assembly features of the microphone. Above approximately 300 Hz, the boundary layer thickness is a negligible fraction of the total enclosure thickness. Thus the adiabatic approximation should give an adequate model of the microphone behavior. Below 300 Hz, the effects of thermal transfer at the enclosure walls may become apparent. Figure 11 shows the low frequency sensitivity measured for the microphone compared with the model of Fig. 10. The midband sensitivity differences in the microphones have been removed in Fig. 11 by normalizing the responses to unity at 1 kHz. The model is calculated in two configurations. The first is using just the adiabatic compliance for the backvolume. The second uses a numerical approximation to the full expression of Eq. (17). Clearly the low frequency behavior is much better modeled when the thermal effects are included.

The change in microphone sensitivity from thermal effects is caused by the change in the compliance of the enclosure at low frequencies. The thermal resistance has little effect on the sensitivity. The thermal resistance could possibly affect the internal noise of the microphone if the noise from this resistance were comparable to or greater than the other thermal noise sources in the microphone. Figure 12 shows the frequency dependence of the total enclosure resistance for the microphone. The spectral density of the acoustic noise pressure from this resistance is[9]

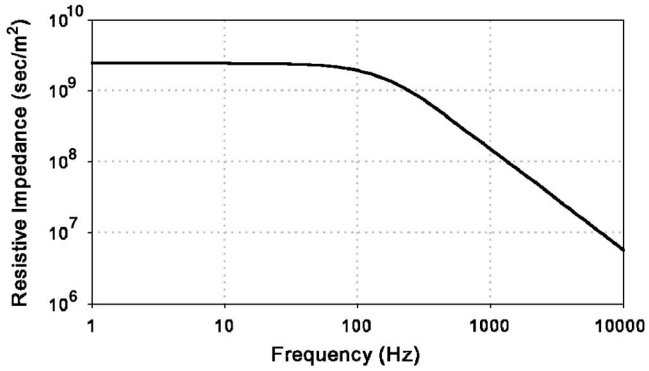


FIG. 12. The calculated enclosure resistance of the backvolume for the EM microphone.

$$N = \sqrt{4kT_0R}, \tag{29}$$

where k is Boltzmann's constant, T_0 is the ambient temperature, and R is the enclosure resistance. The noise voltage from this source is lower by more than an order of magnitude than that measured by Thompson *et al.* for essentially the same microphone. In this case, the thermal resistance of the enclosure has a negligible effect on the total internal noise of the microphone.

It is interesting to note, however, that the shape of the enclosure resistance curve of Fig. 12 is very similar to the shapes of the resistances of all of the other enclosure shapes calculated here. In particular, the low frequency asymptote is horizontal and the high frequency asymptote has the slope $f^{-3/2}$, and consequently the power spectral density of the noise it creates has the same asymptotic behavior. This noise is one of the sources present in any acoustic sensing device in which enclosed air volumes are present. In several studies, Zuckerwar and his colleagues have measured the acoustic noise present in a variety of acoustic sensors. In all cases, they have identified that the power spectral density of the acoustic noise at the low frequency limit of their measurement varies with frequency as f^{-1}. To date, no explanation has been offered for this spectral shape. The authors speculate that the thermal noise in the enclosed air spaces may be at least a part of the explanation for this spectral shape.

VI. CONCLUSIONS

This paper has reviewed the calculation of the acoustic impedance of enclosures and provided a new solution for the finite cylindrical annulus. The frequency-dependent change in acoustical compliance of the enclosure impedance can affect the sensitivity of some acoustic sensing devices for frequencies below that of the adiabatic to isothermal transition. The internal noise due to enclosure resistance is constant with frequency below the transition and falls as $f^{-3/2}$ above the transition.

ACKNOWLEDGMENTS

The authors would like to express deep gratitude to Sabih Hayek for discussions that were instrumental in the solution provided in Sec. IV. There have also been several helpful conversations with Daniel Warren, Steve Garrett, and Tom Gabrielson.

[1]F. B. Daniels, "Acoustical impedance of enclosures," J. Acoust. Soc. Am. **19**, 569–571 (1947).
[2]F. Biagi and R. K. Cook, "Acoustic impedance of a right circular cylindrical enclosure," J. Acoust. Soc. Am. **26**, 506–509 (1954).
[3]M. R. Stinson, "The propagation of plane sound waves in narrow and side circular tubes, and generalization to uniform tubes of arbitrary cross-sectional shape," J. Acoust. Soc. Am. **89**, 550–558 (1991).
[4]S. I. Hayek, *Advanced Mathematical Methods in Science and Engineering* (Dekker, New York, 2001), pp. 314, 336–342.
[5]A. J. Zuckerwar and K. C. Thi Ngo, "Measured $1/f$ noise in the membrane motion of condenser microphones," J. Acoust. Soc. Am. **95**, 1419–1425 (1994).
[6]S. C. Thompson, J. L. LoPresti, E. M. Ring, H. G. Nepomuceno, J. J. Beard, W. J. Ballad, and E. C. Carlson, "Noise in miniature microphones," J. Acoust. Soc. Am. **111**, 861–866 (2002).
[7]A. J. Zuckerwar, T. R. Kuhn, and R. M. Serbyn, "Background noise in piezoresistive, electret condenser, and ceramic microphones," J. Acoust. Soc. Am. **113**, 3179–3187 (2003).
[8]R. Dieme, Gijs Bosman, and Toshikazu Nishida, "Sources of excess noise in silicon piezoresistive microphones," J. Acoust. Soc. Am. **119**, 2710–2720 (2006).
[9]L. L. Beranek, *Acoustical Measurements* (Acoustical Society of America, Melville, 1993), p. 190.

Elastic wave field computation in multilayered nonplanar solid structures: A mesh-free semianalytical approach

Sourav Banerjee[a] and Tribikram Kundu[b]
Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721

(Received 9 August 2007; revised 15 November 2007; accepted 19 November 2007)

Multilayered solid structures made of isotropic, transversely isotropic, or general anisotropic materials are frequently used in aerospace, mechanical, and civil structures. Ultrasonic fields developed in such structures by finite size transducers simulating actual experiments in laboratories or in the field have not been rigorously studied. Several attempts to compute the ultrasonic field inside solid media have been made based on approximate paraxial methods like the classical ray tracing and multi-Gaussian beam models. These approximate methods have several limitations. A new semianalytical method is adopted in this article to model elastic wave field in multilayered solid structures with planar or nonplanar interfaces generated by finite size transducers. A general formulation good for both isotropic and anisotropic solids is presented in this article. A variety of conditions have been incorporated in the formulation including irregularities at the interfaces. The method presented here requires frequency domain displacement and stress Green's functions. Due to the presence of different materials in the problem geometry various elastodynamic Green's functions for different materials are used in the formulation. Expressions of displacement and stress Green's functions for isotropic and anisotropic solids as well as for the fluid media are presented. Computed results are verified by checking the stress and displacement continuity conditions across the interface of two different solids of a bimetal plate and investigating if the results for a corrugated plate with very small corrugation match with the flat plate results.
 [DOI: 10.1121/1.2823258]

PACS number(s): 43.40.At, 43.20.Fn, 43.20.Bi, 43.20.El [RLW] Pages: 1371–1382

I. INTRODUCTION

Elastic wave propagation in a solid medium has been a research topic for over a century. Waves in layered media are comparatively new. Various multilayered structures are being used in layered heat resistant materials, layered insulators, layered cylindrical pipe structures, etc. Ultrasonic, sonic, and other nondestructive evaluations of these materials are needed for assuring integrity of different structural components.

Wave propagation in a layered medium has been a popular problem since mid-19th century.[1–4] Research activity in this area increased considerably after the frequent use of fiber reinforced composite plates in engineering structures.[5–8] Irrespective of the type of material layers (isotropic or anisotropic) and field of applications (aerospace, civil, or electronic industries) the integrity testing needs a better understanding of wave propagation in multilayered structures. For a better understanding and interpretation of the experimental results analytical or numerical modeling of the experimental scenario is also very important. Hence, ultrasonic field modeling in multilayered structures has become popular and important to researchers.

A conventional solution of ultrasonic fields in any material generated by transducers is widely known as the ray tracing technique; it involves solution of the Eikonal equation.[9] Dynamic ray tracing is a major advancement in this field known as the paraxial method or paraxial approximation model.[10] In the above-mentioned literature the elastic wave propagation in multilayered solids has been analyzed by assuming a plane wave striking the solid. In real-life experiments however, the elastic waves generated by ultrasonic transducers of finite dimension have neither plane nor spherical wave front. No researcher has yet modeled the problem of multilayered solids excited by an ultrasonic beam of finite width generated by a transducer of finite dimension that is placed at a finite distance from the solid. However, this is the real problem geometry for most ultrasonic experiments and needs to be thoroughly investigated.

In this article an efficient mesh-free semianalytical tool called Distributed Point Source Method (DPSM) has been adopted to model the ultrasonic field generated by ultrasonic transducers of finite dimension in multilayered structures when both the structure and the transducers are immersed in a fluid (couplant fluid). Thus, it numerically simulates the ultrasonic experiments for multilayered plate inspection. In the bygone years several researchers have developed various techniques to solve the above-mentioned problems. Some of the most popular methods are finite element method (FEM),[11] boundary element method (BEM),[12,13] multi-Gaussian beam model (MGBM),[14–18] charge simulation technique,[19] multiple multipole program (MMP),[20–22] etc. Although the MGBM technique has some computational advantage over other techniques mentioned here, it also has a number of limitations similar to those of other paraxial models. For example, MGBM cannot correctly model the critical

[a]Electronic mail: sourav@email.arizona.edu
[b]Author to whom correspondence should be addressed. Electronic mail: tkundu@email.arizona.edu

reflection phenomenon; it cannot model a transmitted beam at an interface near grazing incidence. This technique also fails if the interface has different curvatures (gradually varying curvature), or when the radius of curvature of the transducer is small, as observed in acoustic microscopy experiments with its tightly focused lens. The multi-Gaussian beam model has not been extended to multilayered structures yet. Therefore, it is important to have a generalized formulation for wave field modeling in multilayered structures without any ray tracing. Such methods should be geometry and material independent. The method should also be capable of handling any geometrical defect or material defect without much difficulty. On the other hand, FEM and BEM packages are very CPU intensive, they require huge amounts of computation memory and time for execution. Similar to MMP another technique was followed by Sanchez-Sesma and Esquivel[23] to solve SH wave scattering problem for ground motion calculation on alluvial valley. They considered plane wave incidence and formulated the problem in terms of a system of Fredholm integral equations of the first kind with integration paths outside the problem boundary. Similar technique for P and SV wave diffraction problem for different surface topographies was reported by Wong[24] and Dravinski and Mossessian.[25] However, none of these studies considered any finite dimensional source for elastic waves as done in this article, also they required Green's functions for solid half-spaces and therefore, these techniques cannot be extended to non-half-space problem geometries.

Distributed point source method (DPSM) is a semianalytical technique, which is different from BEM or MMP. DPSM introduces one layer of point sources distributed close to the problem boundary to model the effect of the boundary, and two layers of point sources next to an interface to model the interface effect. Boundary integral equation (BIE) technique and its numerical version, the BEM do not introduce such fictitious sources. However, mathematical justification of introducing artificial point sources to model the boundary effect can be found in the literature on "indirect" BIE (or IBIE and IBEM).[26–29] The need of introducing point sources for modeling irregular boundaries has been mathematically proven in those publications. In spite of some similarities, it should be pointed out that there are some major differences between DPSM and IBIE. In IBIE, point sources are placed on smooth boundary surfaces. The integral equation thus formed in IBIE has a singular kernel that gives rise to the Fredholm integral equation of the second kind. In the wave propagation literature the IBIE technique has been used to compute the scattered field from an irregular boundary after knowing the incident field in absence of the boundary irregularities. In IBIE formulation the simple incident field is generated by either plane body waves (P, SV, or SH) or surface waves—Rayleigh or Love waves. In DPSM, formulation point sources are placed near the boundary and interfaces but slightly outside the domain of interest to avoid singularities. Avoiding the need of solving singular integral equations by moving the points of singularity outside the domain of interest is an advantage of DPSM. Also, in DPSM formulation the boundary and interfaces can have corners and the incident field can be treated as an unknown field, unlike IBIE where the incident field is assumed to be a predefined known field. A bounded beam generated by a transducer of finite dimension is considered as the incident field in the DPSM-based formulation presented here.

The DPSM technique for ultrasonic field modeling was first developed by Placko and Kundu.[30] They successfully used this technique to model ultrasonic fields in a homogeneous fluid medium. It was then extended to more complicated problems by Banerjee and co-workers such as multilayered fluid structure,[31] solid half space and plate, with and without anomalies,[32–34] wave field modeling in solid plates with nonplanar boundaries[35] and acoustic microscope modeling by Kundu *et al.*[36]

In the DPSM technique it is necessary to have stress and displacement Green's functions for fluid and solid media. Displacement Green's functions in frequency domain for isotropic solids are available in the literature.[37] However, explicit expressions for frequency domain stress Green's functions in isotropic material are only available with some approximations. Explicit expressions of displacement and stress Green's functions for isotropic solids without any approximations have been given by Banerjee and Kundu.[32]

A two-layered copper–aluminum bimetal plate is considered for the present study. Ultrasonic fields are calculated in this structure for both planar and nonplanar boundaries for various angles of incidence of bounded ultrasonic beams. Computed results clearly show how the ultrasonic energy in the plate decays with the distance of propagation in flat and corrugated plates.

II. GENERAL FORMULATION

A multilayered solid structure is considered. The layers in the multilayered solid can consist of isotropic, transversely isotropic, general anisotropic, or a combination of these materials. In this general formulation any type of anisotropy can be considered and both plane and nonplanar interfaces can be modeled. On two sides of the solid structure Fluids 1 and 2 are used as the coupling fluid that transmits ultrasonic waves from the ultrasonic transducers to the solid structure [see Fig. 1(a)]. To model the ultrasonic field inside the multilayered structure and the fluid, the DPSM technique is employed. Let us consider a multilayered solid structure of n different layers made of n different materials. Hence, there are $n+1$ interfaces say $I_1, I_2, \ldots, I_{n+1}$. Following the basics of the DPSM technique two sets of point sources are distributed on two sides of every interface making a total of $2(n+1)$ sets of point sources adjacent to $(n+1)$ interfaces, as shown in Fig. 1(a). Each interface acts as a transmitter as well as a reflector of elastic wave energy generated by the ultrasonic transducers. Point sources are also distributed behind the transducer faces. Transducer sources are denoted as $\mathbf{A}_S$ and $\mathbf{A}_R$ in Fig. 1(a). Two sets of source strength vectors corresponding to the mth interface I_m are denoted by $\mathbf{A}_m$ (for sources located just above the mth interface) and $\mathbf{A}_m^*$ (for sources located just below the mth interface). The sources with source strength $\mathbf{A}_m$ generate the ultrasonic field in the solid below it and the sources with source strength $\mathbf{A}_m^*$ generate the ultrasonic field in the solid above it. The total ultrasonic field in each me-

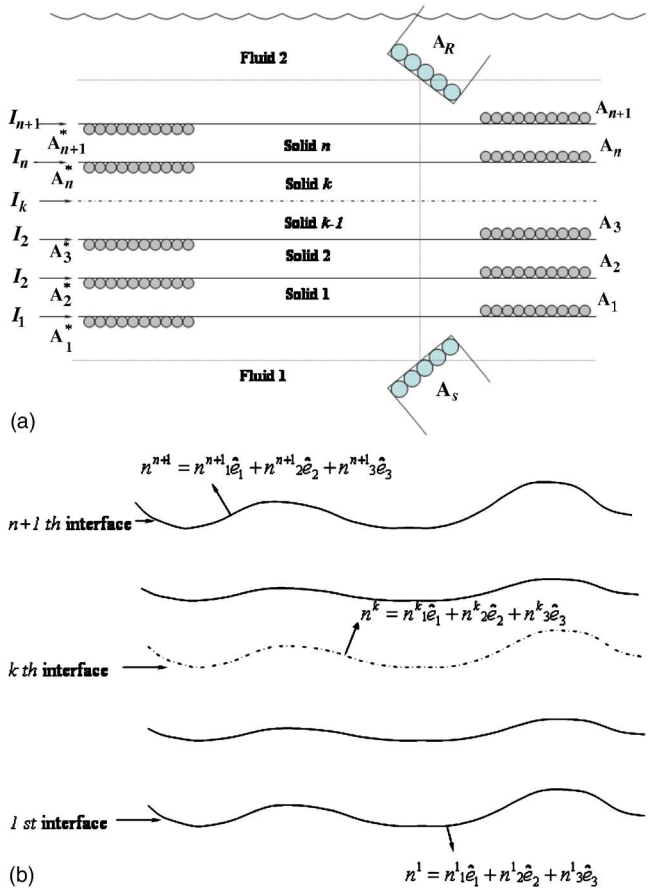


FIG. 1. (Color online) (a) Problem geometry—multilayered solid plate and two ultrasonic transducers immersed in water. Small circles show distributions of point sources, point sources are placed at the centers of these circles. (b) Multilayered solid with nonplanar interfaces and boundaries.

dium is obtained by superimposing the fields generated by two sets of sources. The ultrasonic field generated in Fluid 1 is the summation of fields generated by $\mathbf{A}_S$ and $\mathbf{A}_1$. The ultrasonic field generated in Solid 1 is the summation of fields generated by $\mathbf{A}_1^*$ and $\mathbf{A}_2$. Similarly the ultrasonic field in Solid 2 is the summation of fields generated by $\mathbf{A}_2^*$ and $\mathbf{A}_3$. The field in Solid n is the summation of fields generated by $\mathbf{A}_n^*$ and $\mathbf{A}_{n+1}$, and the field in Fluid 2 is obtained from $\mathbf{A}_{n+1}^*$ and $\mathbf{A}_R$.

A. Matrix formulation

The particle velocity and pressure in fluids at the fluid-solid interfaces can be expressed in matrix form.[18] Let $T1$ and $T2$ be two different sets of target points in the fluid below and above Interfaces 1 and $n+1$, respectively. The velocity at the target points can be written as

$$\mathbf{V}_{T1} = \mathbf{M}_{(T1)S}\mathbf{A}_S + \mathbf{M}_{(T1)1}\mathbf{A}_1, \tag{1}$$

$$\mathbf{V}_{T2} = \mathbf{M}_{(T2)R}\mathbf{A}_R + \mathbf{M}_{(T2)n+1}{}^*\mathbf{A}_{n+1}^*. \tag{2}$$

Similarly, the pressure fields at the target points are

$$\mathbf{PR}_{T1} = \mathbf{PR}_{T1}^s + \mathbf{PR}_{T1}^1 = \mathbf{Q}_{(T1)S}\mathbf{A}_S + \mathbf{Q}_{(T1)1}\mathbf{A}_1, \tag{3}$$

$$\mathbf{PR}_{T2} = \mathbf{PR}_{T2}^s + \mathbf{PR}_{T2}^{n+1}{}^* = \mathbf{Q}_{(T2)R}\mathbf{A}_R + \mathbf{Q}_{(T2)n+1}{}^*\mathbf{A}_{n+1}^*. \tag{4}$$

Elements of the matrices written in Eqs. (1)–(4) are given in Spies.[18] If the boundary surfaces of the interfaces are nonplanar [Fig. 1(b)], normal stress and normal displacement directions at the interface vary from point to point. The direction cosine of the nonplanar interface at any point on the interface can be defined as $n=(n_1e_1+n_2e_2+n_3e_3)$. Projections of unit normal (n) on x_1, x_2, and x_3 axes can be defined according to the problem definition. Point sources needed for modeling solids are different from those used for fluid modeling. Every point source for the solid modeling has three different force components in three mutually perpendicular directions. Stress at point **x** generated by a point source acting at point **y** in a solid can be obtained from stress Green's functions of that material. For a point source acting at **y** in an isotropic solid, the stresses developed at point **x** have been given by Banerjee and Kundu.[38] Displacement Green's functions for transversely isotropic and anisotropic materials have been also presented in the literature.[39–42] However, as the stress Green's function for anisotropic materials are derivatives of displacement Green's functions those need to be calculated numerically in the absence of any closed form expressions. Assuming a point force acting along the x_j direction, stresses at point **x** on the boundary of the interface can be denoted by the stress tensor $\boldsymbol{\sigma}^j=\sigma_{mn}^j$. To obtain stress components that are perpendicular and parallel to a nonplanar interface this stress tensor is to be transformed using the standard stress transformation law $\underline{\boldsymbol{\sigma}'^j}=\mathbf{T}\boldsymbol{\sigma}^j\mathbf{T}^T$, where **T** is the transformation matrix. The transformation matrix for point **x** on the interface depends on the interface geometry. If the normal to the interface does not have an x_3 component then $n_3=0$, and the matrix **T** can be written as

$$\mathbf{T} = \begin{bmatrix} n_2 & -n_1 & 0 \\ n_1 & n_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

To define the boundary conditions at point **x**, the normal stress perpendicular to the boundary surface and two shear stress components, parallel to the boundary surface are needed. Considering a set of M point sources distributed on the interfaces, the normal stress and the shear stress components can be defined as

$$S'_{22} = \sum_{m=1}^{M} \left[\left(\sigma'_{22}{}^1\right)^m P_1^m + \left(\sigma'_{22}{}^2\right)^m P_2^m + \left(\sigma'_{22}{}^3\right)^m P_3^m \right] = \sum_{m=1}^{M} \underline{\mathbf{s}'_{22}{}^m}(\mathbf{P})^m, \tag{6}$$

$$S'_{21} = \sum_{m=1}^{M} \left[\left(\sigma'_{21}{}^1\right)^m P_1^m + \left(\sigma'_{21}{}^2\right)^m P_2^m + \left(\sigma'_{21}{}^3\right)^m P_3^m \right] = \sum_{m=1}^{M} \underline{\mathbf{s}'_{21}{}^m}(\mathbf{P})^m, \tag{7}$$

$$S'_{23} = \sum_{m=1}^{M} \left[\left(\sigma'^{1}_{23}\right)^m P_1^m + \left(\sigma'^{2}_{23}\right)^m P_2^m + \left(\sigma'^{3}_{23}\right)^m P_3^m \right]$$
$$= \sum_{m=1}^{M} \underline{\mathbf{s}'^{m}_{23}} (\mathbf{P})^m, \tag{8}$$

where the definition of **P** is given in Banerjee and Kundu.[38] Displacements at point **x** generated by a point source acting at point **y** in a solid can be obtained from the displacement Green's functions.[38–42] Three displacement components at **x** generated by a point force acting along the x_j direction are denoted by G_{1j}, G_{2j}, and G_{3j}. Considering the same point force along the x_j direction, the normal displacement of the solid surface at **x** can be written as

$$u_n^j = G_{1j} n_1 + G_{2j} n_2 + G_{3j} n_3. \tag{9}$$

If a set of M point sources are distributed on the interface, then the normal displacement at point **x** on the interface can be written as

$$\underline{un}_{=} \sum_{m=1}^{M} \begin{pmatrix} \left(G_{11} n_1 + G_{21} n_2 + G_{31} n_3\right)^m P_1^m + \left(G_{12} n_1 + G_{22} n_2 + G_{32} n_3\right)^m P_2^m + \\ \left(G_{13} n_1 + G_{23} n_2 + G_{33} n_3\right)^m P_3^m \end{pmatrix} = \sum_{m=1}^{M} \underline{\mathbf{Gn}}^m \mathbf{P}^m. \tag{10}$$

Let T be a set of target points in the mth solid layer. Normal displacements at these points (T) on the interface can be written in the following form:

$$\underline{\mathbf{un}}_T = \mathbf{DSn}_{Tm^*} \mathbf{A}_m^* + \mathbf{DSn}_{Tm+1} \mathbf{A}_{m+1}. \tag{11}$$

Similarly transformed normal stress and shear stresses at the target points (T) at the mth solid layer can be written as

$$\mathbf{s22}'_T = \mathbf{S22}'_{Tm^*} \mathbf{A}_m^* + \mathbf{S22}'_{Tm+1} \mathbf{A}_{m+1}, \tag{12a}$$

$$\mathbf{s21}'_T = \mathbf{S21}'_{Tm^*} \mathbf{A}_m^* + \mathbf{S21}'_{Tm+1} \mathbf{A}_{m+1}, \tag{12b}$$

$$\mathbf{s23}'_T = \mathbf{S23}'_{Tm^*} \mathbf{A}_m^* + \mathbf{S23}'_{Tm+1} \mathbf{A}_{m+1}. \tag{12c}$$

Matrices $\mathbf{DSn}_{TS}$, $\mathbf{S22}'_{TS}$, $\mathbf{S21}'_{TS}$, and $\mathbf{S23}'_{TS}$ are given in Banerjee and Kundu.[38] Subscripts T and S denote sets of target and source points, respectively. The displacement components at point **x** generated by a point source at **y** are also available in the literature[38] and are not repeated here.

For a nonplanar corrugated interface using the direction cosines (n_i) of the normal vector to the interface, the displacement component normal to the corrugated interface at point **x** can be written as

$$u_{fn} = u_1 n_1 + u_2 n_2 + u_3 n_3. \tag{13}$$

Therefore, in presence of transducers (see Fig. 1) the displacement of the fluid at Interfaces 1 and $n+1$ can be written as

$$\underline{\mathbf{Un}}_{I1} = \left(\left(\mathbf{DF3}_{(I1)S}\right) n_3 + \left(\mathbf{DF2}_{(I1)S}\right) n_2 + \left(\mathbf{DF1}_{(I1)S}\right) n_1\right) \mathbf{A}_S + \left(\left(\mathbf{DF3}_{(I1)1}\right) n_3 + \left(\mathbf{DF2}_{(I1)1}\right) n_2 + \left(\mathbf{DF1}_{(I1)1}\right) n_1\right) \mathbf{A}_1, \tag{14}$$

$$\underline{\mathbf{Un}}_{In+1} = \left(\left(\mathbf{DF3}_{(In+1)R}\right) n_3 + \left(\mathbf{DF2}_{(In+1)R}\right) n_2 + \left(\mathbf{DF1}_{(In+1)R}\right) n_1\right) \mathbf{A}_R + \left(\left(\mathbf{DF3}_{(In+1)n+1^*}\right) n_3 + \left(\mathbf{DF2}_{(In+1)n+1^*}\right) n_2 + \left(\mathbf{DF1}_{(In+1)n+1^*}\right) n_1\right) \mathbf{A}_{n+1}^* \tag{15}$$

or,

$$\underline{\mathbf{Un}}_{I1} = \mathbf{DFn}_{(I1)S} \mathbf{A}_S + \mathbf{DFn}_{(I1)1} \mathbf{A}_1, \tag{16}$$

$$\underline{\mathbf{Un}}_{In+1} = \mathbf{DFn}_{(In+1)R} \mathbf{A}_R + \mathbf{DFn}_{(In+1)n+1^*} \mathbf{A}_{n+1}^*. \tag{17}$$

Matrix $\mathbf{DFn}_{TS}$ is given in Banerjee and Kundu.[38] Let us consider a set of target points on "Interface k" that is denoted as Ik. The transformed normal stress and shear stress matrices for the referenced target points can be written as

$$\mathbf{s22}'_{Ik} = \mathbf{S22}'_{Ikk^*} \mathbf{A}_k^* + \mathbf{S22}'_{Ik(k+1)} \mathbf{A}_{k+1}, \tag{18a}$$

$$\mathbf{s21}'_{Ik} = \mathbf{S21}'_{Ikk^*} \mathbf{A}_k^* + \mathbf{S21}'_{Ik(k+1)} \mathbf{A}_{k+1}, \tag{18b}$$

$$\mathbf{s23}'_{Ik} = \mathbf{S23}'_{Ikk^*} \mathbf{A}_k^* + \mathbf{S23}'_{Ik(k+1)} \mathbf{A}_{k+1}. \tag{18c}$$

Similarly on Interface $k+1$ the set of target points are denoted as $Ik+1$ and the transformed normal and shear stresses on the interface can be written as

$$\mathbf{s22}'_{Ik+1} = \mathbf{S22}'_{Ik+1(k)^*} \mathbf{A}_k^* + \mathbf{S22}'_{Ik+1(k+1)} \mathbf{A}_{k+1}, \tag{19a}$$

$$\mathbf{s21}'_{Ik+1} = \mathbf{S21}'_{Ik+1(k)^*} \mathbf{A}_k^* + \mathbf{S21}'_{Ik+1(k+1)} \mathbf{A}_{k+1}, \tag{19b}$$

$$\mathbf{s23}'_{Ik+1} = \mathbf{S23}'_{Ik+1(k)^*} \mathbf{A}_k^* + \mathbf{S23}'_{Ik+1(k+1)} \mathbf{A}_{k+1}. \tag{19c}$$

Inside the solid at interfaces Ik and $Ik+1$ three displacement components can be written as

$$\underline{\mathbf{ui}}_{Ik} = \mathbf{DSi}_{(Ik)k^*} \mathbf{A}_k^* + \mathbf{DSi}_{(Ik)k+1} \mathbf{A}_{k+1}, \tag{20}$$

$$\underline{\mathbf{ui}}_{Ik+1} = \mathbf{DSi}_{(Ik+1)k^*}\mathbf{A}_k^* + \mathbf{DSi}_{(Ik+1)k+1}\mathbf{A}_{k+1}, \tag{21}$$

where **i** takes values 1, 2, and 3 to represent three displacement components along x_1, x_2, and x_3 directions, respectively. The previous matrix expressions are valid for all $n+1$ interfaces. Therefore, k can take values from 1 to $n+1$.

B. Boundary and continuity conditions

Across a fluid–solid interface the displacement component normal to the interface should be continuous. Also, at the interface, the transformed negative normal stress ($-\mathbf{s22}'$) in the solid and the pressure in the fluid should be continuous and the shear stresses must vanish. On the other hand, across a solid–solid interface three displacement components, the normal stress and two shear stresses parallel to the interface must be continuous.

Let the normal velocities at the transducer faces be $\mathbf{V}_{S0}$ and $\mathbf{V}_{R0}$, for the lower and upper transducers, respectively. The boundary conditions at the transducer faces are

$$\mathbf{M}_{SS}\mathbf{A}_S + \mathbf{M}_{S1}\mathbf{A}_1 = \mathbf{V}_{S0} \tag{22}$$

$$\mathbf{M}_{R(n+1)^*}\mathbf{A}_{n+1}^* + \mathbf{M}_{RR}\mathbf{A}_R = \mathbf{V}_{R0}. \tag{23}$$

At the fluid–solid interfaces, the continuity conditions can be expressed similar to what Banerjee and Kundu[32] presented. At the kth solid–solid interface the continuity conditions are given by

$$\mathbf{S22}'_{k(k-1)}\mathbf{A}_{k-1}^* + \mathbf{S22}'_{kk}\mathbf{A}_k = \mathbf{S22}'_{kk^*}\mathbf{A}_k^* + \mathbf{S22}'_{k(k+1)}\mathbf{A}_{k+1}, \tag{24}$$

$$\mathbf{S21}'_{k(k-1)^*}\mathbf{A}_{k-1}^* + \mathbf{S21}'_{kk}\mathbf{A}_k = \mathbf{S21}'_{kk^*}\mathbf{A}_k^* + \mathbf{S21}'_{k(k+1)}\mathbf{A}_{k+1}, \tag{25}$$

$$\mathbf{S23}'_{k(k-1)^*}\mathbf{A}_{k-1}^* + \mathbf{S23}'_{kk}\mathbf{A}_k = \mathbf{S23}'_{kk^*}\mathbf{A}_k^* + \mathbf{S23}'_{k(k+1)}\mathbf{A}_{k+1}. \tag{26}$$

Continuity of three displacement components at the kth interface gives

$$\mathbf{DSi}_{k(k-1)^*}\mathbf{A}_{k-1}^* + \mathbf{DSi}_{kk}\mathbf{A}_k = \mathbf{DSi}_{kk^*}\mathbf{A}_k^* + \mathbf{DSi}_{k(k+1)}\mathbf{A}_{k+1}. \tag{27}$$

The boundary and continuity conditions can be written in matrix form[38]:

$$[\mathbf{M}]\{\mathbf{\Psi}\} = \{\mathbf{V}\}. \tag{28}$$

The vectors $\{\mathbf{\Psi}\}$ and $\{\mathbf{V}\}$ are called source strength vector and force vector, respectively, which can be written as

$$\{\mathbf{\Psi}\}^{\text{Transpose}} = \{\mathbf{A}_S\ \mathbf{A}_1\ \mathbf{A}_1^*\ \mathbf{A}_2\ \mathbf{A}_2^*\ \cdots\ \mathbf{A}_n\ \mathbf{A}_n^*\ \mathbf{A}_{n+1}\ \mathbf{A}_{n+1}^*\ \mathbf{A}_R\}^{\text{Transpose}}, \tag{29}$$

$$\{\mathbf{V}\}^{\text{Transpose}} = \{\mathbf{V}_{S0}\ \ \mathbf{0}\ \ \mathbf{0}\ \ \mathbf{0}\ \ \cdots\ \ \mathbf{0}\ \ \mathbf{0}\ \ \mathbf{0}\ \ \mathbf{V}_{R0}\}^{\text{Transpose}}. \tag{30}$$

The matrix **M** is given by

$$[\mathbf{M}] = \begin{bmatrix}
\mathbf{M}_{SS} & \mathbf{M}_{S1} & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{Q}_{1S} & \mathbf{Q}_{11} & \mathbf{S22}'_{11^*} & \mathbf{S22}'_{12} & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mathbf{S21}'_{11^*} & \mathbf{S21}'_{12} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mathbf{S23}'_{11^*} & \mathbf{S23}'_{12} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\mathbf{DFn}_{1S} & \mathbf{DFn}_{11} & -\mathbf{DSn}_{11^*} & -\mathbf{DSn}_{12} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mathbf{S22}'_{21^*} & \mathbf{S22}'_{22} & \mathbf{S22}'_{22^*} & \mathbf{S22}'_{23} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \mathbf{S21}'_{21^*} & \mathbf{S21}'_{22} & \mathbf{S21}'_{22^*} & \mathbf{S21}'_{23} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \mathbf{DSi}_{nn-1^*} & \mathbf{DSi}_{nn} & \mathbf{DSi}_{nn^*} & \mathbf{DSi}_{nn1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \mathbf{S23}'_{nn-1^*} & \mathbf{S23}'_{nn} & \mathbf{S23}'_{nn^*} & \mathbf{S23}'_{nn1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -\mathbf{DSn}_{n1n^*} & -\mathbf{DSn}_{n1n1} & \mathbf{DFn}_{n1n1^*} & \mathbf{DFn}_{n1R} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \mathbf{S23}'_{n1n^*} & \mathbf{S23}'_{n1n1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \mathbf{S21}'_{n1n^*} & \mathbf{S21}'_{n1n1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \mathbf{S22}'_{n1n^*} & \mathbf{S22}'_{n1n1} & \mathbf{Q}_{n1n1^*} & \mathbf{Q}_{n1R} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \mathbf{M}_{Rn1} & \mathbf{M}_{RR}
\end{bmatrix}. \tag{31}$$

The vector of source strengths $\{\Psi\}$ is obtained by inverting Eq. (28). After calculating the source strengths, the pressure, velocity, stress, and displacement at any point can be obtained by placing the target points there and using corresponding source strength vectors as described earlier by Banerjee and Kundu.[38] Thus, DPSM can generate three-dimensional elastic wave field in a multilayered solid structure.

III. GREEN'S FUNCTION

A. Calculation of displacement and pressure Green's functions in fluid

Spherical bulk wave in a fluid can be generated by a point source in an infinite fluid medium.[23] If the point source is harmonic, then it will generate harmonic spherical waves. If a point source is generating bulk waves in a fluid, then the harmonic dirac-delta impulsive force will be the body force. Frequency domain pressure Green's function at a point in an infinite fluid medium at a distance of r from the point of excitation is given in Banerjee and Kundu.[38]

B. Calculation of displacement and stress Green's functions in solid

In DPSM formulation the explicit expressions of both displacement and stress Green's functions are needed. Let the frequency domain displacement Green's function $G_{ij}(\mathbf{x};\mathbf{y},\omega)$ denote the displacement in the ith direction at $\mathbf{y}$ for a point force acting in the jth direction at position $\mathbf{x}$. Therefore, the frequency domain stress Green's tensor for that material can be expressed as $\sigma_{ij}^{q}(\mathbf{x};\mathbf{y},\omega)=\frac{1}{2}C_{ijkl}(G_{kq,l}+G_{lq,k})$, when the source is acting in the qth direction placed at $\mathbf{x}$.

The equation of motion for a solid material can be written as

$$C_{ijkl}(x_n)\frac{\partial^2}{\partial x_j\,\partial x_l}u_k(x_n,t)+F_i(x_n,t)=\rho(x_n)\frac{\partial^2 u_i(x_n,t)}{\partial^2 t}, \quad (32)$$

where $C_{ijkl}(x_n)$ are the material constants, u_i are displacement components, F_i denotes the body force per unit volume and i,j,k,l, and n take values 1, 2, and 3. In the subsequent formulation all subscripts correspond to usual index notation in three-dimensional space, and δ_{ij} is the Kronecker delta symbol. In calculations of the Green's functions, inhomogeneity of the material can be incorporated as done by Manolis *et al.*[43] Let us consider here only homogeneous materials and $C_{ijkl}=C_{jikl}=C_{klij}=C_{ikjl}$ to satisfy the symmetry condition. The external force in Eq. (32) can be considered as an impulsive force at origin with magnitude $\mathbf{P}$:

$$\mathbf{F}(\mathbf{x},t)=\mathbf{P}f(t)\delta(\mathbf{x}) \quad \text{or} \quad F_i=P_i f(t)\delta(x_j) \quad (33)$$

where $\mathbf{P}$ denotes the source strength. From Eqs. (32) and (33) one can write

$$C_{ijkl}\frac{\partial^2}{\partial x_j\,\partial x_l}u_k(x_n,t)+P_i f(t)\delta(x_n)=\rho\frac{\partial^2 u_i(x_n,t)}{\partial^2 t}. \quad (34)$$

The Green's function for different materials can be solved in different ways from the above equation. Expressions of displacement and stress Green's functions are given for isotropic and anisotropic materials as

$$G_{ij}(\mathbf{x};\mathbf{y},\omega)=\frac{1}{4\pi\rho\omega^2}\left[\frac{e^{ik_p r}}{r}\left(k_p^2R_iR_j+(3R_iR_j-\delta_{ij})\left(\frac{ik_p}{r}-\frac{1}{r^2}\right)\right)+\frac{e^{ik_s r}}{r}\left(k_s^2(\delta_{ij}-R_iR_j)-(3R_iR_j-\delta_{ij})\left(\frac{ik_s}{r}-\frac{1}{r^2}\right)\right)\right], \quad R_i=\frac{x_i-y_i}{r},$$
$$r=|\mathbf{x}-\mathbf{y}|. \quad (35)$$

$G_{ij}(\mathbf{x};\mathbf{y},\omega)$ is the displacement Green's function for isotropic materials.

Similarly considering a point force at the origin the displacement Green's function in anisotropic materials can be written as[42]

$$G_{ij}(\mathbf{x};\mathbf{0},\omega)=\frac{i}{8\pi^2}\int_{|\mathbf{n}|=1}\sum_{M=1}^{3}\frac{k_M g_i^M g_j^M}{2\rho c_M^2}\times\exp(i(k_M n_p x_p))d\Omega(n_m)+\frac{1}{8\pi^2}\int_{|\mathbf{n}|=1}\sum_{M=1}^{3}\frac{g_i^M g_j^M}{\rho c_M^2}\delta(n_p x_p)d\Omega(n_m) \quad (36)$$

The stress Green's function in isotropic solids[38]

$$\sigma_{ij}^{q}(\mathbf{x};\mathbf{y},\omega)=(\mu(G_{ik,j}+G_{kj,i})\delta_{kq}+\lambda\delta_{ij}G_{kq,k}) \quad (37)$$

and the stress Green's function in anisotropic materials[42] can be obtained from the literature,

$$\sigma_{ij}^{q}(\mathbf{x};\mathbf{0},\omega)=\int_{-\infty}^{+\infty}\left[\frac{C_{ijkl}H(t)}{8\pi^2}\sum_{M=1}^{3}\int_{|\mathbf{n}|=1}\frac{n_l g_k^M g_q^M}{\rho c_M^3}\frac{\partial^2}{\partial t^2}\times[\delta(c_M t-n_p x_p)]d\Omega(n_m)\right]e^{i\omega t}dt \quad (38)$$

For the previous equation the point force is acting along the q direction.

IV. NUMERICAL RESULTS

Accuracy of the DPSM modeling for the wave field computation in a homogeneous solid with planar and nonplanar boundaries has been shown by Banerjee and Kundu.[35] In this article the results are presented for an inhomogeneous multilayered plate composed of two different isotropic layers. Both the plate and the ultrasonic transducers are immersed in water that acts as the coupling fluid for the ultrasonic signals between the transducers and the plate, see Fig. 1. Results are presented for both planar and nonplanar interfaces. Continuity conditions on stresses and displacements across the interface are checked to indirectly verify the accuracy of the computed results. Two transducers, one above and one below the plate are placed at the horizontal position $x_1=0$. The distance between a transducer face (at the center

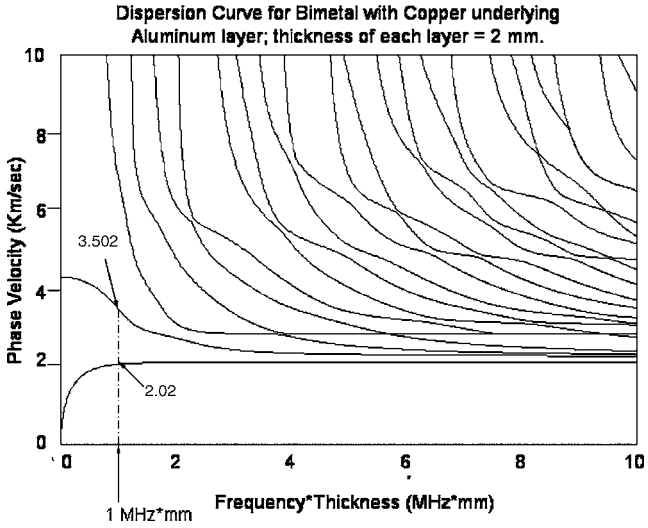


FIG. 2. Dispersion curves for the 4 mm thick bimetal (aluminum–copper) plate.

point) and its nearest fluid–solid interface is 6 mm. The total plate thickness is 4 mm. Therefore, the face to face distance between the two transducers along x_2 axis is 16 mm. The transducers have 4 mm diameter. The plate length and width in the in-plane and the out-of-plane directions, respectively, are much greater than the transducer diameter. The ultrasonic field is computed along a plane which bisects the transducers and the plate and thus forms a plane of symmetry of the problem geometry. Note that we are solving a three-dimensional problem and plotting the field along the vertical plane of symmetry of the problem geometry. One hundred point sources are distributed near each transducer face and additional point sources are placed along the plate boundaries and interfaces, as shown in Fig. 1. The average distance between two neighboring point sources is approximately 0.4 mm. Now the question is how many point sources should be used to model the interfaces and plate boundaries that are extended to infinity in both in-plane and out-of-plane directions. Point sources in the vertical plane of symmetry consisting of several lines of point sources as shown in Fig. 1 are first distributed and the ultrasonic field is computed along the central plane. Then two more planes of point sources are added on the two sides of the central plane and the field is computed again at the central plane. This process of adding two planes of point sources on two sides of the central plane is continued until the computed field at the central plane is converged. Note that the additional planes of point sources on two sides of the central plane are applied only along the plate boundaries and interfaces whose dimensions in the out-of-plane directions are large. For the finite dimension transducers 100 point sources are placed over the total transducer face from the very beginning. Interestingly, the results are found to converge with only three planes of point sources in the out-of-plane direction. However, if one is interested in computing the ultrasonic field at another plane which is not necessarily the plane of symmetry then more planes of point sources might be necessary.

On each side of an interface or a plate boundary 35 point sources are distributed on the central plane. Sources are

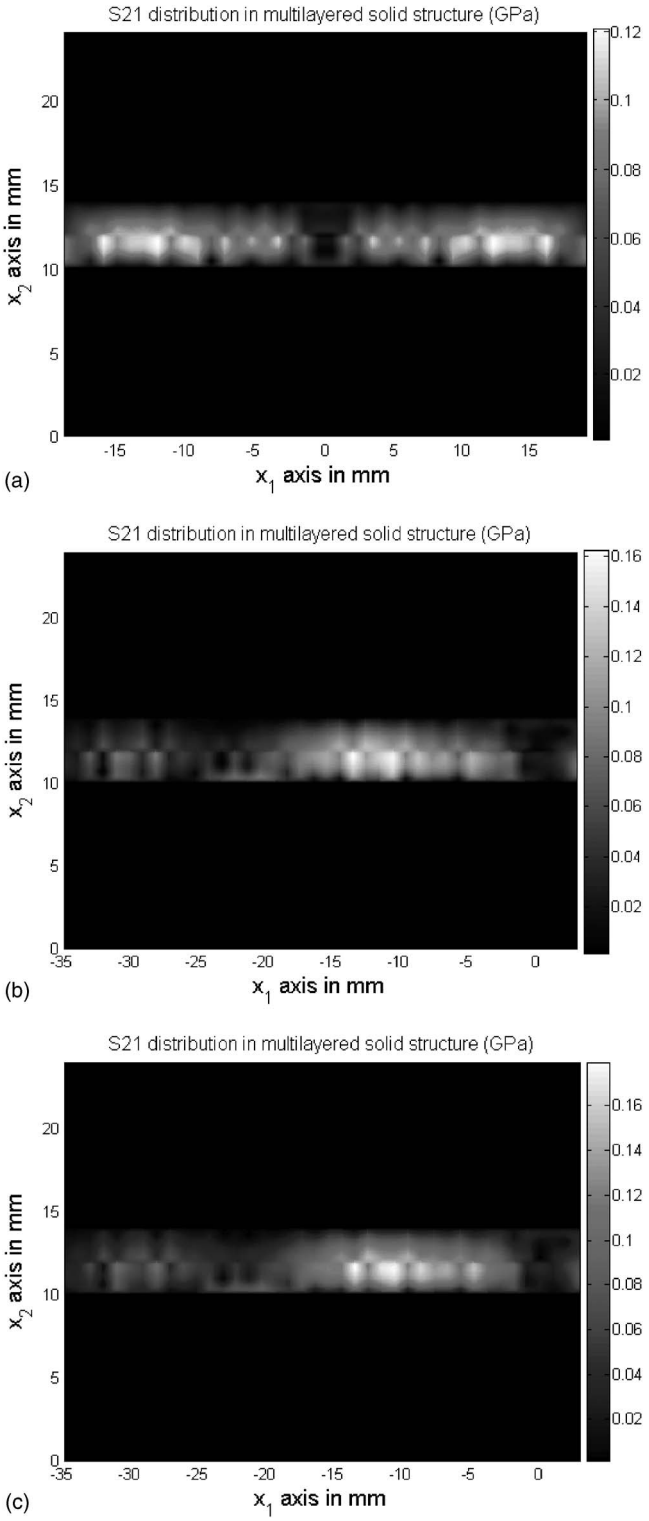


FIG. 3. Shear stress variations in the 4 mm thick bimetal (aluminum–copper) plate generated by two ultrasonic transducers placed on $x_1=0$ line. One transducer is positioned below ($x_2=0$) and the second one is above ($x_2=24$ mm) the plate. Results for normal incidence (top, $\theta=0°$), and inclined incidences (middle, $\theta=25°$ and bottom, $\theta=47.1°$), are shown.

placed along the interfaces and plate boundaries in the illuminated region as well as well beyond this region. A total of 105 sources are then necessary to model the interface with three planes of point sources in the out-of-plane direction. Increasing the number of sources to 175 to construct five

planes of point sources in the out-of-plane direction did not change the computed ultrasonic field in the central plane. The number of point sources taken for the wave field computation is based on the convergence criterion of the DPSM technique.[18] The convergence of the problem solution has been also tested by increasing the number of point sources in one plane of point sources and at the transducer face. When the spacing between two neighboring point sources is less than one-third wavelength then the problem is found to converge. Further increase in the number of point sources did not change the computed results significantly. For most of the results presented in this article the distance between two neighboring point sources has been kept at wavelength/π. Note that 35 point sources in each layer of an interface are distributed over 40 mm length along x_1 axis. Hence, distance between two neighboring point sources is approximately 1.14 mm (less than the shortest wavelength/π), thus the results presented here are well converged. The boundary and continuity conditions are enforced at the apex of the spherical bulbs [little spheres shown in Fig. 1(a)] of the point sources. Apex point is the point where a spherical bulb of a point source touches the problem boundary or interface. Therefore, the number of points where boundary and continuity conditions are enforced is the same as the number of point sources. For example, along the x_1 axis along every line of point sources there are 35 points on the boundary of the plate to enforce continuity conditions. On the transducer surface there are 100 points to enforce boundary conditions. The number of boundary and continuity conditions is equal to the number of unknowns (point source strengths). Wave fields in the layered solid plate are generated for different striking angles of the ultrasonic beams. Results presented in the following show displacement and stress variations along the length and depth of the plate specimen.

A two layered plate composed of aluminum and copper is considered for the analysis. The properties of aluminum are c_p=6.2 km/s, c_s=3.04 km/s, and ρ=2.7 gm/cm^3, where c_p is the P-wave speed, c_s is the S-wave speed, and ρ is the density. The properties of copper are c_p=4.6 km/s, c_s=2.26 km/s, and ρ=8.9 gm/cm^3. It is assumed that the aluminum layer of thickness t=2 mm underlines the copper layer of thickness t=2 mm. Wave fields are generated with 250 kHz transducers. It is well known that in a multilayered plate the wave modes propagate at certain frequencies with certain phase velocities. Wave modes obtained from the dispersion equation are in general dispersive; dispersion curves for this two-layered plate are shown in Fig. 2.

From the dispersion curves one can see that at 250 kHz frequency there are two guided wave modes. For a given frequency for generating a specific guided wave mode one can obtain the critical angle of incidence of the striking ultrasonic beam from the phase velocities (shown along the vertical axis of the dispersion curve plot) using Snell's law,

$$\theta_{\mathrm{cr}} = \sin^{-1}\left(\frac{c_f}{c_{\mathrm{ph}}}\right)$$

where c_f is the acoustic wave speed in the coupling fluid, θ_{cr} is the critical angle of incidence for generating the guided wave mode whose phase velocity is c_{ph}.

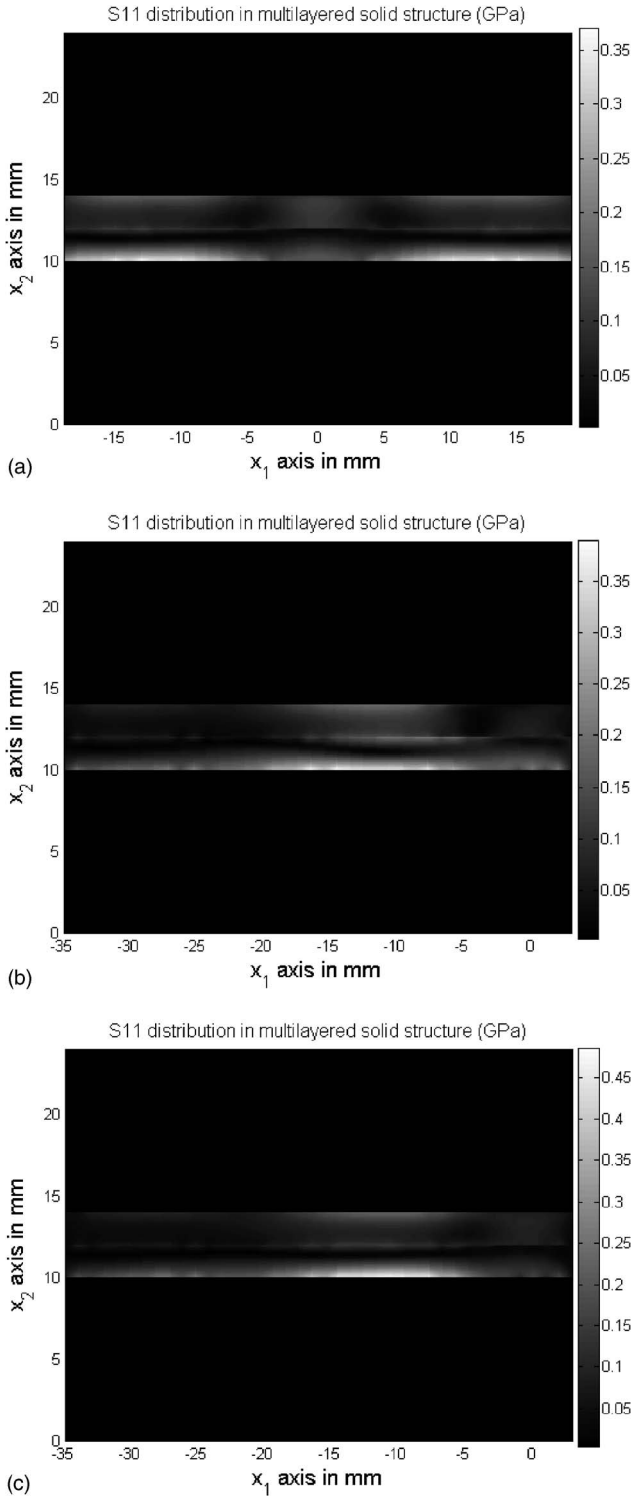


FIG. 4. For the plate and transducer orientations shown in Fig. 1(a) and described in Fig. 3, variations of σ_{11} are shown in the left and right columns, respectively. Three rows show results for normal incidence (top row, θ=0°), and inclined incidences (middle row θ=25° and bottom row θ=47.1°).

Thus, for 250 kHz signal frequency one obtains critical angles 47.1° and 25° for generating Modes 1 and 2, respectively, in the plate specimen. Stress and displacement variations in the plate for the normal incidence (θ=0°) as well as for the two critical angles of incidence (θ=25° and 47.1°) are shown in Figs. 3–5. The shear stress σ_{21} (or S_{21}) variations in the plate for these three angles of incidence are shown in Fig.

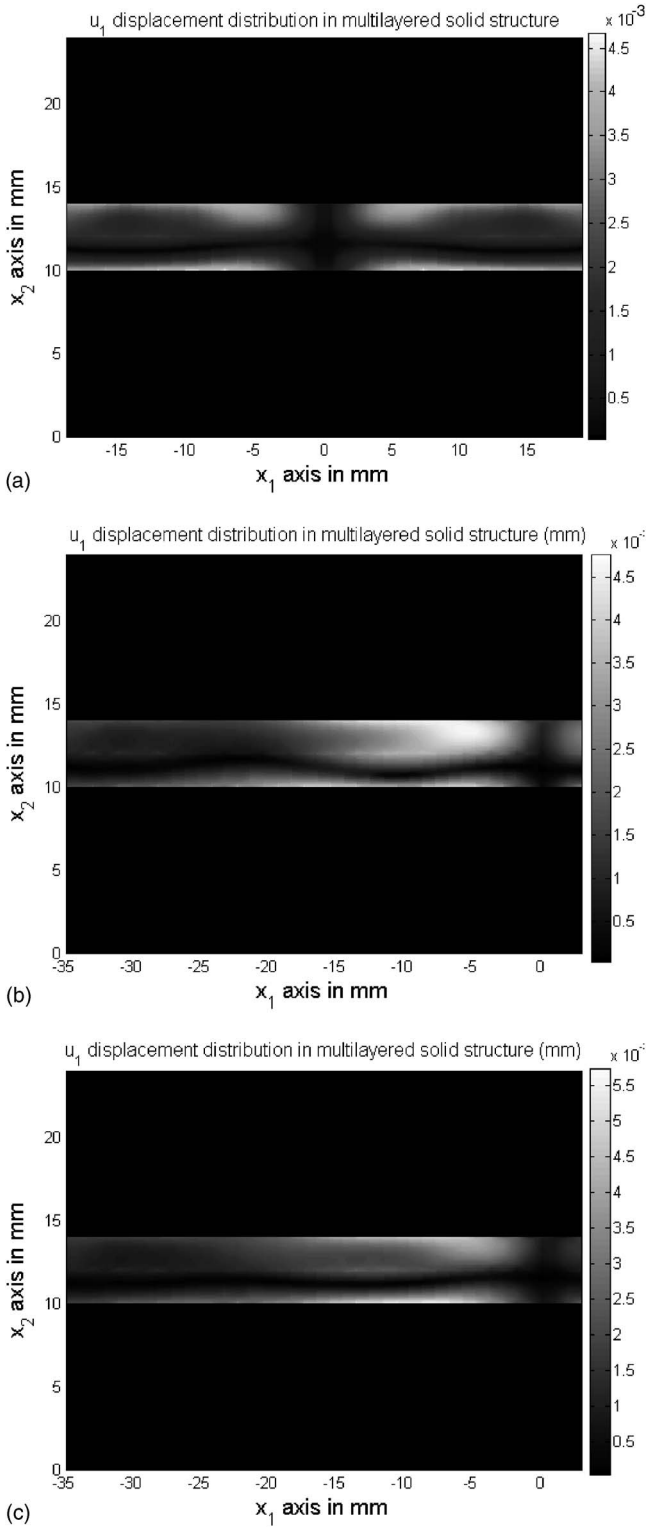


FIG. 5. Horizontal displacements inside the plate for the plate and transducer orientations shown in Fig. 1(a) and described in Fig. 3. Three rows show results for normal incidence (top row, $\theta=0°$), and inclined incidences (middle row $\theta=25°$ and bottom row $\theta=47.1°$).

3, variations of the normal stress component σ_{11} (or S_{11}) are shown in Fig. 4 and that for the displacement component (u_1) are shown in Fig. 5. As expected, for the normal incidence all displacement and stress amplitudes are symmetric about the x_2 axis ($x_1=0$ line) since the ultrasonic beam strikes the plate at $x_1=0$. Such symmetry is not observed for

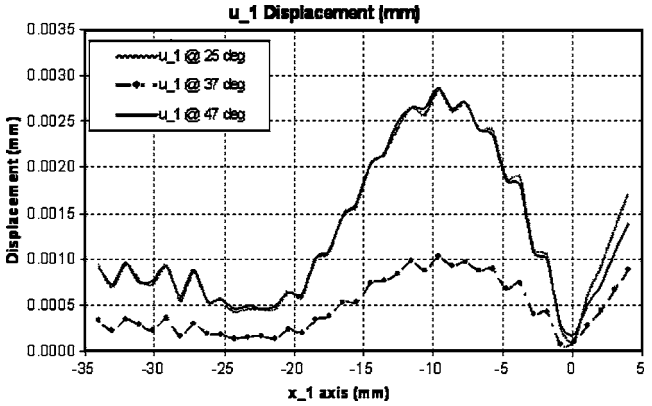


FIG. 6. Horizontal displacement (u_1) variation along the interface of the bimetal plate for three different transducer inclination angles.

the oblique incidence. Figure 6 shows the horizontal displacement component (u_1) along the interface of the plate for three different transducer inclination angles. It can be seen from this figure that when the transducers are inclined at critical angles (47° and 25°) the displacement values at the interface away from the striking beam are significantly higher than that for the noncritical angle ($\theta=37°$) of incidence. The higher displacement values for the critical angle of incidence are expected because of the propagating wave modes. Similar phenomenon is observed for the u_2 displacement component as well.

From the gray scale images of Figs. 3–5 it is not clear if the continuity conditions (continuity of u_1, u_2, σ_{22}, and σ_{12}) across the interface and the boundary condition ($\sigma_{12}=0$) at the top and bottom boundaries are satisfied. These conditions can be verified from the line plot in Figs. 7. Shear stress variation along the plate thickness from the bottom boundary ($x_2=10$ mm) to the top boundary ($x_2=14$ mm) at $x_1=-15$ mm is shown in Fig. 7 for the normal incidence of the ultrasonic beam. Three curves in this figure correspond to real, imaginary and absolute values, respectively. In Fig. 7

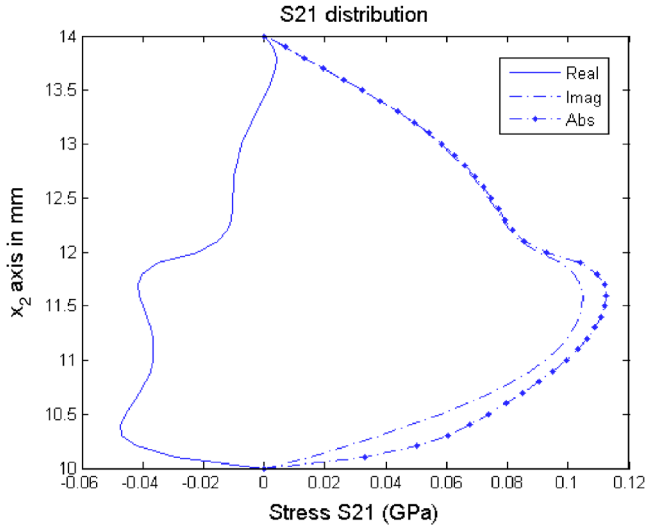


FIG. 7. (Color online) Shear stress variation across the 4 mm thick bimetal plate at $x_1=-15$ mm for normal incidence. Note that it is continuous across the interface at $x_2=12$ mm and zero at the top ($x_2=14$ mm) and bottom ($x_2=10$ mm) boundaries.

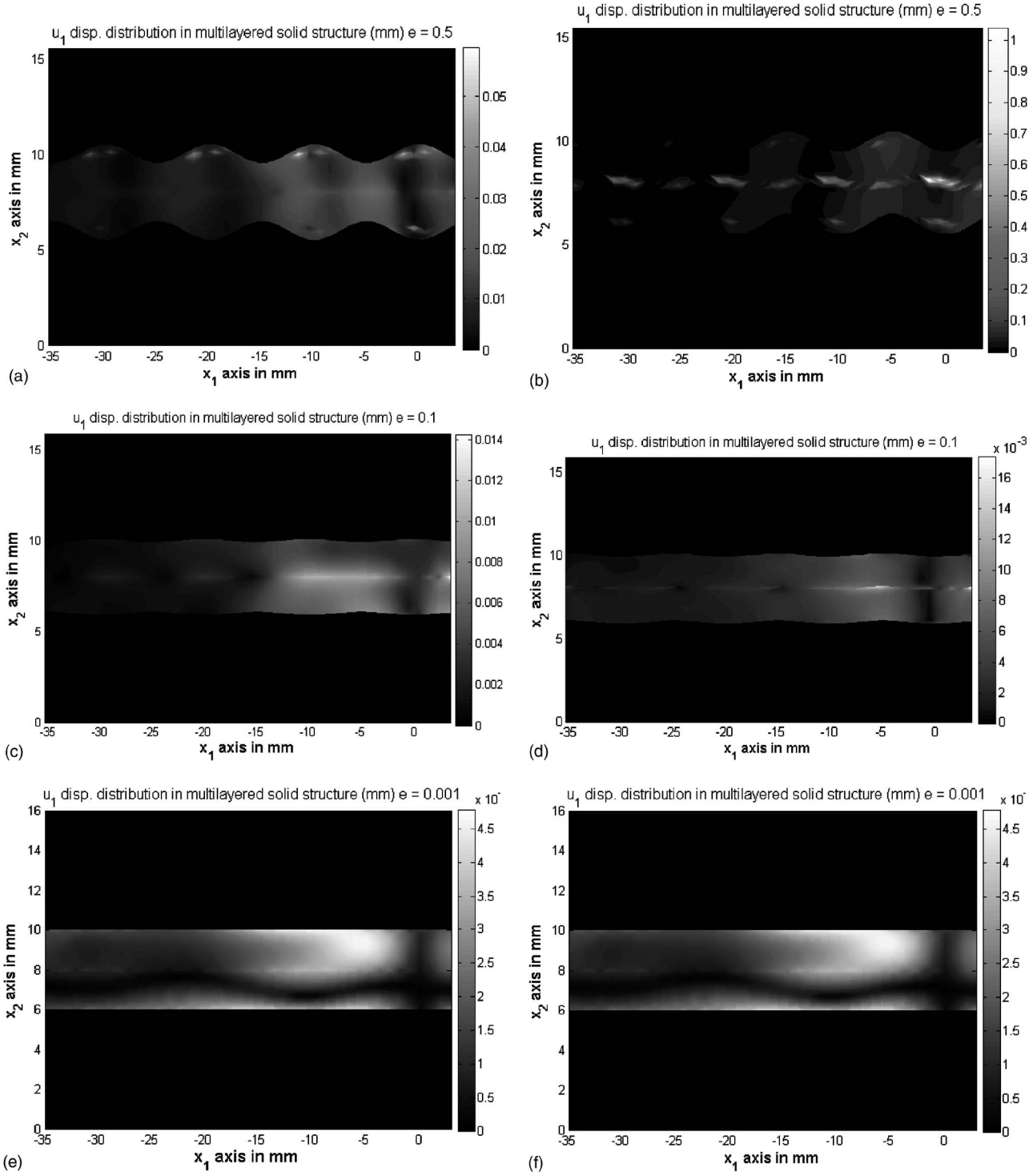


FIG. 8. Horizontal displacements inside the corrugated plate for the plate and transducer orientations shown in Fig. 1 and described in Fig. 3. The left column (a,c,e) is for symmetric corrugation and the right column (b,d,f) is for antisymmetric corrugation. Three rows show results for three different amplitudes of corrugation. Corrugation amplitude (e) is 0.5 mm for the top row (a), 0.1 mm for the middle row (b), and 0.001 for the bottom row (c). Incident angle of the ultrasonic beam is 25°.

one can clearly see that the shear stress is zero at the top and bottom boundaries and continuous across the interface ($x_2 = 12$ mm). Similar plots (not shown) shows discontinuities in σ_{11} at the interface but no such jump at the interface is observed in σ_{22} plot. The displacement components u_1 and u_2 show continuity across the interface and nonzero values at the plate boundaries.

Finally, the effect of the non-planar boundary surfaces and interface of the layered plate is investigated and presented in Figs. 8 and 9. The top and bottom surfaces of the plate are made nonplanar by replacing the flat plate boundaries by sinusoidal boundaries. For the symmetrically corrugated plate the central interface remains flat but for the antisymmetric corrugation the solid–solid interface is also made

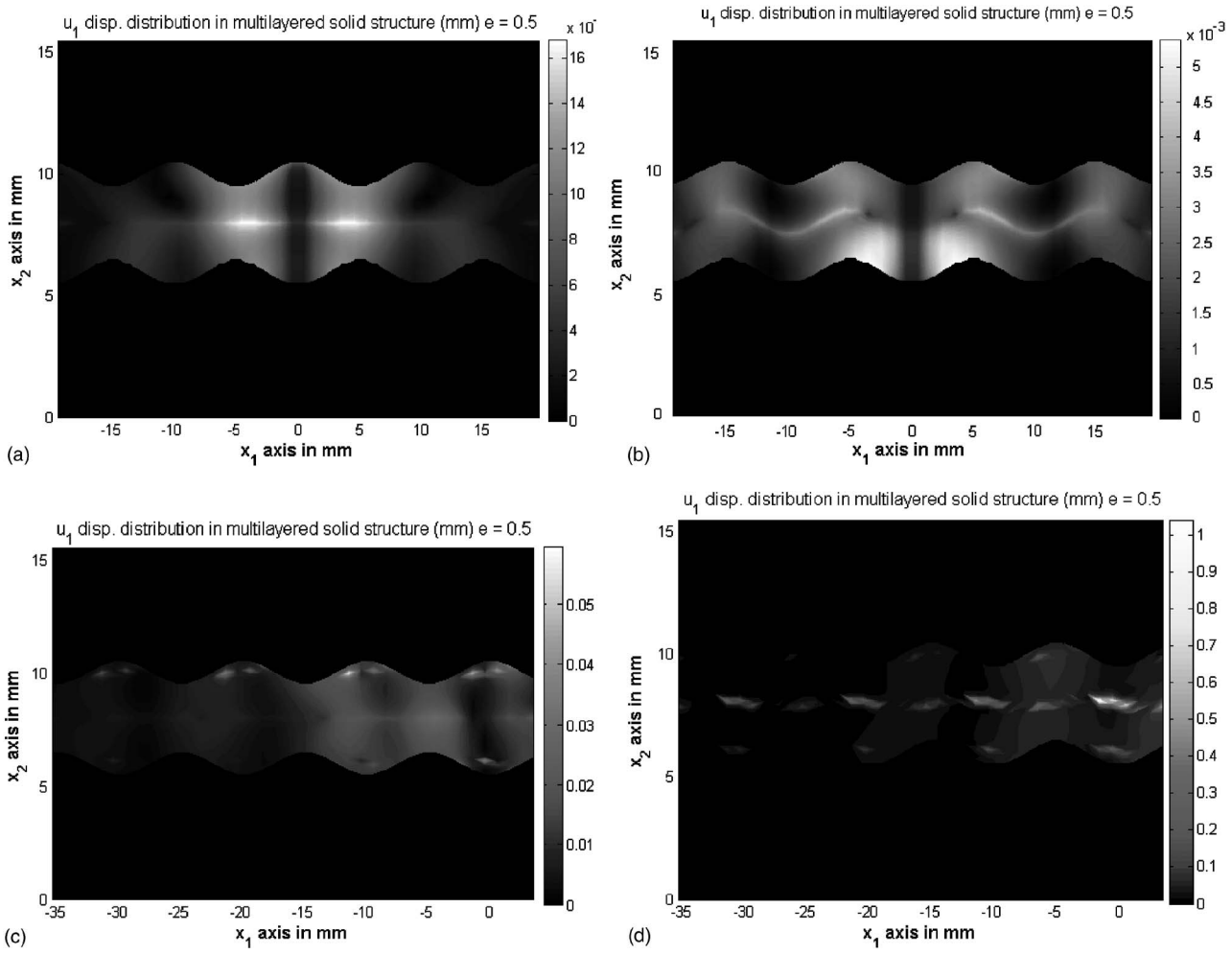


FIG. 9. Horizontal displacements inside the corrugated plate (corrugation amplitude is 0.5 mm) for the plate and transducer orientations shown in Fig. 1 and described in Fig. 3. The left column (a,c,e) is for a symmetric corrugation and the right column (b,d,f) is for antisymmetric corrugation. Two rows show results for normal incidence (top row, θ=0°), and inclined incidences (bottom row θ=25°.

corrugated, as shown in left (a, c, e) and right images (b, d, f) of Figs. 8(a) and 9. The wavelength of the corrugation is 10 mm. The corrugation amplitude (e) is 0.5 mm for the top images [Figs. 8(a) and 8(b)], 0.1 mm for the middle images [Figs. 8(c) and 8(d)] and 0.001 mm for the bottom images [Figs. 8(e) and 8(f)]. Note that the height of the peak and the depth of the dip are both equal to the corrugation amplitude e. Therefore, the vertical distance between the peak and the dip is $2e$. Figure 8 shows the horizontal displacement (u_1) inside the two corrugated plates for 25° angle of strike. Transducer locations are the same as those for the flat plate. Note how the ultrasonic fields change with the corrugation amplitude and type of corrugation (symmetric and antisymmetric). For very small corrugation amplitude (e =0.001 mm), as expected, the ultrasonic fields inside the symmetric and antisymmetric corrugated plates match with that for the flat plate. For high corrugation amplitude (e =0.5) the field is significantly different for symmetric and antisymmetric corrugations. Figure 9 shows the ultrasonic field (u_1 variation) for two different angles of incidence for the striking ultrasonic beam; these are 0° and 25°, shown in top and bottom rows, respectively.

V. CONCLUSIONS

DPSM technique has been generalized in this article for modeling ultrasonic fields inside multilayered solids. Here finite size ultrasonic transducers generate the ultrasonic fields unlike most studies by the earlier investigators where plane waves of infinite width were assumed to be the incident field. Step by step mathematical derivation of the DPSM formulation for multilayered plates, excited by finite-size ultrasonic transducers, has been presented in the article. Ultrasonic fields in bimaterial plates having planar and nonplanar boundaries and interfaces are given and the satisfaction of the boundary and continuity conditions at the interfaces has been checked. Readers who are interested in solving multilayered problems of different geometry can develop their own formulation and computer codes following the mathematical steps outlined in the article. The main objective of this article was to extend the technique to nonhomogeneous solids. In future publications nonhomogenous solids with internal voids and cracks will be modeled.

ACKNOWLEDGMENT

The authors thank Mr. Samik Das, graduate student at the University Arizona, Tucson for generating the dispersion curves for the multilayered solid plate presented in this article. This research was partially supported from NSF Grant Nos. CMS-9901221, CMMI-0226466, CMMI-0443494, CMMI-0530991, and OISE-0352680. The results and conclusions presented here are those of the authors and do not represent the view of NSF

[1]J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. II.
[2]R. Stoneley, "The seismological implications of aeolotropy in continental structure," Mon. Not. R. Astr. Soc. of London, Geophys. Suppl. **5**, 343–349 (1949).
[3]W. T. Thompson, "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89–93 (1950).
[4]N. A. Haskell, "The dispersion of surface waves on multilayered media," Bull. Seismol. Soc. Am. **43**, 17–21 (1953).
[5]T. Kundu and A. K. Mal, "Elastic waves in a multilayered solid due to a dislocation source," Wave Motion **7**, 459–471 (1985).
[6]J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, New York, 1999).
[7]G. R. Liu and Z. C. Xi, *Elastic Waves in Anisotropic Laminates* (CRC Press, Boca Raton, FL, 2002).
[8]*Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization*, edited by T. Kundu (CRC Press, Boca Raton, FL, 2004), Chaps. 2 and 4.
[9]B. Biondi, "Solving the frequency-dependent Eikonal equation," 62nd Annual International Meeting, Society of Exploration of Geophysics, Expanded Abstracts (1992), p. 1315.
[10]V. Cervený and I. Psencik, "Gaussian beams and paraxial ray approximation in three-dimensional elastic inhomogeneous media," J. Geophys. **53**, 1–9 (1983).
[11]*PZFlex Software, Version: 1-j.6* (Weilinger Associates, Inc., Washington, DC, 2001).
[12]X. Zhao and J. L. Rose, "Boundary Element Modeling for Defect Characterization potential in a wave guide," Int. J. Solids Struct. **40**, 2645–2658 (2003).
[13]R. P. Shaw, "Boundary integral equation methods applied to wave problems," *Developments in Boundary Element Methods - 1*, edited by P. K. Banerjee and R. Butterfield (Applied Science Publications, London, 1979), Chap. 6, pp. 121–153.
[14]B. P. Newberry and R. B. Thompson, "A paraxial theory for the propagation of ultrasonic beams in anisotropic solids," J. Acoust. Soc. Am. **85**, 2290–2300 (1989).
[15]J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756 (1988).
[16]M. Spies, "Transducer field modeling in anisotropic media by superposition of Gaussian base functions," J. Acoust. Soc. Am. **105**, 633–638 (1999).
[17]L. W. Schmerr, H. J. Kim, R. Huang, and A. Sedov, "Multi-Gaussian ultrasonic beam modeling," Proceedings of the World Congress of Ultrasonics (WCU 2003), Paris, September 7–10 (2003), Vol. 93.
[18]M. Spies, "Analytical methods for modeling of ultrasonic nondestructive testing of anisotropic media," Ultrasonics **42**, 213–219 (2004).
[19]C. Rajamohan and J. Raamachandran, "Bending of anisotropic plates by charge simulation method," Adv. Eng. Software **30**, 369–376 (1999).
[20]B. Ch. Hafner, "The multiple multipole method (MMP) in electro and magnetostatic problems," IEEE Trans. Magn. **19**, 2367–2370 (1983).
[21]Ch. Hafner, "MMP calculations of guided waves," IEEE Trans. Magn. **21**, 2310–2312 (1985).
[22]M. G. Imhof, "Computing the elastic scattering from inclusions using the multiple multipoles method in three dimensions," Geophys. J. Int. **156**, 287–290 (2004).
[23]F. J. Sanchez-Sesma and J. A. Esquivel, "Ground motion on alluvial valleys under incident planes SH waves," Bull. Seismol. Soc. Am. **69**, 1107–1111 (1979).
[24]H. L. Wong, "Diffraction of P, SV, and Rayleigh waves by surface topography," Bull. Seismol. Soc. Am. **72**, 1167–1171 (1982).
[25]M. Dravinski and T. K. Mossessian, "Scattering of plane harmonic P, SV, and Rayleigh waves by dipping layers of arbitrary shape," Bull. Seismol. Soc. Am. **77**, 212–215 (1987).
[26]F. J. Sanchez-Sesma and M. Campillo, "Diffraction of P, SV and Rayleigh waves by topographic features: A boundary integral formulation," Bull. Seismol. Soc. Am. **81**, 2234–2253 (1991).
[27]F. J. Sanchez-Sesma and M. Campillo, "Topographic effects for incident P, SV and Rayleigh waves," Tectonophysics **218**, 113–125 (1993).
[28]T. Pointer, E. Liu, and J. A. Hudson, "Numerical modeling of seismic waves scattered by hydrofractures: application of the indirect boundary element method," Geophys. J. Int. **135**, 289–303 (1998).
[29]M. Bouchon and F. J. Sanchez-Sesma, "Boundary integral equations and boundary elements methods in elastodynamics," in *Advances in Wave Propagation in Heterogeneous Earth, Advances in Geophysicis*, Vol. 48, edited by R.-S. Wu, V. Maupin, and R. Dmowska (Elsevier-Academic, New York, 2007).
[30]D. Placko and T. Kundu, "A theoretical study of magnetic and ultrasonic sensors: Dependence of magnetic potential and acoustic pressure on the sensor geometry," Proc. SPIE, **4335**, 52–62 (2001).
[31]S. Banerjee, T. Kundu, and D. Placko, "Ultrasonic field modelling in multilayered fluid structures using DPSM technique," ASME J. Appl. Mech. **73**, 598–609 (2006).
[32]S. Banerjee and T. Kundu, "DPSM technique for ultrasonic field modelling near fluid-solid interface," Ultrasonics **46**, 235–250 (2007).
[33]S. Banerjee and T. Kundu, "Semi-analytical modeling of ultrasonic fields in solids with internal anomalies immersed in a fluid," Wave Motion (in press).
[34]S. Banerjee and T. Kundu, "Ultrasonic field modeling in plates immersed in fluids," Int. J. Solids Struct. **44**, 6013–6029 (2007).
[35]S. Banerjee and T. Kundu, "Elastic wave propagation in sinusoidally corrugated wave guides," J. Acoust. Soc. Am. **119**, 2006–2017 (2006).
[36]T. Kundu, J.-P. Lee, C. Blase, and J. Bereiter-Hahn, "Acoustic microscope lens modeling and its application in determining biological cell properties from single and multi-layered cell models," J. Acoust. Soc. Am. **120**, 1646–1654 (2006).
[37]A. K. Mal and S. J. Singh, *Deformation of Elastic Solids* (Prentice Hall, Englewood Cliffs, NJ, 1991).
[38]S. Banerjee and T. Kundu, "Chapter 4: Advanced Applications of Distributed Point Source Method—Ultrasonic Filed Modeling in Solid Media," *DPSM—Distributed Point Source Method for Modelling Engineering Problems*, edited by D. Placko and T. Kundu (Wiley, New York, 2007), pp. 143–229.
[39]A. Tverdokhlebov and J. Rose, "On Green's function for elastic waves in anisotropic media," J. Acoust. Soc. Am. **83**, 118–121 (1988).
[40]C. Y. Wang and J. D. Achenbach, "Elastodynamic fundamental solutions for anisotropic solids," Geophys. J. Int. **118**, 384–392 (1994).
[41]V. K. Tewary, "Computationally efficient representation for elastostatic and elastodynamic Green's functions for anisotropic solids," Phys. Rev. B, **51**, 15695–15702 (1995).
[42]C. Y. Wang and J. D. Achenbach, "A new method to obtain 3-D Green's functions in anisotropic solids," Wave Motion, **18**, 273–289 (1993).
[43]G. D. Manolis, T. V. Rangelov, and P. S. Dineva, "Elastodynamic fundamental solutions for 2D inhomogeneous anisotropic domains," *Proceedings of the 5th GRACM International Congress on Computational Mechanics*, 2005.

Dual mode tuning strategy of a slightly asymmetric ring

Han Gil Park and Yeon June Kang[a)]
Advanced Automotive Research Center, School of Mechanical and Aerospace Engineering, Seoul National University, San 56-1 Shinlim-Dong, Kwanak-Gu, Seoul 151-744, South Korea

Seock Hyun Kim
Division of Mechanical and Mechatronics Engineering, Kangwon National University, Hyoja2-Dong 192-1, Chunchon-Si, Kangwon-Do 200-701, South Korea

(Received 27 July 2007; revised 11 December 2007; accepted 2 January 2008)

Clear beats with proper periods in the first and the second vibration modes are very important factors for the sound of the Korean bell. In this study, the Korean bell is expressed as a circular ring with multiple point masses and a dual mode tuning strategy for clear beat with proper period. For the dual mode tuning, a dual mode equivalent ring model is composed with two point masses, which satisfies the mode pair conditions of the first mode and second mode simultaneously. By adding a suitable amount of additional masses to the dual mode equivalent ring model at appropriate positions, a clear beat with proper period in both of the two important modes is generated. The position and amount of the additional masses are determined analytically. Analytical tuning results are compared and verified with those of the finite element analysis.
 [DOI: 10.1121/1.2836784]

PACS number(s): 43.40.At, 43.40.Ey, 43.40.Rj [JHG] Pages: 1383–1391

I. INTRODUCTION

Korean bells have been cast since the time of the ancient Shilla dynasty (57 B.C.–935 A.D.). They have a unique shape and an exquisite appearance and they produce beautiful sounds.[1] The natural frequencies of bell vibration modes are essentially determined by the main dimensions of the nearly axisymmetric body and the material properties of the bell. Sculptures and figures carved on the bell body and casting irregularities induce a slight asymmetry, and the bell consequently generates beating sounds. The strong beat of a western bell is called a warble and is often considered an undesirable property. In fact, western bell makers try to eliminate this phenomenon or at least maintain it within reasonable limits.[2] On the other hand, a Chinese bell maker will try to enlarge the asymmetry and produce different two tones. Jing analyzed natural frequencies and sound radiation characteristics based on the lens shaped Chinese bell.[3] In Korean bells, however, the intervals between strikings are quite long and, therefore, people hear different sounds from the long-lasting beat of the hum, the fundamental, and the striking sound. The hum and fundamental created by the first and the second vibration modes last much longer than the other higher partials. When the hum has a clear beat with an appropriate period, people hear a lively bell sound, as if the bell were alive and breathing. On the other hand, the strong beat of the fundamental makes the bell sound appear like wail or cry.[4] In most Korean bells, the modes of the hum and the fundamental have monotonous shapes along their heights. They do not exhibit any phase change with the contour along the vertical axis. Therefore, a ring-type structure may be applied as a simplified model for the analysis of the frequency and beat characteristics of low-order vibration modes in these bells.

Much research has been carried out on a ring-type structure with slight asymmetry. Charnley and Perrin investigated the mode splitting of an eccentric ring and examined the warble in a bell with slight asymmetry.[5–7] Allaei and Soedel applied the receptance method to calculate the natural frequencies of a circular ring with local masses and springs.[8] Hong and Lee obtained a precise solution for an asymmetric circular ring by considering the local mass and stiffness deviations as a Heaviside step function.[9] Using the Rayleigh–Ritz method, Fox proposed a theoretical model to calculate the modal parameters of a circular ring with multiple point masses.[10] He and his colleagues showed that the model could be applied to trim *N* pairs of modes simultaneously by removing N trimming masses at particular locations around the ring.[11] When a bell-type structure is struck, the beating in the vibration and sound occur by the interference of the split modes. Using impulse response analysis, Kim *et al.* performed a detailed investigation of the beat distribution characteristics on the circumference of bell-type structures.[12,13]

In this research, we investigate an analytical method to control the period and clarity of the beat in a slightly asymmetric ring. We use the concept of Fox's equivalent imperfection masses. Although Fox and his colleagues proposed this concept with a view to eliminate frequency splitting, we show that their model can be effectively applied to create the required beat conditions. Mode tuning is performed on both 1st and 2nd modes. To identify the beat distribution characteristics, an impulse response under a given striking condition is considered. The nodal line condition for clear beats is investigated and the necessary mass addition and removal are determined to obtain the required beat period and clarity. The validity of the theoretical results is examined by com-

[a)]Author to whom correspondence should be addressed. Electronic mail: yeonjune@snu.ac.kr

paring them with finite element analysis (FEA) results.

II. CONSIDERATIONS IN THE EQUIVALENT IMPERFECTION MASS MODEL

Fox proposed an equivalent imperfection mass model that satisfies the mode splitting condition of a circular ring with multiple point masses.[10] As stated above, in Korean bells, this model, instead of trimming, can be used to obtain the appropriate period and a clear beat. According to Fox's analysis results, the split natural frequency, $\omega_{L,H}$, and the phase angle of the split mode, $\psi_{L,H}$, are determined by the following equations. Here, the subscript L or H indicates the split mode in the nth mode that has a slightly lower or higher frequency, respectively.

$$\tan 2n\psi_n = \frac{\sum_i m_i \sin 2n\phi_i}{\sum_i m_i \cos 2n\phi_i}, \tag{1}$$

$$\omega_{nL,H}^2 = \omega_{0n}^2 \left(\frac{1+\alpha_n^2}{(1+\alpha_n^2) + \sum_i m_i[(1+\alpha_n^2) \mp (1-\alpha_n^2)\cos 2n(\phi_i - \psi_n)]/M_0} \right), \tag{2}$$

where n is mode number, m_i is the ith point mass, ϕ_i is the position of ith point mass, ψ_n is the anti-node of nth L-mode, α_n is the ratio of radial displacement to tangential displacement, and M_0 is mass of perfect circular ring.

As an inverse problem, we consider a circular ring with n imperfection point masses that satisfy the mode condition signified by $\omega_{nL,nH}$ and ψ_n. The mode condition for the analysis is obtained by measurement or by using an asymmetric ring with multiple point masses. We can therefore transform Eqs. (1) and (2) into Eqs. (3) and (4), respectively, by using standard trigonometric identities

$$\sum m_i \sin 2n(\phi_i - \psi_n) = 0 \tag{3}$$

$$\sum m_i \cos 2n(\phi_i - \psi_n) = M\lambda_n,$$

$$\lambda_n = \frac{(\omega_{nL}^2 - \omega_{nH}^2)(1+\alpha_n^2)}{(\omega_{nL}^2 + \omega_{nH}^2)(1-\alpha_n^2)},$$

$$M = M_0 + \sum m_i. \tag{4}$$

From Eqs. (3) and (4), m_i and ϕ_i that satisfy given $\omega_{nL,nH}$ and ψ_n can be calculated. If we consider only the mode condition, we can determine m_{eq} and ϕ_{eq}, which are the magnitude and position of an imperfection mass, respectively, from two constraints of Eqs. (3) and (4). We can thus acquire four constraints that satisfy $n=2$ and $n=3$ mode conditions from Eqs. (3) and (4). In this manner, the magnitude and the position of two point masses m_a, m_b, ϕ_a, and ϕ_b can be calculated. From this information, we can create a dual mode equivalent ring model that simultaneously satisfies the natural frequencies and locations of the anti-node for $n=2$ and $n=3$ modes.

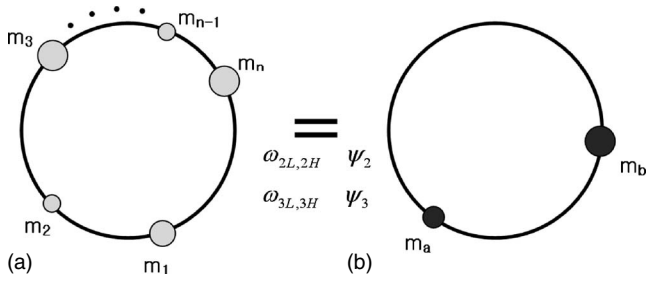


FIG. 1. The concept of dual mode equivalent ring model. (a) Original ring, (b) dual mode equivalent ring.

III. DUAL MODE EQUIVALENT RING MODEL

Let us now determine a dual mode equivalent ring model that satisfies the $n=2$ and $n=3$ mode pair conditions, as shown in Fig. 1. This equivalent model has two imperfection point masses that are not unique. Figure 1(a) shows the original ring model with mode data $\omega_{2L,2H}$, $\omega_{3L,3H}$, ψ_2, and ψ_3. The original ring model is composed of a perfectly symmetric ring model that has a total mass M_0 plus the imperfection masses m_1, m_2,…, and m_n. In a real case, the original ring model is given not by the information on the magnitude and the position of imperfection masses m_1, m_2,…, and m_n but by mode data $\omega_{2L,2H}$, $\omega_{3L,3H}$, ψ_2, and ψ_3. Figure 1(b) shows the dual mode equivalent model that is composed of a symmetric ring model with a total mass M^* plus two imperfection masses m_a and m_b. The dual mode equivalent ring model satisfies the mode conditions that are exactly similar to those of the original ring model in the $n=2$ and $n=3$ modes.

The magnitude and position of two masses must satisfy the given mode data $\omega_{2L,2H}$, $\omega_{3L,3H}$, ψ_2, and ψ_3 when m_a and m_b are applied to Eqs. (3) and (4), as shown below. Here, m_a and m_b indicate the magnitude of two masses and ϕ_a and ϕ_b indicate the positions of two masses, respectively

$$m_a \sin 4(\phi_a - \psi_2) + m_b \sin 4(\phi_b - \psi_2) = 0, \tag{5}$$

$$m_a \sin 6(\phi_a - \psi_3) + m_b \sin 6(\phi_b - \psi_3) = 0, \tag{6}$$

TABLE I. Natural frequency of original ring and equivalent rings.

Specification of axisymmetric ring		M=1702.5 kg, ρ=8700 kg/m^3, R=1.012 m, E=5.6e10 Pa, h=0.203 m, d=0.15 m		
Mode parameter	Mode	ω_1	ω_2	ψ_n
Original ring	n=2	64.57	64.64	14.49
	n=3	182.25	183.05	7.57
n=2 equivalent ring 1	n=2	64.57	64.64	14.49
	n=3	182.50	182.79	14.49
n=3 equivalent ring	n=2	64.50	64.71	7.57
	n=3	182.25	183.05	7.57
Dual mode equivalent ring	n=2	64.57	64.64	14.49
	n=3	182.25	183.05	7.57

$$m_a \cos 4(\phi_a - \psi_2) + m_b \cos 4(\phi_b - \psi_2) = M\lambda_2, \tag{7}$$

$$m_a \cos 6(\phi_a - \psi_3) + m_b \cos 6(\phi_b - \psi_3) = M\lambda_3. \tag{8}$$

Here,

$$\lambda_2 = \frac{(\omega_{2L}^2 - \omega_{2H}^2)(1+\alpha_2^2)}{(\omega_{2L}^2 + \omega_{2H}^2)(1-\alpha_2^2)}, \tag{9}$$

$$\lambda_3 = \frac{(\omega_{3L}^2 - \omega_{3H}^2)(1+\alpha_3^2)}{(\omega_{3L}^2 + \omega_{3H}^2)(1-\alpha_3^2)}. \tag{10}$$

Equations (5) and (6) can be transformed into a matrix form as shown below

$$\begin{bmatrix} \sin 4(\phi_a - \psi_2) & \sin 4(\phi_b - \psi_2) \\ \sin 6(\phi_a - \psi_3) & \sin 6(\phi_b - \psi_3) \end{bmatrix} \begin{bmatrix} m_a \\ m_b \end{bmatrix} = 0. \tag{11}$$

Likewise, Eqs. (7) and (8) can be transformed into the following matrix form:

$$\begin{bmatrix} \cos 4(\phi_a - \psi_2) & \cos 4(\phi_b - \psi_2) \\ \cos 6(\phi_a - \psi_3) & \cos 6(\phi_b - \psi_3) \end{bmatrix} \begin{bmatrix} m_a \\ m_b \end{bmatrix} = \begin{bmatrix} M\lambda_2 \\ M\lambda_3 \end{bmatrix}. \tag{12}$$

From Eq. (12), we obtain m_a and m_b as

$$\begin{bmatrix} m_a \\ m_b \end{bmatrix} = \begin{bmatrix} \cos 4(\phi_a - \psi_2) & \cos 4(\phi_b - \psi_2) \\ \cos 6(\phi_a - \psi_3) & \cos 6(\phi_b - \psi_3) \end{bmatrix}^{-1} \begin{bmatrix} M\lambda_2 \\ M\lambda_3 \end{bmatrix}. \tag{13}$$

By substituting Eqs. (13) in Eq. (11), we obtain

$$\begin{bmatrix} \sin 4(\phi_a - \psi_2) & \sin 4(\phi_b - \psi_2) \\ \sin 6(\phi_a - \psi_3) & \sin 6(\phi_b - \psi_3) \end{bmatrix} \times \begin{bmatrix} \cos 4(\phi_a - \psi_2) & \cos 4(\phi_b - \psi_2) \\ \cos 6(\phi_a - \psi_3) & \cos 6(\phi_b - \psi_3) \end{bmatrix}^{-1} \begin{bmatrix} M\lambda_2 \\ M\lambda_3 \end{bmatrix} = 0, \tag{14}$$

where ϕ_a and ϕ_b can be obtained from Eq. (14), and m_a and m_b can be calculated from Eq. (13).

In order to proceed to beat control, we consider an axisymmetric ring structure as shown in Table I. This structure is designed such that it has the same first frequency (n=2 mode) as that of the famous Divine Bell of King Seong-deok.[13] For inducing the asymmetry, we add five point masses m_i=[3,1,4,6,4] kg to the ring at Φ_i =[0,35,125,260,300]°. The natural frequencies and L-mode anti-nodes of the original asymmetric ring are given in Table I.

By employing Fox's theoretical solutions, we can acquire an n=2 equivalent model and an n=3 equivalent model that only satisfy the mode condition for each n=2 and n=3 mode. The results are as follows:

$n=2$ equivalent ring model: $m_{\text{eq}} = [3.3285]$ kg,

$\phi_{\text{eq}} = [14.49]°$,

$n=3$ equivalent ring model: $m_{\text{eq}} = [9.4143]$ kg,

$\phi_{\text{eq}} = [7.57]°$.

In Table I, the dual mode equivalent ring model can be acquired by substituting n=2 and n=3 mode data in Eqs. (13) and (14) simultaneously. Using Eq. (14), the positions of two masses ϕ_a and ϕ_b that satisfy n=2 and n=3 mode data are located in Fig. 2.

The cross points in Fig. 2 become the positions of the two masses and they are symmetric with respect to a line x=y. The reason for this is that Eq. (14) is complete when we exchange ϕ_a and ϕ_b and the order of the two masses has no effect on the asymmetry. Therefore, nine solutions below the line x=y among the 18 solutions in Fig. 2 are omitted. The

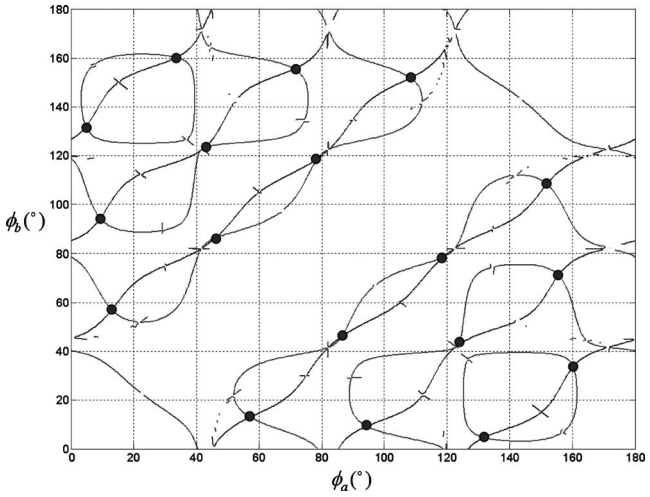


FIG. 2. Position of two imperfection masses

TABLE II. The $n=2$ and $n=3$ equivalent models.

No.	$\phi_a(^\circ)$	$\phi_b(^\circ)$	m_a (kg)	m_b (kg)	M_0 (kg)	ω_{2L}(Hz) ω_{3L}(Hz)	ω_{2H}(Hz) ω_{3H}(Hz)	$\psi_2(^\circ)$ $\psi_3(^\circ)$	$^aT_{2b}$(s) $^bT_{3b}$(s)
1	4.73	131.90	6.03	4.04	1692.4	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
2	33.65	160.27	−4.11	−5.86	1712.5	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
3	9.32	94.28	6.31	−3.40	1699.6	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
4	42.17	122.94	−4.87	5.74	1701.6	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
5	71.24	155.54	3.52	−6.28	1705.3	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
6	13.03	57.08	8.41	5.11	1689.0	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
7	46.09	86.40	−9.83	−8.27	1720.6	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
8	78.62	118.82	8.49	9.83	1684.2	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25
9	108.35	151.87	−5.35	−8.60	1716.5	64.57	64.64	14.49	14.29
						182.25	183.05	7.57	1.25

[a]T_{2b}: beat period of $n=2$ mode.
[b]T_{3b}: beat period of $n=3$ mode.

magnitudes of two masses m_a and m_b can then be obtained by substituting the acquired nine solutions into Eq. (13). Finally, the magnitude and the phase of the dual mode equivalent ring models are listed in Table II. Figure 3 shows the mode shapes and the L-mode anti-nodes of (0,2) and (0,3) modes in the original ring model and dual mode equivalent ring model. The position of L-mode anti-nodes coincides in both the models.

Appropriate dual mode equivalent ring models satisfy given natural frequencies and the location of the low-mode anti-node of the original ring model. Furthermore, all these models exhibit similar changes in natural frequencies and mode shapes when some imperfection masses are added to or subtracted from them. Therefore, if we use any dual mode equivalent model from among the nine models for tuning the beat characteristics and the location of the nodal line, we will obtain similar results.

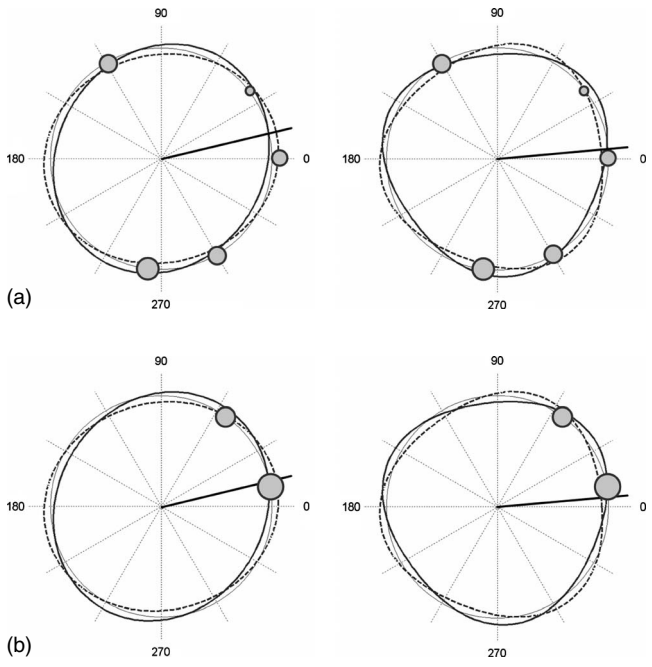


FIG. 3. Mode shapes and L-mode anti-node of (0,2) mode and (0,3) mode in the original ring and equivalent ring models. (a) Original ring with 5 masses, (b) dual mode equivalent ring with two point masses.

Fox's equivalent ring model is based on the thin shell theory. When applying the model to a bell-type structure, the thickness effect may occur. Because the ring used in this research has a similar ratio of thickness to radius of a real bell, we investigated the justification of the dual mode equivalent model using FEA. We used the commercial finite element method (FEM) code ANSYS for the FEA and the element used in the FEA was SOLID 45. The eigenvalue problem formulated within the FEM for the vibration analysis was solved by the reduced subspace method. The theoretical results and FE results are compared in Table III for the original ring model (ring model with five point masses).

The reason why the theoretical natural frequencies of the symmetric ring structure are slightly higher than those obtained in by the FEA result is that we used the thin shell theory. Despite this fact, the low-mode anti-node location, which is an important parameter for evaluating asymmetry, and the differences in natural frequencies between the symmetric and asymmetric ring structures are coincident in both methods. In Fig. 4, we compare the theoretical natural frequencies and the location of low-mode anti-node with those obtained from the FEA results for the dual mode equivalent ring model.

Expectedly, the theoretical location of the low-mode anti-node and that obtained from the FEA results are coincident across the nine dual mode equivalent ring models shown in Figs. 4(a) and 4(b). Figures 4(c) and 4(d) show the beat period, which is the inverse of the difference between

TABLE III. Natural frequencies of original ring.

Mode (n)	ω_{n_TH}(Hz)	ω_{n_FEM}(Hz)	$\omega_{0,n}-\omega_{n\_\mathrm{TH}}$	$\omega_{0,n}-\omega_{n\_\mathrm{FEM}}$	$\psi_{n\_\mathrm{TH}}(^\circ)$	$\psi_{n\_\mathrm{FEM}}(^\circ)$
2-0	64.95	60.91	•	•	•	•
2-*L*	64.57	60.52	−0.38	−0.39	14.49	14.5
2-*H*	64.64	60.59	−0.31	−0.32	−30.51	−30.5
3-0	183.62	167.79	•	•	•	•
3-*L*	182.25	166.39	−1.37	−1.4	7.57	7.5
3-*H*	183.05	167.18	−0.57	−0.61	−22.43	−22.5

the low- and high-mode natural frequencies. In the original symmetric ring, some error occurs due to the difference between the theoretical and FEA results; however, the beat periods in the $n=2$ and $n=3$ modes are very similar in both methods. This research focuses not on the expectation of the exact natural frequencies but on the change in the natural frequencies and location of the low-mode anti-node by asymmetry, parameters that are most important for determining beat clarity and period. In conclusion, it would be acceptable to move on to the tuning procedure by accepting that differences between the theoretical results and the FEA results do exist.

IV. DUAL MODE TUNING PROCEDURE

By the impulse response analysis,[12] a slightly asymmetric ring exhibits a beat in each mode as per the following equation:

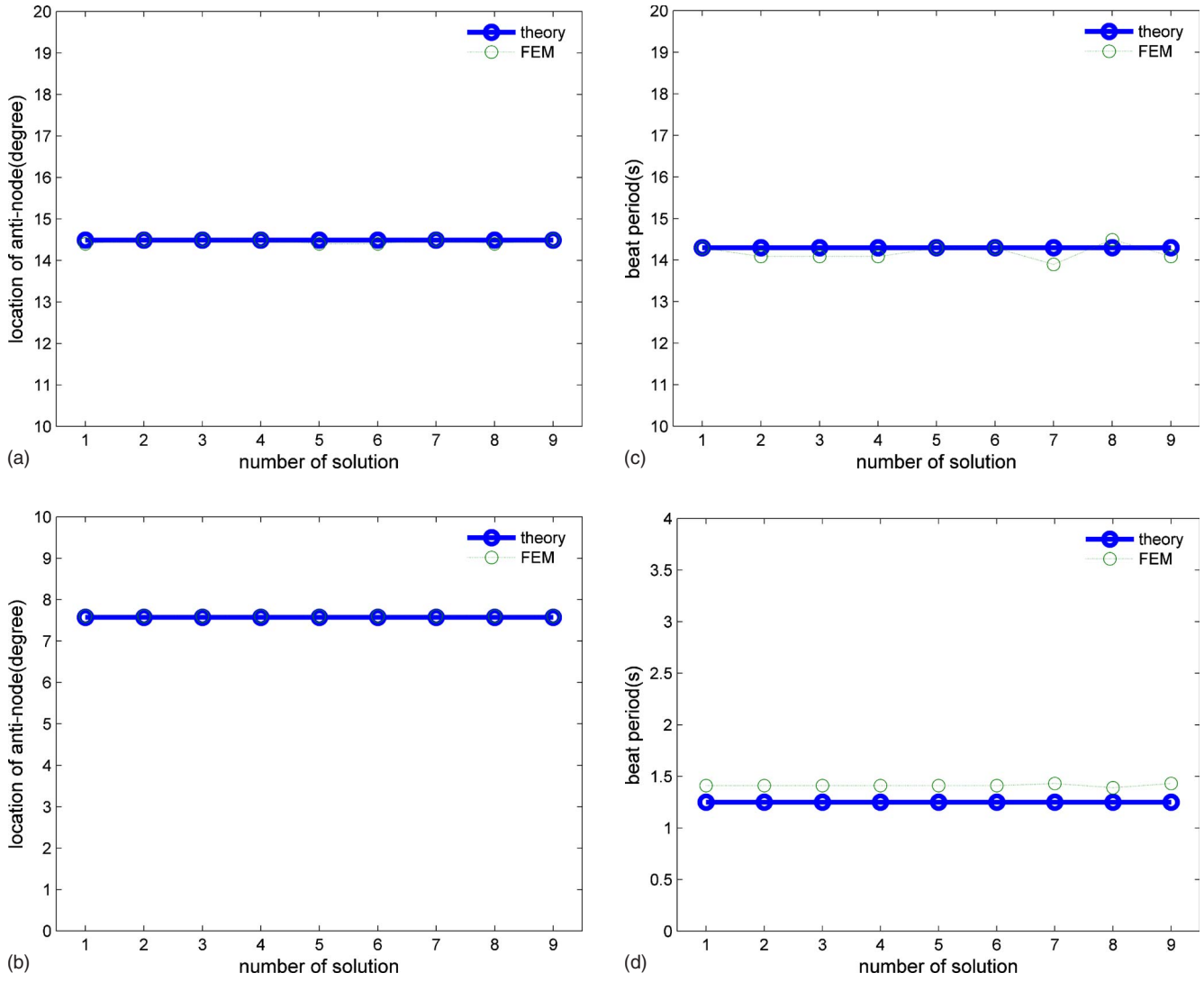


FIG. 4. (Color online) Comparison theoretical results to FE results in natural frequencies and location of low mode anti-node. (a) The $n=2$ mode low mode anti-node, (b) $n=3$ mode low mode anti-node, (c) $n=2$ mode beat period, and (d) $n=3$ mode beat period.

$$\ddot{u}_n(x,\theta,t) = -e^{-\zeta_{na}\frac{\omega_{nL}+\omega_{nH}}{2}t}[\cos n(\theta^* - \phi_L)\cos n(\theta - \phi_L)\sin(\omega_{nL}t) + \cos n(\theta^* - \phi_H)\cos n(\theta - \phi_H)\sin(\omega_{nH}t)], \quad (15)$$

where ζ_{na} is average value of damping, ω_{nH} is natural frequency in high mode, ω_{nL} is natural frequency in low mode, θ^* is position of striking point, ϕ_L is position of anti-node in low mode, and ϕ_H is position of anti-node in high mode.

When the striking point is placed at the center of low- and high-mode anti-nodes, the low and high modes are equally excited and a clear beat is generated. It is, therefore, desirable to modify a Korean bell after casting so that the anti-node in the $n=2$ mode is located at 22.5 or −22.5° and that in the $n=3$ mode, at 15 or −15° from the striking point. The beat period is also as important a factor as beat clarity. When the beat period is too long, people will not hear the beating sound. A short beat period makes the sound appear like a warble. People naturally experience a sound filled with vitality when the beat period in the $n=2$ mode is 3–4 s, such as the period of human breathing.[1] We tuned the original ring so that it had a clear beat with an appropriate period. The original ring does not exhibit a very good mode condition for a clear beat because the $n=2$ mode anti-node is located at 14.49° and the $n=3$ mode is located at 7.57°. The $n=2$ and $n=3$ low-mode anti-nodes should be shifted to 22.5° and 15°, respectively, so that a clear beat is obtained. Furthermore, the beat period in the $n=2$ mode of the original ring model was 14.29 s, which is too long, and that in the $n=3$ mode was 1.25 s, which is too short. We tuned the original ring model so that its beat period in the $n=2$ and $n=3$ modes was 4 and 7 s, respectively, with reference to the beat period of the Divine Bell of King Seong-deok. Similar results are obtained when any dual mode equivalent ring models are used. Here, we use the sixth dual mode equivalent ring model shown in Table II (m_i=[8.42,5.11] kg, Φ_i =[13.03,57.08]°).

Tuning objective: T_{2b}=4 s, T_{3b}=7 s, ψ_2^*=22.5°, ψ_3^* =15.0° (T_{nb}: beat period of nth mode, ψ_n^*: target position of the nth mode anti-node).

We apply constraints to λ_2 and λ_3 in Eqs. (9) and (10) to satisfy the target beat period in the manner shown below:

$$\lambda_n = \frac{(\omega_{nL}^2 - \omega_{nH}^2)(1+\alpha_n^2)}{(\omega_{nL}^2 + \omega_{nH}^2)(1-\alpha_n^2)} = \frac{(\omega_{nL}+\omega_{nH})(\omega_{nL}-\omega_{nH})(1+\alpha_n^2)}{(\omega_{nL}^2 + \omega_{nH}^2)(1-\alpha_n^2)} \propto (\omega_{nL}-\omega_{nH}) = \frac{1}{T_{nb}}. \quad (16)$$

We use $1/k_n\lambda_n$ instead of λ_n to increase the beat period by k_n times because λ_n is nearly inversely proportional to T_{nb}. When we add m_c and m_d to ϕ_c and ϕ_d and develop the dual mode equivalent ring model to satisfy the target beat period and the location of the low-mode anti-node, the following equations are obtained:

$$m_a \sin 4(\phi_a - \psi_2^*) + m_b \sin 4(\phi_b - \psi_2^*) + m_c \sin 4(\phi_c - \psi_2^*) + m_d \sin 4(\phi_d - \psi_2^*) = 0, \quad (17)$$

TABLE IV. Tuning results in beat and location of low-mode anti-node.

Specification of axisymmetric ring		M=1702.5 kg, ρ=8700 kg/m^3, R=1.012 m, E=5.6e10 Pa, h=0.203 m, d=0.15 m				
No.	Magnitude and position of 3rd and 4th masses	Mode	ω_{nL}(Hz)	ω_{nH}(Hz)	Beat period(s)	ψ_n(°)
1	m=[12.53,6.23] kg	$n=2$	64.133	64.379	4.07	−22.5
	Φ=[31.97,176.14]°	$n=3$	181.582	181.726	6.93	−15.0
2	m=[8.71,−3.81] kg	$n=2$	64.389	64.638	4.02	22.49
	Φ=[31.92,140.12]°	$n=3$	182.311	182.453	7.0	14.97
3	m=[−7.74,−1.58] kg	$n=2$	64.658	64.910	3.97	22.49
	Φ=[67.98,175.46]°	$n=3$	183.075	183.219	6.94	−15.0
4	m=[6.37,3.81] kg	$n=2$	64.290	64.538	4.04	22.5
	Φ=[31.85,104.11]°	$n=3$	182.031	182.172	7.06	15.0
5	m=[−8.72,1.58] kg	$n=2$	64.616	64.868	3.97	22.50
	Φ=[67.92,139.46]°	$n=3$	182.957	183.100	6.96	15.14
6	m=[7.74,−6.38] kg	$n=2$	64.456	64.705	4.00	22.5
	Φ=[103.96,175.85]°	$n=3$	182.5	182.643	7.01	15.14
7	m=[2.54,−6.17] kg	$n=2$	64.55	64.801	3.99	22.47
	Φ=[31.48,68.08]°	$n=3$	182.768	182.911	6.98	15.03
8	m=[−10.30,−2.56] kg	$n=2$	64.725	64.978	3.95	22.49
	Φ=[67.85,103.46]°	$n=3$	183.266	183.411	6.92	15.01
9	m=[14.13,10.31] kg	$n=2$	64.025	64.270	4.09	22.47
	Φ=[103.89,139.84]°	$n=3$	181.278	181.418	7.15	15.26
10	m=[−12.53,−14.20] kg	$n=2$	64.992	65.248	3.91	22.60
	Φ=[139.96,175.94]°	$n=3$	184.023	184.169	6.82	15.66

$$m_a \cos 4(\phi_a - \psi_2^*) + m_b \cos 4(\phi_b - \psi_2^*) + m_c \cos 4(\phi_c - \psi_2^*) + m_d \cos 4(\phi_d - \psi_2^*) = M\frac{1}{k_2}\lambda_2, \tag{18}$$

$$m_a \sin 6(\phi_a - \psi_3^*) + m_b \sin 6(\phi_b - \psi_3^*) + m_c \sin 6(\phi_c - \psi_3^*) + m_d \sin 6(\phi_d - \psi_3^*) = 0, \tag{19}$$

$$m_a \cos 6(\phi_a - \psi_3^*) + m_b \cos 6(\phi_b - \psi_3^*) + m_c \cos 6(\phi_c - \psi_3^*) + m_d \cos 6(\phi_d - \psi_3^*) = M\frac{1}{k_3}\lambda_3. \tag{20}$$

Here m_a, m_b, ϕ_a, ϕ_b, λ_2, and λ_3, provides information on the $n=2$ and $n=3$ equivalent ring models, and the values of these parameters are already known. Therefore, by equating Eqs. (17)–(20) to calculate m_c, m_d, ϕ_c, ϕ_d, the values of which are not known, we obtain the following equations:

$$\begin{bmatrix} \sin 4(\phi_c - \psi_2^*) & \sin 4(\phi_d - \psi_2^*) \\ \sin 6(\phi_d - \psi_3^*) & \sin 6(\phi_d - \psi_3^*) \end{bmatrix} \begin{bmatrix} m_c \\ m_d \end{bmatrix} = \begin{bmatrix} -m_a \sin 4(\phi_a - \psi_2^*) - m_b \sin 4(\phi_b - \psi_2^*) \\ -m_a \sin 6(\phi_a - \psi_3^*) - m_b \sin 6(\phi_b - \psi_3^*) \end{bmatrix}, \tag{21}$$

$$\begin{bmatrix} \cos 4(\phi_c - \psi_2^*) & \cos 4(\phi_d - \psi_2^*) \\ \cos 6(\phi_c - \psi_3^*) & \cos 6(\phi_d - \psi_3^*) \end{bmatrix} \begin{bmatrix} m_c \\ m_d \end{bmatrix} = \begin{bmatrix} -m_a \cos 4(\phi_a - \psi_2^*) - m_b \cos 4(\phi_b - \psi_2^*) + M\frac{1}{k_2}\lambda_2 \\ -m_a \cos 6(\phi_a - \psi_3^*) - m_b \cos 6(\phi_b - \psi_3^*) + M\frac{1}{k_3}\lambda_3 \end{bmatrix}. \tag{22}$$

From Eqs. (21) and (22), we obtain Eq. (23):

$$\begin{bmatrix} \cos 4(\phi_c - \psi_2^*) & \cos 4(\phi_d - \psi_2^*) \\ \cos 6(\phi_c - \psi_3^*) & \cos 6(\phi_d - \psi_3^*) \end{bmatrix} \begin{bmatrix} \sin 4(\phi_c - \psi_2^*) & \sin 4(\phi_d - \psi_2^*) \\ \sin 6(\phi_d - \psi_3^*) & \sin 6(\phi_d - \psi_3^*) \end{bmatrix}^{-1} \begin{bmatrix} -m_a \sin 4(\phi_a - \psi_2^*) - m_b \sin 4(\phi_b - \psi_2^*) \\ -m_a \sin 6(\phi_a - \psi_3^*) - m_b \sin 6(\phi_b - \psi_3^*) \end{bmatrix} = \begin{bmatrix} -m_a \cos 4(\phi_a - \psi_2^*) - m_b \cos 4(\phi_b - \psi_2^*) + M\frac{1}{k_2}\lambda_2 \\ -m_a \cos 6(\phi_a - \psi_3^*) - m_b \cos 6(\phi_b - \psi_3^*) + M\frac{1}{k_3}\lambda_3 \end{bmatrix}. \tag{23}$$

Here, information on the sixth dual mode equivalent ring model listed in Table II is as follows:

$$m_a = 8.42 \text{ kg}, \quad m_b = 5.11 \text{ kg}, \quad \phi_a = 13.03^\circ,$$

$$\phi_b = 57.08^\circ, \quad \lambda_2 = 0.002, \quad \lambda_3 = 0.0055,$$

$$T_{2b} = 13.28s, \quad T_{3b} = 1.25s, \quad \psi_2^* = \pm 22.5^\circ,$$

$$\psi_3^* = \pm 15^\circ.$$

If we wish to change the beat period of the original ring model by 4 s in the $n=2$ mode and by 7 s in the $n=3$ mode, $k_2=4/13.28$ and $k_3=7/1.25$ should be set.

By substituting the above data in Eq. (23), we obtain

$$\begin{bmatrix} \cos 4(\phi_c - \psi_2^*) & \cos 4(\phi_d - \psi_2^*) \\ \cos 6(\phi_c - \psi_3^*) & \cos 6(\phi_d - \psi_3^*) \end{bmatrix} \times \begin{bmatrix} \sin 4(\phi_c - \psi_2^*) & \sin 4(\phi_d - \psi_2^*) \\ \sin 6(\phi_d - \psi_3^*) & \sin 6(\phi_d - \psi_3^*) \end{bmatrix}^{-1} \begin{bmatrix} 1.766 \\ 6.600 \end{bmatrix} = \begin{bmatrix} 8.2302 \\ -5.0155 \end{bmatrix}. \tag{24}$$

From Eq. (24), the attachment positions of the two masses ϕ_c and ϕ_d can be determined by generating a similar one to Fig. 2 and finding the cross points of the two curves. By substituting ϕ_c and ϕ_d into Eq. (22), we can then calculate m_c and m_d, which are the magnitudes of the two masses. Note that from 20 solutions that are symmetric with respect to a line $x=y$, only ten solutions are selected based on symmetry and the results are summarized in Table IV.

We can verify that the location of the $n=2$ mode anti-node has shifted to $\psi_2 = \pm 22.5^\circ$ and that of the $n=3$ mode has shifted to $\psi_3 = \pm 15^\circ$ in all the ten tuning results. The beat period in the $n=2$ mode has changed to about 4 s and that in the $n=3$ mode has changed to about 7 s; these values are in agreement with our target tuning objective.

Figure 5 shows the comparison of the theoretical results and the FEA results after tuning. In Figs. 5(a) and 5(b), the locations of the low-mode anti-node in both the methods are coincident for the ten dual mode equivalent ring models. In Fig. 5(a), the $n=2$ low-mode anti-node in the first model is located at -22.5°, which is different from that in the other models, and it does not influence beat clarity. A similar behavior is observed in the case of the $n=3$ mode in Fig. 5(b). In Figs. 5(c) and 5(d), the theoretical beat period is slightly greater than that obtained from the FEA because of the initial error due to the thickness effect. However, the beat periods in all the ten models are very similar and this justifies the theoretical analysis. In conclusion, an equivalent model of an asymmetric ring can be constructed from given $n=2$ and n

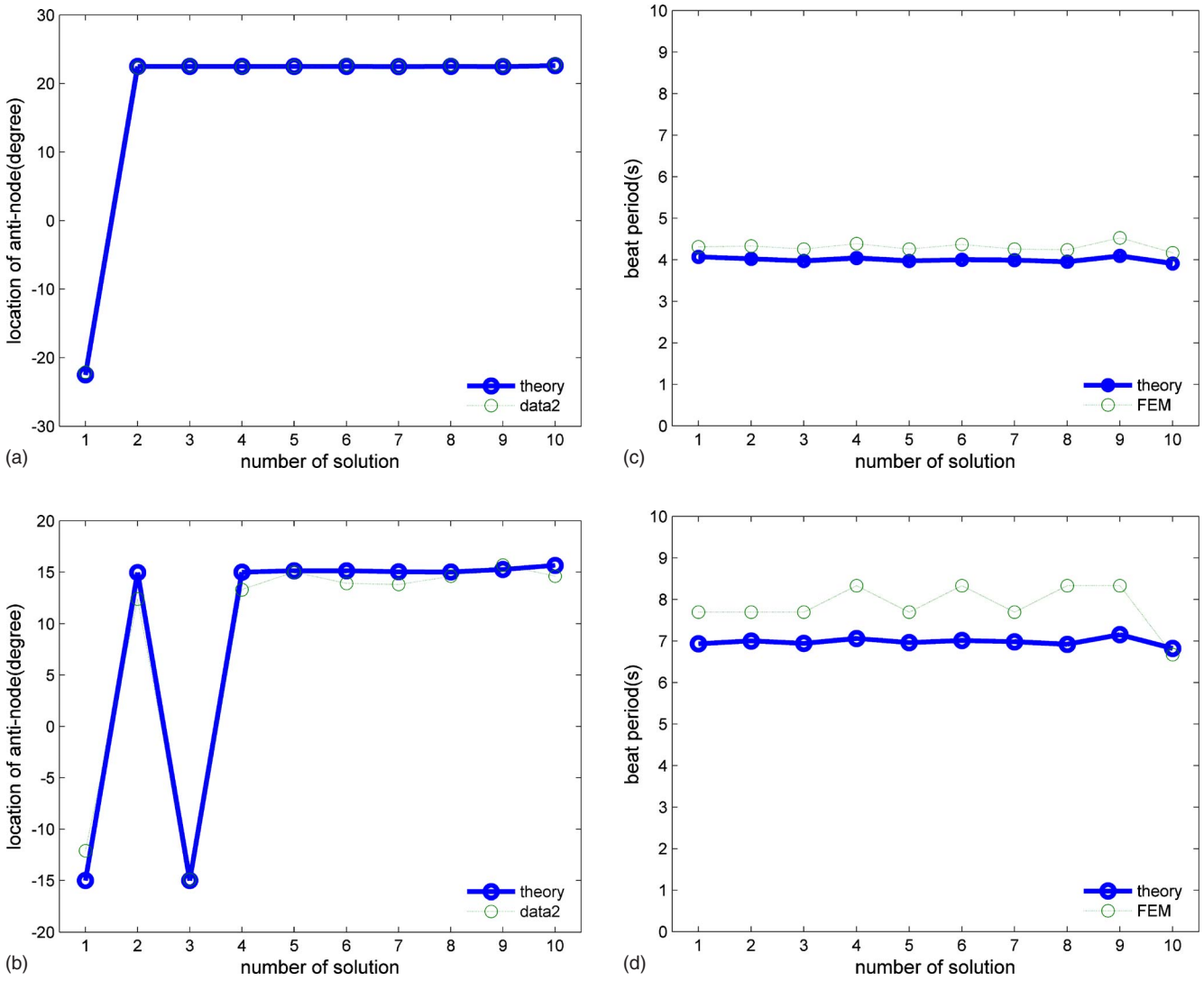


FIG. 5. (Color online) Beat period and location of low-mode anti-node after tuning. (a) The $n=2$ mode anti-node, (b) $n=3$ mode anti-node, (c) $n=2$ mode beat period and (d) $n=3$ mode beat period.

$=3$ mode data and the beat characteristics can be tuned by calculating the magnitude and position of additional masses.

In reality, however, it is a more frequent practice to introduce cuts in the structure to modify it rather than add masses. The cut simultaneously diminishes the mass and stiffness, and Hong showed that the stiffness reduction effect is considerably stronger than the mass reduction effect.[9] Since the stiffness effect is much stronger than the mass effect, the anti-node of the *L*-mode passes the location of the cut; this result is the same as that obtained in the case of an equivalent mass located at the anti-node of the *L*-mode. These phenomena were verified by comparison with the FEA results, which were obtained for a real Korean bell structure.[14,15] Since it is very difficult to compose an equivalent model that reflects the exact cut effect, an equivalent model using imperfection masses could serve as a substitute. If we can determine the relationship between the cut effect and the mass effect from measurement or by FEA, we can improve the usefulness of the suggested tuning method that employs imperfection masses.

V. CONCLUSION

There were some differences between the theoretical results and the FEA results with regard to the natural frequencies; however, the changes in the natural frequencies and location of the low-mode anti-node were nearly coincident. As the analytical results suggested, the equivalent ring model could be used not only to trim the split mode but also to tune the beat characteristics and locations of the nodal line in a special mode or two modes simultaneously.

ACKNOWLEDGMENTS

This research was supported by the Institute of Advanced Machinery and Design at Seoul National University.

[1]Y. H. Yum, A study on the Korean bell, Research Institute of Korean Spirits and Culture Research Report No. 84-14 (1984).
[2]R. Perrin, "A group theoretical approach to warble in ornamented bells,"J. Sound Vib. **52**, 307–313 (1977).
[3]M. Jing, "A theoretical study of the vibration and acoustics of ancient Chinese bell," J. Acoust. Soc. Am. **114**(3), 1622-1628 (2003).
[4]Kyungjoo National Museum, Comprehensive Research Reports on the

King Seong-deok Divine Bell (1999).
[5]T. Charnley and R. Perrin, "Studies with eccentric bell," J. Sound Vib. **58**, 517–525 (1978).
[6]R. Perrin, T. Charnley, and H. Bandu, "Increasing the lifetime of warble-suppressed bells," J. Sound Vib. **80**, 298–303 (1982).
[7]R. Perrin, T. Charnley, and G. M. Swallowe, "On the tuning of church and carillon bells," Appl. Opt. **46**, 83–101 (1995).
[8]D. Allaei, W. Soedel, and T. Y. Yang, "Vibration analysis of non-axisymmetric tires," J. Sound Vib. **122**, 11–29 (1988).
[9]J. S. Hong and J. M. Lee, "Vibration of circular rings with local deviation," J. Appl. Mech. **61**, 317–322 (1994).
[10]C. H. J. Fox, A simple theory for the analysis and correction of frequency splitting in slightly imperfect rings," J. Sound Vib. **142**(2), 227–243 (1990).
[11]A. K. Rourke, S. McWilliam, and C. H. J. Fox, "Multi-mode trimming of imperfect rings," J. Sound Vib. **248**(4), 695–724 (2001).
[12]S. H. Kim, W. Sodel, and J. M. Lee, "Analysis of the beating response of bell type structures," J. Sound Vib. **173**(4), 517–536 (1994).
[13]S. H. Kim, C. W. Lee, and J. M. Lee, "Beat characteristics and beat maps of the King Seong-deok Divine Bell," J. Sound Vib. **281**, 21–44 (2005).
[14]S. H. Cheon, J. M. Lee, S. H. Kim, and Y. H. Yum, "A study on the vibration property of Korean bell," Trans. Kor. Soc. Mech. **13**, 397–403 (1989).
[15]J. M. Lee, S. H. Kim, S. J. Lee, J. D. Jeong, and H. G. Choi, "A study on the vibration charactereistics of a large size Korean bell," J. Sound Vib. **257**(4), 779–790 (2002).

Inclusion of localized forces due to turbulent boundary layer convected pressure at junctions of coplanar structural sections

M. L. Rumerman[a)]
Signatures Department (Code 7204), Naval Surface Warfare Center, Carderock Division, 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5700, USA

(Received 25 June 2007; revised 23 November 2007; accepted 8 December 2007)

Previous researchers showed that the concept of turbulent boundary layer (TBL) "edge forces," which are related to the convection peak of the TBL pressure wave number spectral density, is useful and important when using a finite element model to compute the bending response to TBL excitation of a plate that has a free edge. This paper examines the applicability of more general localized forces to the junction of two distinct coplanar sections that are joined along a line transverse to the flow. A formalism is developed for introducing these forces and applying them to a structure. The approach is made plausible by using two string sections as a simple structural model and showing that the sum of the localized force response and that due to the low wave number excitation matches the separately computed response corresponding to the full wave number spectral density. These localized forces can have a noticeable effect when the connection between sections is resilient.
[DOI: 10.1121/1.2831771]

PACS number(s): 43.40.Dx, 43.40.Hb, 43.40.At [DF] Pages: 1392–1398

I. INTRODUCTION

The mean squared response of a structure driven by a turbulent boundary layer (TBL) pressure field that is stationary and homogeneous can be calculated with a finite element model (FEM) by forming the complex conjugate product of all pairs of computed transfer functions to a particular response point and using the TBL spatial coherence function to weight the contributions. In order to keep the FEM computationally manageable, the element size is chosen no smaller than that required to resolve the smallest vibration wavelength of interest, usually the bending wavelength. This presents a practical problem for many applications of steel plates in water due to the nature of TBL pressure wave number spectral density for the flow direction, which is sketched in Fig. 1 at zero transverse wave number. If the convection wavelength is smaller than the bending wavelength, as is typical for such applications, the FEM cannot properly respond to the large pressures at the convection peak, and use of the full spatial coherence function, which is equivalent to the wave number spectral density, will produce incorrect results. Hwang[1] and Hwang and Hambric[2] considered this difficulty and proposed a solution based on the recognition that, for many structures, the bending response is controlled by that portion of the wave number spectral density in the neighborhood of the structure's bending wave number, k_b. Because, in many practical underwater applications, the wave number spectral density tends to be constant in this region, it can be represented as the band-limited spectral density given by Eq. (1), in which $\Phi_p(\omega)$ is the point pressure frequency spectral density and A is a factor that is independent of the two-dimensional wave vector γ_1, γ_3, where the number 1 axis is parallel to the direction of flow and the number 3 axis is transverse to the flow,

$$S(\gamma_1,\gamma_3;\omega)=\Phi_p(\omega)A, \quad |\gamma|<k_e, \tag{1a}$$

$$S(\gamma_1,\gamma_3;\omega)=0, \quad |\gamma|>k_e. \tag{1b}$$

The nonzero range must include the structural wave number, and for plates in bending, it is sufficient to set $k_e=1.5k_b$. Hwang and Hambric computed the coherence function corresponding to Eq. (1), which is compatible with the resolution of a standard FEM for bending, and used it in a finite element analysis of a clamped rectangular plate set in an infinite baffle subjected to a stationary and homogeneous TBL pressure field. They showed that this approach produced the same results as did the full coherence function applied to a finer-mesh FEM that resolved the convection wavelength.

A TBL-driven structure that has a free edge transverse to the flow direction is an important exception to the rule that the full and band-limited wave number spectral densities (or their corresponding coherence functions) yield equivalent responses. For such a boundary condition, analysis indicates that a significant part of the mean squared response can be attributed to a virtual line force normal to the structure applied along that edge. Chandiramani[3] showed how such "edge forces" were related to the convection peak of the wave number spectral density. Hwang and Maidanik[4] examined the response of rectangular plates to the convection peak and demonstrated that plates with free edges have larger responses than do plates with simply supported edges, which in turn respond better than clamped plates. Hambric *et al.*[5] applied an edge force in the FEM calculation of the TBL response of a baffled rectangular plate with a free trailing edge and three clamped edges and showed that it could account for a large fraction of the response levels calculated using the full coherence function and a high resolution FEM.

The purpose of this paper is to show how general localized line forces can be applied along transverse lines that are

a)Electronic mail: melvyn.rumerman@navy.mil.

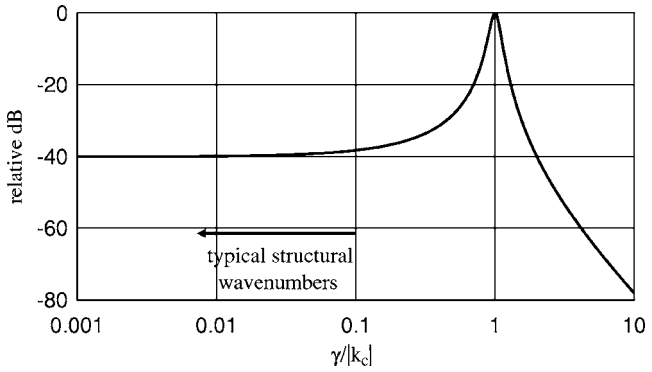


FIG. 1. Representative TBL pressure wave number spectral density.

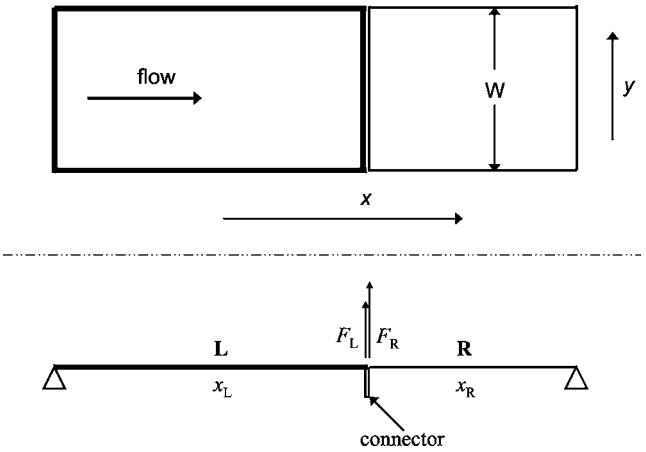


FIG. 2. Coplanar connected sections.

not at upstream or downstream structural terminations, but at transverse junctions of connected sections. It will be shown that such localized forces are relevant when sections are connected through a resilient element, so that the input admittance is discontinuous across the junction, and that the sum of the responses due to the localized force and the low wave number excitation is a good approximation to the response due to the complete excitation spectral density. In Sec. II, the TBL pressure model is described. Section III develops a formalism for developing localized forces at structural junctions. Section IV contains specific illustrations for a simplified structure, and the analysis assumes that the structure is placed in a baffle so that the TBL pressure field can be considered stationary and homogeneous. Real structures require special consideration of their interactions with the flow at leading and trailing edges, and some of these aspects are discussed by Howe.[6]

II. TBL PRESSURE MODEL

The analysis and calculations will be based on the "modified Corcos" TBL model recommended by Hwang,[1] in which the statistics in the flow direction are independent of those in the transverse direction and the levels are proportional to the mean squared point pressure frequency spectral density. The frequency-wave-number spectral density is given in Eq. (2a) as the product of the point pressure frequency spectral density and the normalized wave number spectral densities for the flow and transverse directions given in Eqs. (2b) and (2c). Each of these normalized spectral densities forms a Fourier transform pair with the respective coherence functions in Eqs. (2d) and (2e),

$$S(\gamma_1,\gamma_3;\omega)=\Phi_p(\omega)\bar{S}_1(\gamma_1)\bar{S}_3(\gamma_3), \tag{2a}$$

$$\bar{S}_1(\gamma_1)=\frac{1}{|k_c|}\frac{2\alpha_1^3k_c^4}{\pi[(\gamma_1-k_c)^2+(\alpha_1k_c)^2]^2}, \tag{2b}$$

$$\bar{S}_3(\gamma_3)=\frac{1}{|k_c|}\frac{\alpha_3k_c^2}{\pi[\gamma_3^2+(\alpha_3k_c)^2]}, \tag{2c}$$

$$\Gamma_1(v_1)=(1+\alpha_1|k_cv_1|)\exp(-\alpha_1|k_cv_1|)\exp(+ik_cv_1), \tag{2d}$$

$$\Gamma_3(v_3)=\exp(-\alpha_3|k_cv_3|). \tag{2e}$$

Hwang quotes values for α_1 and α_3 in the ranges of 0.11–0.12 and 0.7–1.2, respectively.

Hwang and Hambric[2] developed their plate analysis as a two-dimensional problem. If the dependence of the structural response on the coordinate transverse to the flow can be represented in terms of Fourier modes, then, for each of these modes, the problem can be considered one dimensional in the flow direction. The corresponding coherence function for a band-limited spectral density model is then given by Eq. (3), in which A is obtained by setting γ_1 equal to zero in Eq. (2b),

$$\Gamma_1(v_1)=A\int_{-k_e}^{k_e}\exp(iv_1\gamma_1)d\gamma_1=\frac{2A\sin(k_ev_1)}{v_1}, \tag{3a}$$

$$A=\frac{1}{|k_c|}\frac{2\alpha_1^3}{\pi[1+\alpha_1^2]^2}\approx\frac{2\alpha_1^3}{\pi|k_c|}. \tag{3b}$$

This will be referred to as the band-limited coherence function.

III. LOCALIZED FORCE FORMALISM

Consider the "structure" sketched in Fig. 2. It consists of two sections, labeled L and R, each of which is of width W in the y direction transverse to the flow and constrained to have zero displacement at its outer transverse edge. They are joined at their inner adjacent edges by an arbitrary uniform line connector. Although Fig. 2 shows the sections slightly separated for clarity, there is no intended gap between the adjacent edges. Assume that the edges along $y=0$ and $y=W$ are simply supported, so that the y dependence can be expressed in terms of Fourier modes, which have modal wave numbers $m\pi/W$. Consider the mth mode and let known modal line forces per unit length, F_L and F_R, be applied to the inner edges of sections L and R, respectively, and let $h_L(x_L)$ and $h_R(x_L)$ be the transfer function relating each modal line force to the mth modal structural response, $w(x_L)$ in section L. (The transfer functions vary with the transverse mode number m, which is omitted.) The square of the magnitude of the modal response to these forces is then given by

$$|w(x_L)|^2 = |F_L h_L(x_L) + F_R h_R(x_L)|^2 = |F_L|^2 |h_L(x_L)|^2 + F_L F_R^* h_L(x_L) h_R^*(x_L) + F_L^* F_R h_L^*(x_L) h_R(x_L) + |F_R|^2 |h_R(x_L)|^2. \quad (4a)$$

If these are TBL localized forces, the expected values of the terms are needed,

$$\langle |w(x_L)|^2 \rangle = \langle |F_L|^2 \rangle |h_L(x_L)|^2 + \langle F_L F_R^* \rangle h_L(x_L) h_R^*(x_L) + \langle F_L^* F_R \rangle h_L^*(x_L) h_R(x_L) + \langle |F_R|^2 \rangle |h_R(x_L)|^2. \quad (4b)$$

The modal mean squared line forces $\langle |F_L|^2 \rangle$ and $\langle |F_R|^2 \rangle$ are equal and given by

$$\langle |F_L|^2 \rangle = \langle |F_R|^2 \rangle = \frac{2\pi}{W} \bar{S}_3(m\pi/W) \frac{\Phi_p(\omega)}{k_c^2}. \quad (5)$$

This result can be obtained by considering the convected pressure wave $P_c(y;\omega)\exp[i(k_c x - \omega t)]$ with flow in the positive x direction from negative infinity toward $x=0$. The uncanceled pressure at the downstream edge of section L at $x=0$ is given by Eq. (6), where the subscript "c" stands for convective,

$$\int_{-\infty}^{0} P_{cL}(y;\omega)\exp(ik_c x)dx = \frac{P_{cL}(y;\omega)}{ik_c}. \quad (6)$$

Consideration of another point along the edge yields $P_{cL}^*(y+v_3;\omega)/-ik_c$ for the complex conjugate. Because, for the modified Corcos spectral density, almost the entire mean squared point pressure is contained within a narrow wave number interval across the convection peak, the expected value of the product of the two terms is $\langle |F(y)F^*(y+v_3|\rangle = [\Phi_p(\omega)/k_c^2]\Gamma_3(v_3)$. Apart from a factor of $\pi/2$, which may be due to a normalization, this is the edge force used by Hambric *et al.*[5] in their FEM calculations. Transformation to the transverse wave number domain replaces Γ_3 with $2\pi\bar{S}_3/W$ giving Eq. (5).

The expected value $\langle F_L F_R^* \rangle$ is obtained by a similar approach. Consider the convected pressure wave flowing from the edge of section R at $x=0$ toward positive infinity,

$$F_R = \int_0^{\infty} P_{cR}(y+v_3;\omega)\exp(ik_c x)dx = -\frac{P_{cR}(y+v_3;\omega)}{ik_c}, \quad (7a)$$

$$F_L F_R^* = -\frac{P_{cL}(y;\omega)P_{cR}^*(y+v_3;\omega)}{k_c^2}. \quad (7b)$$

Upon forming the expected value and transforming to the transverse wave number domain, the following is obtained:

$$\langle F_L F_R^* \rangle = -\frac{2\pi}{W} \bar{S}_3(m\pi/W) \frac{\Phi_p(\omega)}{k_c^2}. \quad (7c)$$

$\langle F_L^* F_R \rangle$ is the complex conjugate of $\langle F_L F_R^* \rangle$ and, because each is real, they are equal.

The factor $(2\pi/W)\bar{S}_3$ and explicit frequency dependence will now be suppressed. Substitution of Eqs. (5) and (7c) into Eq. (4b) results in

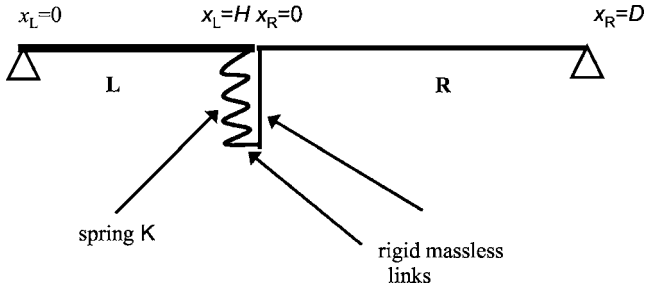


FIG. 3. Collinear strings connected by spring.

$$\langle |w(x_L)|^2 \rangle = \frac{\Phi_p}{k_c^2}\{|h_L(x_L)|^2 + |h_R(x_L)|^2 - 2\,\mathrm{Re}[h_L(x_L)h_R^*(x_L)]\}. \quad (8a)$$

Similarly, the mean squared response in section R is given by

$$\langle |w(x_R)|^2 \rangle = \frac{\Phi_p}{k_c^2}\{|h_L(x_R)|^2 + |h_R(x_R)|^2 - 2\,\mathrm{Re}[h_L(x_R)h_R^*(x_R)]\}. \quad (8b)$$

Two observations may be made regarding Eq. (8). First, if there is no structural connection between sections L and R, and coupling through the water is neglected, then $h_L(x_R)=0$ and $h_R(x_L)=0$, and

$$\langle |w(x_L)|^2 \rangle = \frac{\Phi_p}{k_c^2}|h_L(x_L)|^2, \quad (9a)$$

$$\langle |w(x_R)|^2 \rangle = \frac{\Phi_p}{k_c^2}|h_R(x_R)|^2. \quad (9b)$$

If there is no other constraint on either edge, these are the responses for edge forces applied to free ends. Second, if the connection between sections L and R can be considered rigid, so that $h_L(x_L)=h_R(x_L)$ and $h_L(x_R)=h_R(x_R)$, then $\langle |w(x_L)|^2 \rangle = \langle |w(x_R)|^2 \rangle = 0$. This is expected because the structural response characteristics are continuous across the junction, and the convected pressure is coherent across the junction and has no uncanceled component.

IV. ILLUSTRATIVE EXAMPLE

In this section, the effectiveness of combining the response due to the localized force with that due to the band-limited coherence will be tested for a one-dimensional structural model by comparing their sum to the response due to the complete coherence. The simple string structure shown in Fig. 3 will be used for illustrative purposes. Fluid loading is not included and is not necessary for testing the localized force model. Two string sections, L and R, are fixed at their outer ends and connected at their adjacent inner ends by a spring attached to the end of section L and by horizontal and vertical rigid massless links from the end of the spring to the end of section R. This arrangement keeps the strings collinear and avoids a step to the flow, which is directed from L to R. Because a string is being considered, rather than a membrane or plate, Φ_p will represent a mean squared force per

unit length rather than a mean squared pressure. The subscript "1," explicitly identifying the flow direction, will be omitted.

The strings are specified by a tension T, a mass per unit length ρ, and a loss factor ϵ, which are distinguished for each string by the subscript L or R. The length of string L is H and that of string R is D. The transfer functions, $Y_{L,R}$, that relate the displacement response of each unconnected string (with interior end free) to concentrated normal forces at arbitrary locations $x_L=\xi$ and $x_R=\eta$ are given in the following:

$$Y_L(x_L,\xi)=\frac{1}{\omega\rho_L c_L}\frac{\cos[k_L(H-\xi)]\sin(k_L x_L)}{\cos(k_L H)},\quad x_L<\xi, \tag{10a}$$

$$Y_L(x_L,\xi)=\frac{1}{\omega\rho_L c_L}\frac{\cos[k_L(H-x_L)]\sin(k_L \xi)}{\cos(k_L H)},\quad x_L>\xi, \tag{10b}$$

$$Y_R(x_R,\eta)=\frac{1}{\omega\rho_R c_R}\frac{\sin[k_R(D-x_R)]\cos(k_R \eta)}{\cos(k_R D)},\quad x_R>\eta, \tag{11a}$$

$$Y_R(x_R,\eta)=\frac{1}{\omega\rho_R c_R}\frac{\sin[k_R(D-\eta)]\cos(k_R x_R)}{\cos(k_R D)},\quad x_R<\eta. \tag{11b}$$

The responses of the unconnected strings to a force per unit length excitation, $P(x_{L,R})$, are given by

$$W_{L0}(x_L)=\int_0^H P(\xi)Y_L(x_L,\xi)d\xi, \tag{12a}$$

$$W_{R0}(x_R)=\int_0^D P(\eta)Y_R(x_R,\eta)d\eta. \tag{12b}$$

The solution to the connected system can now be found by noting that the force in the spring is the product of the spring constant and the difference of the total displacements at its top and bottom, that the total displacement of the bottom of the spring is equal to the total displacement of the inner end of string R (because the link is rigid), and that the force that the spring exerts on the end of string L is equal and opposite to the force that the vertical link exerts on the end of string R (because the horizontal and vertical links are massless). The resulting total displacements in each string are given by

$$w_L(x_L)=W_{L0}(x_L)-\frac{K[W_{L0}(H)-W_{R0}(0)]}{1+K[Y_L(H,H)+Y_R(0,0)]}Y_L(x_L,H), \tag{13a}$$

$$w_R(x_R)=W_{R0}(x_R)+\frac{K[W_{L0}(H)-W_{R0}(0)]}{1+K[Y_L(H,H)+Y_R(0,0)]}Y_R(x_R,0). \tag{13b}$$

For a completely rigid connection, $K\to\infty$ and $w_L(H)=w_R(0)$, as required.

A. Mean squared responses

The square of the absolute value of the response is found by multiplying the right-hand side of each equation of Eqs. (13) by its complex conjugate. If the driving force per unit length, P, represents a TBL-like excitation, its statistical representation is included as $\langle P(x)P^*(x-\upsilon)\rangle=\Phi_p\Gamma(\upsilon)$. This leads to a series of double integrals, some of which are illustrated in the following:

$$\langle|W_{L0}(x_L)|^2\rangle=\Phi_p\int_0^H d\xi\int_0^H d\eta\Gamma(\xi-\eta)Y_L(x_L,\xi)Y_L^*(x_L,\eta), \tag{14a}$$

$$\langle|W_{R0}(0)|^2\rangle=\Phi_p\int_0^D d\xi\int_0^D d\eta\Gamma(\xi-\eta)Y_R(0,\xi)Y_R^*(0,\eta), \tag{14b}$$

$$\langle W_{L0}(H)W_{R0}^*(0)+W_{L0}^*(H)W_{R0}(0)\rangle=2\Phi_p\,\mathrm{Re}\int_0^H d\xi\int_0^D d\eta\Gamma(\xi-\eta)Y_L(H,\xi)Y_R^*(0,\eta), \tag{14c}$$

$$\langle W_{R0}(x_R)W_{L0}^*(H)+W_{R0}^*(x_R)W_{L0}(H)\rangle=2\Phi_p\,\mathrm{Re}\int_0^D d\xi\int_0^H d\eta\Gamma(\xi-\eta)Y_R(x_R,\xi)Y_L^*(H,\eta). \tag{14d}$$

B. Response due to band-limited coherence

The band-limited coherence function is found from

$$\Gamma_{bl}(\upsilon)=\frac{4\alpha^3}{\pi|k_c|}\frac{\sin(k_e\upsilon)}{\upsilon}. \tag{15}$$

Use of Eq. (15) in calculations produces only the response to low wave number excitation. The solution is completed by separately calculating the response due to the localized forces.

C. Response due to localized forces

To compute the mean squared response due to the localized forces, consider two concentrated forces, F_L and F_R applied to the inner ends of string L and string R, respectively. The responses of these strings, when unconnected, are $W_{L0}=F_L Y_L(x_L,H)$ and $W_{R0}=F_R Y_R(x_R,0)$. These may be substituted into Eq. (13) to find the coupled response,

$$w_L(x_L) = F_L Y_L(x_L,H) - \frac{K[F_L Y_L(H,H) - F_R Y_R(0,0)]}{1 + K[Y_L(H,H) + Y_R(0,0)]} Y_L(x_L,H) = F_L \frac{[1 + KY_R(0,0)]Y_L(x_L,H)}{1 + K[Y_L(H,H) + Y_R(0,0)]} + F_R \frac{[KY_R(0,0)]Y_L(x_L,H)}{1 + K[Y_L(H,H) + Y_R(0,0)]}, \quad (16a)$$

$$w_R(x_R) = F_R Y_R(x_R,0) + \frac{K[F_L Y_L(H,H) - F_R Y_R(0,0)]}{1 + K[Y_L(H,H) + Y_R(0,0)]} Y_R(x_R,0) = F_L \frac{[KY_L(H,H)]Y_R(x_R,0)}{1 + K[Y_L(H,H) + Y_R(0,0)]} + F_R \frac{[1 + KY_L(H,H]Y_R(x_R,0)}{1 + K[Y_L(H,H) + Y_R(0,0)]}. \quad (16b)$$

In terms of the transfer functions for the connected structure defined in Sec. III, the coefficients of F_L and F_R in Eq. (16a) correspond to $h_L(x_L)$ and $h_R(x_L)$, respectively, and the coefficients of F_L and F_R in Eq. (16b) correspond to $h_L(x_R)$ and $h_R(x_R)$, respectively. The mean squared response due to the localized forces can be found by inserting these expressions into Eq. (8).

D. Calculations and comparisons

Calculations were performed for the above-described string structure using the following parameters:

$$\rho_L \,\mathrm{Re}(c_L) = \rho_R \,\mathrm{Re}(c_R) = \rho c_0,$$

$$\mathrm{Re}(c_R) = 0.5\,\mathrm{Re}(c_L), \quad \mathrm{Re}(k_R) = 2\,\mathrm{Re}(k_L),$$

$$H = 3, \quad D = 4, \quad \epsilon_L = 0.01, \quad \epsilon_R = 0.05,$$

$$k_L = (\omega/c_L) = \mathrm{Re}(k_L)[1 + 0.5i\epsilon_L],$$

$$k_R = (\omega/c_R) = \mathrm{Re}(k_R)[1 + 0.5i\epsilon_R], \quad \alpha = 0.1,$$

$$|k_c| = 10\,\mathrm{Re}(k_R),$$

The units are arbitrary but consistent so that $k_L H$, for example, is independent of the units chosen. Three values of the spring constant were used: $K=0$ (no connectivity), $K=\omega\rho c_0$ (intermediate connectivity), and $K \gg \omega\rho c_0$ (full connectivity).

Results for the three connectivities are presented in Figs. 4–6 for the displacement responses at the two inner ends of the strings, $x_L=H$ and $x_R=0$. The displacements are normalized with respect to $\Phi_p/(\omega\rho c_0)^2$. Each figure includes the response due to the band-limited coherence, the sum of the band-limited response and the localized force response and the response due to the full coherence function. For full connectivity, the effective localized force vanishes, and the band-limited and full coherence functions yield the same results, as shown in Fig. 4. For intermediate connectivity (Fig. 5) and zero connectivity (Fig. 6), there are observable localized force contributions that combine with the low wave number contributions to match the results obtained with the full coherence function, but the importance of these contributions varies with frequency and the other parameters.

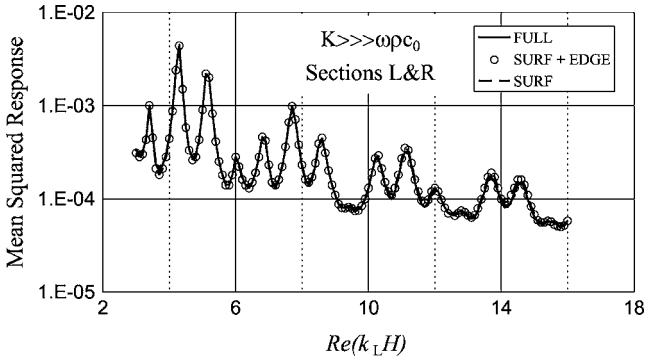


FIG. 4. Mean squared response for full connectivity: Section L ($x_L=H$) and section R ($x_R=0$).

V. DISCUSSION

This paper has attempted to show that the concept of a TBL localized force is relevant not only to plate-like structures having terminating transverse edges that are free to move, but to structures that have sections that are connected by a resilient connection transverse to the flow (so that the input admittance is discontinuous across the connection). In Sec. III it was shown how such localized line forces could be applied, and the formalism developed was used to compute that contribution to the response of two string sections connected by a spring, which was then added to the response

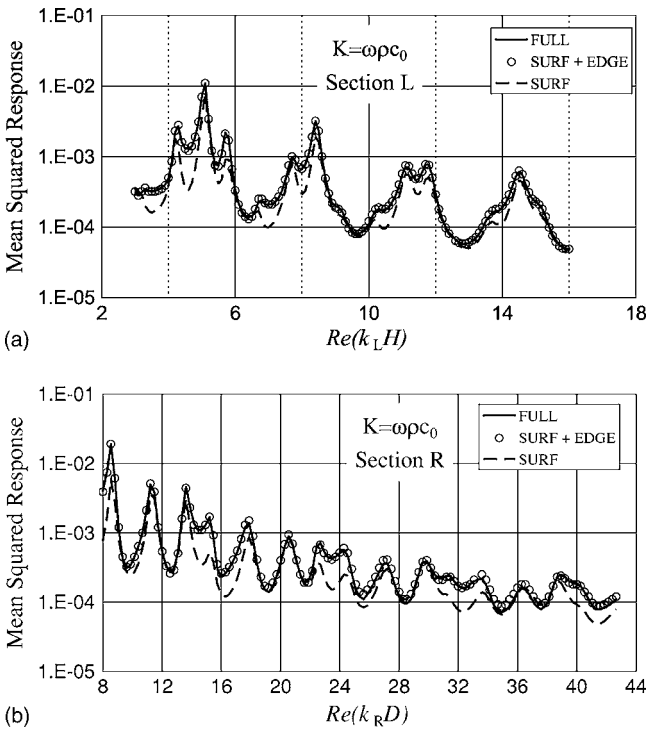
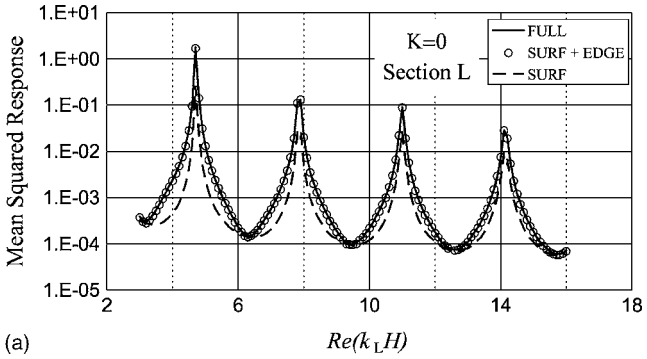
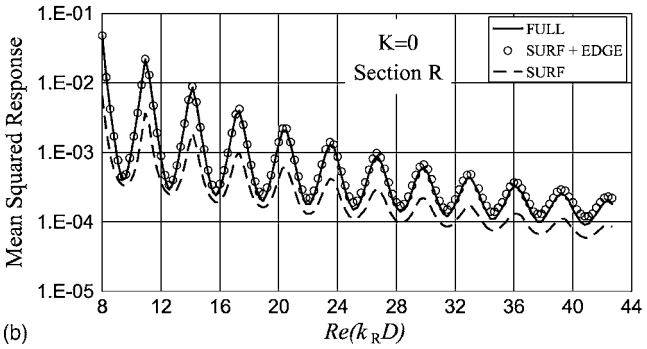


FIG. 5. Mean squared response for intermediate connectivity: (a) Section L ($x_L=H$) and (b) Section R ($x_R=0$).



(a)



(b)

FIG. 6. Mean squared response for no connectivity: (a) Section L ($x_L=H$) and (b) Section R ($x_R=0$).

that was calculated using the approximate band-limited coherence function. The sum is in agreement with the result obtained using the full coherence function.

Although addition of the localized force response to that due to the low wave number excitation produces a better overall result, its relative contribution varies even for a free edge. A feel for what parameters are important in determining the significance of the localized force can be obtained by considering the string with one fixed end and one free end, such as string L with $K=0$. Its total response at the free end is given by Eq. (17a), in which the subscript L has been omitted. The second term on the right-hand side is the contribution of the localized force,

$$\langle|w^2(H)|\rangle \approx \frac{2A\Phi_p}{|\omega\rho c\cos(kH)|^2}\int_0^H d\xi\int_0^H d\eta\frac{\sin[k_e(\xi-\eta)]}{(\xi-\eta)} \times|\sin(k\xi)\sin(k^*\eta)| + \frac{\Phi_p}{k_c^2}\left|\frac{\tan(kH)}{\omega\rho c}\right|^2. \quad (17a)$$

In using a FEM for the computation, the bandwidth, k_e, of the approximate spectral density must remain finite, but when using an exact solution of the string equation it can increase without limit. This allows $\sin[k_e(\xi-\eta)]/(\xi-\eta)$ to be replaced by $\pi\delta(\xi-\eta)$, and Eq. (17a) is replaced by

$$\langle|w^2(H)|\rangle \approx \frac{2\pi A\Phi_p}{|\omega\rho c\cos(kH)|^2}\int_0^H d\xi|\sin(k\xi)|^2 + \frac{\Phi_p}{k_c^2}\left|\frac{\tan(kH)}{\omega\rho c}\right|^2. \quad (17b)$$

The integral can be evaluated, and the result is given by

$$|w^2(H)| \approx \frac{\pi A\Phi_p}{2|\omega\rho c\cos(kH)|^2}\left[\frac{\sinh(\epsilon kH)}{\epsilon k}-\frac{\sin(kH)}{k}\right]+\frac{\Phi_p}{k_c^2}\left|\frac{\tan(kH)}{\omega\rho c}\right|^2$$
$$\approx \frac{\pi A\Phi_p H}{2|\omega\rho c\cos(kH)|^2}\left[1-\frac{\sin(kH)}{kH}\right]+\frac{\Phi_p}{k_c^2}\left|\frac{\sin(kH)}{\omega\rho c\cos(kH)}\right|^2, \quad \epsilon kH \ll 1$$
$$\approx \frac{\pi A\Phi_p}{\epsilon k|\omega\rho c|^2}+\frac{\Phi_p}{k_c^2|\omega\rho c|^2}, \quad \epsilon kH \gg 1. \quad (18)$$

For the modified Corcos model, $A=2\alpha^3/\pi|k_c|$; therefore, the two limiting cases of Eq. (18) result in Eq. (19), with $\alpha=0.1$,

$$\frac{|w^2(H)|_{\text{edge}}}{|w^2(H)|_{\text{surface}}} \approx \frac{2|\sin(kH)|^2}{\pi A k_c^2 H}=\frac{|\sin(kH)|^2}{\alpha^3|k_c|H} \approx \left(\frac{1000|\sin(kH)|^2}{|k_c|H}\right), \quad \epsilon kH \ll 1 \text{and } kH>5, \quad (19a)$$

$$\frac{|w^2(H)|_{\text{edge}}}{|w^2(H)|_{\text{surface}}} \approx \frac{\epsilon k}{\pi A k_c^2}=\frac{\epsilon k}{2\alpha^3|k_c|} \approx \left(\frac{500\epsilon k}{|k_c|}\right), \quad \epsilon kH \gg 1. \quad (19b)$$

For a lightly damped string and $|\sin(kH)|^2>0.5$, the localized contribution will be at least 10 dB greater than the low wave number contribution if $k_cH<50$. For a strongly damped string with ϵ as large as 0.1, the localized contribution would be 10 dB greater than the low wave number con-

tribution only if $|k_c| < 5k$, which is not typical for underwater steel plates and typical flow speeds.

Reversing the flow direction changes the sign of k_c, but the low wave number response is independent of the direction of flow because it is even in k_c. The approximate localized force response is also even in k_c and independent of the flow direction. The flow direction does matter, however, when the full coherence function, which has a nonsymmetric wave number spectral density, is used and the structure is dissipative. This behavior can be illustrated by considering a string of length L, fixed at $x=0$, free at $x=L$, and driven by a deterministic convected load with wave number k_c. Let the excitation vary as $\exp[i(k_c x - \omega t)]$. The wave number k_c is positive (negative) if convection in the positive (negative) x direction. The solution at $x=L$ is given by

$$w(L) = -\frac{W_0}{\cos(kL)}\left[1 - \frac{k_c}{k}\sin(kL)\sin(k_c L) - \cos(kL)\cos(k_c L)\right] - \frac{iW_0}{\cos(kL)}\left[\frac{k_c}{k}\sin(kL)\cos(k_c L) - \cos(kL)\sin(k_c L)\right]. \quad (20)$$

In Eq. (20), k is the free wave number of the string, real if the string is lossless and complex if dissipative, and W_0 is the response of a similar string of infinite length. If k is real, the expression in each set of square brackets is real, and the square of the absolute value is even in k_c and independent of flow direction. For complex k, consider the special case of $\sin(k_c L)=0$ and $\cos(k_c L)=1$, which leads to

$$w(L) = -2W_0\frac{\sin(0.5kL)}{\cos(kL)}\left[\sin(0.5kL) - i\frac{k_c}{k}\cos(0.5kL)\right]. \quad (21)$$

In this case, the absolute value is dependent on the sign of k_c, and the result will depend on the direction of flow.

Given this dependence on flow direction, the application of the localized force model would be problematic were it not for the fact that, in many steel-water applications of interest, $|k_c/k| \gg 1$. For a frequency-plate thickness product as low as 100 Hz cm, the bending wave speed is approximately 9500 cm/s and increases with frequency and thickness. For a convection speed as high as 1000 cm/s, $|k_c/k| = 9.5$, and in most instances will be larger. Therefore, unless $|\sin(kL)|$ is very small, Eq. (20) can be approximated by

$$w(L) \approx -iW_0 \tan(kL)\frac{k_c}{k}\exp(ik_c L). \quad (22)$$

The absolute value of the right-hand side is even in k_c, and the localized force model can be used.

VI. CONCLUSION

This paper has attempted to demonstrate the relevance and application of TBL localized forces, which are related to the convection ridge portion of the TBL pressure wave number spectral density, to the calculation of structural response to TBL excitation. A model of two connected strings of finite length was used to show that the sum of the response due to equivalent localized forces and the response due to low wave number excitation give a good approximation to the response due to the complete excitation spectral density. Actual calculations were performed using mathematically equivalent spatial coherence functions. The numerical results for the example considered show that the localized force contribution *can* be significant relative to that of the low wave number excitation, but it is difficult to assess, *a priori*, whether it *will* form a significant portion of the overall response. In finite element analyses, however, the response transfer functions required to include localized force contributions will already have been calculated because they are needed to find the low wave number response. Therefore, localized force responses can be included in finite element analyses with small fractional increase in computational cost.

APPENDIX: NOTE ON THE STRING MODEL

The string is not intended to represent any realistic structure, but it is a mathematically well-posed model. The boundary conditions applied to the ends of the string sections are proper. Tension can be maintained in a string with a free end by attaching a filament of vanishing cross-sectional area (and vanishing mass per unit length) to that end and applying tension to the system. The impedance that the filament presents to the string end approaches zero because it is proportional to the square root of the filament's vanishing mass per unit length; therefore, the string end is free.

[1] Y. F. Hwang, "A discrete model of turbulence loading function for computation of flow-induced vibration and noise," *Proceedings of IMECE, Noise Control and Acoustics Division*, NCA-**25**, 389–395 (1998).
[2] Y. F. Hwang and S. A. Hambric, "Forcing function models for structures excited by low-speed flow," Noise-Con 2000, Newport Beach, CA, 2000.
[3] K. L. Chandiramani, "Response of underwater structures to convective component of flow noise," J. Acoust. Soc. Am. **73**, 835–839 (1983).
[4] Y. F. Hwang and G. Maidanik, "A wavenumber analysis of the coupling of a structural mode and flow turbulence," J. Sound Vib. **142**, 135–152 (1990).
[5] S. A. Hambric, Y. F. Hwang, and W. K. Bonness, "Vibrations of plates with clamped and free edges excited by low-speed turbulent boundary layer flow," J. Fluids Struct. **19**, 93–110 (2004).
[6] M. S. Howe, *Acoustics of Fluid-Structure Interactions* (Cambridge University Press, Cambridge, 1998).

Diffuse wave density and directionality in anisotropic solids

Andrew N. Norris[a)]
Mechanical and Aerospace Engineering, Rutgers University, Piscataway, New Jersey 08854

(Received 6 July 2007; revised 5 November 2007; accepted 27 December 2007)

Several general results are derived for diffuse waves in anisotropic solids, including concise expressions for the modal density per unit volume $d(\omega)$, and for the participation factor matrix $\mathbf{G}$. The latter is a second-order tensor which describes the orientational distribution of diffuse wave or reverberant energy, and reduces to the identity $\mathbf{I}$ under isotropy. Calculations of $\mathbf{G}$ for a variety of example materials show significant deviation from $\mathbf{I}$ even under moderate levels of anisotropy.
 [DOI: 10.1121/1.2836755]

PACS number(s): 43.40.Hb, 43.55.Cs [PEB] Pages: 1399–1408

I. INTRODUCTION

We consider how material anisotropy effects the directional partition of reverberant or diffuse wave energy. Diffuse waves in solids are the long time response when multiple scattering has equilibrated the energy distribution among modes. Preferential orientation of the root mean square particle velocity does not arise in isotropic materials but is a characteristic of anisotropy. Our objective is to describe this orientation effect and to quantify it in real materials. An ability to determine, directly or by inference, the orientational distribution of kinetic energy density in a solid allows one to essentially "hear" the texture of a crystal. We will demonstrate that the key quantity that needs to be measured is the autocorrelation function, or the Green's function evaluated at its source. By deriving an explicit formula for the autocorrelation, or the admittance matrix, we can completely describe the directional distribution of the diffuse wave energy.

We introduce two quantities for the description of reverberant energy in the presence of anisotropy: the participation tensor $\mathbf{G}$ and the modal spectral density per unit volume, $d(\omega)$. The two are in fact intimately related as we will see. Under steady state time harmonic conditions the total energy of a body is equally divided between potential and kinetic. The latter is $\frac{1}{2}\omega^2\int dV\rho|\bar{\mathbf{u}}|^2$ where $|\bar{\mathbf{u}}|$ is the root mean square particle displacement, and assuming a uniform spatial distribution, the total energy is $E=V\rho\omega^2|\bar{\mathbf{u}}|^2$. This may be inverted to express the mean square displacement. Let $\bar{u}_i=|\bar{\mathbf{u}}\cdot\mathbf{e}_i|$ where $\mathbf{e}_i$, $i=1,2,3$ is an orthonormal triad. Since $\bar{u}_1^2+\bar{u}_2^2+\bar{u}_3^2=|\bar{\mathbf{u}}|^2$ we may write

$$\bar{u}_i^2=\frac{E}{3V\rho\omega^2}\bar{G}_i,\quad \bar{G}_i=\mathbf{e}_i\cdot\mathbf{G}\cdot\mathbf{e}_i, \tag{1}$$

for i=1,2,3 (no sum) where $\mathbf{G}$ is a second-order symmetric tensor satisfying

$$\mathrm{tr}\,\mathbf{G}=3. \tag{2}$$

For isotropic materials $\mathbf{G}$ is simply the unit matrix or identity (second-order) tensor. Deviations from this can occur under three general situations: (i) If the field point is near a surface or boundary. This was considered in detail by Weaver,[1] who found expressions for the components of $\mathbf{G}$ at a free surface in terms of simple integrals, see also Egle.[2] (ii) By analogy, $\mathbf{G}$ will be influenced by local inhomogeneity in the material, for instance if the field point is close to a rigid inclusion, or a void. We will not discuss this further here. (iii) Material anisotropy can also influence $\mathbf{G}$. Here we consider the simplest case of a field point in a homogeneous material of infinite extent. It is expected that $\mathbf{G}$ displays the symmetries appropriate to the degree of anisotropy. Thus, it is characterized by a single parameter for materials with isotropic and cubic symmetries, and by two or three parameters for materials with lower symmetry.

The spectral density of modes D at frequency ω in a volume V is $D(\omega)=Vd(\omega)$. It can be estimated as $D=\partial N/\partial\omega\approx V\omega^2/c^3$ by noting the total number of modes scales as $N(\mathbf{k})\approx Vk^3$ where $k=\omega/c$ is typical wave number. A more precise counting yields, for isotropic bodies, the well-known result[3]

$$d(\omega)=\frac{\omega^2}{2\pi^2}\left(\frac{2}{c_t^3}+\frac{1}{c_l^3}\right), \tag{3}$$

where c_l and c_t are the longitudinal and transverse elastic wave speeds.

The objective is to derive analogous expressions of $d(\omega)$ and $\mathbf{G}$ for anisotropic elastic materials. This will be achieved by explicit calculation of the admittance tensor $\mathbf{A}$, defined in Sec. II, combined with a general relation between $d(\omega)$, $\mathbf{G}$, and $\mathbf{A}$. The spectral density and the participation tensor in the presence of material anisotropy do not appear to have received much attention. Some work on the related issue of admittance in bounded anisotropic thin plate systems has appeared.[4] Weaver[5] considered isotropic plates of finite thickness and infinite lateral extent. Tewary *et al.*[6] derived an expression for the admittance at the free surface of an anisotropic half space as a double integral. Here the focus is on infinite systems, and the modal density per unit volume in this limit. Finite structures, such as plates both thin and of finite thickness, will be considered in a separate paper.

a)Electronic mail: norris@rutgers.edu

TABLE I. The form of the participation tensor **G** for the different material symmetries. TI, tet, and trig are abbreviations for transverse isotropy, tetragonal, and trigonal symmetries, respectively. The **e** unit vectors are defined by the symmetry, while **a**, **b**, and **c** result from averaging. The positive numbers α, β, and γ are constrained as indicated in order to satisfy Eq. (2).

G	Material symmetry
$\mathbf{I}$	Isotropic, cubic
$\alpha\mathbf{e}\otimes\mathbf{e}+\beta(\mathbf{I}-\mathbf{e}\otimes\mathbf{e})$	TI, tet, trig $\alpha+2\beta=3$
$\alpha\mathbf{e}_1\otimes\mathbf{e}_1+\beta\mathbf{e}_2\otimes\mathbf{e}_2+\gamma\mathbf{e}_3\otimes\mathbf{e}_3$	Orthotropic $\alpha+\beta+\gamma=3$
$\alpha\mathbf{e}\otimes\mathbf{e}+\beta\mathbf{a}\otimes\mathbf{a}+\gamma\mathbf{b}\otimes\mathbf{b}$	Monoclinic $\alpha+\beta+\gamma=3$
$\alpha\mathbf{a}\otimes\mathbf{a}+\beta\mathbf{b}\otimes\mathbf{b}+\gamma\mathbf{c}\otimes\mathbf{c}$	Triclinic $\alpha+\beta+\gamma=3$

Our principal results are that the modal spectral density per unit volume and the participation tensor are given by

$$d(\omega)=\frac{\omega^2}{2\pi^2}\langle \mathrm{tr}\,\mathbf{Q}^{-3/2}\rangle, \tag{4a}$$

$$\mathbf{G}=3\frac{\langle \mathbf{Q}^{-3/2}\rangle}{\langle \mathrm{tr}\,\mathbf{Q}^{-3/2}\rangle}, \tag{4b}$$

where $\mathbf{Q}(\mathbf{n})$ is the acoustical or Christoffel tensor for plane waves propagating in the direction $\mathbf{n}$, and $\langle f\rangle$ is the orientation average of a function that depends on the direction,

$$\langle f\rangle\equiv\frac{1}{4\pi}\int_{4\pi}\mathrm{d}\Omega(\mathbf{n})f(\mathbf{n}). \tag{5}$$

In an isotropic solid Eq. (4a) reduces to Eq. (3) and **G** is simply the identity **I**. After deriving Eq. (4), the remainder of the paper will explore its implications, in particular the form of **G** is investigated, and the parameters in Table I deduced. It is interesting to note that the material constant that determines the density of states of diffuse waves, $\mathrm{tr}\langle\mathbf{Q}^{-3/2}\rangle$, also defines the Debye temperature Θ of a crystal. Thus (see Chap. 9 of Ref. 17),

$$\Theta=\frac{h}{k}\left(\frac{18\pi^2}{V_a\mathrm{tr}\langle\mathbf{Q}^{-3/2}\rangle}\right)^{1/3}, \tag{6}$$

where h is Planck's constant, k is Boltzmann's constant, and V_a is the volume per atom or lattice site. Fedorov[17] provides a detailed discussion of $\mathrm{tr}\langle\mathbf{Q}^{-3/2}\rangle$ in this context. The emphasis in this paper is on the more general tensor $\langle\mathbf{Q}^{-3/2}\rangle$ although connections with Fedorov's analysis will be mentioned later.

The outline of the paper is as follows. The admittance tensor **A** is defined and calculated in Sec. II, from which the main result (4) follows. Several alternative representations of the fundamental quantity $\mathbf{Q}^{-3/2}$ are developed in Sec. III. In particular it is shown that **G** for transverse isotropy can be evaluated as a single integral. Weak anisotropy is considered in Sec. IV and numerical examples are presented in Sec. V.

II. DERIVATION OF *d* AND G

A. Admittance tensor

The admittance **A** is a second-order tensor defined by the average power radiated by a time harmonic point force **F** according to

$$\Pi=\mathbf{F}\cdot\mathbf{A}\cdot\mathbf{F}. \tag{7}$$

Alternatively, **A** is equal to the power expended at the source point—which is the more conventional definition of admittance, as the the inverse of drive point impedance. The admittance is clearly related to the autocorrelation of the Green's function, and as such is a special case of the two-point cross correlation of the Green's function.[7] The important connection for the present purposes is the relation between the radiation from a point force and the diffuse wave density.[8,9] In the present notation this becomes

$$\mathbf{A}=\frac{\pi}{12\rho}d(\omega)\mathbf{G}. \tag{8}$$

A short derivation of Eq. (8) is given in Appendix A. The admittance of isotropic bodies is simply determined from Eq. (3) and $\mathbf{G}=\mathbf{I}$. Our objective here is to calculate **A** for anisotropic solids, and then to use the result to determine $d(\omega)$ and **G**.

The central result for **A** is the following: The second-order symmetric admittance tensor of Eq. (7) that determines the total power radiated to infinity from the point source averaged over a period is

$$\mathbf{A}=\frac{\omega^2}{8\pi\rho}\langle\mathbf{Q}^{-3/2}\rangle, \tag{9}$$

where $\mathbf{Q}(\mathbf{n})$ is the acoustical tensor,

$$Q_{ik}(\mathbf{n})=c_{ijkl}n_jn_l \text{ with } c_{ijkl}=\frac{1}{\rho}C_{klij}. \tag{10}$$

The elastic moduli (stiffness) C_{ijkl} have the symmetries $C_{ijkl}=C_{klij}$ and $C_{ijkl}=C_{jikl}$, and thus have at most 21 independent elements. Note that **A** has dimensions of admittance (inverse impedance). We next derive Eq. (9) by explicitly calculating the admittance for a time harmonic point force.

B. Radiation from a point force

The displacement resulting from a point force $\mathbf{F}\cos\omega t$ at the origin is $\mathbf{u}(\mathbf{x},t)=\mathrm{Re}\,\tilde{\mathbf{u}}(\mathbf{x},\omega)e^{-i\omega t}$ where $\tilde{\mathbf{u}}$ satisfies

$$C_{ijkl}\tilde{u}_{k,jl}+\rho\omega^2\tilde{u}_i=-F_i\delta(\mathbf{x}),\quad -\infty\leqslant x_1,x_2,x_3\leqslant\infty.$$

Here ρ is the mass density and $\delta(\mathbf{x})$ is the three-dimensional Dirac delta function. The equation of motion may be written

$$\mathbf{Q}(\nabla)\tilde{\mathbf{u}}+\omega^2\tilde{\mathbf{u}}=-\frac{1}{\rho}\delta(\mathbf{x})\mathbf{F}, \tag{11}$$

and the problem definition is completed by the requirement that the energy radiates away from the point source.

The solution to Eq. (11) in a solid of infinite extent is well known. For our purpose we will find the following representation from Norris [Ref. 10, Eq. (3.22)] useful for determining the admittance:

$$\tilde{\mathbf{u}} = \frac{1}{8\pi^2\rho|\mathbf{x}|}\oint \mathrm{d}\theta(\mathbf{n})\sum_{j=1}^{3}\frac{\mathbf{q}_j\otimes\mathbf{q}_j}{\lambda_j}\mathbf{F} + \frac{1}{16\pi^2\rho}\int_{4\pi}\mathrm{d}\Omega(\mathbf{n})\sum_{j=1}^{3}\frac{ik_j}{\lambda_j}\mathbf{q}_j\otimes\mathbf{q}_j\mathbf{F}e^{ik_j\mathbf{n}\cdot\mathbf{x}}. \quad (12)$$

Here $\lambda_1,\lambda_2,\lambda_3$ are the eigenvalues and $\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3$ the eigenvectors of $\mathbf{Q}(\mathbf{n})$, which then has the spectral decomposition

$$\mathbf{Q}(\mathbf{n}) = \lambda_1\mathbf{q}_1\otimes\mathbf{q}_1 + \lambda_2\mathbf{q}_2\otimes\mathbf{q}_2 + \lambda_3\mathbf{q}_3\otimes\mathbf{q}_3. \quad (13)$$

Also, $k_j=\omega/\lambda_j^{1/2}$ are the wave numbers of the three distinct branches of the slowness surface defined by the eigenvectors. The first integral in Eq. (12) is around the unit circle formed by the intersection of the plane $\mathbf{n}\cdot\mathbf{x}=0$ with the unit $\mathbf{n}$ sphere. This is just the static Green's function of elasticity.[10] The important dynamic quantity is the second integral which is evaluated over the sphere $\{|\mathbf{n}|=1\}$. In order to make this more apparent, we rewrite Eq. (12) as

$$\tilde{\mathbf{u}} = \tilde{\mathbf{u}}|_{\omega=0} + \frac{i\omega}{4\pi\rho}\sum_{j=1}^{3}\left\langle e^{ik_j\mathbf{n}\cdot\mathbf{x}}\frac{\mathbf{q}_j\otimes\mathbf{q}_j}{\lambda_j^{3/2}}\right\rangle\mathbf{F}, \quad (14)$$

and note for future reference that the first term on the right-hand side is real valued.

The average power radiated per period is equal to the power expended by the force

$$\Pi = \lim_{\mathbf{x}\to 0}\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\mathrm{d}t\,\cos\omega t\mathbf{F}\cdot\mathbf{v}(\mathbf{0},t), \quad (15)$$

where $\mathbf{v}(\mathbf{x},t)=\mathrm{Re}(-i\omega\tilde{\mathbf{u}}(\mathbf{x},\omega)e^{-i\omega t})$ is the particle velocity. Thus,

$$\Pi = \frac{\omega^2}{8\pi\rho}\sum_{j=1}^{3}\left\langle\frac{1}{\lambda_j^{3/2}}(\mathbf{q}_j\cdot\mathbf{F})^2\right\rangle. \quad (16)$$

The spectral decomposition (13) implies that

$$\lambda_1^{-3/2}\mathbf{q}_1\otimes\mathbf{q}_1 + \lambda_2^{-3/2}\mathbf{q}_2\otimes\mathbf{q}_2 + \lambda_3^{-3/2}\mathbf{q}_3\otimes\mathbf{q}_3 = \mathbf{Q}^{-3/2},$$

which together with Eq. (7) proves the main result (9).

The scalar $d(\omega)$ and the tensor $\mathbf{G}$ are defined such that their product is $12\rho/\pi$ times the admittance $\mathbf{A}$, see Eqs. (3), (4), (8), and (9). This defines d and $\mathbf{G}$ to within a constant, which is determined uniquely by the constraint $\mathrm{tr}\,\mathbf{G}=3$. We therefore obtain the general results of Eq. (4). As discussed, d is the generalization of the classical density of states per unit volume, Eq. (3) for isotropic solids, and the participation factor tensor $\mathbf{G}$ describes the directional distribution of the energy at a point. While it is convenient to consider them separately, d and $\mathbf{G}$ are both defined by the averaged tensor $\langle\mathbf{Q}^{-3/2}\rangle$, which will be the focus of the remainder of the paper.

Before considering the properties of d and $\mathbf{G}$ we note that the isotropic modal density of states follows immediately from Eq. (4a). Starting with the acoustical tensor for an isotropic solid,

$$\mathbf{Q}(\mathbf{n}) = c_l^2\mathbf{n}\otimes\mathbf{n} + c_t^2(\mathbf{I}-\mathbf{n}\otimes\mathbf{n}) \quad \text{isotropy}, \quad (17)$$

we have $\mathbf{Q}^{-3/2}=c_l^{-3}\mathbf{n}\otimes\mathbf{n}+c_t^{-3}(\mathbf{I}-\mathbf{n}\otimes\mathbf{n})$. Then using the fact that $\langle\mathbf{n}\otimes\mathbf{n}\rangle=\frac{1}{3}\mathbf{I}$ it follows that

$$\langle\mathbf{Q}^{-3/2}\rangle = \tfrac{1}{3}(c_l^{-3}+2c_t^{-3})\mathbf{I}. \quad (18)$$

Hence, the density of states per unit volume is $d=(\omega^2/2\pi^2)\times(c_l^{-3}+2c_t^{-3})^{-1}$, in agreement with the well known identity (3), and $\mathbf{G}=\mathbf{I}$, as expected.

III. $\mathbf{Q}^{-3/2}$ AND RELATED QUANTITIES

The key quantity is the tensor $\mathbf{Q}^{-3/2}$ and its directional average. In practice, this may be evaluated numerically without difficulty. It is however useful to examine semiexplicit forms for the tensor, both for general anisotropy and for specific symmetries, particularly the case of transverse isotropy. We begin with two alternative and general formulations based on the spectral properties and the invariants of the acoustical tensor.

A. General representations for arbitrary anisotropy

1. A method based on invariants

Functions of a positive definite tensor can be simplified using the Cayley–Hamilton formula for the tensor, which for $\mathbf{Q}$ is

$$\mathbf{Q}^3 - I_1\mathbf{Q}^2 + I_2\mathbf{Q} - I_3\mathbf{I} = 0. \quad (19)$$

The principal positive invariants of $\mathbf{Q}$ are

$$I_1 = \mathrm{tr}\,\mathbf{Q}, \quad I_2 = \tfrac{1}{2}(\mathrm{tr}\,\mathbf{Q})^2 - \tfrac{1}{2}\mathrm{tr}\,\mathbf{Q}^2, \quad I_3 = \det\mathbf{Q}. \quad (20)$$

Based on these fundamental properties, it can be shown that

$$\mathbf{Q}^{-3/2} = [(I_1I_3 + i_1i_3I_2 + i_2I_3)(\mathbf{Q}^2 - I_1\mathbf{Q} + I_2\mathbf{I}) + i_1i_3I_3(\mathbf{Q}-I_1\mathbf{I}) - I_3^2\mathbf{I}]/[(i_1i_2-i_3)I_3^2], \quad (21)$$

where i_1, i_2, and i_3 are the positive invariants of $\mathbf{Q}^{1/2}$ which can be expressed as functions of the invariants I_1,I_2, and I_3, see the following. Details of the derivation of Eq. (21) are given in Appendix B.

The appealing feature of Eq. (21) for $\mathbf{Q}^{-3/2}(\mathbf{n})$ is that it only involves powers of $\mathbf{Q}$, its three invariants, and the additional invariants i_1, i_2, and i_3. These are related to I_1, I_2, and I_3 by[11,12]

$$i_1^2 - 2i_2 = I_1, \quad i_2^2 - 2i_1i_3 = I_2, \quad i_3^2 = I_3. \quad (22)$$

The last implies $i_3=I_3^{1/2}$, while expressions for i_1 and i_2 are given by Hoger and Carlson[11] and by Norris.[12] For instance,[12]

$$i_1 = \sqrt{I_1 - \beta + 2\sqrt{I_3/\beta}} + \sqrt{\beta}, \quad (23a)$$

$$i_2 = \sqrt{I_2 - I_3/\beta + 2\sqrt{I_3\beta}} + \sqrt{I_3/\beta}, \quad (23b)$$

$$i_3 = \sqrt{I_3}, \quad (23c)$$

where β is any eigenvalue of $\mathbf{Q}$, e.g.,

$$\beta = \tfrac{1}{3}(I_1 + [(\xi + \sqrt{\xi^2 - (I_1^2-3I_2)^3}\,]^{1/3} + [(\xi - \sqrt{\xi^2 - (I_1^2-3I_2)^3}\,]^{1/3}), \quad (24a)$$

$$\xi = \tfrac{1}{2}(2I_1^3 - 9I_1I_2 + 27I_3). \quad (24b)$$

Note that[13] $i_1 i_2 - i_3 = \det(i_1 \mathbf{I} - \mathbf{Q}^{1/2}) > 0$.

Taking the trace of Eq. (21) gives

$$\operatorname{tr} \mathbf{Q}^{-3/2} = \frac{(I_1 + i_2) I_2 I_3 + (I_2^2 - 2 I_1 I_3) i_1 i_3 - 3 I_3^2}{(i_1 i_2 - i_3) I_3^2}. \tag{25}$$

This quantity, when averaged over all orientations, gives the density of states function $d(\omega)$ of Eq. (4a). Hence d can be calculated from the invariants $\mathbf{Q}$ and the derived invariants i_1, i_2, i_3.

2. A spectral representation

The second form for $\mathbf{Q}^{-3/2}$ is based on the spectral decomposition (17). The latter can be expressed in a form that does not explicitly involve the eigenvectors,

$$\mathbf{Q}^{-3/2} = \lambda_1^{-3/2} \mathbf{N}(\lambda_1) + \lambda_2^{-3/2} \mathbf{N}(\lambda_2) + \lambda_3^{-3/2} \mathbf{N}(\lambda_3). \tag{26}$$

The second-order tensors $\mathbf{N}(\lambda_j)$, which are alternative expressions for the dyadics formed by the eigenvectors, $\mathbf{q}_j \otimes \mathbf{q}_j$, can be expressed in terms of $\mathbf{Q}$ using Sylvester's formula

$$\mathbf{N}(\lambda, \mathbf{n}) = \frac{\lambda \mathbf{Q}^2 + (\lambda - I_1)\lambda \mathbf{Q} + I_3 \mathbf{I}}{\lambda^3 + (\lambda - I_1)\lambda^2 + I_3}. \tag{27}$$

The identity (26) is derived in Appendix B.

Calculation of Eq. (26) requires knowledge of the three eigenvalues, which are zeros of the characteristic polynomial defined by Eq. (19),

$$p(\lambda) = \lambda^3 - I_1 \lambda^2 + I_2 \lambda - I_3. \tag{28}$$

The eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ can be expressed in terms of the invariants as

$$\left\{\beta, \tfrac{1}{2}(I_1 - \beta) \pm \tfrac{1}{2}\sqrt{(I_1 - \beta)^2 - 4 I_3/\beta}\right\}, \tag{29}$$

where β is defined in Eq. (24a). Every[14] derived alternate closed-form expressions based on the trigonometric solution of the characteristic cubic. The alternative version of Eq. (25) is

$$\operatorname{tr} \mathbf{Q}^{-3/2} = \lambda_1^{-3/2} + \lambda_2^{-3/2} + \lambda_3^{-3/2}, \tag{30}$$

which is the starting point for Fedorov's calculation[17] of the trace.

B. Transverse isotropy

Transverse isotropy or hexagonal symmetry is an important class of anisotropy. It occurs in many practical circumstances, whether from layering in the earth to laminated composite materials, or from underlying crystal structure. It is the highest symmetry for which the participation factor tensor is not the identity, since $\mathbf{G} = \mathbf{I}$ under isotropy and cubic material symmetry. We now demonstrate that the evaluation of d and $\mathbf{G}$ may be reduced to the evaluation of two single integrals, one for $\langle \operatorname{tr} \mathbf{Q}^{-3/2} \rangle$ and one for the parameter α that defines $\mathbf{G}$, see Table I.

Transversely isotropic solids have five independent moduli: $c_{11} = c_{22}$, c_{33}, c_{12}, $c_{13} = c_{23}$, $c_{44} = c_{55}$, $c_{66} = \frac{1}{2}(c_{11} - c_{12})$. Let $\mathbf{e}$ be the axis of symmetry. The SH slowness decouples to give

$$\mathbf{Q} = \lambda_3(\mathbf{n} \cdot \mathbf{e}) \mathbf{q}_3 \otimes \mathbf{q}_3 + \mathbf{Q}_\perp, \tag{31}$$

where (Ref. 15, p. 95)

$$\lambda_3(\mathbf{n} \cdot \mathbf{e}) = c_{66} + (c_{44} - c_{66})(\mathbf{n} \cdot \mathbf{e})^2, \tag{32}$$

and $\mathbf{q}_3 = \mathbf{e} \wedge \mathbf{n} / |\mathbf{e} \wedge \mathbf{n}|$. The two-dimensional symmetric tensor $\mathbf{Q}_\perp$ is[15]

$$\begin{aligned} \mathbf{Q}_\perp = {} & [c_{44} + (c_{33} - c_{44})(\mathbf{n} \cdot \mathbf{e})^2] \mathbf{e} \otimes \mathbf{e} \\ & + [c_{11} + (c_{44} - c_{11})(\mathbf{n} \cdot \mathbf{e})^2] \mathbf{d} \otimes \mathbf{d} \\ & + (c_{13} + c_{44}) \mathbf{n} \cdot \mathbf{e} \sqrt{1 - (\mathbf{n} \cdot \mathbf{e})^2} [\mathbf{d} \otimes \mathbf{e} + \mathbf{e} \otimes \mathbf{d}], \end{aligned}$$

where $\mathbf{d} = \mathbf{e} \wedge \mathbf{q}_3$. Replacing $\mathbf{n} \cdot \mathbf{e}$ by the integration parameter ξ, it follows that

$$\langle \lambda_3^{-3/2} \mathbf{q}_3 \otimes \mathbf{q}_3 \rangle = \frac{1}{2} \int_0^1 d\xi \lambda_3^{-3/2}(\xi) \mathbf{I}_\perp, \tag{33}$$

where $\mathbf{I}_\perp$ projects onto the plane perpendicular to $\mathbf{e}$,

$$\mathbf{I}_\perp = \mathbf{I} - \mathbf{e} \otimes \mathbf{e}. \tag{34}$$

It remains to consider the orientational average of $\mathbf{Q}_\perp^{-3/2}$.

The tensor $\mathbf{Q}_\perp$ satisfies a quadratic Cayley–Hamilton equation

$$\mathbf{Q}_\perp^2 - J_1 \mathbf{Q}_\perp + J_2 \mathbf{I}_\perp = 0, \tag{35}$$

with $J_1 = \operatorname{tr} \mathbf{Q}_\perp = \lambda_1 + \lambda_2$ and $J_2 = \det \mathbf{Q}_\perp = \lambda_1 \lambda_2$. Similarly, the Cayley–Hamilton equation for the square root is

$$(\mathbf{Q}_\perp^{1/2})^2 - j_1 \mathbf{Q}_\perp^{1/2} + j_2 \mathbf{I}_\perp = 0, \tag{36}$$

where $j_1 = \operatorname{tr} \mathbf{Q}_\perp^{1/2}$ and $j_2 = \det \mathbf{Q}_\perp^{1/2}$ satisfy $J_1 = j_1^2 - 2 j_2$, $J_2 = j_2^2$, and are therefore related to J_1 and J_2 by $j_1 = \sqrt{J_1 + 2\sqrt{J_2}}$, $j_2 = \sqrt{J_2}$. Using Eqs. (35) and (36), respectively, leads to the identities

$$\mathbf{Q}_\perp^{-2} = J_2^{-2}[(J_1^2 - J_2)\mathbf{I}_\perp - J_1 \mathbf{Q}_\perp], \tag{37a}$$

$$\mathbf{Q}_\perp^{1/2} = j_1^{-1}(\mathbf{Q}_\perp + j_2 \mathbf{I}_\perp). \tag{37b}$$

Multiplication of these and further use of Eq. (35) leads to

$$\mathbf{Q}_\perp^{-3/2} = \frac{1}{j_1 j_2 J_2}[(J_1 + j_2)(J_1 \mathbf{I}_\perp - \mathbf{Q}_\perp) - J_2 \mathbf{I}_\perp]. \tag{38}$$

Again using $\xi = \mathbf{n} \cdot \mathbf{e}$, we have

$$\langle \operatorname{tr} \mathbf{Q}^{-3/2} \rangle = \int_0^1 d\xi [J_2^{-3/2}(J_1 - \sqrt{J_2}) \sqrt{J_1 + 2\sqrt{J_2}} + \lambda_3^{-3/2}(\xi)],$$

and from Table I,

$$\alpha = \frac{3}{\langle \operatorname{tr} \mathbf{Q}^{-3/2} \rangle} \int_0^1 d\xi \frac{(J_1 + \sqrt{J_2})(J_1 - \mathbf{e} \cdot \mathbf{Q}_\perp \cdot \mathbf{e}) - J_2}{J_2^{3/2} \sqrt{J_1 + 2\sqrt{J_2}}}.$$

The modal density parameter $\langle \operatorname{tr} \mathbf{Q}^{-3/2} \rangle$ and the scalar α that defines the participation tensor can therefore be expressed as single integrals, which follow from the above-presented results and Eqs. (31)–(33), as

$$\langle \mathrm{tr}\,\mathbf{Q}^{-3/2}\rangle = \int_0^1 \mathrm{d}\xi \left[\frac{(a+b\xi^2-\sqrt{d+e\xi^2+f\xi^4})\sqrt{a+b\xi^2+2\sqrt{d+e\xi^2+f\xi^4}}}{(d+e\xi^2+f\xi^4)^{3/2}} + \frac{1}{[c_{66}+(c_{44}-c_{66})\xi^2]^{3/2}} \right], \tag{39a}$$

$$\alpha = \frac{3}{\langle \mathrm{tr}\,\mathbf{Q}^{-3/2}\rangle} \int_0^1 \mathrm{d}\xi \left[\frac{(c_{11}+c\xi^2)(a+b\xi^2+\sqrt{d+e\xi^2+f\xi^4})-(d+e\xi^2+f\xi^4)}{(d+e\xi^2+f\xi^4)^{3/2}\sqrt{a+b\xi^2+2\sqrt{d+e\xi^2+f\xi^4}}} \right], \tag{39b}$$

where

$$a = c_{11}+c_{44}, \quad b = c_{33}-c_{11},$$

$$c = c_{44}-c_{11}, \quad d = c_{11}c_{44},$$

$$e = c_{11}c_{33}-c_{13}^2-2c_{44}(c_{11}+c_{13}),$$

$$f = -c_{11}c_{33}+c_{13}^2+c_{44}(c_{11}+c_{33}+2c_{13}).$$

IV. WEAK ANISOTROPY

Although the general expressions for the modal density d and the participation tensor $\mathbf{G}$ are not difficult to compute, it is often the case that the medium is to a first approximation isotropic, and appropriate approximations can be made. The state of small or *weak anisotropy* is defined relative to a background isotropic medium, and it is important to select the latter properly. In this section we calculate d and $\mathbf{G}$ in the presence of weak anisotropy. Fedorov[17] provides a detailed analysis of the expansion of $\mathrm{tr}\langle\mathbf{Q}^{-3/2}\rangle$ to arbitrary orders in the perturbation parameter. Our emphasis is more on obtaining estimates of the tensor $\langle\mathbf{Q}^{-3/2}\rangle$, which is not discussed explicitly by Fedorov. We begin with a description of the comparison isotropic moduli and then proceed to calculate the first two terms in a perturbation series for d and $\mathbf{G}$.

A. Background isotropic moduli

Regardless of the level of the anisotropy it is always possible to define a unique set of isotropic moduli which minimize the Euclidean distance between the exact set of moduli and the equivalent isotropic moduli.[16] This procedure is equivalent to requiring that the mean square Euclidean difference in the slowness surfaces is minimal.[16,17] Thus, let the background isotropic moduli be

$$c^{(0)}_{ijkl} = c_l^2\delta_{ij}\delta_{kl} + c_t^2(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}-2\delta_{ij}\delta_{kl}), \tag{40}$$

where c_l and c_t are the effective longitudinal and transverse wave speeds. These are defined by simultaneously minimizing the quantity $\langle|\mathbf{Q}-\mathbf{Q}_0|^2\rangle$ with respect to both c_l and c_t, where $\mathbf{Q}_0(\mathbf{n})$ is defined by the moduli $c^{(0)}_{ijkl}$. The unique solution is

$$c_l^2 = \tfrac{1}{3}\mathrm{tr}\,\mathbf{C}_l, \quad c_t^2 = \tfrac{1}{3}\mathrm{tr}\,\mathbf{C}_t, \tag{41}$$

where the second-order tensors of reduced moduli are

$$C_{l,ij} = \tfrac{2}{5}c_{ikjk} + \tfrac{1}{5}c_{ijkk}, \quad C_{t,ij} = \tfrac{3}{10}c_{ikjk} - \tfrac{1}{10}c_{ijkk}. \tag{42}$$

The background Lamé moduli λ and μ are obtained using $c_l^2=(\lambda+2\mu)/\rho$ and $c_t^2=\mu/\rho$. The elements of $\mathbf{C}_l$ and $\mathbf{C}_t$ follow from

$$c_{ijkk} = \begin{pmatrix} c_{11}+c_{12}+c_{13} & c_{16}+c_{26}+c_{36} & c_{15}+c_{25}+c_{35} \\ c_{16}+c_{26}+c_{36} & c_{12}+c_{22}+c_{23} & c_{14}+c_{24}+c_{34} \\ c_{15}+c_{25}+c_{35} & c_{14}+c_{24}+c_{34} & c_{13}+c_{23}+c_{33} \end{pmatrix},$$

$$c_{ikjk} = \begin{pmatrix} c_{11}+c_{55}+c_{66} & c_{16}+c_{26}+c_{45} & c_{15}+c_{46}+c_{35} \\ c_{16}+c_{26}+c_{45} & c_{22}+c_{44}+c_{66} & c_{24}+c_{34}+c_{56} \\ c_{15}+c_{46}+c_{35} & c_{24}+c_{34}+c_{56} & c_{33}+c_{44}+c_{55} \end{pmatrix}.$$

B. Perturbation analysis

Let

$$c_{ijkl} = c^{(0)}_{ijkl} + \varepsilon c^{(1)}_{ijkl}, \tag{43}$$

where the nondimensional parameter ε is introduced only to simplify the perturbation analysis. In practice ε is set to unity on the assumption that the additional moduli $c_{ijkl}-c^{(0)}_{ijkl}$ are small in comparison with the isotropic background.

We seek expansions in powers of the small parameter ε. The key quantity $\mathbf{Q}^{-3/2}$ will be determined as the product of $\mathbf{Q}^{-2}$ and $\mathbf{Q}^{1/2}$. Based on Eq. (43), the acoustical tensor is

$$\mathbf{Q} = \mathbf{Q}_0 + \varepsilon\mathbf{Q}_1, \tag{44}$$

and simple perturbation gives

$$\mathbf{Q}^{-2} = \mathbf{Q}_0^{-2} - \varepsilon(\mathbf{Q}_0^{-2}\mathbf{Q}_1\mathbf{Q}_0^{-1} + \mathbf{Q}_0^{-1}\mathbf{Q}_1\mathbf{Q}_0^{-2}) + O(\varepsilon^2).$$

Let

$$\mathbf{Q}^{1/2} = \mathbf{Q}_0^{1/2} + \varepsilon\mathbf{S}_1 + O(\varepsilon^2),$$

then $\mathbf{S}_1$ satisfies

$$\mathbf{Q}_0^{1/2}\mathbf{S}_1 + \mathbf{S}_1\mathbf{Q}_0^{1/2} = \mathbf{Q}_1. \tag{45}$$

In order to calculate $\mathbf{Q}^{-2}$ and also the square root of $\mathbf{Q}$, we now use the fact that the leading order moduli $c^{(0)}_{ijkl}$ are isotropic. The explicit form of $\mathbf{Q}_0^{1/2}$ follows from Eq. (17) and the identity

$$\mathbf{Q}_0^m = c_l^{2m}\mathbf{n}\otimes\mathbf{n} + c_t^{2m}\mathbf{P}, \tag{46}$$

where m is any real number and $\mathbf{P}=\mathbf{I}-\mathbf{n}\otimes\mathbf{n}$. Equation (45) can be solved by observing that $\mathbf{Q}_1$ may be partitioned $\mathbf{Q}_1=(\mathbf{Q}_1^{(1)}+\mathbf{Q}_1^{(2)}+\mathbf{Q}_1^{(3)}$ where $\mathbf{Q}_1^{(1)}=\mathbf{n}\cdot\mathbf{Q}_1\cdot\mathbf{n})\mathbf{n}\otimes\mathbf{n}$, $\mathbf{Q}_1^{(2)}=\mathbf{P}\mathbf{Q}_1\mathbf{P}$, and $\mathbf{Q}_1^{(3)}=\mathbf{P}\mathbf{Q}_1\cdot\mathbf{n}\otimes\mathbf{n}+\mathbf{n}\otimes\mathbf{P}\mathbf{Q}_1\cdot\mathbf{n}$. Assuming a solution of

the form $\mathbf{S}_1 = p_1\mathbf{Q}_1^{(1)} + p_2\mathbf{Q}_1^{(3)} + p_3\mathbf{Q}_1^{(3)}$, the coefficients can be determined easily from Eq. (45), i.e.,

$$\mathbf{S}_1 = \frac{1}{2c_l}\mathbf{Q}_1^{(1)} + \frac{1}{2c_t}\mathbf{Q}_1^{(2)} + \frac{1}{c_l + c_t}\mathbf{Q}_1^{(3)}. \tag{47}$$

Combining the asymptotic expansions for $\mathbf{Q}^{-2}$ and $\mathbf{Q}^{1/2}$ gives

$$\mathbf{Q}^{-3/2} = \mathbf{Q}_0^{-3/2} + \varepsilon\mathbf{V}_1 + O(\varepsilon^2), \tag{48}$$

where

$$\begin{aligned}\mathbf{V}_1 &= \mathbf{Q}_0^{-2}\mathbf{S}_1 - \mathbf{Q}_0^{-2}\mathbf{Q}_1\mathbf{Q}_0^{-1/2} - \mathbf{Q}_0^{-1}\mathbf{Q}_1\mathbf{Q}_0^{-3/2} \\ &= -\frac{3}{2c_t^5}\mathbf{Q}_1 - \left[\frac{(c_l^2 + c_t^2 + c_l c_t)}{c_l^3 c_t^3 (c_l + c_t)} - \frac{3}{2c_t^5}\right] \\ &\quad \times [\mathbf{Q}_1 \cdot \mathbf{n} \otimes \mathbf{n} + \mathbf{n} \otimes \mathbf{Q}_1 \cdot \mathbf{n}] \\ &\quad + \left[2\frac{(c_l^2 + c_t^2 + c_l c_t)}{c_l^3 c_t^3 (c_l + c_t)} - \frac{3}{2c_l^5} - \frac{3}{2c_t^5}\right] \\ &\quad \times (\mathbf{n} \cdot \mathbf{Q}_1 \cdot \mathbf{n})\mathbf{n} \otimes \mathbf{n}.\end{aligned}$$

The orientational average $\langle \mathbf{Q}^{-3/2} \rangle$ can then be effected using the identities

$$\langle n_i n_j n_k n_l \rangle = \tfrac{1}{15}(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \equiv K_{ijkl},$$

$$\begin{aligned}\langle n_i n_j n_k n_l n_p n_q \rangle = \tfrac{1}{7}(&\delta_{ij}K_{klpq} + \delta_{ik}K_{jlpq} + \delta_{il}K_{kjpq} + \delta_{ip}K_{kljq} \\ &+ \delta_{iq}K_{klpj}).\end{aligned}$$

The resulting expressions for $\langle \mathbf{Q}^{-3/2} \rangle$ is

$$\begin{aligned}\langle \mathbf{Q}^{-3/2} \rangle_{ij} = &\frac{1}{3}\left(\frac{2}{c_t^3} + \frac{1}{c_l^3}\right)\delta_{ij} + \varepsilon\left\{-\frac{1}{2c_t^5}c_{ikjk}^{(1)} - \frac{2}{15}\right. \\ &\times \left[\frac{(c_l^2 + c_t^2 + c_l c_t)}{c_l^3 c_t^3 (c_l + c_t)} - \frac{3}{2c_t^5}\right](c_{ijkk}^{(1)} + 2c_{ikjk}^{(1)}) \\ &+ \frac{1}{105}\left[2\frac{(c_l^2 + c_t^2 + c_l c_t)}{c_l^3 c_t^3 (c_l + c_t)} - \frac{3}{2c_l^5} - \frac{3}{2c_t^5}\right] \\ &\left.\times [\delta_{ij}(c_{kkll}^{(1)} + 2c_{klkl}^{(1)}) + 4(c_{ijkk}^{(1)} + 2c_{ikjk}^{(1)})]\right\} \\ &+ O(\varepsilon^2).\end{aligned}$$

We note that both $c_{iijj}^{(1)}$ and $c_{ijij}^{(1)}$ vanish by virtue of the choice of the background isotropic moduli. This implies that the trace of $\langle \mathbf{Q}^{-3/2} \rangle$ differs from the isotropic approximant only at the second order of anisotropic perturbation,

$$\mathrm{tr}\langle \mathbf{Q}^{-3/2} \rangle = \frac{2}{c_t^3} + \frac{1}{c_l^3} + O(\varepsilon^2). \tag{49}$$

This is in agreement with Fedorov,[17] who also provides explicit forms for the higher order terms; for instance, the expansion for cubic crystals up to fourth order in the perturbation is given by Eqs. (50.12)–(50.14) of Ref. 17. The leading order approximation of Eq. (49) when combined with the identity (4b), gives

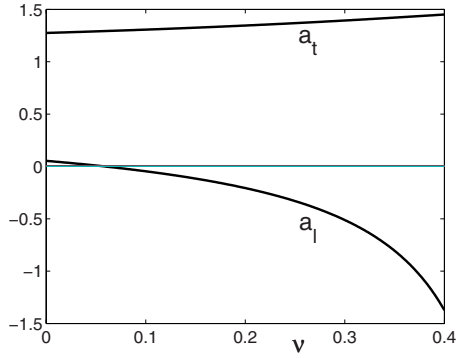


FIG. 1. (Color online) The nondimensional parameters a_l and a_t as a function of the Poisson's ratio ν.

$$\begin{aligned}G_{ij} = \delta_{ij} - \varepsilon\left(\frac{2}{c_t^3} + \frac{1}{c_l^3}\right)^{-1}&\left\{\frac{3}{2c_t^5}c_{ikjk}^{(1)} + \frac{3}{35}\left[2\frac{(c_l^2 + c_t^2 + c_l c_t)}{c_l^3 c_t^3 (c_l + c_t)}\right.\right. \\ &\left.\left. + \frac{2}{c_l^5} - \frac{5}{c_t^5}\right] \times (c_{ijkk}^{(1)} + 2c_{ikjk}^{(1)})\right\} + O(\varepsilon^2).\end{aligned}$$

Ignoring terms of order ε^2 and then setting $\varepsilon \to 1$ yields the leading order approximation to the participation tensor as

$$\mathbf{G} \approx \mathbf{I} + a_l(\mathbf{I} - c_l^{-2}\mathbf{C}_l) + a_t(\mathbf{I} - c_t^{-2}\mathbf{C}_t), \tag{50}$$

where the nondimensional coefficients are

$$a_l = \frac{6}{7(2 + \kappa^{-3})}\left(\frac{1}{\kappa^3} + \frac{1}{\kappa} - \frac{1}{\kappa + 1} + 1 - \frac{3}{4}\kappa^2\right), \tag{51a}$$

$$a_t = \frac{3}{2 + \kappa^{-3}}, \tag{51b}$$

and

$$\kappa \equiv \frac{c_l}{c_t}. \tag{52}$$

Figure 1 shows a_l and a_t as functions of the Poisson's ratio ν, using $\kappa^2 = 2(1-\nu)/(1-2\nu)$. Note that $1.27\ldots < a_t < 3/2$ for $0 < \nu < 1/2$ while $a_l \approx -\frac{9}{28}(1-2\nu)^{-1}$ as $\nu \to 1/2$.

C. Transversely isotropic materials

As an example of the general perturbation approach, we consider the particular case of TI materials. We take the axis of symmetry ($\mathbf{e}$ in Sec. III) in the 3 direction so that

$$c_{ijkk} = \begin{pmatrix} c_{11} + c_{12} + c_{13} & 0 & 0 \\ 0 & c_{11} + c_{12} + c_{13} & 0 \\ 0 & 0 & c_{33} + 2c_{13} \end{pmatrix},$$

$$c_{ikjk} = \begin{pmatrix} c_{11} + c_{44} + c_{66} & 0 & 0 \\ 0 & c_{11} + c_{44} + c_{66} & 0 \\ 0 & 0 & c_{33} + 2c_{44} \end{pmatrix},$$

where $c_{66} = \frac{1}{2}(c_{11} - c_{12})$. The wave speeds in the background isotropic medium are then

$$c_l^2 = \tfrac{1}{15}(8c_{11} + 3c_{33} + 4c_{13} + 8c_{44}), \tag{53a}$$

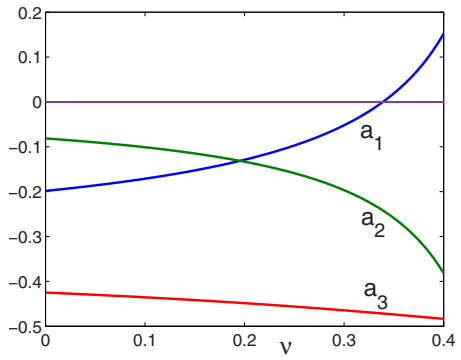


FIG. 2. (Color online) The nondimensional parameters a_1, a_2, and a_3 for weak transverse isotropy as a function of the Poisson's ratio ν of the background medium.

$$c_t^2 = \tfrac{1}{30}(2c_{11} + 2c_{33} - 4c_{13} + 12c_{44} + 10c_{66}). \tag{53b}$$

According to Table I the participation tensor is defined by a single parameter, α, which to leading order is unity. Let

$$\alpha = 1 - 2\beta, \tag{54}$$

so that

$$\mathbf{G} = \begin{pmatrix} 1+\beta & 0 & 0 \\ 0 & 1+\beta & 0 \\ 0 & 0 & 1-2\beta \end{pmatrix}. \tag{55}$$

Applying the general perturbation theory we find that the leading order correction to the isotropic participation tensor is given by

$$\beta = \frac{a_l}{15c_l^2}(-4c_{11} + 3c_{33} + c_{13} + 2c_{44}) + \frac{a_t}{30c_t^2}(-c_{11} + 2c_{33} - c_{13} + 3c_{44} - 5c_{66}), \tag{56}$$

where a_l and a_t are defined in Eq. (51a).

Thomsen's anisotropy parameters[18] ϵ, γ, δ provide a means to characterize weakly anisotropic TI materials. The parameters are defined $\epsilon=(c_{11}-c_{33})/(2c_{33})$, $\delta=[(c_{13}+c_{44})^2 - (c_{33}-c_{44})^2]/[2c_{33}(c_{33}-c_{44})]$, $\gamma=(c_{66}-c_{44})/(2c_{44})$, and are commonly used in geophysical applications to describe rock properties. The correction term β can be expressed in terms of the Thomsen parameters as

$$\beta \approx a_1\epsilon + a_2\delta + a_3\gamma, \tag{57}$$

where the coefficients a_1, a_2, and a_3 are (see Fig. 2)

$$a_1 = -\frac{8a_l}{15} - \frac{\kappa^2 a_t}{15}, \quad a_2 = \frac{a_l}{15} - \frac{\kappa^2 a_t}{30}, \quad a_3 = -\frac{a_t}{3}. \tag{58}$$

V. EXAMPLES AND DISCUSSION

The participation matrix was computed for many anisotropic solids. Table II summarizes the results for a selection of materials with anisotropy ranging from weak to strong. Table II provides the numerical values of diagonal elements of $\mathbf{G}$ (there are no off-diagonal elements for the symmetries considered). In each case the elements sum to three, $G_{11}+G_{22}+G_{33}=3$, although the individual numbers can differ markedly from unity.

In order to quantify the level of anisotropy, Table II also shows the number dist. This is a nondimensional positive measure of the degree of anisotropy of a set of anisotropic elastic constants. dist is chosen here as the log-Euclidean distance or length from isotropy,[19,20] although other measures are possible, see Norris[19] for a comparative discussion. The log-Euclidean distance has the advantage that it is invariant regardless of whether the compliance or stiffness tensor are considered. We use dist as a convenient and simple measure of the degree of anisotropy. Appendix C provides a little more detail on its exact definition, including a short MATLAB script to compute dist.

Large deviations from the isotropic participation tensor are apparent. Consider the ratio R of the largest to smallest

TABLE II. The participation matrix $\mathbf{G}$ for a variety of anisotropic materials. Sym denotes material symmetry: transversely isotropic (TI), tetragonal (Tet), or orthotropic (Orth). The Frobenius ($p=2$) norm is used to compare $\mathbf{G}$ with the isotropic result ($\mathbf{I}$) and with the perturbation approximation $\tilde{\mathbf{G}}$ defined by Eq. (50). dist is a nondimensional and invariant measure of the anisotropy (Ref. 19), equal to zero for isotropy. dist$\geqslant$1 signifies considerable anisotropy.

Material	Sym	G_{11}	G_{22}	G_{33}	$\lvert\mathbf{G}-\mathbf{I}\rvert$	$\lvert\mathbf{G}-\tilde{\mathbf{G}}\rvert$	dist
Beryllium[a]	TI	1.05	1.05	0.89	0.13	0.00	0.22
Sulphur[a]	Ort	0.95	1.32	0.73	0.42	0.11	0.95
Cadmium[a]	TI	0.73	0.73	1.55	0.67	0.10	1.02
Barium titanate[b]	Tet	0.81	0.81	1.39	0.48	0.01	1.11
Rochelle salt[a]	Ort	1.38	0.65	0.97	0.52	0.09	1.16
Zinc[a]	TI	0.71	0.71	1.58	0.71	0.14	1.17
Graphite/Epoxy[c]	TI	1.38	1.38	0.25	0.92	0.81	2.35
Tellurium dioxide[d]	Tet	1.30	1.30	0.40	0.74	0.72	2.87
Mercurous iodide[d]	Tet	1.37	1.37	0.26	0.91	0.14	3.02
Spruce[a]	Ort	1.35	1.63	0.02	1.22	1.30	5.59

Elastic moduli from Ref. 15.
From Ref. 21.
From Ref. 22.
From Ref. 23.

element of **G**. Even for small to moderate anisotropy, such as cadmium we see that $R=G_{33}/G_{11}>2$. The ratio becomes much larger for the more anisotropic materials considered. Spruce is included because of its enormous ratio, $R \geqslant \approx 80$. These ratios can be compared with the results for the relative partition of the diffuse wave energy at the free surface of an isotropic solid. If $\mathbf{e}_3$ is the normal to the surface, then the calculations of Weaver[8] indicate that $1 \leqslant G_{33}/G_{11} < 1.25$ where the lower (upper) bound is reached as ν approaches 1/2 (0). The upper bound ≈ 1.25 is approximate and based on Fig. 3 of Ref. 8.

The numbers in Table II indicate that the perturbation approximation is adequate for small anisotropy. This can be characterized loosely as $0 < \text{dist} \leqslant 1$, and strong anisotropy is $\text{dist} \geqslant 2$, roughly. The examples in Table II suggest that the weak anisotropy approximation is not useful in the presence of strong anisotropy. This is evident from the fact that the errors $|\mathbf{G}-\mathbf{I}|$ and $|\mathbf{G}-\tilde{\mathbf{G}}|$ are of the same order of magnitude for the strongly anisotropic materials, whereas $|\mathbf{G}-\tilde{\mathbf{G}}|$ is much less than $|\mathbf{G}-\mathbf{I}|$ for weak anisotropy.

We note that for all materials considered the numerical calculations show Eq. (49) underestimating $\text{tr}\langle \mathbf{Q}^{-3/2} \rangle$. However, the more refined perturbation expansion of $\text{tr}\langle \mathbf{Q}^{-3/2} \rangle$ by Fedorov[17] suggests that this is not a universal result.

The dependence of **G** and $d(\omega)$ on the moduli is obviously complicated by virtue of the averages required in Eq. (4). However, the formula (50) for **G** for weak anisotropy illustrates the dependence more explicitly. The form of the matrices $\mathbf{C}_l$ and $\mathbf{C}_t$ imply that only 12 combinations of the 21 independent anisotropic moduli enter into the first term in the perturbation expansion. For orthotropic materials, with 9 independent moduli, this number reduces to 6, and the matrices $\mathbf{C}_l$ and $\mathbf{C}_t$ are then diagonal. In the case of weak TI only two combinations of moduli influence **G**, see Eq. (56).

The nondimensional tensor **G** also has important implications for radiation from a point source. The connection follows from the relation (8) between **G** and **A**, combined with the correspondence between the drive point admittance tensor and the radiation efficiency in Eq. (7). Thus, the direction in which a force must be applied to most efficiently radiate power is the principal direction of **G** with the largest element. Conversely, the least amount of power is radiated if the force is directed along the principal direction with the smallest element. For instance, Table II indicates that a point force of given magnitude will radiate most power in cadmium if the force is directed along the axis of hexagonal symmetry. The situation is reversed for aligned graphite/epoxy, where forcing along the fiber direction produces the least amount of total radiated power.

The inverse problem of determining anisotropy from measurements of **G** is clearly ill-posed. However, possible measurement could be advantageous in particular circumstances. Consider for instance, three-component measurement of the displacement downhole in a borehole environment. Assuming the frequency is such that the wavelengths are large compared with the bore radius, the three-component data are sufficient to compute the autocorrelation and hence **G**. The principal directions of **G** and the relative magnitude of its diagonal elements provide significant information about the local geostratigraphy and formation properties.

VI. CONCLUSION

We have derived general formulas for diffuse waves in anisotropic solids. The main results are concise expressions for the modal density per unit volume and frequency, $d(\omega)$ of Eq. (4a), and the participation tensor **G** of Eq. (4b). The latter is a material constant with one or two independent constants, and with principal axes dictated by the material symmetry. In the absence of symmetry the participation tensor defines principal axes for diffuse wave energy distribution, and for radiation efficiency. Calculation of $d(\omega)$ and **G** requires, in general, averaging over the surface of the unit sphere. Single integrals suffice for transverse isotropy, with the important quantities given in Eq. (39). In the case of weak anisotropy, a perturbation scheme produces explicit formulas, Eqs. (49) and (50). The main quantity in all cases is the second-order averaged tensor $\langle \mathbf{Q}^{-3/2} \rangle$. We have illustrated the results through calculations for several materials. These display the main effects that would occur in all anisotropic solids. In particular, the deviation **G** from the unit identity tensor can be significant. Ratios of 2 or more for the relative magnitude of diffuse wave energy in different directions in crystals can occur under moderate levels of anisotropy, with far larger ratios possible in realistic materials.

ACKNOWLEDGMENTS

The author would like to thank the anonymous reviewer who pointed out relevant work by Fedorov.

APPENDIX A: DERIVATION OF EQ. (8)

We use an argument based on a modal representation[8] for the solution to the point force problem,

$$\left(\rho \frac{\partial^2}{\partial t^2} - L\right)\mathbf{u} = \mathbf{F}\delta(\mathbf{x}-\mathbf{x}_0)\cos\,\omega t, \tag{A1}$$

where L is a second-order differential operator. The resulting velocity $\mathbf{v}=\partial\mathbf{u}/\partial t$ may be found by standard means as

$$\mathbf{v} = \frac{1}{\rho}\text{Re}\sum_m \frac{-i\omega\mathbf{F}\cdot\mathbf{u}_m(\mathbf{x}_0)\mathbf{u}_m(\mathbf{x})}{\omega_m^2-\omega^2-i0}e^{-i\omega t},$$

where the modes $\mathbf{u}_m(\mathbf{x})e^{-i\omega_m t}$ are solutions of the homogeneous equation (A1), with the properties

$$\delta(\mathbf{x}-\mathbf{x}_0)\mathbf{I} = \sum_m \mathbf{u}_m(\mathbf{x})\mathbf{u}_m(\mathbf{x}_0),$$

$$\int_V d\mathbf{x}\mathbf{u}_m(\mathbf{x})\cdot\mathbf{u}_m(\mathbf{x}) = 1.$$

The power output averaged over a cycle is therefore

$$\Pi(\mathbf{x}_0,\omega)=\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\mathrm{d}t\,\cos\,\omega t\mathbf{F}\cdot\mathbf{v}(\mathbf{x}_0,t)$$

$$=\frac{1}{2\rho}\sum_m[\mathbf{F}\cdot\mathbf{u}_m(\mathbf{x}_0)]^2\mathrm{Re}\frac{-i\omega}{\omega_m^2-\omega^2-i0}. \tag{A2}$$

The strict nondissipative limit of $\mathrm{Re}[-i\omega(\omega_m^2-\omega^2-i0)^{-1}]$ is $\pi\omega\delta(\omega_m^2-\omega^2)=\frac{1}{2}\pi\delta(\omega_m-\omega)$ where δ is the Dirac delta function. However, modal overlap in the presence of nonzero dissipation spreads the influence over many modes. The effect is to make $\mathrm{Re}[-i\omega(\omega_m^2-\omega^2-i0)^{-1}]\to\frac{1}{2}\pi f(\omega_m-\omega)$ where $f(\nu)$ is smooth with bounded support in $\nu\in\{-\Omega,\Omega\}$, say, and unit sum:

$$\sum_{\omega_m}{}'f(\omega_m-\omega)=1. \tag{A3}$$

Here Σ'_{ω_m} indicates the sum over modal frequencies $\omega_m\in\{\omega-\Omega,\omega+\Omega\}$. Using the density of modes, $Vd(\omega_m)$, to replace the sum over modes in Eq, (A2) by a sum over modal frequencies, gives

$$\Pi(\mathbf{x}_0,\omega)=\frac{\pi V}{4\rho}\sum_{\omega_m}{}'d(\omega_m)f(\omega_m-\omega)[\mathbf{F}\cdot\mathbf{u}_m(\mathbf{x}_0)]^2. \tag{A4}$$

We now make the assumption that the support of $f(\nu)$ is small enough that the modal density function, $d(\omega_m)$, may be replaced by $d(\omega)$. This is perfectly reasonable based on known forms for $d(\omega)$, e.g., Eq. (3). At the same time, we assume that the support of $f(\nu)$ is sufficiently large that we may use the equipartition of energy among modes to make the replacement [see Eq. (1)]

$$V\sum_{\omega_m}{}'f(\omega_m-\omega)\mathbf{u}_m\otimes\mathbf{u}_m\to V\frac{\rho\omega^2}{E}\bar{\mathbf{u}}\otimes\bar{\mathbf{u}}=\frac{1}{3}\mathbf{G}. \tag{A5}$$

Hence,

$$\Pi(\mathbf{x}_0,\omega)=\frac{\pi}{12\rho}d(\omega)\mathbf{F}\cdot\mathbf{G}\cdot\mathbf{F}, \tag{A6}$$

and since $\mathbf{F}$ is arbitrary, the admittance $\mathbf{A}$ follows from the definition of Π in Eq. (7). This completes the derivation of the identity (8).

APPENDIX B: DERIVATION OF EQS. (21) AND (26)

The Cayley–Hamilton relation for $\mathbf{Q}$ is $p(\mathbf{Q})=0$, where p is the characteristic cubic polynomial defined in Eq. (28), and $I_1(\mathbf{n})$, $I_2(\mathbf{n})$, $I_3(\mathbf{n})$ are the invariants defined in Eq. (20). Thus,

$$I_1=\lambda_1+\lambda_2+\lambda_3,\quad I_2=\lambda_1\lambda_2+\lambda_2\lambda_3+\lambda_3\lambda_1,$$

$$I_3=\lambda_1\lambda_2\lambda_3,$$

and since $\lambda_\alpha=v_\alpha^2$, it follows that the invariants are all positive, $I_1>0$, $I_2>0$ and $I_3>0$. Multiplying Eq. (19) by $\mathbf{Q}^{-1}$ and $\mathbf{Q}^{-2}$ yields equations for the same quantities:

$$\mathbf{Q}^{-1}=I_3^{-1}\mathbf{Q}^2-I_1I_3^{-1}\mathbf{Q}+I_2I_3^{-1}\mathbf{I}, \tag{B1a}$$

$$\mathbf{Q}^{-2}=I_3^{-1}\mathbf{Q}-I_1I_3^{-1}\mathbf{I}+I_2I_3^{-1}\mathbf{Q}^{-1}. \tag{B1b}$$

Eliminating $\mathbf{Q}^{-1}$ gives an equation for $\mathbf{Q}^{-2}$:

$$\mathbf{Q}^{-2}=I_3^{-2}[I_2\mathbf{Q}^2-(I_1I_2-I_3)\mathbf{Q}+(I_2^2-I_1I_3)\mathbf{I}].$$

We next derive a similar type of equation for $\mathbf{Q}^{1/2}$ using a method due to Hoger and Carlson.[11] The product of this with $\mathbf{Q}^{-2}$, combined with the Cayley–Hamilton equation (19) yields the desired relation (21).

First we note the general expression

$$(\mathbf{Q}-\lambda\mathbf{I})^{-1}=\frac{1}{p(\lambda)}[-\mathbf{Q}^2+(I_1-\lambda)\mathbf{Q}-(\lambda^2-I_1\lambda+I_2)\mathbf{I}], \tag{B2}$$

where p is the characteristic polynomial for $\mathbf{Q}$, from Eq. (28). The identity (B2) may be checked by direct multiplication and use of Eq. (19). The square root tensor $\mathbf{R}\equiv\mathbf{Q}^{1/2}$ satisfies the Cayley–Hamilton equation

$$\mathbf{R}^3-i_1\mathbf{R}^2+i_2\mathbf{R}-i_3\mathbf{I}=0, \tag{B3}$$

where i_1, i_2, and i_3 are related to the invariants of $\mathbf{Q}$ by

$$I_1=i_1^2-2i_2,\quad I_2=i_2^2-2i_1i_3,\quad I_3=i_3^2. \tag{B4}$$

Explicit formulas for i_1, i_2, and i_3 are given in Eq. (23a). Rearranging Eq. (B3) as $\mathbf{R}(\mathbf{R}^2+i_2\mathbf{I})=i_1\mathbf{R}^2+i_3\mathbf{I}$ and using $\mathbf{R}^2=\mathbf{Q}$ gives

$$\mathbf{R}=(i_1\mathbf{Q}+i_3\mathbf{I})(\mathbf{Q}+i_2\mathbf{I})^{-1}. \tag{B5}$$

Application of Eq. (B2) along with some simplifications using Eq. (B4), such as $p(-i_2)=-(i_3-i_1i_2)^2$, yields

$$\mathbf{Q}^{1/2}=(i_3-i_1i_2)^{-1}[\mathbf{Q}^2+(i_2-i_1^2)\mathbf{Q}-i_1i_3\mathbf{I}]. \tag{B6}$$

Combining Eqs. (B2) and (B6) gives Eq. (21). Alternatively,

$$\mathbf{Q}^{-3/2}=a\mathbf{Q}^2+b\mathbf{Q}+c\mathbf{I}, \tag{B7}$$

where

$$a=\frac{I_3(i_2-i_1^2)-I_2i_1i_3}{I_3^2(i_3-i_1i_2)},$$

$$b=\frac{I_1I_3(i_1^2-i_2)+(I_1I_2-I_3)i_1i_3}{I_3^2(i_3-i_1i_2)},$$

$$c=\frac{I_3^2+I_2I_3(i_2-i_1^2)+(I_1I_3-I_2^2)i_1i_3}{I_3^2(i_3-i_1i_2)}. \tag{B8}$$

The second form (26) for $\mathbf{Q}^{-3/2}$ is based on the identity (17). The tensor products of eigenvectors for λ_i satisfy

$$\mathbf{q}_i\otimes\mathbf{q}_i=\frac{(\mathbf{Q}-\lambda_j\mathbf{I})(\mathbf{Q}-\lambda_k\mathbf{I})}{(\lambda_i-\lambda_j)(\lambda_i-\lambda_k)},$$

$$i\neq j\neq k\neq i\quad(\text{no sum}).$$

This follows, for example, by eliminating the other two tensor products using the spectral expressions for $\mathbf{I}$, $\mathbf{Q}$, and $\mathbf{Q}^2$. The dependence on λ_j and λ_k can be removed in favor of λ_j and the invariants I_1 and I_3, and hence Eq. (27). Note that the latter can be expressed

$$\mathbf{N}(\lambda,\mathbf{n}) = \frac{1}{\lambda p'(\lambda)}[\lambda\mathbf{Q}^2 + (\lambda - I_1)\lambda\mathbf{Q} + I_3\mathbf{I}], \tag{B9}$$

where $p'(x)$ is the derivative of the characteristic polynomial. This indicates that the general expression (27) is invalid at double roots where the slowness surface exhibits degeneracy, and proper limits are required. The possibility of such points does not present a practical impediment to numerical integration.

APPENDIX C: THE LOG-EUCLIDEAN DISTANCE

The procedure[19] is to first calculate an effective isotropic set of moduli analogous to $c^{(0)}_{ijkl}$ of Eq. (40) but for the matrix logarithm of the six-dimensional Voigt matrix of moduli C_{IJ}. Some matrix factors are required to convert from the Voigt notation. The following MATLAB lines compute dist if C is the 6×6 Voigt matrix.

```
J = 1/3 * [1 1 1 0 0 0]' * [1 1 1 0 0 0];

K = eye(6)-J;

T = diag([1 1 1 sqrt(2) * [1 1 1]]);

L = logm(T * C * T);

dist = norm(logm(J * exp(trace(J * L))
       + K * exp(1/5 *  trace(K * L))) – L, ' fro ' );
```


[1] R. L. Weaver, "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609 (1982).
[2] D. M. Egle, "Diffuse wave fields in solid media," J. Acoust. Soc. Am. **70**, 476–480 (1981).
[3] C. Kittel, *Elementary Statistical Physics* (Wiley, New York, 1958).
[4] R. S. Langley, "The modal density of anisotropic structural components," J. Acoust. Soc. Am. **99**, 3481–3487 (1996).
[5] R. L. Weaver, "Diffuse waves in finite plates," J. Sound Vib. **94**, 319–335 (1984).
[6] V. K. Tewary, M. Mahapatra, and C. M. Fortunko, "Green's function for anisotropic half-space solids in frequency space and calculation of mechanical admittance," J. Acoust. Soc. Am. **100**, 2960–2963 (1996).
[7] K. Wapenaar, "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301-254304 (2004).
[8] R. L. Weaver, "Diffuse elastic waves at a free surface," J. Acoust. Soc. Am. **78**, 131–136 (1985).
[9] P. J. Shorter and R. S. Langley, "On the reciprocity relationship between direct field radiation and diffuse reverberant loading," J. Acoust. Soc. Am. **117**, 85–95 (2005).
[10] A. N. Norris, "Dynamic Green's functions in anisotropic piezoelectric, thermoelastic and poroelastic solids," Proc. R. Soc. London, Ser. A **447**, 175–188 (1994).
[11] A. Hoger and D. E. Carlson, "Determination of the stretch and rotation in the polar decomposition of the deformation gradient," Q. Appl. Math. **42**, 113–117 (1984).
[12] A. N. Norris, "Invariants of $\mathbf{c}^{1/2}$ in terms of the invariants of $\mathbf{c}$," J. Mech. Mater. Struct. **2**(9), 1805–1812 (2007).
[13] M. M. Carroll, "Derivatives of the rotation and stretch tensors," Math. Mech. Solids **9**, 543–553 (2004).
[14] A. G. Every, "General closed-form expressions for acoustic waves in elastically anisotropic solids," Phys. Rev. B **22**, 1746–1760 (1980).
[15] M. J. P. Musgrave, *Crystal Acoustics* (Acoustical Society of America, New York, 2003).
[16] A. N. Norris, "Elastic moduli approximation of higher symmetry for the acoustical properties of an anisotropic material," J. Acoust. Soc. Am. **119**, 2114–2121 (2006).
[17] F. I. Fedorov, *Theory of Elastic Waves in Crystals* (Plenum, New York, 1968).
[18] L. Thomsen, "Weak elastic anisotropy," Geophysics **51**, 1954–1966 (1986).
[19] A. N. Norris, "The isotropic material closest to a given anisotropic material," J. Mech. Mater. Struct. **1**, 223–238 (2006).
[20] M. Moakher and A. N. Norris, "The closest elastic tensor of arbitrary symmetry to an elasticity tensor of lower symmetry," J. Elast. **85**, 215–263 (2006).
[21] R. Truell, C. Elbaum, and B. Chick, *Ultrasonic Methods in Solid State Physics* (Academic, New York, 1969).
[22] B. Vandenbossche, R. Kriz, and T. Oshima, "Stress-wave displacement polarizations and attenuation in unidirectional composites: Theory and experiment," Res. Nondestruct. Eval. **8**, 101–124 (1996).
[23] A. Cazzani and M. Rovati, "Extrema of Young's modulus for elastic solids with tetragonal symmetry," Int. J. Solids Struct. **42**, 1713–1744 (2005).

Perturbations of the seismic reflectivity of a fluid-saturated depth-dependent poroelastic medium

Louis de Barros[a] and Michel Dietrich[b]
Laboratoire de Géophysique Interne et Tectonophysique (LGIT), CNRS, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France

(Received 30 May 2007; revised 8 December 2007; accepted 20 December 2007)

Analytical formulas are derived to compute the first-order effects produced by plane inhomogeneities on the point source seismic response of a fluid-filled stratified porous medium. The derivation is achieved by a perturbation analysis of the poroelastic wave equations in the plane–wave domain using the Born approximation. This approach yields the Fréchet derivatives of the *P–SV*- and *SH*-wave responses in terms of the Green's functions of the unperturbed medium. The accuracy and stability of the derived operators are checked by comparing, in the time–distance domain, differential seismograms computed from these analytical expressions with complete solutions obtained by introducing discrete perturbations into the model properties. For vertical and horizontal point forces, it is found that the Fréchet derivative approach is remarkably accurate for small and localized perturbations of the medium properties which are consistent with the Born approximation requirements. Furthermore, the first-order formulation appears to be stable at all source–receiver offsets. The porosity, consolidation parameter, solid density, and mineral shear modulus emerge as the most sensitive parameters in forward and inverse modeling problems. Finally, the amplitude-versus-angle response of a thin layer shows strong coupling effects between several model parameters. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2835419]

PACS number(s): 43.40.Ph, 43.20.Gp, 43.20.Jr, 43.20.Bi [RAS] Pages: 1409–1420

I. INTRODUCTION

The evaluation of the sensitivity of a seismic wave field to small perturbations of the material properties is a classical issue of seismology which arises in the solution of forward and inverse scattering problems (Aki and Richards, 1980; Tarantola, 1984). Sensitivity operators are mainly useful for the optimal design of field or laboratory experiments, for the interpretation of time-lapse monitoring surveys, and for the development of imaging and linearized inversion techniques. In particular, sensitivity operators play a central role in least-square inversion schemes using gradient techniques. All these applications, especially the latter, call for fast and effective numerical computation methods of the sensitivity operators. Indeed, the most intuitive approach to compute the perturbational or differential seismograms in a structure described by N parameters is to use a finite perturbation scheme that demands $N+1$ forward modeling computations. This approach rapidly becomes prohibitive as the number of parameters increases. The problem can be solved more elegantly by first deriving the so-called Fréchet derivatives of the seismic wave fields with respect to material properties. Although restricted to first-order effects only, this procedure makes it possible to efficiently predict small changes of the seismic response resulting from slight modifications of the material properties. Furthermore, this solution can be implemented by solving only one large forward problem.

One of the most significant contributions in this area is the work of Tarantola (1984), who applied a first-order perturbation analysis to the elastodynamic wave equation to derive a series of general formulas relating the scattered wave field to heterogeneities in an arbitrarily complex elastic medium. The same approach was subsequently used in a more restrictive sense for layered media in the plane–wave domain by Pan *et al.* (1988) in the acoustic case and by Dietrich and Kormendi (1990) in the *P–SV* case. The Fréchet derivatives obtained in these different cases are all expressed as combinations of the incident wave field generated by the seismic source at the scatterer location and the Green's functions between the scatterer and receiver locations. The structure of these expressions underlines the fact that the scattered waves due to perturbations in the material properties (usually density and acoustic or elastic parameters) can be interpreted as a wave field generated by a set of secondary body forces coincident with the heterogeneities, and determined by complex interactions between the incident waves and the medium perturbations. This structure is met with all problems of acoustic, seismic, or electromagnetic wave propagation in weakly inhomogeneous media, and will also be found in more complex situations such as anisotropic or poroelastic media.

The poroelastic model, which is the subject of this paper, involves more parameters than the viscoelastic case, but on the other hand, the wave velocities, attenuation, and dispersion characteristics are computed from the medium's intrinsic properties without having to resort to empirical relationships. Since the pioneering work of Biot (1956), many authors (Auriault *et al.*, 1985; Dutta and Odé, 1979; Johnson *et al.*, 1994, for example) have introduced improvements of the poroelastodynamic equations, either by averaging or by integrating techniques. The Biot (1956) theory and its appli-

[a]Electronic mail: louis.debarros@ujf-grenoble.fr
[b]Now at Institut Français du Pétrole, 92852 Rueil-Malmaison, France.

cations is still a field of active research, as demonstrated by the large number of current publications devoted to the subject (see, e.g., Trifunac, 2006). The forward problem, i.e., the computation of synthetic seismograms in poroelastic media has been solved in different configurations and with several techniques (Carcione, 1996; Dai *et al.*, 1995; Garambois and Dietrich, 2002; Haartsen and Pride, 1997). However, the inverse problem has only been rarely addressed, and to our knowledge, it has never been without first estimating the wave velocities (Berryman *et al.*, 2002; Chotiros, 2002; Spikes *et al.*, 2006). Yet, inversion algorithms can provide useful information on the material properties, notably permeability and porosity which are the most important parameters to characterize porous media.

The main objective of this work is to extend the methodology used in the elastic case (Dietrich and Kormendi, 1990) to obtain the Fréchet derivatives for stratified poroelastic media. We consider here a depth-dependent, fluid-saturated porous medium representing reservoir rocks or sedimentary layers. The computation of the point source seismic response of the layered structure is carried out by combining the generalized reflection and transmission matrix method (Kennett, 1983) with the discrete wave number technique (Bouchon, 1981). This combination was already implemented by Garambois and Dietrich (2002) for the numerical simulation of the coupled seismic and electromagnetic wave propagation in porous media by using the work of Pride (1994). In the following sections, we first present the governing equations and constitutive parameters for porous materials before expressing the wave propagation equations for depth-dependent media. Next, we develop the analytical computation of the Fréchet derivatives in the frequency-ray parameter domain for the *P–SV*- and *SH*-wave cases. Finally, we check the accuracy of the sensitivity operators obtained in the time–distance domain, both in an infinite medium and in a complex seismic model. We conclude with the sensitivity of the seismic wave forms with respect to the different model parameters.

II. WAVE PROPAGATION IN STRATIFIED POROUS MEDIA

A. Governing equations

Assuming a $e^{-i\omega t}$ dependence, where t is time and ω the angular frequency, Pride (1994, 2005) rewrote Biot's (1956) equations of poroelasticity in the form

$$\begin{aligned}
&\nabla \cdot \boldsymbol{\tau} = -\omega^2(\rho \mathbf{u} + \rho_f \mathbf{w}),\\
&\boldsymbol{\tau} = [K_U \nabla \cdot \mathbf{u} + C \nabla \cdot \mathbf{w}]\mathbf{I} + G[\nabla \mathbf{u} + (\nabla \mathbf{u})^T - 2/3(\nabla \cdot (\mathbf{u}\mathbf{I}))],\\
&-P = C \nabla \cdot \mathbf{u} + M \nabla \cdot \mathbf{w},\\
&-\nabla P = -\omega^2 \rho_f \mathbf{u} - \omega^2 \tilde{\rho} \mathbf{w},
\end{aligned} \tag{1}$$

where $\mathbf{u}$ and $\mathbf{w}$ are the solid average displacement and the relative fluid-to-solid displacement, respectively. More precisely, defining $\mathbf{u}_s$ and $\mathbf{u}_f$ as the displacements of the solid and fluid phases of a porous continuum, we can write $\mathbf{u} \simeq \mathbf{u}_s$ and $\mathbf{w} = \phi(\mathbf{u}_s - \mathbf{u}_f)$, where ϕ is the porosity. P represents the interstitial pressure and $\boldsymbol{\tau}$ is the 3×3 stress tensor. ρ denotes the density of the porous medium which is related to the fluid density ρ_f, solid density ρ_s, and porosity ϕ via the relationship

$$\rho = (1 - \phi)\rho_s + \phi \rho_f. \tag{2}$$

The undrained bulk modulus K_U is defined under the condition $\mathbf{w} = \mathbf{0}$. G is the shear modulus of a drained or an undrained medium as it is independent of the fluid properties (Gassmann, 1951). The fluid storage coefficient M represents the amount of fluid a sample can accumulate at constant sample volume. Biot's C modulus is a mechanical parameter describing the variation of the fluid pressure due to a change of the sample volume in an undrained medium. At low frequencies, these parameters as well as the Lamé parameter λ_u defined in the following are real, frequency-independent, and can be expressed in terms of the drained modulus K_D, porosity ϕ, mineral modulus of the grains K_s, and fluid modulus K_f (Gassmann, 1951):

$$\begin{aligned}
&K_U = \frac{\phi K_D + \left[1 - (1 + \phi)\frac{K_D}{K_s}\right]K_f}{\phi(1 + \Delta)}, \quad \lambda_U = K_U - \frac{2}{3}G,\\
&C = \frac{\left[1 - \frac{K_D}{K_s}\right]K_f}{\phi(1 + \Delta)}, \quad M = \frac{K_f}{\phi(1 + \Delta)}\\
&\text{with} \quad \Delta = \frac{1 - \phi}{\phi}\frac{K_f}{K_s}\left[1 - \frac{K_D}{(1 - \phi)K_s}\right].
\end{aligned} \tag{3}$$

It is also possible to link the bulk properties K_D and G to the porosity and constitutive mineral properties via empirical relationships derived from experimental results (Bemer *et al.*, 2004; Pride, 2005):

$$K_D = K_s \frac{1 - \phi}{1 + c_s \phi}, \quad G = G_s \frac{1 - \phi}{1 + 3c_s \phi/2}. \tag{4}$$

Equation (4) has the merit of being very simple and introduces only two additional parameters, namely, the shear modulus of the grains G_s and the consolidation parameter c_s. The latter mainly depends on the cementing properties of the grains, but also on the pore shape. The value of the consolidation parameter c_s is typically between 2 and 20 in a consolidated medium, and can be very much greater than 20 in an unconsolidated soil.

Finally, the wave attenuation is explained by Darcy's law, which uses a complex, frequency-dependent dynamic permeability (Johnson *et al.*, 1994):

$$\tilde{\rho} = i\frac{\eta}{\omega k(\omega)} \quad \text{with } k(\omega) = k_0 \Big/ \left[\sqrt{1 - i\frac{4}{n_J}\frac{\omega}{\omega_c}} - i\frac{\omega}{\omega_c}\right]. \tag{5}$$

The dynamic permeability $k(\omega)$ tends toward the hydrogeological (dc) permeability k_0 at low frequencies where viscous losses are dominant. It includes a correction accounting for the inertial effects at higher frequencies. These two domains are separated by the relaxation frequency

$$\omega_c = \frac{\eta}{\rho_f F k_0}, \tag{6}$$

where η is the viscosity of the fluid. Archie's law ($F=\phi^{-m}$) expresses the electrical formation factor F in terms of the porosity ϕ and cementing exponent m whose value is between 1 and 2 depending on the pore topology. Parameter n_J is considered constant and equal to 8 to simplify the equations. We refer the reader to the work of Pride (2005) for more information on the parameters used in this study.

B. Coupled second-order equations for plane waves

The horizontally layered model lends itself to a number of analytical developments if one performs a plane–wave decomposition of the wave fields represented by Eq. (1). This procedure involves a series of changes of variables and integral transforms which are described in detail by Kennett (1983) in the elastic case. As in the elastic case, the introduction of new variables (Hudson, 1969) leads to a useful decomposition into $P-SV$- and SH-wave systems in cylindrical coordinates. By applying the whole sequence of transformations and arranging the terms, we find that the governing equations of the $P-SV$-wave system in depth-dependent poroelastic media reduce to the following system of second-order equations in the angular frequency ω and ray parameter p domain:

$$\begin{aligned}
F_{1z} &= \frac{\partial}{\partial z}\left[(\lambda_U + 2G)\frac{\partial U}{\partial z} + C\frac{\partial W}{\partial z} - \omega p(\lambda_U V + CX)\right] \\
&\quad - \omega p G\frac{\partial V}{\partial z} + \omega^2[\rho U - p^2 G U + \rho_f W], \\
F_{1r} &= \frac{\partial}{\partial z}\left[G\frac{\partial V}{\partial z} + \omega p G U\right] + \omega p\left[\lambda_U\frac{\partial U}{\partial z} + C\frac{\partial W}{\partial z}\right] \\
&\quad + \omega^2[\rho V - p^2(\lambda_U + 2G)V - p^2 C X], \\
F_{2z} &= \frac{\partial}{\partial z}\left[C\left(\frac{\partial U}{\partial z} - \omega p V\right) + M\left(\frac{\partial W}{\partial z} - \omega p X\right)\right] + \omega^2[\rho_f U \\
&\quad + \tilde{\rho} W], \\
F_{2r} &= \omega p\left[C\frac{\partial U}{\partial z} + M\frac{\partial W}{\partial z}\right] + \omega^2[\rho_f V + \tilde{\rho} X - p^2(CV \\
&\quad + MX)].
\end{aligned} \tag{7}$$

In Eq. (7), $U=U(z_R,\omega;z_S)$ and $V=V(z_R,\omega;z_S)$ respectively denote the vertical and radial components of the solid displacements. Similarly, $W=W(z_R,\omega;z_S)$ and $X=X(z_R,\omega;z_S)$ respectively denote the vertical and radial components of the relative fluid-to-solid displacements. Variables z_R and z_S stand for the receiver and seismic source depths. Equation (7) is valid in the presence of body forces $\mathbf{F}_1 = [F_{1z}(z_S,\omega), F_{1r}(z_S,\omega)]^T$ and $\mathbf{F}_2 = [F_{2z}(z_S,\omega), F_{2r}(z_S,\omega)]^T$ defined by their vertical (z index) and radial (r index) components: Force $\mathbf{F}_1$ is applied on an average volume of porous medium and represents a stress discontinuity, while force $\mathbf{F}_2$ is derived from the pressure gradient in the fluid.

We can then cast Eq. (7) in the form of a matrix differential equation as

$$\mathbf{L}^{PSV}\mathbf{Q}^{PSV} = \mathbf{F}^{PSV},$$

$$\text{where } \mathbf{Q}^{PSV} = \begin{bmatrix} U \\ V \\ W \\ X \end{bmatrix}, \quad \mathbf{F}^{PSV} = \begin{bmatrix} F_{1z} \\ F_{1r} \\ F_{2z} \\ F_{2r} \end{bmatrix}. \tag{8}$$

$\mathbf{L}^{PSV}$ is a differential operator given by

$$\begin{aligned}
\mathbf{L}^{PSV} &= \frac{\partial}{\partial z}\left(\mathbf{M}_1^{PSV}\frac{\partial}{\partial z} + \omega p\mathbf{M}_2^{PSV}\right) - \omega p[\mathbf{M}_2^{PSV}]^T\frac{\partial}{\partial z} \\
&\quad + \omega^2(\mathbf{M}_3^{PSV} - p^2\mathbf{M}_4^{PSV})
\end{aligned} \tag{9}$$

where the $\mathbf{M}_i^{PSV}$, $i=1,\ldots,4$ are 4×4 matrices defined by

$$\mathbf{M}_1^{PSV} = \begin{bmatrix} \lambda_U + 2G & 0 & C & 0 \\ 0 & G & 0 & 0 \\ C & 0 & M & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_2^{PSV} = \begin{bmatrix} 0 & -\lambda_U & 0 & -C \\ G & 0 & 0 & 0 \\ 0 & -C & 0 & -M \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M}_3^{PSV} = \begin{bmatrix} \rho & 0 & \rho_f & 0 \\ 0 & \rho & 0 & \rho_f \\ \rho_f & 0 & \tilde{\rho} & 0 \\ 0 & \rho_f & 0 & \tilde{\rho} \end{bmatrix}, \tag{10}$$

$$\mathbf{M}_4^{PSV} = \begin{bmatrix} G & 0 & 0 & 0 \\ 0 & \lambda_U + 2G & 0 & C \\ 0 & 0 & 0 & 0 \\ 0 & C & 0 & M \end{bmatrix}.$$

Apart from the dimensions of the matrices, we may note that Eqs. (9) and (10) are very similar to the expressions obtained in the elastic case.

Formally, Eq. (8) admits an integral solution for the displacement fields in terms of the Green's functions of the problem. For example, the vertical displacement U at depth z_R and frequency ω is given by

$$\begin{aligned}
U(z_R,\omega) = \int_{\mathcal{M}} &[G_{1z}^{1z}(z_R,\omega;z')F_{1z}(z',\omega) \\
&+ G_{1z}^{2z}(z_R,\omega;z')F_{2z}(z',\omega) \\
&+ G_{1z}^{1r}(z_R,\omega;z')F_{1r}(z',\omega) \\
&+ G_{1z}^{2r}(z_R,\omega;z')F_{2r}(z',\omega)]dz',
\end{aligned} \tag{11}$$

where $G_{ij}^{kl}(z_R,\omega;z_S)$ is the Green's function corresponding to the displacement at depth z_R of phase i (the values $i=1,2$ correspond to solid and relative fluid-to-solid motions, respectively) in direction j (z or r) generated by a harmonic point force $F_{kl}(z_S,\omega)$ ($k=1,2$) at depth z_S in direction l (z or r). A total of 16 different Green's functions are needed to express the 4 displacements U, V, W, and X in the $P-SV$-wave system (4 displacements $\times$ 4 forces).

The integrals of Eq. (11) are taken over the depths z' of a region $\mathcal{M}$ including the body forces $\mathbf{F}_1$ and $\mathbf{F}_2$. In the case of a vertical point force at depth z_S, the expressions of the forces become

$$F_{1z}(z_S,\omega)=\delta(z-z_S)S_1(\omega),\quad F_{2z}(z_S,\omega)=\delta(z-z_S)S_2(\omega),$$

$$F_{1r}(z_S,\omega)=0,\quad F_{2r}(z_S,\omega)=0, \tag{12}$$

where $S_1(\omega)$ and $S_2(\omega)$ are the Fourier transforms of the source time functions associated with forces F_1 and F_2. Assuming that the amplitudes of both forces are similar ($\|\mathbf{F}_1\| \simeq \|\mathbf{F}_2\|$) (Garambois and Dietrich, 2002), we take $S(\omega)=S_1(\omega)=S_2(\omega)$. The displacement fields for a vertical point force can then be written in simple forms with the Green's functions:

$$U(z_R,\omega;z_S)=[G_{1z}^{1z}(z_R,\omega;z_S)+G_{1z}^{2z}(z_R,\omega;z_S)]S(\omega),$$

$$V(z_R,\omega;z_S)=[G_{1r}^{1z}(z_R,\omega;z_S)+G_{1r}^{2z}(z_R,\omega;z_S)]S(\omega),$$

$$W(z_R,\omega;z_S)=[G_{2z}^{1z}(z_R,\omega;z_S)+G_{2z}^{2z}(z_R,\omega;z_S)]S(\omega),$$

$$X(z_R,\omega;z_S)=[G_{2r}^{1z}(z_R,\omega;z_S)+G_{2r}^{2z}(z_R,\omega;z_S)]S(\omega). \tag{13}$$

The displacement fields corresponding to a horizontal point force and to an explosive point source are similarly defined. Explosions would be represented by Green's functions G_{ij}^{kE} ($i=1,2$; $j=z,r$; $k=1,2$) representing the radiation of an explosive point source E.

The SH case is treated in exactly the same way as the $P-SV$ case. The corresponding second-order differential equations of motion are

$$F_{1t}=\frac{\partial}{\partial z}\left[G\frac{\partial T}{\partial z}\right]+\omega^2[-p^2GT+\rho T+\rho_f Y],$$

$$F_{2t}=\omega^2[\rho_f T+\tilde{\rho}Y], \tag{14}$$

where $T=T(z_R,\omega;z_S)$ and $Y=Y(z_R,\omega;z_S)$ stand for the transverse solid and fluid-to-solid displacements; $F_{1t}(z_S,\omega)$ and $F_{2t}(z_S,\omega)$ are the transverse components of body forces $\mathbf{F}_1$ and $\mathbf{F}_2$.

As before, we can rewrite Eq. (14) in matrix form as $\mathbf{L}^{SH}\mathbf{Q}^{SH}=\mathbf{F}^{SH}$,

$$\text{where}\quad \mathbf{Q}^{SH}=\begin{bmatrix}T\\ Y\end{bmatrix},\quad \mathbf{F}^{SH}=\begin{bmatrix}F_{1t}\\ F_{2t}\end{bmatrix}. \tag{15}$$

Here, $\mathbf{L}^{SH}$ is a linear operator defined as

$$\mathbf{L}^{SH}=\frac{\partial}{\partial z}\left(\mathbf{M}_1^{SH}\frac{\partial}{\partial z}\right)+\omega^2(\mathbf{M}_2^{SH}-p^2\mathbf{M}_1^{SH}), \tag{16}$$

where the $\mathbf{M}_i^{SH}$, $i=1,2$, are 2×2 matrices defined by

$$\mathbf{M}_1^{SH}=\begin{bmatrix}G & 0\\ 0 & 0\end{bmatrix},\quad \mathbf{M}_2^{SH}=\begin{bmatrix}\rho & \rho_f\\ \rho_f & \tilde{\rho}\end{bmatrix}. \tag{17}$$

III. FRÉCHET DERIVATIVES OF THE PLANE–WAVE REFLECTIVITY

A. Statement of the problem

The Fréchet derivatives are usually introduced by considering the forward problem of the wave propagation, in which a set of synthetic seismograms $\mathbf{d}$ is computed for an earth model $\mathbf{m}$ using the nonlinear relationship $\mathbf{d}=\mathbf{f}(\mathbf{m})$. Tarantola (1984) uses a Taylor series expansion to relate a small perturbation $\delta\mathbf{m}$ in the model parameters to a small perturbation $\delta\mathbf{f}$ in the wave field

$$\mathbf{f}(\mathbf{m}+\delta\mathbf{m})=\mathbf{f}(\mathbf{m})+\mathbf{D}\,\delta\mathbf{m}+\mathbf{o}(\|\delta\mathbf{m}\|^2)\quad \text{or}\quad \delta\mathbf{f}=\mathbf{D}\,\delta\mathbf{m}, \tag{18}$$

where $\mathbf{D}=\partial\mathbf{f}/\partial\mathbf{m}$ is the matrix of Fréchet derivatives.

Our aim is to compute the various Fréchet derivatives corresponding to slight modifications of the model parameters at a given depth. Considering for instance the density at depth z, this problem reduces, in the $P-SV$ case, to finding analytical expressions for the quantity

$$\mathcal{A}_1^{PSV}(z_R,\omega;z_S|z)=\frac{\partial U(z_R,\omega;z_S)}{\partial\rho(z)}. \tag{19}$$

We can similarly define the Fréchet derivatives $\mathcal{A}_2^{PSV}$, $\mathcal{A}_3^{PSV}$, and $\mathcal{A}_4^{PSV}$ for the displacements V, W and X. In the same way, $\mathcal{B}_i^{PSV}$, $\mathcal{C}_i^{PSV}$, $\mathcal{D}_i^{PSV}$, $\mathcal{E}_i^{PSV}$, $\mathcal{F}_i^{PSV}$ and $\mathcal{G}_i^{PSV}$, $i=1,\ldots,4$ are the derivatives for model parameters ρ_f, $\tilde{\rho}$, C, M, λ_U, and G. This is the natural choice of parameters to carry out a perturbation analysis because of the linear dependence of these parameters with the wave equations. We also introduce the set of Fréchet derivatives $\hat{\mathcal{A}}_i^{PSV}$, $\hat{\mathcal{B}}_i^{PSV}$, $\hat{\mathcal{C}}_i^{PSV}$, $\hat{\mathcal{D}}_i^{PSV}$, $\hat{\mathcal{E}}_i^{PSV}$, $\hat{\mathcal{F}}_i^{PSV}$, $\hat{\mathcal{G}}_i^{PSV}$, and $\hat{\mathcal{H}}_i^{PSV}$, $i=1,\ldots,4$, corresponding to model parameters ρ_s, ρ_f, k_0, ϕ, K_s, K_f, G_s, and c_s, which we will use in a second stage, and which are more convenient to use as physical parameters of the problem. The sensitivity operators are derived by following the procedure presented in Dietrich and Kormendi (1990) for the elastic case.

B. Perturbation analysis

We first present, with some detail, the perturbation analysis for the $P-SV$ case before addressing the simpler SH case. We consider small changes in the model parameters at

a given depth z that result in small perturbations $\mathbf{\Delta Q}^{PSV}=[\delta U,\delta V,\delta W,\delta X]^T$ of the seismic wave field and in a modified seismic response $\mathbf{Q'}^{PSV}=[U',V',W',X']^T$. The latter can be written as

$$\mathbf{Q'}^{PSV}=\mathbf{Q}^{PSV}+\mathbf{\Delta Q}^{PSV} \tag{20}$$

by assuming that the magnitudes of the scattered waves $\mathbf{\Delta Q}^{PSV}$ are much smaller than those of the primary waves $\mathbf{Q}^{PSV}$. Considering for instance the first component of the above-presented vector equation

$$U'(z_R,\omega;z_S)=U(z_R,\omega;z_S)+\delta U(z_R,\omega;z_S), \tag{21}$$

we can write the scattered displacement δU as

$$\begin{aligned}\delta U(z_R,\omega;z_S)=\int_{\mathcal{M}}[&\mathcal{A}_1^{PSV}(z_R,\omega;z_S|z)\,\delta\rho(z)\\&+\mathcal{B}_1^{PSV}(z_R,\omega;z_S|z)\,\delta\rho_f(z)\\&+\mathcal{C}_1^{PSV}(z_R,\omega;z_S|z)\,\delta\tilde{\rho}(z)\\&+\mathcal{D}_1^{PSV}(z_R,\omega;z_S|z)\,\delta C(z)\\&+\mathcal{E}_1^{PSV}(z_R,\omega;z_S|z)\,\delta M(z)\\&+\mathcal{F}_1^{PSV}(z_R,\omega;z_S|z)\,\delta\lambda_U(z)\\&+\mathcal{G}_1^{PSV}(z_R,\omega;z_S|z)\,\delta G(z)]dz.\end{aligned} \tag{22}$$

With the model parametrization adopted (i.e., a linear dependence of the model parameters with the wave equation), the perturbation analysis can mainly be done in symbolic form. Indeed, the wave operator $\mathbf{L}^{PSV}$ in the perturbed medium can be written as

$$\mathbf{L'}^{PSV}=\mathbf{L}^{PSV}+\mathbf{\Delta L}^{PSV}, \tag{23}$$

so that Eq. (8) becomes (Hudson and Heritage, 1981)

$$(\mathbf{L}^{PSV}+\mathbf{\Delta L}^{PSV})(\mathbf{Q}^{PSV}+\mathbf{\Delta Q}^{PSV})=\mathbf{F}^{PSV}. \tag{24}$$

We then use the single scattering (or Born) approximation to solve Eq. (24) for $\mathbf{\Delta Q}^{PSV}$ under the assumption (already used earlier) that $\|\mathbf{\Delta Q}^{PSV}\|\ll\|\mathbf{Q}^{PSV}\|$. Combining this approximation with Eq. (8), we obtain

$$\mathbf{L}^{PSV}\mathbf{\Delta Q}^{PSV}\simeq-\mathbf{\Delta L}^{PSV}\mathbf{Q}^{PSV}\equiv\mathbf{\Delta}F^{PSV}. \tag{25}$$

This matrix equation shows that the scattered waves $\mathbf{\Delta Q}^{PSV}$ due to perturbations of the material properties can be interpreted as a wave field generated by secondary body forces $\mathbf{\Delta F}^{PSV}$ defined by the interaction of the incident waves with the heterogeneities (term $\mathbf{\Delta L}^{PSV}\,\mathbf{Q}^{PSV}$). Moreover, this wave field propagates in the unperturbed medium represented by the wave operator $\mathbf{L}^{PSV}$. Consequently, Eq. (25) has for each of its components a solution similar to Eq. (11), by substituting $\mathbf{\Delta F}^{PSV}$ for $\mathbf{F}^{PSV}$ and $\mathbf{\Delta Q}^{PSV}$ for $\mathbf{Q}^{PSV}$. Considering again the scattered displacement δU, we have

$$\begin{aligned}\delta U(z_R,\omega;z_S)=\int_{\mathcal{M}}[&G_{1z}^{1z}(z_R,\omega;z')\,\delta F_{1z}(z',\omega)\\&+G_{1z}^{2z}(z_R,\omega;z')\,\delta F_{2z}(z',\omega)\\&+G_{1z}^{1r}(z_R,\omega;z')\,\delta F_{1r}(z',\omega)\\&+G_{1z}^{2r}(z_R,\omega;z')\,\delta F_{2r}(z',\omega)]dz',\end{aligned} \tag{26}$$

where the secondary Born sources δF_{1z}, δF_{1r}, δF_{2z}, δF_{2r} are obtained from Eq. (7):

$$\mathbf{\Delta F}^{PSV}=\begin{pmatrix}\delta F_{1z}\\ \delta F_{1r}\\ \delta F_{2z}\\ \delta F_{2r}\end{pmatrix}=\begin{pmatrix}-\frac{\partial}{\partial z}\left[(\delta\lambda_U+2\delta G)\frac{\partial U}{\partial z}+\delta C\frac{\partial W}{\partial z}-\omega p(\delta\lambda_U V+\delta CX)\right]+\omega p\,\delta G\frac{\partial V}{\partial z}-\omega^2[\delta\rho U-p^2\delta GU+\delta\rho_f W]\\ -\frac{\partial}{\partial z}\left[\delta G\frac{\partial V}{\partial z}+\omega p\,\delta GU\right]-\omega p\left[\delta\lambda_U\frac{\partial U}{\partial z}+\delta C\frac{\partial W}{\partial z}\right]-\omega^2[\delta\rho V-p^2(\delta\lambda_U+2\delta G)V-p^2\delta CX]\\ -\frac{\partial}{\partial z}\left[\delta C\left(\frac{\partial U}{\partial z}-\omega pV\right)+\delta M\left(\frac{\partial W}{\partial z}-\omega pX\right)\right]-\omega^2[\delta\rho_f U+\delta\tilde{\rho}W]\\ -\omega p\left[\delta C\frac{\partial U}{\partial z}+\delta M\frac{\partial W}{\partial z}\right]-\omega^2[\delta\rho_f V+\delta\tilde{\rho}X-p^2(\delta CV+\delta MX)]\end{pmatrix}. \tag{27}$$

By inserting these expressions into Eq. (26) and integrating by parts in order to separate the contributions in $\delta\rho$, $\delta\rho_f$, $\delta\tilde{\rho}$, δC, δM, $\delta\lambda_U$, and δG, we obtain an integral representation of the scattered wave field δU that we can directly identify with Eq. (22) to get the Fréchet derivatives $\mathcal{A}_1^{PSV}$, $\mathcal{B}_1^{PSV}$, $\mathcal{C}_1^{PSV}$, $\mathcal{D}_1^{PSV}$, $\mathcal{E}_1^{PSV}$, $\mathcal{F}_1^{PSV}$, and $\mathcal{G}_1^{PSV}$ corresponding to displacement U:

$$\mathcal{A}_1^{PSV}=-\omega^2[UG_{1z}^{1z}+VG_{1z}^{1r}],$$

$$\mathcal{B}_1^{PSV}=-\omega^2[WG_{1z}^{1z}+UG_{1z}^{2z}+VG_{1z}^{2r}],$$

$$\mathcal{C}_1^{PSV}=-\omega^2[WG_{1z}^{2z}+XG_{1z}^{2r}],$$

$$\mathcal{D}_1^{PSV}=\left[\frac{\partial W}{\partial z}-\omega pX\right]\left[\frac{\partial G_{1z}^{1z}}{\partial z}-\omega pG_{1z}^{1r}\right]+\left[\frac{\partial U}{\partial z}-\omega pV\right]\times\left[\frac{\partial G_{1z}^{2z}}{\partial z}-\omega pG_{1z}^{2r}\right],$$

$$\mathcal{E}_1^{PSV} = \left[\frac{\partial W}{\partial z} - \omega p X\right]\left[\frac{\partial G_{1z}^{2z}}{\partial z} - \omega p G_{1z}^{2r}\right],$$

$$\mathcal{F}_1^{PSV} = \left[\frac{\partial U}{\partial z} - \omega p V\right]\left[\frac{\partial G_{1z}^{1z}}{\partial z} - \omega p G_{1z}^{1r}\right],$$

$$\mathcal{G}_1^{PSV} = \left[\frac{\partial V}{\partial z} + \omega p U\right]\left[\frac{\partial G_{1z}^{1r}}{\partial z} + \omega p G_{1z}^{1z}\right] + 2\left[\frac{\partial U}{\partial z}\frac{\partial G_{1z}^{1z}}{\partial z} + \omega^2 p^2 V G_{1z}^{1r}\right]. \quad (28)$$

Here, $U=U(z,\omega;z_S)$, $V=V(z,\omega;z_S)$, $W=W(z,\omega;z_S)$, and $X=X(z,\omega;z_S)$ denote the incident wave fields at the model perturbations. These wave fields can be expressed in terms of Green's functions using Eq. (13) for a vertical point force, and similarly for a horizontal point force or for an explosive point source. Expressions $G_{ij}^{kl}=G_{ij}^{kl}(z_R,\omega;z)$ represent the Green's functions conveying the scattered wave fields from the inhomogeneities to the receivers, as noted before. A total of 32 Green's functions (16 for upgoing waves and another 16 for downgoing waves) are to be computed to completely solve these equations.

The Fréchet derivatives for the radial displacement V are easily deduced from expression (28) by changing G_{1z}^{kl} to G_{1r}^{kl}. In the same way, the Fréchet derivatives corresponding to the vertical and radial relative fluid-to-solid displacements are obtained by substituting G_{2j}^{kl} for G_{1j}^{kl}. As a first verification of the formulas, we note that the expressions corresponding to perturbations of parameters ρ, λ_U, and G ($\mathcal{A}_i^{PSV}$, $\mathcal{F}_i^{PSV}$, and $\mathcal{G}_i^{PSV}$, respectively) are in perfect agreement with the Fréchet derivatives formulas computed for the density ρ and Lamé parameters λ and μ in the elastic case (Dietrich and Kormendi, 1990).

C. Fréchet derivatives for relevant parameters

As mentioned earlier, the set of seven parameters (ρ, ρ_f, $\tilde{\rho}$, C, M, λ_U, and G) used in the perturbation analysis was primarily chosen to considerably simplify the derivation of expression (28). In practice, and as with the Lamé parameters in the elastic case, it is more convenient to consider model parameters which are easier to measure or estimate. We introduce here a new set of eight parameters, namely ρ_s, ρ_f, k_0, ϕ, K_f, K_s, G_s, and c_s which are more naturally related to the solid and fluid phases. Furthermore, these parameters are independent from each other in terms of their mechanical or hydrological meaning. They are all real and frequency independent, contrary to $\tilde{\rho}$. Fluid viscosity is not considered here as the Fréchet derivatives with respect to the fluid viscosity and to the permeability are strictly proportional. We concentrate on the permeability by stressing that the comments made for the permeability will also be valid for the fluid viscosity.

To obtain the Fréchet derivatives corresponding to the new set of parameters from expression (28), we construct the 8×7 Jacobian matrix $\mathbf{J}$ whose coefficients are formally defined by

$$J_{ij} = \frac{\partial p_i}{\partial \hat{p}_j}, \quad (29)$$

where $p_i, i=1,\ldots,7$, stands for one of the parameters ρ, ρ_f, $\tilde{\rho}$, C, M, λ_U, or G, and where $\hat{p}_j, j=1,\ldots,8$, represents one of the parameters ρ_s, ρ_f, k_0, ϕ, K_s, K_f, G_s, or c_s. The elements of matrix $\mathbf{J}$ are determined from Eqs. (2)–(6).

The Fréchet derivatives with respect to the new set of parameters for the vertical component of the solid displacement U are then defined by

$$[\hat{\mathcal{A}}_i^{PSV}\ \hat{\mathcal{B}}_i^{PSV}\ \hat{\mathcal{C}}_i^{PSV}\ \hat{\mathcal{D}}_i^{PSV}\ \hat{\mathcal{E}}_i^{PSV}\ \hat{\mathcal{F}}_i^{PSV}\ \hat{\mathcal{G}}_i^{PSV}\ \hat{\mathcal{H}}_i^{PSV}]^T = \mathbf{J}[\mathcal{A}_i^{PSV}\ \mathcal{B}_i^{PSV}\ \mathcal{C}_i^{PSV}\ \mathcal{D}_i^{PSV}\ \mathcal{E}_i^{PSV}\ \mathcal{F}_i^{PSV}\ \mathcal{G}_i^{PSV}]^T. \quad (30)$$

The combinations involved in Eq. (30) significantly complicate the expressions of the Fréchet derivatives corresponding to the new set of parameters. For sake of clarity and simplification, we introduce the following quantities:

$$\Omega = \sqrt{n_J - 4i\frac{\omega}{\omega_c}},$$

$$a = (1+c_s\phi)K_s + c_s(1-\phi)K_f,$$

$$b = \phi^2\{4(2+3c_s)G_s a^2 - 3(2+3c_s\phi)^2 K_s(1+c)[K_s^2 - c_s K_f^2 + (c_s-1)K_s K_f]\},$$

$$c = \phi^2[(2+3c_s\phi)^2 K_s(K_s-K_f)^2 - 4a^2 G_s]. \quad (31)$$

With these parameters, the final expressions of the Fréchet derivatives $\hat{\mathcal{A}}_1^{PSV}$, $\hat{\mathcal{B}}_1^{PSV}$, $\hat{\mathcal{C}}_1^{PSV}$, $\hat{\mathcal{D}}_1^{PSV}$, $\hat{\mathcal{E}}_1^{PSV}$, $\hat{\mathcal{F}}_1^{PSV}$, $\hat{\mathcal{G}}_1^{PSV}$, and $\hat{\mathcal{H}}_1^{PSV}$ with respect to model parameters ρ_s, ρ_f, k_0, ϕ, K_s, K_f, G_s, and c_s are

$$\hat{\mathcal{A}}_1^{PSV} = (1-\phi)\mathcal{A}_1^{PSV},$$

$$\hat{\mathcal{B}}_1^{PSV} = \phi\mathcal{A}_1^{PSV} + \mathcal{B}_1^{PSV} + \frac{2+\Omega}{\Omega}F\mathcal{C}_1^{PSV},$$

$$\hat{\mathcal{C}}_1^{PSV} = -\frac{\left(in_J + 2\frac{\omega}{\omega_c}\right)\eta}{k_0^2\Omega\omega}\mathcal{C}_1^{PSV},$$

$$\hat{\mathcal{D}}_1^{PSV} = \frac{1}{a^2\phi^2(2+3c_s\phi)^2}\Big\{a^2\phi^2(\rho_s-\rho_f)(2+3c_s\phi)^2\mathcal{A}_1^{PSV} - 2a^2\phi^2(2+3c_s)G_s\mathcal{G}_1^{PSV} - (2+3c_s\phi^2)[(1+c_s\phi)^2K_s + c_s(1-2\phi-c_s\phi^2)K_f]K_sK_f\mathcal{E}_1^{PSV} - \phi^2c_s(1+c_s)(2+3c_s\phi)^2(K_s-K_f)K_sK_f\mathcal{D}_1^{PSV} + \frac{1}{3}b\mathcal{F}_1^{PSV}\Big\},$$

$$\hat{\mathcal{E}}_1^{PSV} = \frac{1-\phi}{a^2\phi}\Big\{\phi c_s(1+c_s)K_f^2\mathcal{D}_1^{PSV} + c_s(1+c_s\phi)K_f^2\mathcal{E}_1^{PSV} + \phi[(1+c_s\phi)K_s^2 + c_s(c_s+\phi)K_f^2 + 2c_s(1-\phi)K_sK_f]\mathcal{F}_1^{PSV}\Big\},$$

$$\hat{\mathcal{F}}_1^{PSV} = \frac{1}{a^2\phi}\{(1+c_s\phi)K_s^2[\phi(1+c_s)\mathcal{D}_1^{PSV} + (1 + c_s\phi)\mathcal{E}_1^{PSV}] + [\phi(c_s+\phi)(1+c_s\phi) - c_s(1-\phi)^2]K_s^2\mathcal{F}_1^{PSV}\},$$

$$\hat{\mathcal{G}}_1^{PSV} = 2\frac{1-\phi}{2+3c_s\phi}\left[-\frac{2}{3}\mathcal{F}_1^{PSV} + \mathcal{G}_1^{PSV}\right],$$

$$\hat{\mathcal{H}}_1^{PSV} = \frac{1-\phi}{a^2\phi(2+3c_s\phi)^2}\{(2+3c_s\phi)^2K_sK_f[\phi(K_s - K_f)\mathcal{D}_1^{PSV} - K_f\mathcal{E}_1^{PSV}] - c\mathcal{F}_1^{PSV} - 6a^2\phi^2G_s\mathcal{G}_1^{PSV}\}. \quad (32)$$

We note that the Fréchet derivatives $\hat{\mathcal{B}}_1^{PSV}$ and $\hat{\mathcal{C}}_1^{PSV}$ with respect to fluid density ρ_f and permeability k_0 are complex due to the role of these parameters in the attenuation and dispersion of seismic waves.

In addition, formula (32) can be further simplified if source and receivers are located at the same depth $z_0 = z_R = z_S$. In this case, we can take advantage of the representation of the incident wave fields U, V, W, and X in terms of the Green's functions [see Eq. (13)], and use the reciprocity theorem:

$$G_{ij}^{kl}(z_R,\omega;z_S) = G_{kl}^{ij}(z_S,\omega;z_R). \quad (33)$$

The number of Green's functions required to describe the wave propagation then reduces from 32 to 16. These simplifications are straightforward and are not developed here.

D. *SH* case

We follow the same procedure as used earlier to derive the Fréchet derivatives of the solid displacement T and relative fluid-to-solid displacement Y in the *SH* case. We denote by $\mathcal{A}_i^{SH}$, $\mathcal{B}_i^{SH}$, $\mathcal{C}_i^{SH}$, $\mathcal{D}_i^{SH}$, $\mathcal{E}_i^{SH}$, $\mathcal{F}_i^{SH}$, and $\mathcal{G}_i^{SH}$, respectively, the Fréchet derivatives with respect to model parameters ρ, ρ_f, $\tilde{\rho}$, M, K_U, G, and C, where subscript $i=1$ refers to the solid displacement T and subscript $i=2$ refers to the relative fluid-to-solid displacement Y. We also introduce the notations $\hat{\mathcal{A}}_i^{SH}$, $\hat{\mathcal{B}}_i^{SH}$, $\hat{\mathcal{C}}_i^{SH}$, $\hat{\mathcal{D}}_i^{SH}$, $\hat{\mathcal{E}}_i^{SH}$, $\hat{\mathcal{F}}_i^{SH}$, $\hat{\mathcal{G}}_i^{SH}$, and $\hat{\mathcal{H}}_i^{SH}$ for the Fréchet derivatives relative to our alternative set of model parameters ρ_s, ρ_f, k_0, ϕ, K_s, K_f, G_s, and c_s. The perturbation analysis leads to the following expressions of the Fréchet derivatives of the transverse solid displacement T:

$$\mathcal{A}_1^{SH} = -\omega^2 T G_{1t}^{1t},$$

$$\mathcal{B}_1^{SH} = -\omega^2[YG_{1t}^{1t} + TG_{1t}^{2t}],$$

$$\mathcal{C}_1^{SH} = -\omega^2 YG_{1t}^{2t},$$

$$\mathcal{D}_1^{SH} = 0,$$

$$\mathcal{E}_1^{SH} = 0,$$

$$\mathcal{F}_1^{SH} = 0,$$

$$\mathcal{G}_1^{SH} = \frac{\partial T}{\partial z}\frac{\partial G_{1t}^{1t}}{\partial z} + \omega^2 p^2 T G_{1t}^{1t}. \quad (34)$$

As before, the displacements T and Y can be expressed in terms of the Green's functions G_{it}^{jt}, where subscripts $i=1$ and $j=1$ relate to solid displacement and force, whereas subscripts $i=2$ and $j=2$ relate to relative fluid-to-solid displacement and force. Subscript t stands for the tangential displacement or force. The transformation of these expressions with the Jacobian matrix $\mathbf{J}$ defined in Eq. (29) finally yields

$$\hat{\mathcal{A}}_1^{SH} = (1-\phi)\mathcal{A}_1^{SH},$$

$$\hat{\mathcal{B}}_1^{SH} = \phi A_1^{SH} + \mathcal{B}_1^{SH} + \frac{2+\Omega}{\Omega}F\mathcal{C}_1^{SH},$$

$$\hat{\mathcal{C}}_1^{SH} = -\frac{\left(in_J + 2\dfrac{\omega}{\omega_c}\right)\eta}{k_0^2\Omega\omega}\mathcal{C}_1^{SH},$$

$$\hat{\mathcal{D}}_1^{SH} = (\rho_s - \rho_f)\mathcal{A}_1^{SH} - 2G_s\frac{2+3c_s}{(2+3c_s\phi)^2}\mathcal{G}_1^{SH},$$

$$\hat{\mathcal{E}}_1^{SH} = 0,$$

$$\hat{\mathcal{F}}_1^{SH} = 0,$$

$$\hat{\mathcal{G}}_1^{SH} = 2\frac{1-\phi}{2+3c_s\phi}\mathcal{G}_1^{SH},$$

$$\hat{\mathcal{H}}_1^{SH} = -6G_s\phi\frac{1-\phi}{(2+3c_s\phi)^2}\mathcal{G}_1^{SH}. \quad (35)$$

We note that the Fréchet derivatives with respect to K_s and K_f are zero since these parameters have no influence on shear waves.

IV. NUMERICAL SIMULATIONS AND ACCURACY TESTS

A. Fréchet derivatives versus discrete perturbations

In our numerical applications, the Fréchet derivatives are first calculated in the frequency–wave number domain before being transformed into the time–distance domain with the discrete wave number integration method (Bouchon, 1981).

In order to test our analytical formulations and assess their limitations, we compare the differential seismograms computed with the first-order Fréchet derivative approach with the seismograms obtained by introducing discrete perturbations in the medium properties. Thus, considering for instance the vertical component of the solid displacement in the $P-SV$ case, the partial derivative $\partial U/\partial p_j$ with respect to parameter p_j can be approximated by the following finite difference expression:

TABLE I. Model parameters of the infinite medium used for the numerical tests of the Fréchet derivatives.

ϕ ( )	k_0 (m^2)	ρ_f (kg/m^3)	ρ_s (kg/m^3)	K_s (GPa)	G_s (GPa)	K_f (GPa)	c_s ( )
0.20	10^{-12}	1000	2700	35	25	2.2	50

$$\frac{\partial U(z_R,\omega;z_S)}{\partial p_j} \simeq \frac{U^{p_j+\Delta p_j}(z_R,\omega;z_S) - U^{p_j}(z_R,\omega;z_S)}{\Delta p_j}, \quad (36)$$

where Δp_j represents a small perturbation of parameter p_j.

The similarity between the seismograms computed with the two approaches is evaluated from the correlation coefficients between the traces.

B. Uniform medium

We first consider the simple case of small perturbations $\delta\rho_s$,$\delta\rho_f$, δk_0, $\delta\phi$, δK_s, δK_f, δG_s, and δc_s within a thin slab embedded in an infinite uniform medium. The slab thickness is of the order of one-twentieth of the dominant wavelength of the P waves (i.e., 1 m). The amplitude of the relative perturbations $\Delta p_j/p_j$ is 10% for each of the parameters considered. Source and receivers are located at the same depth, 50 m above the model perturbation. The parameters of the uniform model are listed in Table I.

The simulations shown in Fig. 1 include P_{fast}, P_{slow}, and S waves whose computed velocities are respectively equal to 2250, 130, and 750 m/s at a frequency of 85 Hz. The slow P waves are not visible, but three reflected waves (compressional PP, converted PS and SP, and shear SS) are easily identified in the four sections displayed in Fig. 1. It is seen that a small perturbation of the fluid modulus K_f has no influence on shear waves. The same behavior is observed for the solid modulus K_s (not shown). On the contrary, slight changes in the other parameters mainly generate SS reflections, as noted in particular for the porosity ϕ, mineral shear modulus G_s, and permeability k_0 in Fig. 1. We also observed that the differential seismograms are very similar for the following pairs of perturbations: (i) consolidation parameter and porosity; (ii) fluid and solid moduli; and (iii) fluid and solid densities. In addition, we found that the correlation coefficients between the Fréchet derivative and discrete perturbation seismograms are greater than 99% for most model parameters at all source–receiver offsets. The only exception concerns the Fréchet derivative with respect to the permeability which shows correlation coefficients between 60% and 95% in the P–SV case, depending on the source–receiver offset. However, this operator appears more accurate in the SH case. By and large, the tests performed in a uniform medium validate our analytical expressions derived in Secs. III C and III D.

We further check the accuracy and stability of the first-order sensitivity operators by modifying the amplitude of the discrete perturbations, with the following results: (i) the Fréchet derivatives with respect to parameters that only influence P waves are more accurate than the Fréchet derivatives with respect to parameters that influence both P and S waves. (ii) Strong perturbations of the solid and fluid moduli K_s and K_f do not produce any distortion of the wave forms, but merely result in a global increase of the amplitudes of the discrete perturbation seismograms. (iii) For strong amplitude perturbations, the Fréchet derivatives are more stable at near offsets (i.e., for small angles of incidence) than at large offsets. We interpret this observation as being due to the nonlinearity inherent to large offsets where the wave fields interact more strongly with the subsurface structure. However, as an exception to this rule, the Fréchet derivatives with respect to permeability appear more stable at large source–receiver offsets. This may be explained by the fact that a perturbation of the permeability mainly influences the wave attenuation and dispersion and therefore has a stronger effect for longer travel paths. The deterioration of the seismograms correlation with decreasing k_0 is another indication of the weak

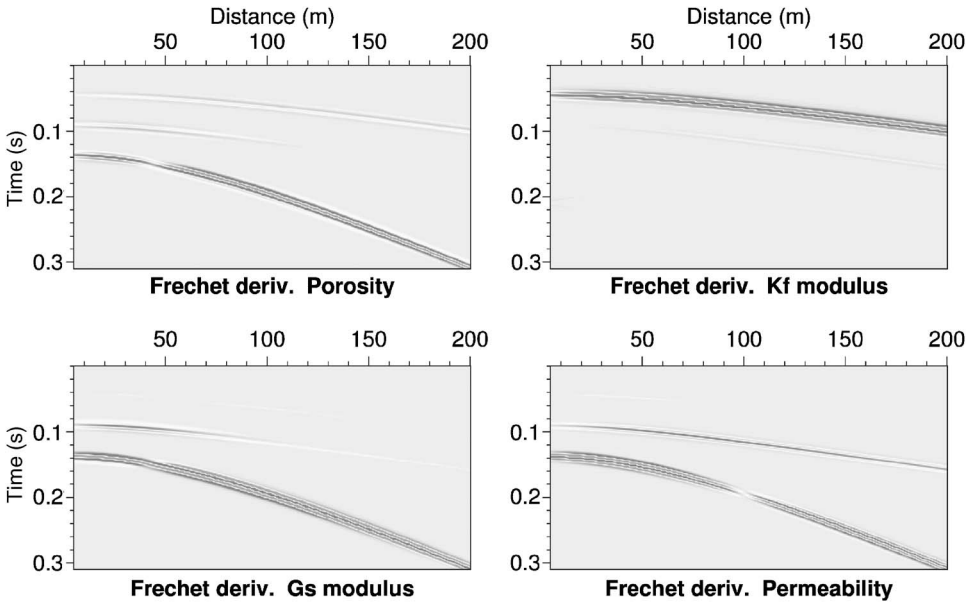


FIG. 1. Seismic sections corresponding to the first-order perturbation of the vertical displacement with respect to porosity ϕ, fluid modulus K_f, mineral shear modulus G_s, and permeability k_0 in the uniform and infinite medium described in Table I. The seismic excitation is a vertical point force.

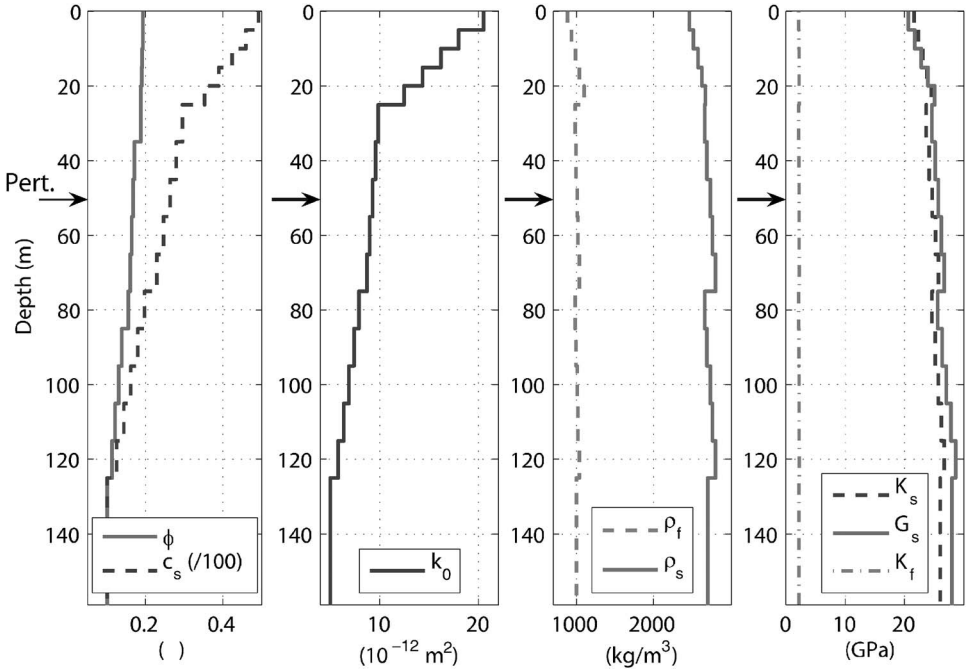


FIG. 2. Sixteen-layer model used for the numerical tests of the Fréchet derivative formulation in a stratified medium. The distributions of the eight model parameters as a function of depth are shown.

influence of the permeability on the seismograms. (iv) As a general rule, the first-order approximations appear remarkably accurate for amplitude perturbations up to 20% in absolute value.

We now consider the robustness of the Fréchet derivative seismograms with respect to the thickness of the perturbed layer. In our uniform model, the wavelengths λ_P and λ_S corresponding to P and S waves are respectively equal to 26 and 9 m at the dominant frequency of the Ricker wavelet used in the simulation. Our computations show that the Fréchet derivative seismograms are very well correlated with the discrete perturbation seismograms until the thickness of the perturbed layer reaches about 20% of the dominant wavelength λ_P (that is, $\simeq$5 m) for parameters K_s and K_f, and 20% of the dominant wavelength λ_S (that is, $\simeq$2 m) for all other parameters that influence both P and S waves. Thus, we observe that the Fréchet derivatives with respect to K_s and K_f are more robust than the other expressions with respect to departures from the "small and localized perturbation" assumption of the Born approximation.

C. Complex model

We used the 16-layer model shown in Fig. 2 to numerically check the stability and accuracy of the Fréchet derivative formulas in a more complex structure. In this model, the perturbed layer is at a depth of 50 m, source and receivers being located near the surface. Figure 3 presents the seismic sections obtained with the Fréchet derivative and discrete perturbation methods for slight changes of the solid density ρ_s. On the whole, the comparison of the wave forms obtained with both methods is very satisfactory.

Figure 4 shows that the Fréchet derivative seismograms are remarkably accurate as long as the perturbation amplitude remains small. In the case considered, the maximum acceptable perturbation amplitude is approximately 10% of the model parameter value. When the perturbation amplitude is increased beyond this limit, the P wave forms remain practically unchanged whereas the S wave forms are distorted.

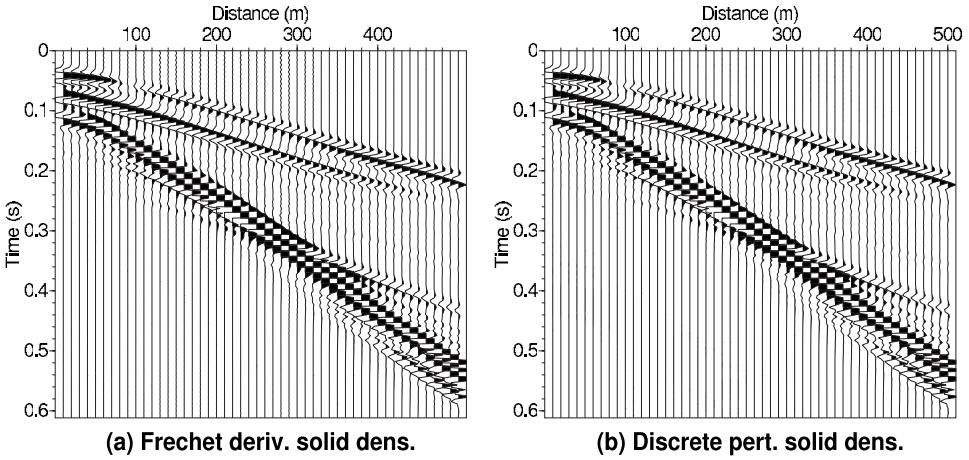


FIG. 3. (a), (b) Seismic sections obtained with the Fréchet derivatives and with the discrete perturbation methods for a perturbation of the solid density ρ_s at z=50 m in the 16-layer model depicted in Fig. 2. The seismograms represent the vertical displacement generated by a vertical force (P-SV wave system).

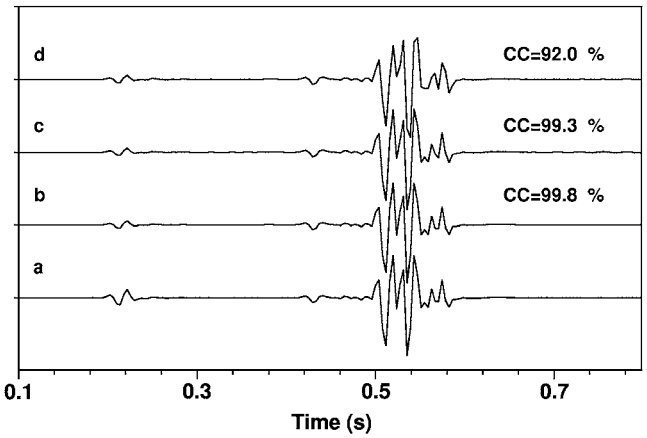


FIG. 4. Detailed comparison of individual seismic traces obtained with the Fréchet derivative and with the discrete perturbation methods for various perturbation amplitudes. The model used for the computations is the 16-layer model presented in Fig. 2. The seismic source is a vertical point force. The perturbation of the vertical displacement with respect to the consolidation parameter *cs* for a source–receiver offset of 500 m is shown. The seismogram labeled (a) corresponds to the Fréchet derivative formulation. The traces labeled (b), (c), and (d) are respectively associated with discrete perturbations of the model parameter *cs* with relative amplitude contrasts of 1%, 10%, and 50%. The similarity between the Fréchet derivative and discrete perturbation seismograms is quantified by their correlation coefficient, which is indicated above the traces.

We also checked the behavior of the Fréchet derivative operators relative to the thickness of the perturbation layer. As for the uniform medium investigated before, the first-order approximations remain very accurate as long as the layer thickness layer is lower than $\lambda_P/5$ or $\lambda_S/5$ depending on the model parameter considered.

V. SENSITIVITY STUDY

A. Relative influence of the model parameters

In this section, we assess the relative influence of small modifications of the model parameters on the seismic wave field. For this, we determine the maximum amplitude of the seismograms computed with the discrete perturbation method for all pairs of vertical and horizontal forces and displacements. These computations are done with the model parameters of Table I, aside from the consolidation parameter c_s which is given a value of 20 in Fig. 5 and 100 in Fig. 6. Perturbation depth, source and receiver locations, and maximum offset are identical to those used in Sec. IV B. The amplitudes thus obtained are multiplied by the parameter variation (Δp) to obtain the displacement change ΔU of Eq. (36), and are normalized with respect to the maximum value found. The same quantities were computed with the Fréchet derivative approach to check the agreement between the two computation techniques. Figures 5 and 6 show that the seismograms are essentially sensitive to porosity ϕ in the uniform medium considered. The seismograms are also strongly influenced by perturbations of the consolidation parameter c_s, mineral density ρ_s, and shear modulus G_s. On the other hand, changes in fluid density ρ_f, mineral modulus K_s, fluid modulus K_f, and permeability k_0 have only a weak influence on the wave amplitudes. We also note that the influence of

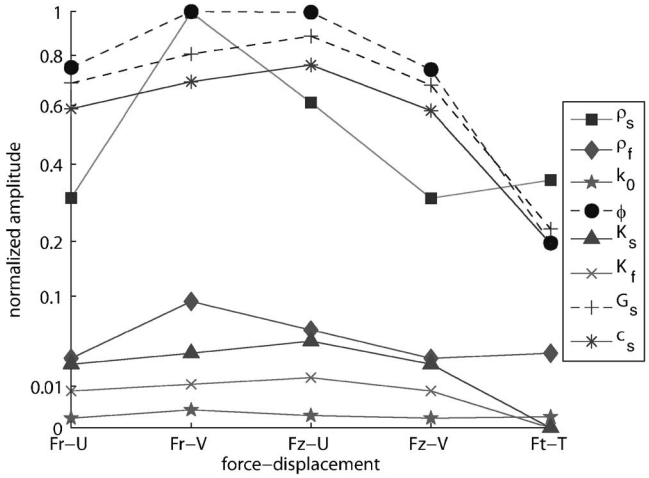


FIG. 5. Normalized maximum amplitude of the Fréchet derivative seismograms for all solid force and displacement pairs. F_z, F_r, and F_t respectively denote vertical, horizontal radial, and horizontal tangential point forces. U, V, and T are the vertical, horizontal radial, and horizontal tangential components of the solid displacement. The model parameters are those given in Table I except for the consolidation parameter c_s, which is equal to 20. The amplitude scale is nonlinear for sake of readability.

parameters c_s and G_s on the one hand, and ρ_s and ρ_f on the other hand, are very similar for the various force-displacement pairs.

For the model with the lowest value of the consolidation parameter (corresponding to the most consolidated material, Fig. 5), the solid modulus K_s has a stronger influence on the seismograms than the fluid modulus K_f. On the contrary, for an unconsolidated medium (Fig. 6), the seismograms are mainly influenced by the fluid properties. With the model parameters used in this study, we find that the transition between these two regimes occurs for a consolidation parameter of 35. We also verified, as suggested by our observations in Sec. IV B, that the seismogram amplitudes in a high porosity medium depend more strongly on the fluid modulus than on the solid modulus, and vice versa for a medium with low porosity. This confirms the field observations that the P waves are strongly influenced by the fluid properties when they propagate in a fluid-saturated and poorly consolidated medium with high porosity. This makes it possible to deter-

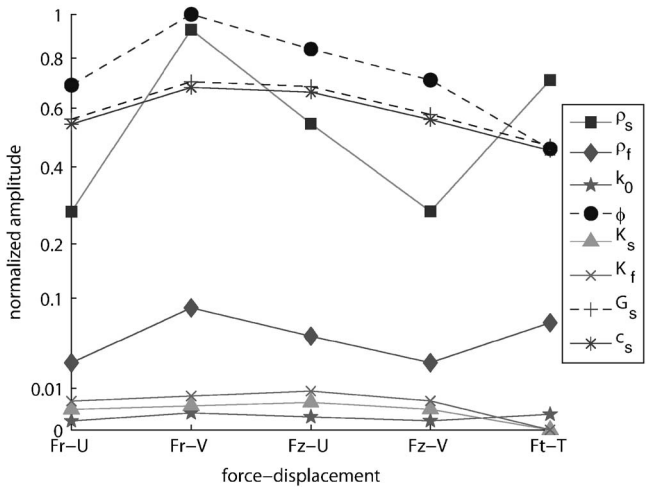


FIG. 6. Same as Fig. 5 with consolidation parameter c_s equal to 100.

mine the fluid characteristics from the seismic wave forms if these favorable conditions are met. Another consequence of the results shown in Figs. 5 and 6 is that ϕ and c_s are the most attractive parameters to invert for in an inversion procedure, assuming that ρ_s and G_s can be estimated independently.

B. Amplitude of the perturbation seismograms versus angle of incidence

The previous figures already stressed the interdependence (or coupling) of some model parameters. Parameter coupling means that small perturbations of two or more parameters result in similar modifications of the seismic response. An obvious consequence of parameter coupling is that it becomes difficult or even impossible to reliably estimate the model parameters in an inversion procedure (Tarantola, 1986).

To look into this problem, we computed the plane–wave responses corresponding to 10% perturbations of the model properties in the infinite medium described in Table I. Figure 7 shows the reflected energy for the eight model parameters as a function of angle of incidence at the perturbed layer, both for the *PP* and *SS* reflections. The smooth aspect of the curves is due to the intrinsic attenuation of the seismic waves propagating in the porous medium. The peaks and troughs of the curves are explained by the strong variations of the reflection and transmission coefficients, as shown for instance by de la Cruz *et al.* (1992). Some of these rapid variations are seen on the seismograms of Fig. 1 which were computed with the same model and source–receiver configuration.

The magnitudes of the seismic responses shown in Fig. 7 are consistent with the study presented in Sec. V A. *SS* reflections are about twice as large as *PP* reflections for perturbations of the solid and fluid densities. For *SS* reflections, the maximum value is reached for a perturbation in porosity ϕ, whereas G_s, c_s, and ρ_s produce perturbations of the same magnitude. For *PP* reflections, the most influential parameters are the solid density ρ_s and porosity ϕ.

We also see in Fig. 7 that the radiation patterns associated with perturbations in ρ_s, ρ_f, and k_0 on the one hand, and K_s and K_f on the other hand are exactly the same for the *PP* reflections. The same resemblance is observed for the *SS* reflections for the ρ_s, ρ_f, and k_0 group of parameters, and for the G_s, ϕ, and c_s group of parameters. We note in particular that permeability k_0 and densities ρ_s and ρ_f behave similarly despite their different roles in the constitutive equations. In all cases, the backscattered energy is maximum at normal incidence because of the shorter wave path and corresponding minimal attenuation.

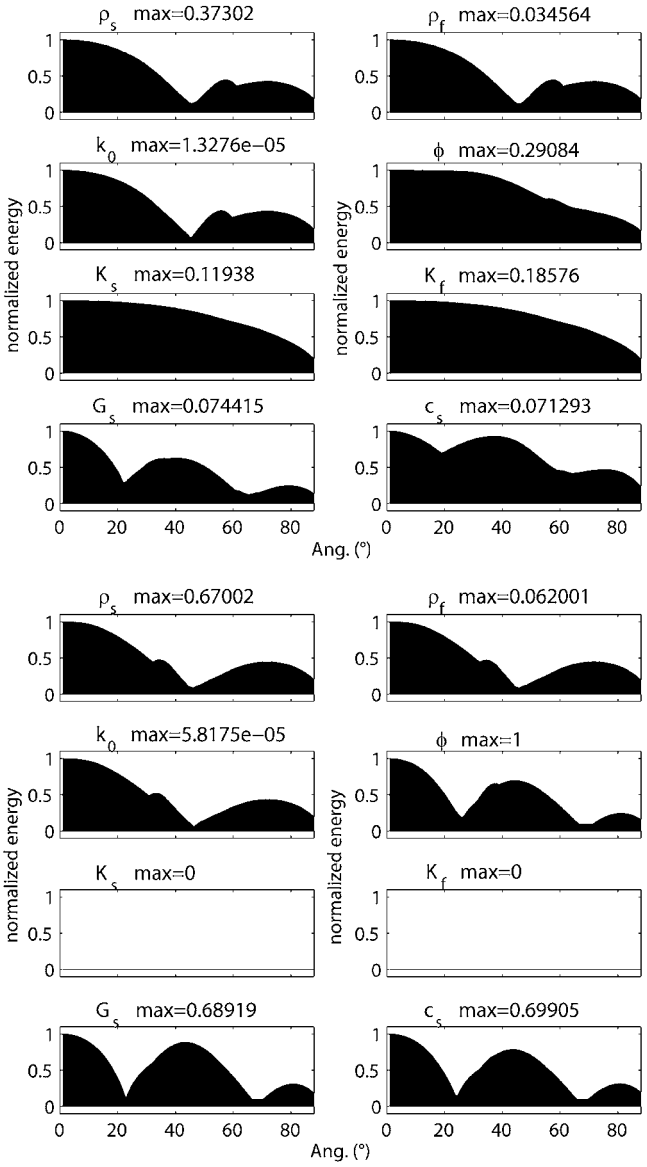


FIG. 7. Energy of plane waves reflected from perturbations in ρ_s, ρ_f, k_0, ϕ, K_s, K_f, G_s, and cs, as a function of incidence angle. The eight upper panels and eight lower panels respectively correspond to *PP* and *SS* reflections. The curves are normalized with respect to the maximum value indicated above each panel.

VI. CONCLUSIONS

We derived the Fréchet derivatives of the seismic response of a depth-dependent porous medium. The Fréchet derivatives are analytically expressed in terms of the Green's functions of the propagation medium through a perturbation analysis of the poroelastic wave equations expressed in the plane wave domain. Started with a primary set of seven model parameters chosen because of their linear relationship with the wave equations, the derivation was carried on with a secondary set of eight model parameters more convenient to use as physical parameters of the problem. The eight model parameters considered in our analysis are related to the fluid properties (density, bulk modulus), to the mineral properties (density, bulk modulus, shear modulus) and to the arrangement of the porous material (porosity, permeability, and consolidation parameter). In the *P*–*SV* case, we derived Fréchet derivatives for three different sources (horizontal and vertical point forces and explosive point source), four displacement components, and eight model parameters. In the *SH* case, we obtained twelve expressions for one horizontal point force, two displacement components, and six model parameters leading to nonzero Fréchet derivatives.

We checked the accuracy of these sensitivity operators in the time–distance domain by comparing the wave forms

computed from the first-order expressions with seismograms obtained by introducing discrete perturbations in the medium properties. The numerical tests were carried out both in a homogeneous and in a more complex earth model excited by oriented point forces. By and large, we found that our analytical expressions of the Fréchet derivatives are remarkably accurate as long as the Born approximation assumptions are satisfied, that is, as long as the perturbations of the model parameters are weak and localized. However, as in other studies relying on the Born approximation, we showed that the first-order operators are robust enough to model parameter perturbations up to 20% and layer thicknesses up to one-fifth of the dominant wavelength. Furthermore, our formulation appears to be stable at all source–receiver offsets.

Due to their analytical formulation, the sensitivity operators derived in this paper will be especially useful in full wave form inversion algorithms implemented with gradients techniques. As a first step toward such an application, we evaluated the sensitivity of the seismic response of a poroelastic medium with respect to each model parameter. We showed that the porosity and consolidation parameter are the most attractive parameters to invert for, whereas the permeability appears to be the most difficult parameter to determine. The wave fields are more sensitive to the fluid bulk modulus than to the mineral bulk modulus, or inversely, according to porosity and consolidation parameter values. Our sensitivity study also showed that a multiparameter inversion of backscattered energy looks challenging because of the strong coupling of several model parameters in a wide range of angles of incidence. Nevertheless, the formulas obtained should prove useful for the interpretation of time-lapse monitoring surveys and for checking solutions (yet to come) accounting for three-dimensional heterogeneities in the propagation medium.

ACKNOWLEDGMENTS

We are grateful to Stéphane Garambois, Michel Bouchon, Jean-Louis Auriault, Patrick Rasolofosaon, and Helle Pedersen for many helpful discussions in the course of this work. The numerical applications were performed by using the computer facilities of the Grenoble Observatory. We thank the Associate Editor and two anonymous reviewers for their useful comments.

Aki, K., and Richards, P. (**1980**). *Quantitative Seismology* (Freeman, New York).

Auriault, J. L., Borne, L., and Chambon, R. (**1985**). "Dynamics of porous saturated media, checking of the generalized law of Darcy," J. Acoust. Soc. Am. **77**, 1641–1650.

Bemer, E., Vincké, O., and Longuemare, P. (**2004**). "Geomechanical log deduced from porosity and mineralogical content," Oil Gas Sci. Technol. **59**, 405–426.

Berryman, J. G., Berge, P., and Bonner, B. (**2002**). "Estimating rock porosity and fluid saturation using only seismic velocities," Geophysics **67**, 391–404.

Biot, M. A. (**1956**). "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range, II. Higher frequency range," J. Acoust. Soc. Am. **28**, 168–191.

Bouchon, M. (**1981**). "A simple method to calculate Green's functions for elastic layered media," Bull. Seismol. Soc. Am. **71**, 959–971.

Carcione, J. M. (**1996**). "Wave propagation in anisotropic, saturated porous media: Plane wave theory and numerical simulation," J. Acoust. Soc. Am. **99**, 2655–2666.

Chotiros, N. (**2002**). "An inversion for Biot parameters in a water-saturated sand," J. Acoust. Soc. Am. **112**, 1853–1868.

Dai, N., Vafidis, A., and Kanasewich, E. (**1995**). "Wave propagation in heteregeneous porous media: A velocity-stress, finite-difference method," Geophysics **60**, 327–340.

de la Cruz, V., Hube, J., and Spanos, T. (**1992**). "Reflection and transmission of seismic waves at the boundaries of porous media," Wave Motion **16**, 323–338.

Dietrich, M., and Kormendi, F. (**1990**). "Perturbation of the plane-wave reflectivity of a depth-dependent elastic medium by weak inhomogeneities," Geophys. J. Int. **100**, 2003–214.

Dutta, N., and Odé, H. (**1979**). "Attenuation and dispersion of compressional waves in fluid-filled porous rocks with partial gas saturation (White model). I. Biot theory," Geophysics **44**, 1777–1788.

Garambois, S., and Dietrich, M. (**2002**). "Full waveform numerical simulations of seismoelectromagnetic wave conversions in fluid-saturated stratified porous media," J. Geophys. Res. **107**, 2148–2165.

Gassmann, F. (**1951**). "On the elasticity of porous media (Über die elastizität poröser medien)," Vierteljahrsschr. Natforsch. Ges. Zur. **96**, 1–23.

Haartsen, M., and Pride, S. (**1997**). "Electroseismic waves from point sources in layered media," J. Geophys. Res. **102**, 745–769.

Hudson, J. (**1969**). "A quantitative evaluation of seismic signals at teleseismic distances. I. Radiation from point sources," Geophys. J. R. Astron. Soc. **18**, 233–249.

Hudson, J., and Heritage, J. (**1981**). "The use of the Born approximation in seismic scattering problems," Geophys. J. R. Astron. Soc. **66**, 221–240.

Johnson, D. L., Plona, T., and Kojima, H. (**1994**). "Probing porous media with first and second sound," J. Appl. Phys. **76**, 104–125.

Kennett, B. (**1983**). *Seismic Wave Propagation in Stratified Media* (Cambridge University Press, Cambridge).

Kennett, B., and Kerry, N. (**1979**). "Seismic waves in a stratified half space," Geophys. J. R. Astron. Soc. **57**, 557–583.

Pan, G., Phinney, R., and Odom, R. (**1988**). "Full-waveform inversion of plane-wave seismograms in stratified acoustic media: Theory and feasability," Geophysics **53**, 21–31.

Pride, S. R. (**1994**). "Governing equations for the coupled electromagnetics and acoustics of porous media," Phys. Rev. B **50**, 15678–15696.

Pride, S. R. (**2005**). *Relationships between Seismic and Hydrological Properties* (Water Science and Technology Library, Springer, Berlin), Chap. 8.

Spikes, K., Dvorkin, J., and Mavko, G. (**2006**). "Rock physics model-based inversion," 76th Meeting of the Society of Exploration Geophysicists, New Orleans, 1–6 October, pp. 1645–1649.

Tarantola, A. (**1984**). "Inversion of seismic reflection data in the acoustic approximation," Geophysics **49**, 1259–1266.

Tarantola, A. (**1986**). "A strategy for nonlinear elastic inversion of seismic reflection data," Geophysics **51**, 1893–1903.

Trifunac, M., ed. (**2006**). "Biot Centennial - Earthquake engineering," Soil Dyn. Earthquake Eng. **26**, 483–724.

Rectangular plate with velocity feedback loops using triangularly shaped piezoceramic actuators: Experimental control performance

Yohko Aoki,[a] Paolo Gardonio,[b] and Stephen J. Elliott[c]
Institute of Sound and Vibration, Highfield, SO17 1BJ, Southampton, United Kingdom

(Received 6 August 2007; revised 13 December 2007; accepted 26 December 2007)

This paper presents experimental results on the implementation of decentralized velocity feedback control on a new smart panel in order to produce active damping. The panel is equipped with 16 triangularly shaped piezoceramic patch actuators along its border and accelerometer sensors located at the top vertex of the triangular actuators. The primary objective of this paper is to demonstrate the vibration and sound radiation control using the new smart panel. Narrow frequency band experimental results highlight that the 16 control units can produce reductions up to 15 dB at resonance frequencies between 100 and 700 Hz in terms of both structural vibration and sound power radiation. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2835663]

PACS number(s): 43.40.Vn, 43.50.Ki [KAC] Pages: 1421–1426

I. INTRODUCTION

Active control of sound radiation from vibrating structures has been an area of intensive research in the past decades. Active structural acoustic control (ASAC) aims to minimize the sound radiation by modifying the response of the structure through structural inputs rather than by exciting the acoustic medium.[1,2] The appeal of ASAC systems is that sensor and actuator transducers are embedded to the radiating structure, so that a lightweight and compact smart structure is achieved. However, so far, the complexity of the sensor and actuator transducers necessary to implement ASAC has prevented the development of a practical smart panel.

During the past five years, an alternative approach has been suggested, where the sensor and actuator transducer are arranged in closely located pairs for the implementation of decentralized velocity feedback control loops.[3] In this case, the system achieves active vibration control at the error sensor positions. However, provided moderated feedback control gains are implemented, active damping can be generated on the whole panel, which reduces the response and sound radiation of well separated resonances of low order modes. Thus, the arrays of decentralized velocity feedback loops can be effectively used to reduce the vibration and sound radiation induced by broadband disturbances at low audio frequencies.

Recently Gardonio *et al.*[4] have carried out analytical and experimental studies regarding the suppression of the structural vibration and sound radiation from a flat panel using 16 square piezoceramic actuators and closely located accelerometer sensor pairs distributed over the whole panel, as shown in Fig. 1(a). These 16 sensor–actuator pairs are used to implement decentralized velocity feedback control loops, which significantly reduce the vibration and sound radiation of the panel at low frequency resonances.

This paper presents experimental work on a novel configuration[5] using 16 triangularly shaped piezoceramic actuators arranged along the borders of the panel, and accelerometer sensors on the top vertices of the actuators, as illustrated in Fig. 1(b). This arrangement has the advantage in that it is not invasive, and thus this configuration could be used in a wider class of applications, including the windows of transportation vehicles, such as aircrafts, helicopters, cars, trucks, coaches, etc. A theoretical study[6] has shown that this sensor–actuator pair enables the implementation of stable feedback control loops with rather high gain.

This paper is organized in four sections. Following this introduction section, Sec. II describes the experimental setup used to assess the control performance of the new smart panel with triangular piezoceramic actuators. The stability properties of the control systems are discussed using the generalized Nyquist criteria. Section III presents the experimental results of both vibration and sound radiation reductions. Finally in Sec. IV, the principal characteristics of the control performance are reviewed to highlight the advantages of the proposed control arrangement.

II. EXPERIMENTAL SETUP

A. Multi-input–multi-output system

The smart panel built for this study consists of a rectangular aluminum panel with 16 control units. As shown in Fig. 2, the smart panel is clamped by a rigid aluminum frame, which is mounted on the open side of a Perspex box in order to measure the sound radiation by the panel. The box is composed of 30 mm thick Perspex plates, so that the sound transmission through the lateral walls of the box is much lower than that through the smart panel in the frequency range of interest.[4] The smart panel is excited either by an acoustic field generated by a loudspeaker placed in the box (see Fig. 2 and Table I), or by a point force generated by a shaker (LDS v201) acting to the aluminum panel (see Fig. 1 and Table I).

[a]Electronic mail: ya@isvr.soton.ac.uk
[b]Electronic mail: pg@isvr.soton.ac.uk
[c]Electronic mail: sje@isvr.soton.ac.uk

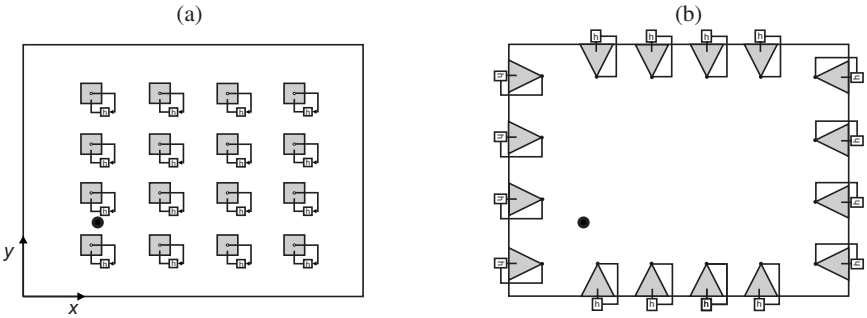


FIG. 1. Smart panels with 16 decentralized velocity feedback control units using (a) square piezoceramic patch actuators and (b) triangularly shaped piezoceramic patch actuators. The position of the primary force excitation is shown as a solid circle.

Each control unit is composed of an isosceles triangular piezoceramic patch actuator and a high sensitivity inductively coupled plasma accelerometer error sensor (PCB Piezotronics, model 352C67). All 16 piezoceramic patches were made of Pz27 material, produced by E.P. Electronic Component LTD. The patch actuators are bonded to the inner side of the panel via a thin layer of nonconductive glue to ensure that the piezoceramic transducers and the plate are not connected electrically. The error sensors are located on the opposite side of the panel at the top vertices of the triangular actuators. As illustrated in Fig. 1(b), these 16 piezoceramic patches are evenly distributed along the perimeter of the panel, with the base edges aligned along the border of the panel. The physical properties and geometry of the test rig and the piezoceramic actuators are summarized in Tables I and II, respectively.

When the feedback control loops are closed, the error sensor output signals are fed to analog integrators with an identical amplification gain in order to provide the feedback controllers with velocity signals. These integrated signals are then again amplified by charge amplifiers with an identical amplification gain in order to drive the piezoceramic actuators with the magnified velocity signals. As shown in Fig. 3, this decentralized multiple-input–multiple-output (MIMO) control system can be formulated in terms of a classic disturbance rejection block diagram. Provided that the control system is stable, the vector of the spectra for the residual signals at the sensors output $\mathbf{y}(j\omega)=[y_1 y_2 \cdots y_{16}]^T$ is given by

$$\mathbf{y}(j\omega) = [\mathbf{I} + \mathbf{G}(j\omega)\mathbf{H}(j\omega)]^{-1}\mathbf{G}_p(j\omega)f_p(j\omega), \quad (1)$$

where ω is the circular frequency, $j=\sqrt{-1}$, and $\mathbf{I}$ is the 16 by 16 identity matrix. $\mathbf{G}(j\omega)$ is a fully populated matrix with the transfer functions between the sensor signals $\mathbf{y}(j\omega)|_{fp=\mathbf{0}}$, whereas no primary excitation is applied to the panel, and the actuator control signals $\mathbf{u}(j\omega)=[u_1 u_2 \cdots u_{16}]^T$. $\mathbf{G}_p(j\omega)$ is a vector of the transfer functions between the sensor outputs $\mathbf{y}(j\omega)|_{H=\mathbf{0}}$, whereas no control excitation is applied to the panel, and the primary disturbance input signal $f_p(j\omega)$, which is either the driving voltage of the loudspeaker V_{in} or the applied force to the panel generated by the shaker F_{in}.

For decentralized control, $\mathbf{H}(j\omega)$ is a diagonal matrix of fixed control gains for each control unit, including the transfer function of the analog controller. In this study the decentralized controller has the same gain h_0 in each control loop. Thus, the feedback transfer matrix $\mathbf{H}(j\omega)$ is given by $\mathbf{H}(j\omega)=h_0 H_T(j\omega)\mathbf{I}$, where $H_T(j\omega)$ is the transfer function of the analog controller. The identical control gain h_0 is chosen to be the highest possible value for which the system remains stable.

B. MIMO stability

Velocity feedback control using collocated and dual sensor–actuator is unconditionally stable, even for multiple channels.[3,7] In this case, the fixed feedback gains can, in principle, be increased without limit, so that the signals from the error sensors can be driven to zero. However, as discussed by Gardonio *et al.*,[5] the triangularly shaped piezoceramic actuator and the velocity sensor pairs considered in this study are strictly not dual or collocated, such that the MIMO decentralized control system is only conditionally stable. Therefore, the control performance is limited by the maximum stable control gain that can be implemented in the feedback loops.

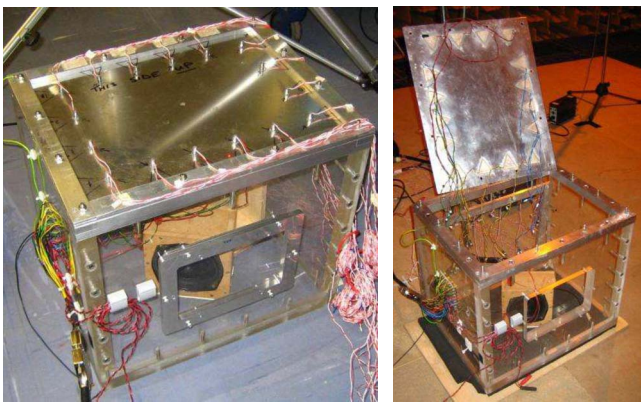


FIG. 2. (Color online) Smart panel mounted on a Perspex box.

TABLE I. Geometry and physical properies of the smart panel and the Perspex box.

Parameter	Values
Dimension (mm)	414×314
Thickness (mm)	1.0
Density (kg/m^3)	2700
Young's modulas (GPa)	70
Poisson's ratio	0.33
Loss factor[a]	0.05
Loudspeaker location	Box
Shaker position (x,y) (mm)	(86.1, 111.4)
Inner dimension (mm)	414×314×400
Wall thickness (mm)	30

Reference 12.

The stability of MIMO decentralized control system can be assessed using the generalized Nyquist criteria.[8] Assuming that the open loop system is stable, this criteria states that the closed loop system depicted in Fig. 3 is stable if and only if the locus of the determinant of the measured return matrix, $\det[\mathbf{I}+\mathbf{G}(j\omega)\mathbf{H}(j\omega)]$, does not encircle the origin, nor passes through the origin, as the angular frequency ω varies between $-\infty$ and $+\infty$.

Figure 4 shows the locus of $\det[\mathbf{I}+\mathbf{G}(j\omega)\mathbf{H}(j\omega)]$ between 25 Hz and 42 kHz. The locus between 25 and 700 Hz is plotted by a faint line, and the locus for frequencies above 700 Hz is plotted by a dotted line. The locus is characterized by a series of large circles at low frequency, which are determined by the resonant response of the plate. At low frequencies, the circles start close to the point $(1, 0j)$. As the frequency rises, the circles tend to drift away from $(1, 0j)$, and around 700 Hz the locus moves into the left-hand side of the vertical axis passing through the point $(1, 0j)$ due to propagation delays between the sensor and the actuator pairs. This plot indicates that the decentralized feedback control system under study is only conditionally stable, and generates positive feedback effects at high frequencies. In general the positive feedback are generated by the noncollocation and nonduality properties of each sensor–actuator pair, which originates both self-control spillover effects on each feedback loop and cross control spillover effects between neighbor feedback loops.

III. CONTROL IMPLEMENTATION

This section presents the experimental results for the control performance of the new smart panel. In Sec. III A, the control performance is discussed with reference to the vibration level of the panel in the frequency range 0–1 kHz. Sections III B and III C present the reduction in terms of the total radiated sound power in the same frequency range.

For both cases, the control performance is discussed in two stages: the passive effect produced by the sensor–actuator transducers, and the active control effect of the 16 decentralized control loops. The passive control performance is assessed by comparing the responses and sound radiation produced by the panel without transducers (faint lines), and the panel with transducers but no feedback control (thick solid lines). The active control effect is assessed by comparing the responses and sound radiation produced by the panel with transducers but no feedback control (thick solid lines) and the panel with transducers and decentralized velocity feedback control (dotted lines).

TABLE II. Geometry and physical properties of the triangularly shaped piezoceramic actuators.

Parameter		Values
Base and height (mm)		40×40
Thickness (mm)		1.0
Density (kg/m^3)		7700
Poisson's ratio		0.39
Elastic compliances	s_{11}^E	17
(10^{-12} m^2/N)	s_{33}^E	23
Piezoelectric charge	d_{31}	170
coefficient (10^{-12} m/V)	d_{33}	425

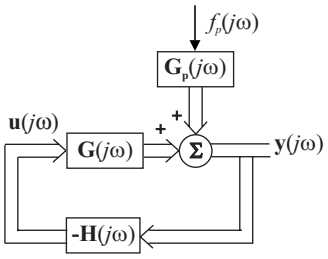


FIG. 3. Disturbance rejection block diagram of the velocity feedback control system.

A. Structural vibration

Figure 5 shows the measured narrow band spatially averaged spectra of the out-of-plane vibration levels of the panel per unit acoustic primary excitation input V_{in} [Fig. 5(a)] and structural primary excitation input F_{in} [Fig. 5(b)]. The spectra have been derived by remotely measuring the velocity of the panel at 165 points evenly distributed over the panel surface using a scanning laser vibrometer.

Comparing the two plots in Fig. 5, it can be observed that, when the panel is excited by a point force generated by a shaker, the response of the panel is characterized by a rather large number of resonant peaks. In contrast, when the panel is excited by acoustic field, the spectra show fewer sharp resonant peaks, mainly at the first, third, and fifth resonance frequencies, which correspond to the odd modes, (1,1), (3,1), and (1,3) of the panel, respectively. The peaks of the second and fourth resonances, which are controlled by (2,1) and (1,2) vibration modes, are relatively low. This is due to the fact that the shaker excites nearly all structural

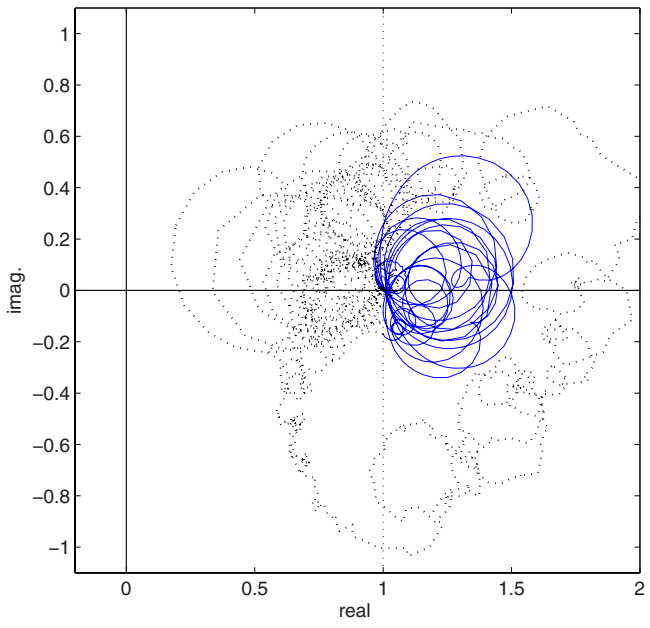


FIG. 4. (Color online) Locus of $\det[\mathbf{I}+\mathbf{G}(j\omega)\mathbf{H}(j\omega)]$ between 25 and 700 Hz (faint line) and between 700 Hz and 42 kHz (dotted line).

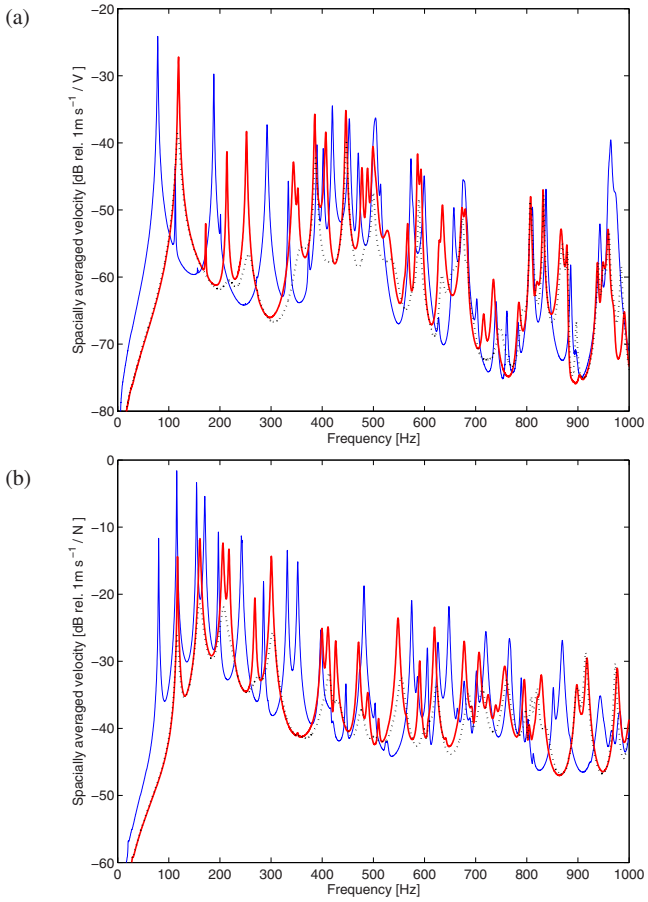


FIG. 5. (Color online) Narrow band spectra of (a) the spatially averaged velocity of the panel, measured between 0 and 1 kHz, per unit voltage driving the loudspeaker and (b) per unit force generated by the shaker; measured using the panel (1) without transducers (faint line), (2) with transducers but no feedback control (thick solid line), and (3) with transducers and feedback control (dotted line).

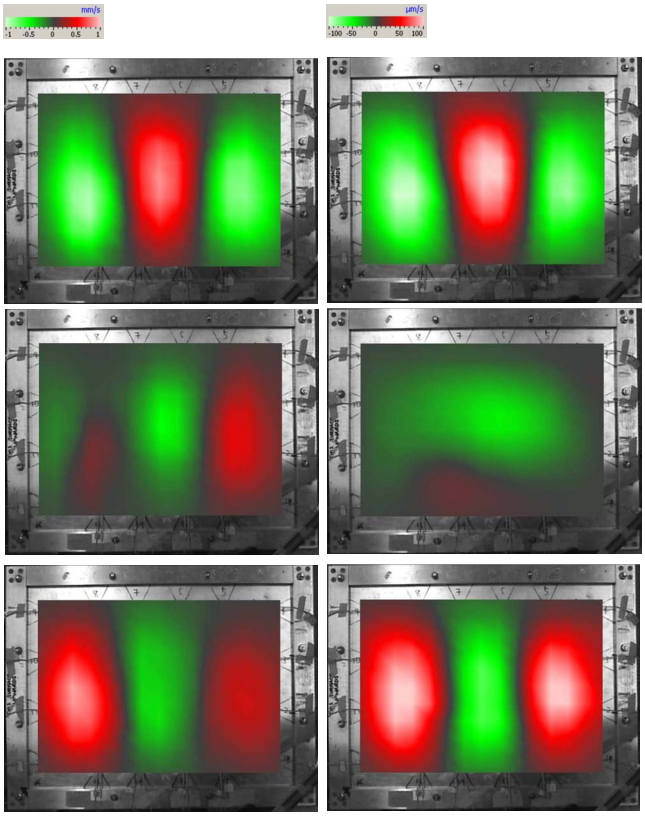


FIG. 6. (Color online) Vibratory fields of the panel excited by the loudspeaker at the 258 Hz resonance frequency, where the response is controlled by (3,1) natural mode of the smart panel; without the control (left column) and with control (right column) at phases $\varphi=0°$ (top), $\varphi=120°$ (middle), and $\varphi=240°$ (bottom).

modes of the panel, whereas the acoustic field generated by the loudspeaker efficiently excites only specific structural modes of the panel that are well coupled to the acoustic modes of the cavity underneath the panel.[1]

The faint and thick solid lines in Fig. 5 show that, when a set of 16 control units is bonded on the panel, the resonance frequencies are shifted toward higher frequencies, and the amplitudes of the peaks are reduced by approximately 3 dB with acoustic excitation, and by a maximum 10 dB with structural excitation. A significant increase in the natural frequencies is generated by piezoceramic actuators, which generate a local stiffness effect on the panel especially at low frequencies.[9] The response at resonant frequencies is reduced, because the bonding layer between the actuator and the panel adds damping to the panel. The sensor–actuator control units also produce mass loading effects on the panel in the mass controlled frequency zone, i.e., high frequencies, which are above the frequency range considered in this study.[9]

The thick solid and dotted lines in Fig. 5 show that the active control system smooths the resonant peaks between 100 and 700 Hz by approximately 6–18 dB, except for the resonance peaks at around 390, 445, and 500 Hz. These noncontrolled peaks are dominated by the resonances of the cavity beneath the panel. In fact, the first three predicted natural frequencies of the rigid walled cavity underneath the panel are calculated as 414.3, 476.4, and 546.2 Hz, which correspond to the (1,0,0), (0,0,1), and (0,1,0) acoustic modes, respectively.[10] The measured resonant frequencies are slightly lower than the predicted ones, because the cavity does not have perfectly rigid walls, especially due to the smart panel on the top side. This cavity–plate coupling[11] problem is particularly strong with this test rig setup. Therefore, it is expected that in applications where the panel has no backing cavity, and is excited by an acoustic disturbance that comes from a free field or a very large enclosure with diffuse acoustic fields, the control performances would be higher than those shown in Fig. 5. At frequencies higher than 700 Hz, the response of the panel slightly increases when control loops are closed, because of the positive feedback control effect mentioned in Sec. II B.

When active control is implemented, the spectrum of the vibration level has fewer peaks, as active damping can effectively reduce the response at resonances of the panel. This is confirmed by comparing the vibratory field of the panel at resonant frequency with and without control. Figure 6 shows the vibratory field of the smart panel excited by the acoustic field produced by the loudspeaker in the cavity at 258 Hz, which is the resonant frequency corresponding to the (3,1) natural mode of the panel. The pictures on the left-hand side show the vibratory fields of the panel without control at

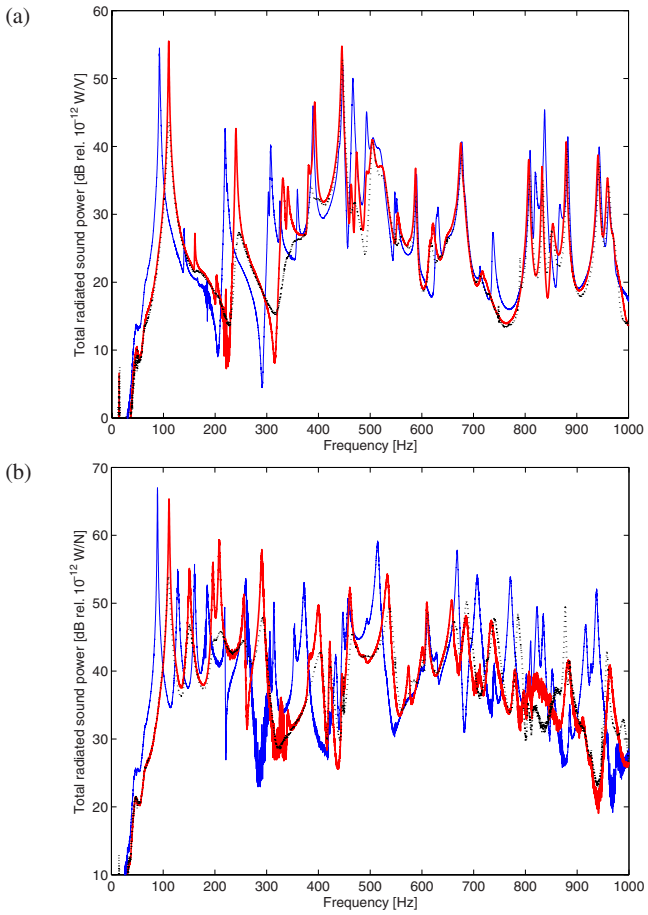


FIG. 7. (Color online) Narrow band spectra of the total radiated sound power, measured between 0 and 1 kHz, (a) per unit voltage driving the loudspeaker and (b) per unit force generated by the shaker; measured using the panel (1) without transducers (faint line), (2) with transducers but no feedback control (thick solid line), and (3) with transducers and feedback control (dotted line).

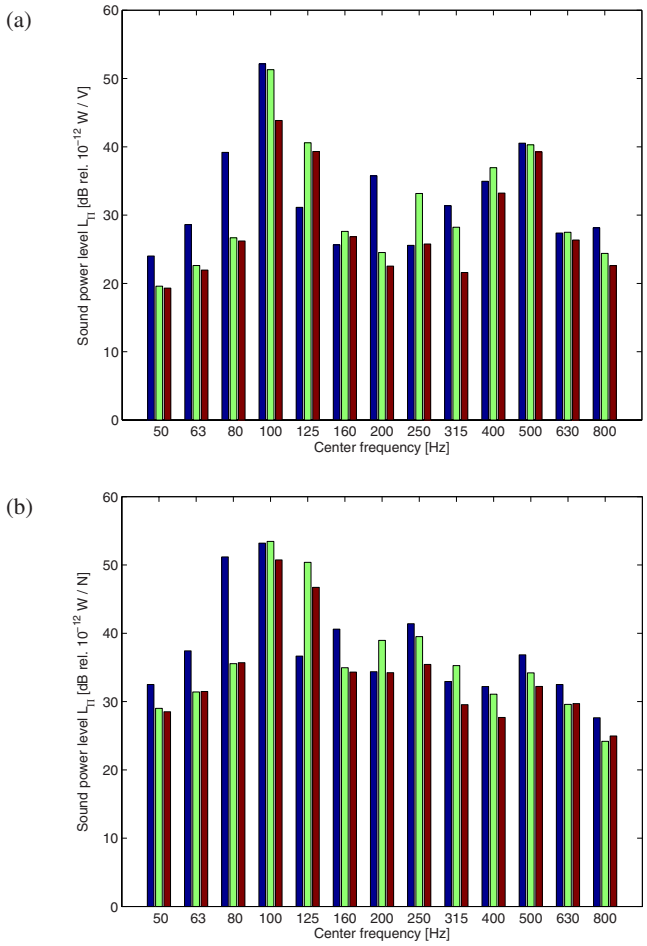


FIG. 8. (Color online) Total radiated sound power in one-third octave bands between 50 and 800 Hz center bands (a) per unit voltage driving the loudspeaker and (b) per unit force generated by the shaker; Measured using the panel without transducers (first column), with transducers but no feedback control (middle column), and with transducers and feedback control (last column).

phases $\varphi=0°$ (top), $\varphi=120°$ (middle), and $\varphi=240°$ (bottom). The right-hand side pictures show the vibratory fields with decentralized MIMO velocity feedback control with a 10 times small scale. The left-hand side pictures indicate that the response of the plate at 258 Hz is dominated by the (3,1) natural mode shape of the panel. In contrast, the right-hand side pictures highlight that when active damping is applied to the panel, the response is not characterized by the (3,1) natural mode shape. Although the (3,1) natural mode shape is still present, the residual effects from the neighboring modes are observed especially in the middle picture in the right-hand column of Fig. 6.

B. Total radiated sound power—narrow band analysis

Figure 7 shows the measured narrow band total radiated sound power per unit acoustic primary excitation input V_{in} [Fig. 7(a)] and structural primary excitation input F_{in} [Fig. 7(b)]. The spectra have been calculated using the far field sound pressure levels measured in a large anechoic chamber by nine free-field microphones (Bruel & Kjaer, Type 4165) placed around the test rig, according to ISO 3744.[1] The test rig has been placed on the solid floor in order to get a baffling effect. As even modes of the panel do not radiate sound efficiently,[11] the sound power responses are characterized by a smaller number of resonances compared with the spectra for the structural response shown in Fig. 5.

The thick solid lines in Fig. 7 again show that, when the 16 transducer pairs are attached to the panel, the resonances are shifted up in frequency, and the amplitudes of the spectra are reduced. The dotted lines in Fig. 7 show that the active control system reduces the resonant peaks between 100 and 500 Hz by a maximum 15 dB, for both acoustic and structural excitations. At higher resonance frequencies, between 500 and 700 Hz, although the control system suppresses the vibration of the structure, the sound power reduction is relatively poor. At frequencies higher than around 700 Hz, no active control effect is observed.

C. Total radiated sound power—one-third octave band analysis

The previous subsection presented a narrow band frequency analysis, which is useful in understanding the control performance. When acoustic noise is measured, however, it is customary to analyze the continuous spectrum sound in constant percentage bandwidths, because the hearing system

is responsive to frequency bands rather than individual frequencies. Therefore, in this section, one-third octave spectra[10] is used to evaluate the control performance in terms of the total radiated sound power per unit acoustic primary excitation input V_{in} [Fig. 8(a)] and structural primary excitation input F_{in} [Fig. 8(b)]. The primary excitation sources are driven by a pink noise signal between 0 and 1 kHz in order to give equal energy in all octave bands.

The two plots in Fig. 8 highlight that when the 16 transducers are attached to the panel the radiated sound power is significantly reduced, by 5–15 dB at the first three one-third octave bands for both acoustic and structural primary excitation. These reductions are generated by the shift of the resonant frequencies due to the increased stiffness of the panel. When active control is implemented, the radiated sound power is reduced by 1–7 dB for the fourth to the eleventh one-third octave band in both plots. No control effect is found in the first three one-third octave bands, where the response of the panel is not characterized by resonances, and thus cannot be controlled by implementing active damping. At higher frequencies, very little or no control effect is observed.

IV. CONCLUSIONS

This paper summarizes the experimental results of a study to assess the control effectiveness of a smart panel using triangularly shaped piezoceramic actuator and accelerometer sensor pairs. Sixteen decentralized velocity feedback control loops have been implemented in order to generate active damping and to reduce both the response and the sound radiation via the panel at resonant frequencies.

The experimental results have highlighted that, when the control transducers are attached to the thin aluminum panel, its natural frequencies are shifted up due to the passive stiffening effect of the piezoceramic patch actuators, and the amplitudes of the resonant peaks are slightly reduced due to passive damping effects generated by the bonding layer and the dissipative effects induced by the short circuited piezoceramic actuators. The closed loop experimental results have highlighted that significant control effects are obtained in the frequency band between 100 and 700 Hz with reductions of up to 15 dB at the resonant peaks in both structural vibration and sound radiation, except for some resonance frequencies controlled by resonances of the cavity under the panel. At frequencies higher than around 700 Hz, the vibration level and the radiated sound power slightly increases.

These results have been obtained from a prototype smart panel. It is believed that the control performance can be further enhanced by properly designing the size and shape of the triangular piezoceramic actuators in such a way as to improve the limit of the stable control gain and extend the frequency range where the response of the panel can be reduced by the decentralized velocity feedback control.

ACKNOWLEDGMENTS

This project has been carried out with joint support from InMAR European Commission project and RENAULT. One of the authors (Y.A.) was also supported by Nakajima Foundation, Japan.

[1] F. J. Fahy and P. Gardonio, *Sound and Structural Vibration*, 2nd ed. (Academic, London, 2006).
[2] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *The Active Control of Vibration* (Academic, New York, 1997).
[3] S. J. Elliott, P. Gardonio, T. C. Sors, and M. J. Brennan, "Active vibroacoustic control with multiple local feedback loops," J. Acoust. Soc. Am. **111**, 908–915 (2002).
[4] P. Gardonio, E. Bianchi, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission, Part I: Theoretical predictions," J. Sound Vib. **274**, 163–192 (2004); "Smart panel with multiple decentralized units for the control of sound transmission, Part II: Design of the decentralised control units," J. Sound Vib. **274**, 193–213 (2004); "Smart panel with multiple decentralized units for the control of sound transmission, Part III: Control system implementation," J. Sound Vib. **274**, 215–232 (2004).
[5] P. Gardonio and S. J. Elliott, "Smart panels with velocity feedback control systems using triangularly shaped strain actuators," J. Acoust. Soc. Am. **117**, 2046–2064 (2005).
[6] J. M. Sullivan, J. E. Hubbard, Jr., and S. E. Burke, "Modeling approach for two-dimensional distributed transducers of arbitrary spatial distribution," J. Acoust. Soc. Am. **99**, 2965–2974 (1996).
[7] M. J. Balas, "Feedback-control of flexible system," IEEE Trans. Autom. Control **23**, 673–679 (1978).
[8] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control; Analysis and Design*, 2nd ed. (Wiley, New York 2005).
[9] Y. Aoki, P. Gardonio, and S. J. Elliott, "Modeling of piezoceramic patch actuator for velocity feedback control," Smart Mater. Struct. **17**, 015052 (2008).
[10] D. A. Bies and C. H. Hansen, *Engineering Noise Control; Theory and Practice*, 3rd ed. (E & FN Spon, London, 2003).
[11] I. L. Ver, *Noise and Vibration Control Engineering: Principles and Applications* (Wiley, New York 2005).
[12] S. S. Rao, *Mechanical Vibrations*, 3rd ed. (Addison-Wesley, Reading, MA, 1995).

Reduction of turbulent boundary layer induced interior noise through active impedance control

Paul J. Remington,[a] Alan R. D. Curtis, Ronald B. Coleman, and J. Scott Knight[b]
BBN Technologies, 10 Moulton Street, Cambridge, Massachusetts 02138

(Received 8 May 2007; revised 29 October 2007; accepted 26 December 2007)

The use of a single actuator tuned to an optimum impedance to control the sound power radiated from a turbulent boundary layer (TBL) excited aircraft panel into the aircraft interior is examined. An approach to calculating the optimum impedance is defined and the limitations on the reduction in radiated power by a single actuator tuned to that impedance are examined. It is shown that there are too many degrees of freedom in the TBL and in the radiation modes of the panel to allow a single actuator to control the radiated power. However, if the panel modes are lightly damped and well separated in frequency, significant reductions are possible. The implementation of a controller that presents a desired impedance to a structure is demonstrated in a laboratory experiment, in which the structure is a mass. The performance of such a controller on an aircraft panel is shown to be effective, if the actuator impedance is similar to but not the same as the desired impedance, provided the panel resonances are well separated in frequency and lightly damped.
 [DOI: 10.1121/1.2835608]

PACS number(s): 43.50.Ki, 43.40.Vn, 43.40.Rj, 43.55.Rg [KAC] Pages: 1427–1438

I. INTRODUCTION

Turbulent boundary layer (TBL) generated interior noise in commercial turbofan powered aircraft is a well-recognized problem, threatening the in-flight comfort of passengers and disrupting communication among crew members. At many locations in the cabin it is the dominant source of noise. Passive approaches to its control are often limited because of the weight penalty associated with incorporating damping or high transmission loss layers in the fuselage design. The general subject of this paper is the reduction of sound power transmission through structures excited by complex excitation fields with a focus on the reduction of the interior noise of aircraft due to TBL excitation of the fuselage. The approach examined is to attach to the fuselage panels a limited number of actuators, whose impedance is, then, controlled so as to reduce optimally the sound power transmitted into the aircraft cabin.

The approach is illustrated in Fig. 1. There a small inertial actuator that is attached to a panel with a force gauge to measure the force exerted by the actuator and an accelerometer to measure the response of the structure. Figure 1 shows a separate accelerometer attached to the panel near the force gauge. In actuality the accelerometer would be integrated with the force gauge as is the case with most commercially available impedance heads. By appropriate signal processing the impedance of the actuator as seen by the panel can be controlled. The approach is to select that impedance in such a way as to minimize the transmission of sound power through the panel and to do so with only a single actuator.

The idea of an impedance controller is not new. Guicking *et al.*[1] were the first to study the use of an actuator to present a prescribed impedance to a structure. They employed an electrodynamic shaker, a force gauge, and an accelerometer connected to a beam that was driven by a second shaker. Because they utilized analog circuitry for the controller they were only able to operate the system effectively at individual frequencies, but they recognized that by using digital filters broadband performance could likely be achieved. Although there have been numerous studies examining the modification of the acoustic impedance of systems to achieve noise reduction or performance improvements[2–4] or to enhance absorption,[5–10] there have been many fewer studies of approaches to modifying the mechanical impedance of a control device to achieve a desired effect. Bras *et al.*[11] applied the concept in a very interesting way. They used active control to adjust the mechanical impedance in a test rig for testing actuators, so that a chosen actuator could be presented with a desired drive point impedance. Of more direct relevance to this study is a paper by Benassi and Elliott[12] who investigated strategies for suppression of plate vibration using a single inertial actuator. Basing the control law on minimizing the total power injected into the plate from a primary point force and a secondary control force, they determined the optimum impedance that needed to be presented by the control force. Although this impedance was found to be noncausal, they were able to find a suboptimal impedance that provided good performance and was implemented in the laboratory.

Others have studied the suppression of vibration and sound radiation from plates. Most have involved the use of multiple secondary sources. Deffayet and Nelson[13] looked at the control of radiated sound from a panel using secondary acoustic sources; Snyder *et al.*[14] examined the use of both acoustic and point force control sources to minimize the radiated sound power from a simply supported panel; Pan

[a]Presently at Raytheon, Integrated Defense Systems, Missile Defense Center, 225 Presidential Way, Woburn, MA 01801-5143.
[b]Presently at Ball Aerospace and Technology Inc., 1600 Commerce Street, Boulder, CO 80301.

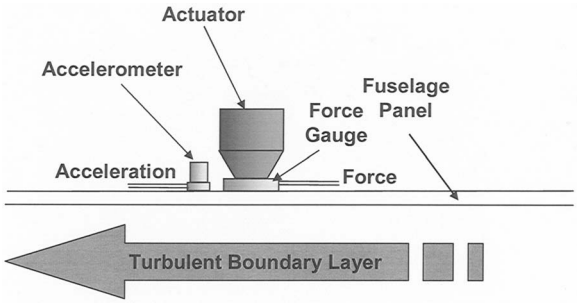


FIG. 1. Impedance controller arrangement.

et al.[15] considered the active control of a baffled panel excited by airborne or structure-borne sources using acoustic or point force actuators; Benassi and Elliott[16] studied the active suppression of plate vibration when the total power transmitted to a plate from both primary and secondary point force sources is minimized. Finally, there have been studies of active control approaches to improve the transmission loss of panels. Pan and Bao[17] examined a number of strategies for increasing the transmission loss of a double walled panel; Wang *et al.*[18] investigated the use of multiple piezoelectric and point force actuators to reduce sound transmission through plates; Pan *et al.*[19] studied analytically the use of point force actuators to control the transmission of noise through a panel into a cavity.

The concept examined in this paper extends the work described in the previous references in a number of ways. First, TBL excitation is the mechanism considered for driving the panel. The TBL has, as will be shown, many more degrees of freedom than a single force or incident plane wave. Increasing the number of degrees of freedom in the excitation can result in a more difficult control problem. Second, the problem is formulated to include the excitation, the panel response, and panel sound radiation, while restricting the control system to a single actuator. The restriction to a single actuator can significantly limit the effectiveness of the control approach. Consequently, to provide a physical understanding of those limitations the paper examines the requirements for effective control of sound radiation, if a single actuator is used to control either the excitation from the TBL, the panel response or the sound radiation from the panel. Finally, the effectiveness of the control algorithms used for driving the actuator to provide a desired impedance are examined for different dynamic characteristics of the panel and actuator and for different desired impedances in order to provide a physical understanding of the limitations on actively generating a desired impedance.

One might question the limitation to a single actuator; however, minimizing the number of actuators is important, because the fuselages of most aircraft are broken up into many panels separated by longitudinal and circumferential stiffeners. Multiple actuators on each panel would quickly result in an impossibly large number of actuators with the potentially significant cost and weight penalties. In addition, if multiple actuators are used on a single panel, they will in most cases be well coupled and require a multiple-input-multiple–output controller for each panel. With only one actuator per panel there is a good chance that the coupling will be weak enough that each actuator could be controlled independently of the actuators on other panels. This approach is appealing, because, if it can be made to perform well, one could imagine attaching small impedance control actuators to individual panels, tuned to the impedance best able to suppress sound transmission through that panel.

In the next section the mathematical model is presented which includes the TBL model for the excitation of the panel, the modal model of a simply supported panel, and the coupling of that panel to the acoustic space on the side opposite the TBL excitation. A critical issue is the determination of the desired impedance to implement in the actuators. That process based on minimizing the sound power radiated by the panel from its interior side is discussed in Sec. III. To provide some physical insight into the problem, Sec. IV discusses under what circumstances perfect control could be achieved with a single actuator. Section V presents the results of laboratory measurements where a simple impedance was implemented using a commercially available, small shaker. Section VI presents estimates of the performance for implementation of the concept on a test panel using measured panel and actuator impedances as well as simple models of alternate actuator impedances. Conclusions are presented in Sec. VII.

II. FORMULATION OF THE PROBLEM

The details of the formulation of the mathematics coupling the TBL excitation to the sound power radiated by the panel can be found in Ref. 20. Here only a broad outline of the formulation is presented. The model consists of three parts: source, panel response, and panel radiation. The sources are the TBL pressures acting on the panel. These are assumed to be random processes and are characterized by the cross-spectral density matrix of the panel modal forces generated by these pressures.

$$\mathbf{S}_{ff} = \langle \vec{f}\vec{f}^{H} \rangle, \tag{1}$$

where $\vec{f}$ is the Fourier transform of the vector of modal forces on the panel, the pair of angular brackets means expected value, and the superscript H means the conjugate transpose of the vector. The TBL cross-spectral density matrix is derived based on the usual Corcos Model[21] with the power spectral density of the TBL pressures estimated using Efimtsov's formula.[22] It turns out that the resulting integrals of the mode shape functions for a simply supported panel and the cross-spectral density function from Corcos and Efimtsov are analytically tractable and can be solved in closed form.[20]

The panel is characterized by its modal admittance

$$\vec{v} = \mathbf{G}_0 \vec{f}, \tag{2}$$

where $\mathbf{G}_0$ is a diagonal matrix, the elements of which are given by $G_{o,mn} = j\omega/\rho[\omega_{mn}^2(1+j\eta)-\omega^2]$ where ρ is the panel area density, η is the loss factor, and ω_{mn} is the panel natural frequency of mode (m,n). The natural frequencies are estimated for a simply supported panel with tension to account for internal pressurization of the aircraft fuselage. Internal

pressurization is included because it can dramatically increase the natural frequencies of the panel.[23]

The radiation is based on the modal radiation impedance of the panel

$$\vec{p} = \mathbf{Z}\vec{v}, \tag{3}$$

where $\vec{p}$ is the vector of modal pressures on the baffled panel due to the vector of panel modal velocities $\vec{v}$. The radiation impedance matrix, $\mathbf{Z}$, is derived in Appendix A of Ref. 20. Note that fluid loading in Eq. (2) can be accommodated by substituting for $\mathbf{G}_0$ the expression $\mathbf{G}=(\mathbf{I}+\mathbf{G}_0\mathbf{Z}_T)^{-1}\mathbf{G}_0$, where $\mathbf{Z}_T$ is the total modal radiation impedance for both sides of the panel, and $\mathbf{G}$ is no longer diagonal. However the modal coupling implied by $\mathbf{G}$ is under most circumstances not significant in air[20] and is not included in the calculations to follow.

For random excitation the power radiated by the panel is given by

$$\Pi(\omega) = \langle(\vec{p}^H\vec{v} + \vec{v}^H\vec{p})\rangle/4 = \langle\vec{v}^H(\mathbf{Z}+\mathbf{Z}^H)\vec{v}\rangle/4 = \langle\vec{v}^H\mathbf{Z}_R\vec{v}/2\rangle$$
$$= \mathrm{tr}\{\mathbf{Z}_R\langle\vec{v}\vec{v}^H\rangle\}/2, \tag{4}$$

where $\mathbf{Z}_R$ is the real part of the radiation impedance matrix, tr { } means the trace of the matrix or the sum of the diagonal terms and $\Pi(\omega)$ is the scalar spectrum of the radiated power. Note that the circulant rule ($\mathrm{tr}\{\mathbf{ABC}\}=\mathrm{tr}\{\mathbf{CAB}\}=\mathrm{tr}\{\mathbf{BCA}\}$) for the trace of a product of matrices has been used to obtain the last equation in Eq. (4). Substituting Eqs. (1) and (2) into Eq. (4), yields

$$\Pi(\omega) = \mathrm{tr}\{\mathbf{G}^H\mathbf{Z}_R\mathbf{G}\mathbf{S}_{\mathrm{ff}}\}/2. \tag{5}$$

A Hermitian matrix has the property that it can be subdivided into the product of a matrix and its conjugate transpose through singular value decomposition. The matrix $\mathbf{S}_{ff}$ is a cross-spectral density matrix and, hence, is Hermitian, and the matrix $\mathbf{Z}_R$ is real and symmetric, and, hence, also hermitian. As a consequence, one can write for the source $\mathbf{S}_{ff}=\mathbf{S}\mathbf{S}^H$ and for the real part of the impedance matrix $\mathbf{Z}_R=\mathbf{R}^H\mathbf{R}$. The implication is that one can define a set of radiation modes $\vec{r}=\mathbf{R}\vec{v}$ and for the source a set of sources defined by $\vec{f}=\mathbf{S}\vec{s}$. The vector $\vec{r}$ is a transformation of the modal velocities of the panel into what are normally referred to as radiation modes and $\vec{s}$ is a vector of unit amplitude uncorrelated sources in the TBL. If the matrix products are substituted into Eq. (5) and the products rearranged using the aforementioned circulant rule, the result can be written

$$2\Pi(\omega) = \vec{r}^H\vec{r} = \mathrm{tr}\{(\mathbf{RGS})^H(\mathbf{RGS})\}, \tag{6}$$

where $\vec{r}=\mathbf{RGS}\vec{s}$. This last equation leads to a convenient modeling formalism in which the vector $\vec{r}$ becomes the residual to be minimized. Figure 2(a) illustrates the baseline model relating the radiation mode vector $\vec{r}$ to the TBL source vector $\vec{s}$ and Fig. 2(b) shows the baseline model with an actuator mounted on the panel applying a desired impedance. In Fig. 2 the vector, $\vec{\phi}$, for a single impedance is a column vector, the elements of which are the amplitudes of the panel mode shape functions of the important panel modes at the actuator mounting point. Implicit here is the fact that a finite number of panel modes have been chosen that adequately characterize the panel dynamics. For multiple actuators $\vec{\phi}$ would become a matrix with the number of columns equal to the number of actuators. From here on only a single actuator will be considered.

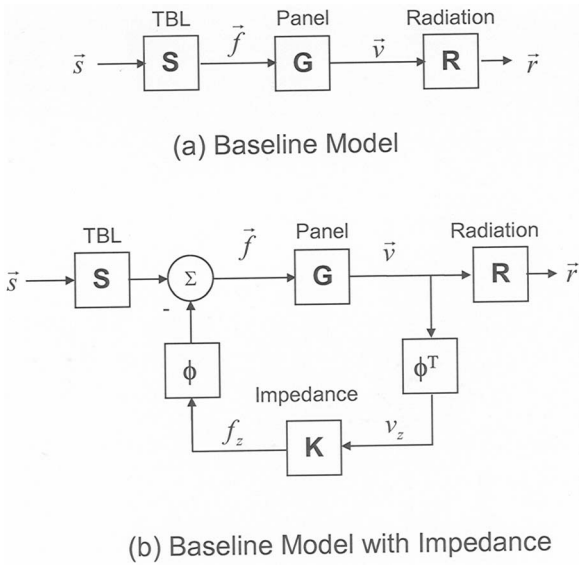


FIG. 2. Block diagram of the TBL radiation model.

III. DETERMINING THE OPTIMUM IMPEDANCE

To determine the optimum impedance the block diagram of Fig. 2 has been modified as indicated in Fig. 3. In Fig. 3 the Youla transform or "inner model" approach[24] has been used to model the impedance which is now given by

$$K = (1 - Q\vec{\phi}^T\mathbf{G}\vec{\phi})^{-1}Q, \tag{7}$$

where K is the impedance and Q is a filter that will be determined to minimize the power radiated by the panel.

It can be easily shown that the feedforward block diagram at the bottom of Fig. 3 is the same as the feedback block diagram at the top of Fig. 3, provided the quantity $\vec{\phi}^T\mathbf{G}\vec{\phi}$ in the feedback loop in the impedance block accu-

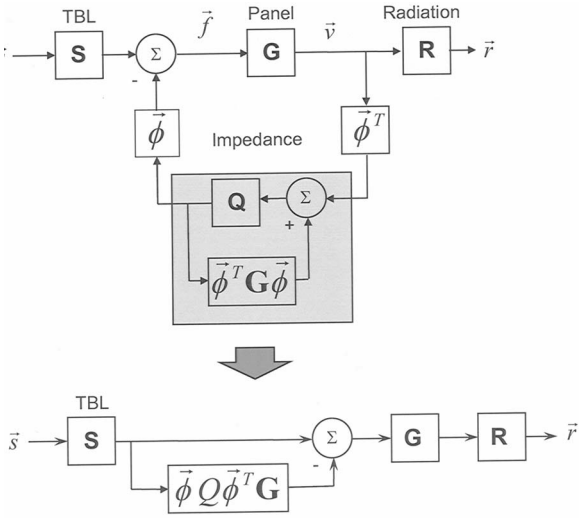


FIG. 3. Youla transform model of the impedance.

TABLE I. Perfect control with a single impedance.

Mechanism	Condition	Physics	Effect	Comments
TBL	$\mathbf{S}=s\vec{\phi}$ s=any scalar	• Single dominant source in TBL • modal force vector from TBL and actuator parallel		• There is an optimum actuator location • Many dofs in TBL make a single dominant source unlikely
Radiation	$\mathbf{R}=r\vec{\phi}^T$ r=any scalar	• Single dominant radiation mode • TBL and actuator drive dominant radiation mode proportionally	$Q\approx Z_{\text{panel}}$ $K\approx\infty$ $v_{\text{drive pt}}\approx 0$	• There is an optimum actuator location • Many dofs in radiation make a single dominant radiation mode unlikely
Panel	$G=g\vec{b}\vec{b}^T$ g=any scalar $\vec{b}$=any vector $\vec{b}^T\vec{\phi}\neq 0$	• Single dominant mode in panel • Actuator not at node of dominant mode		• No optimum actuator location • Good control possible near resonances of lightly damped panel

rately approximates the panel admittance at the actuator attachment point. It is interesting to note that even if Q is stable, Eq. (7) does not guarantee that K will be stable; however, the closed loop system will be stable if the above-mentioned requirement (that $\vec{\phi}^T\mathbf{G}\vec{\phi}$ in the feedback loop in the impedance block accurately approximates the panel admittance) is met.

The problem can now be formulated to compute the filter Q so as to minimize the mean square of the residual $\text{tr}\{\vec{r}\vec{r}^H\}$. From the feedforward block diagram in Fig. 3 one can write

$$\text{tr}\{\langle\vec{r}\vec{r}^H\rangle\}=|\vec{r}|^2=\text{tr}\{\mathbf{RG}(\mathbf{I}-\vec{\phi}Q\vec{\phi}^T\mathbf{G})\langle\mathbf{S}\vec{s}\vec{s}^H\mathbf{S}^H\rangle(I-\mathbf{G}^H\vec{\phi}Q^*\vec{\phi}^T)\mathbf{G}^H\mathbf{R}^H\}, \tag{8}$$

where $\vec{s}$ a vector of unit amplitude independent sources in the TBL and $\langle\vec{s}\vec{s}^H\rangle=I$, the unit matrix. Carrying out the multiplication in Eq. (8) and using the circulant rule for the trace of a product of matrices yields the residual

$$|\vec{r}(\omega)|^2=\text{tr}\{\mathbf{A}(\omega)-B(\omega)Q(\omega)-B^*(\omega)Q^*(\omega)+C(\omega)D(\omega)|Q^2(\omega)|\}, \tag{9}$$

where

$$\mathbf{A}=\mathbf{RGSS}^H\mathbf{R}^H\mathbf{G}^H,$$

$$B=\vec{\phi}^T\mathbf{GSS}^H\mathbf{G}^H\mathbf{R}^H\mathbf{RG}\vec{\phi},$$

$$C=\vec{\phi}^T\mathbf{G}^H\mathbf{R}^H\mathbf{RG}\vec{\phi},$$

$$D=\vec{\phi}^T\mathbf{G}^H\mathbf{SS}^H\mathbf{G}\vec{\phi}.$$

Equation (9) is a quadratic equation in the scalar Q. Taking $\partial|\vec{r}^2|/\partial Q=0$ and solving the equation for Q, yields a noncausal Q that minimizes the power radiated by the panel. As for a single actuator B, C, and D are scalar quantities, the noncausal optimum Q can be written

$$Q=\frac{\vec{\phi}^T\mathbf{G}^H\mathbf{R}^H\mathbf{RGSS}^H\mathbf{G}^H\vec{\phi}}{\{\vec{\phi}^T\mathbf{G}^H\mathbf{R}^H\mathbf{RG}\vec{\phi}\}\{\vec{\phi}^T\mathbf{GSS}^H\mathbf{G}^H\vec{\phi}\}}=\frac{B^H}{CD}, \tag{10}$$

which on substitution into Eq. (9) yields the residual radiated sound power

$$|\vec{r}|^2=\text{tr}\left\{\mathbf{RGSS}^H\mathbf{G}^H\mathbf{R}^H-\frac{|\vec{\phi}^T\mathbf{G}^H\mathbf{R}^H\mathbf{RGSS}^H\mathbf{G}^H\vec{\phi}|^2}{\{\vec{\phi}^T\mathbf{G}^H\mathbf{R}^H\mathbf{RG}\vec{\phi}\}\{\vec{\phi}^T\mathbf{GSS}^H\mathbf{G}^H\vec{\phi}\}}\right\}=\text{tr}\{\mathbf{A}\}-\left\{\frac{|B^2|}{CD}\right\}. \tag{11}$$

IV. IMPLICATIONS FOR PERFECT CONTROL

It is instructive to examine Eq. (11) to determine under what conditions perfect control can be achieved with a single actuator. Consider first the source. If $\mathbf{S}=s\vec{\varphi}$, where s is a scalar source strength and $\vec{\varphi}$ is a vector, and the location of the actuator is chosen such that the vector $\vec{\phi}$ is the same as $\vec{\varphi}$, then, Eq. (11) will go to zero, implying no radiated power from the panel. Further, if this expression for $\mathbf{S}$ is substituted into Eq. (10), the optimum Q is found to be the panel impedance at the point of attachment of the actuator $Q=1/\vec{\phi}^T\mathbf{G}\vec{\phi}$ and substituting this into Eq. (7) shows that the optimum impedance, K, is infinity, effectively pinning the panel at the actuator location. This simple expression for $\mathbf{S}$ implies that there is a single independent source in the boundary layer exciting the panel, an unlikely situation as will be shown later, and that the actuator location has been properly "tuned." By tuned we mean that an actuator location is chosen such that the magnitude of the modal shape functions at that location forms a vector $\vec{\phi}$ that is approximately equal to $\vec{\varphi}$. A similar simple model for the radiation impedance will also result in zero radiated power. If $\mathbf{R}^H$

$=r\vec{\varphi}$ (with the actuator location chosen such that $\vec{\phi}=\vec{\varphi}$) is substituted into Eqs. (10) and (11), the residual radiated power will go to zero, $Q=1/\vec{\phi}^T\mathbf{G}\vec{\phi}$, and the desired impedance, K, will go to infinity, again pinning the panel at the actuator location. The implication here is that there is only one significant radiation mode and in such a case perfect control is possible with a single impedance if the actuator location is properly tuned. Finally if the panel modal admittance is given by $\mathbf{G}=g\vec{b}\vec{b}^T$ where g is a any nonzero scalar and $\vec{b}$ is any vector with $\vec{b}^T\vec{\phi}\neq 0$, the residual radiated power in Eq. (11) will go to zero. The optimum Q is again, $Q=1/g\vec{\phi}^T\vec{b}\vec{b}^T\vec{\phi}=1/\vec{\phi}^T\mathbf{G}\vec{\phi}$, the panel impedance at the actuator attachment point, and the desired impedance, K, once again becomes infinity, pinning the panel at the actuator. The implication here is that good control is possible near a resonance, if the panel has lightly damped well separated resonant frequencies, and the location of the impedance is not at a node of a mode to be controlled. These observations are summarized in Table I.

It is useful to examine the likelihood of any of the above conditions existing in a TBL excited panel. In Fig. 4 the normalized singular values of $\mathbf{S}_{ff}$, $\mathbf{G}$, and $\mathbf{Z}$ are shown versus frequency for typical aircraft panel and TBL parameters (see the caption of Fig. 4). The graph at the top of Fig. 4 shows the normalized singular values of the cross-spectral density matrix of the modal forces in the TBL, $\mathbf{S}_{ff}$. If there were one dominant singular value, then, $\mathbf{S}\approx\sigma^{1/2}\vec{\varphi}$ where σ is the largest singular value and $\vec{\varphi}$ the corresponding singular vector. If the location of the actuator were chosen such that $\vec{\phi}$ and $\vec{\varphi}$ were approximately parallel, then, the conditions for perfect control $\mathbf{S}=s\vec{\phi}$ would be approximately satisfied and good control would be achieved with a single actuator. Unfortunately, Fig. 4 indicates that there is not one dominant singular value, and the TBL is composed of multiple degrees of freedom.

A similar conclusion applies to the radiation impedance, $\mathbf{Z}$. Above the first resonant frequency of the panel at 320 Hz, as shown in the bottom graph in Fig 4, there are at least two singular values within about 6 dB of one another. On the other hand, below 100 Hz one singular value dominates and a single actuator could, in principle, control the radiation

For the panel admittance the singular values near the resonant frequencies (the troughs in the middle graph in Fig. 4) show that one singular value dominates all others by at least 15 dB. Consequently, one would expect to achieve good control in the vicinity of the resonant frequencies of the panel provided the actuator was not placed at a node of a resonant mode that was to be controlled. Note that the dominance of the largest singular values as well as the reduction in radiated power would increase if the panel were more lightly damped.

In Fig. 5(a) the result is shown of adding a single active impedance to the panel of Fig. 4. The dash-dot curve in Fig. 5(a) is for a noncausal filter Q as defined by Eqs. (10) and (7). Figure 5(a) shows that good control is achieved in the vicinity of the resonant frequencies of the panel. As the panel modes are well separated and only modestly damped, good reductions at resonance are expected based on the perfect

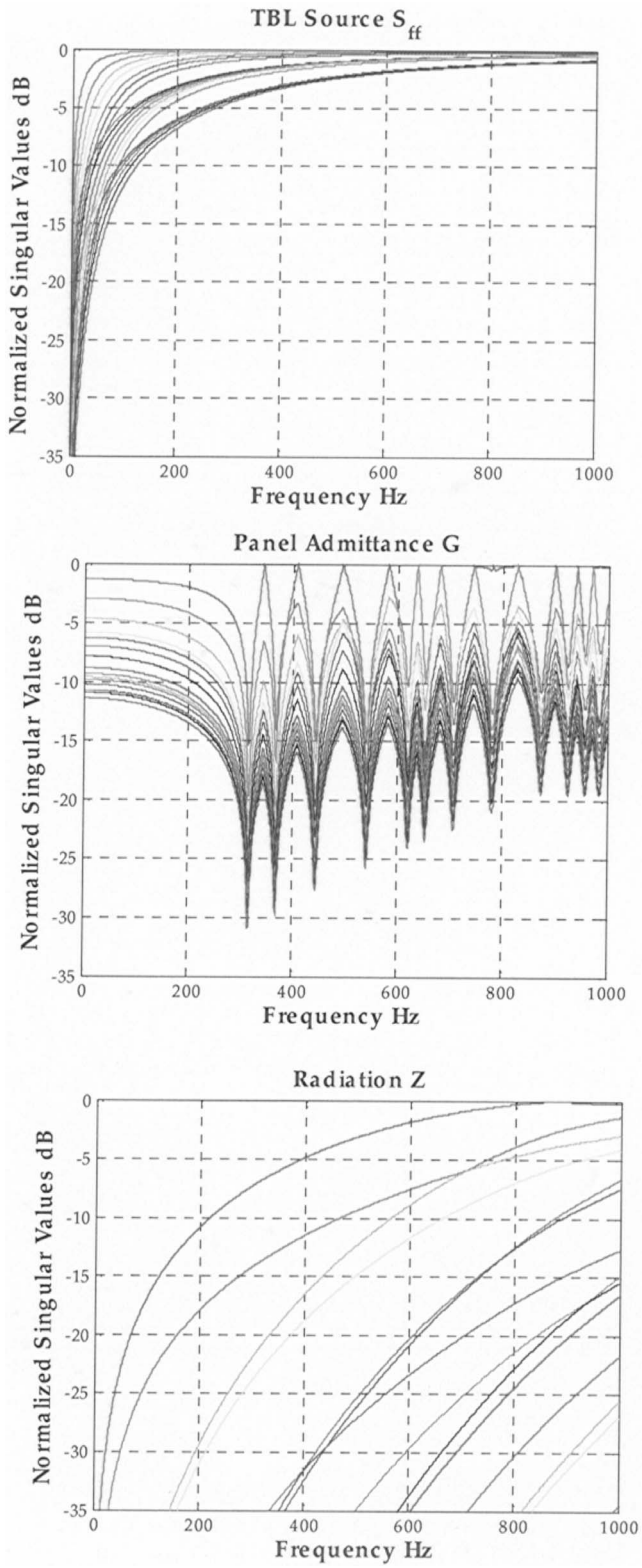


FIG. 4. Singular values of the source, radiation impedance, and panel admittance matrices normalized by the largest singular value for a simply supported flat aluminum panel 0.5 m×0.25 m×1 mm thick with loss factor 0.01; 3×10^4 N/m axial and 6×10^4 N/m circumferential tension; flow in the axial direction at Mach 0.7 with a TBL displacement thickness of 3 cm.

control arguments given earlier. Note that no reduction is achieved at 653 and 780 Hz. At each of those two frequencies there are two panel resonances that are at very nearly the

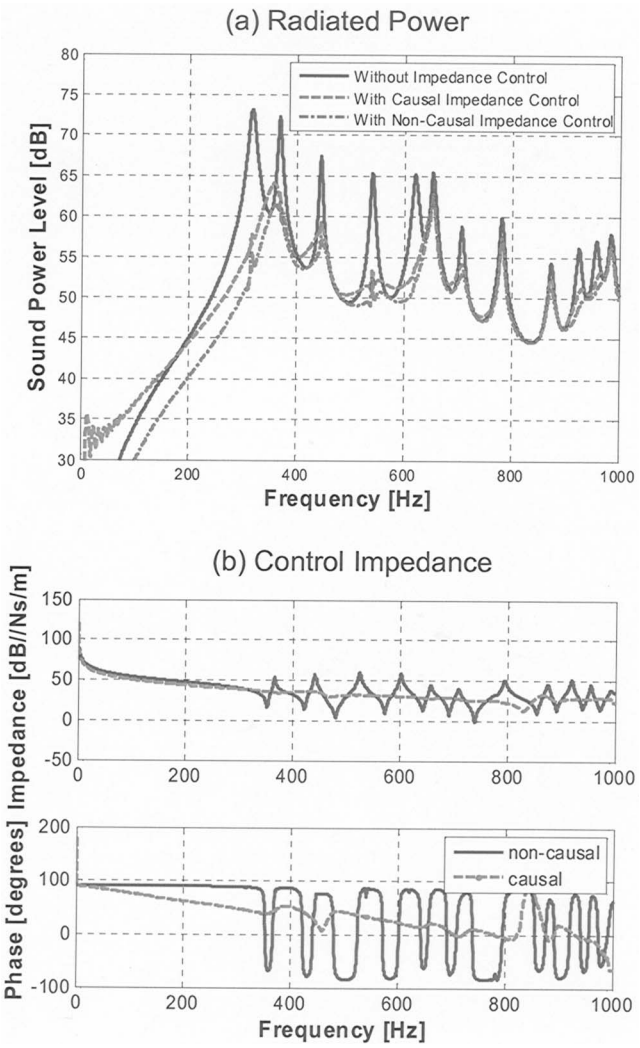


FIG. 5. Reduction in radiated power from the TBL excited panel of Fig. 4 with an optimum impedance located at 0.19 m in the axial (flow) direction from the end of the panel and 0.06 m in the circumferential direction from the side of the panel.

same frequency. As a result a single actuator would not be expected to provide good control, because G is not dominated by a single singular vector. Note that on careful examination Fig. 4 confirms this result.

Also shown for comparison in Fig. 5(a) is the result when the filter Q is constrained to be causal. The procedure for calculating the control filter begins with constraining Q to be a finite impulse response (FIR) filter. The control filter coefficients are, then, obtained using a linear least squares procedure that seeks to minimize the mean square of the residual given in Eq. (9). An expression for an N tap FIR filter is substituted for the control filter, Q, into Eq. (9):

$$Q(\omega) = \sum_{n=1}^{N} Q_n e^{-j\omega(n-1)\Delta T}, \tag{12}$$

where $1/\Delta T$ is the sampling frequency and the Q_n are the filter coefficients. Equation (9) is now integrated over the frequency band of interest,

$$\int_{\omega_1}^{\omega_2} |\vec{r}^2(\omega)| d\omega = \mathrm{tr}\left\{ \sum_n \int_{\omega_1}^{\omega_2} \mathbf{A}(\omega) e^{-j\omega(n-1)\Delta T} d\omega \right\} - \sum_n Q_n \left[\int_{\omega_1}^{\omega_2} B(\omega) e^{-j\omega(n-1)\Delta T} d\omega - \int_{\omega_1}^{\omega_2} B^*(\omega) e^{j\omega(n-1)\Delta T} d\omega \right] + \sum_n Q_n^2 \int_{\omega_1}^{\omega_2} C(\omega) D(\omega) d\omega, \tag{13}$$

the derivative is taken with respect to each Q_n, and each result is set equal to zero. From that set of linear equations the unknown filter coefficients Q_n that minimize the mean square residual can be determined.

The dashed curve in Fig. 5 is the result when Q is a 400 tap FIR filter at a sampling rate of 2 kHz. Although no latency due to antialiasing or reconstruction filters, half sample delays due to digital/analog converters, computational delays, or finite impedance of the actuator are included in the calculation, the result is promising especially when one considers the differences in the causal and noncausal impedances shown in Fig. 5(b). The causal impedance is a considerable simplification of the noncausal yet the resulting performance is not seriously degraded.

Considering that the very different causal and noncausal impedances provide similar radiated power reductions, it is instructive to fit a very simple impedance to the optimum causal impedance and examine the resulting performance. That result is shown in Fig. 6. There a simple spring–mass–damper impedance has been added to the panel at the same location as in Fig. 5. The parameters of the spring–mass–damper, which were determined by a parameter search, are: 0.065 kg mass, 5×10^4 N/m spring, and 87 N s/m damper. Figure 6(a) shows similar reductions in radiated power for the simple impedance compared to the causal optimum impedance even though the agreement between the two impedances in amplitude and phase in Fig. 6(b) is at best modest.

V. LABORATORY IMPLEMENTATION OF A PRESCRIBED IMPEDANCE

In the previous section causal and noncausal impedances were defined to optimally reduce the power radiated by a TBL excited panel. In addition a simple mass–spring–damper impedance was defined by trial and error that performed equally well. This section focuses on the implementation of that impedance. From here on it will be more convenient to use the term "dynamic mass" rather than impedance. The term dynamic mass means the inverse of the accelerance, the force divide by the acceleration. We use this terminology here for want of a better term because in the implementation of the desired impedance the control system will utilize accelerometers for sensors, and, as will be shown, it is more convenient in that context to refer to dynamic mass rather than impedance. In Fig. 7 a block diagram is shown for implementing the dynamic mass, I_d. The approach uses a Youla type or "inner model" structure to implement the con-

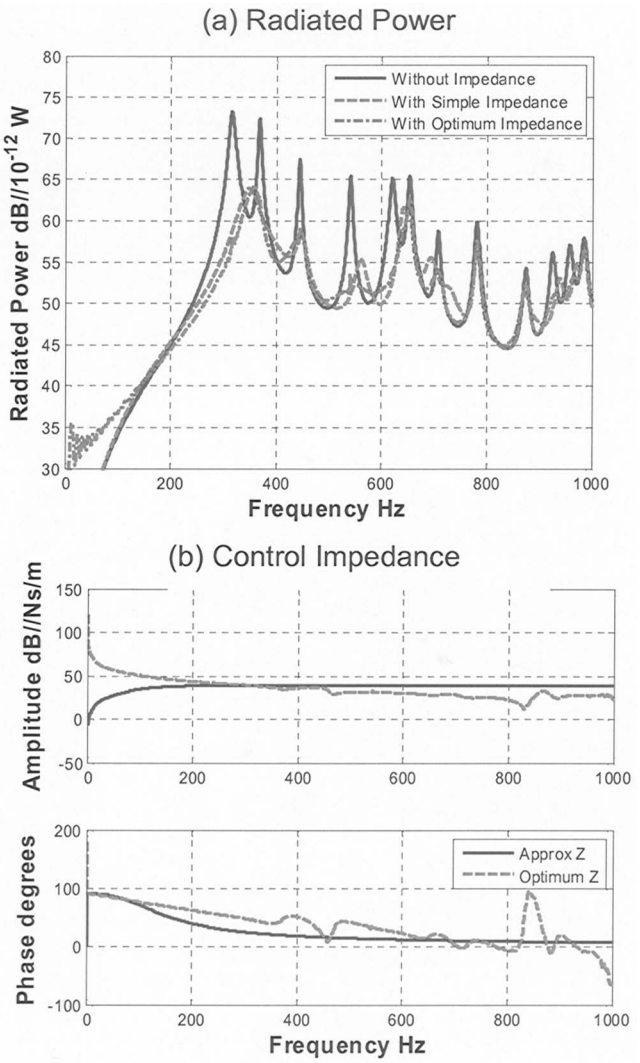


FIG. 6. Reduction in radiated power from the TBL excited panel of Fig. 4 with a *suboptimum* impedance at the same location as in Fig. 5.

troller. The quantity $\tilde{P}$ in the Fig. 7 is an estimate of the plant transfer function. The plant transfer function relates the signal at the controller input to a signal applied to the physical system at the controller output. The true plant is given by

$$P = \frac{1 + I_d A_{\mathrm{STR}}}{1 + I_{\mathrm{ACT}} A_{\mathrm{STR}}}, \tag{14}$$

where I_{ACT} is the actuator dynamic mass, I_d is the desired dynamic mass that the controller is designed to implement, and A_{STR} is the accelerance of the structure to be controlled, in this case a mass.

Note that the stability of the feedback system is determined by the poles of $\{\boldsymbol{I}+\boldsymbol{K}(\tilde{\boldsymbol{P}}-\boldsymbol{P})\}^{-1}$, where P is the true plant transfer function relating controller input to the controller output and $\tilde{P}$ is an estimate of the plant transfer function implemented in the controller usually as a FIR filter. The system will generally be stable if $\tilde{\boldsymbol{P}} \approx \boldsymbol{P}$ and in fact it is guaranteed to be stable if $\tilde{\boldsymbol{P}} = \boldsymbol{P}$.

The approach in Fig. 7 seeks to minimize the residual

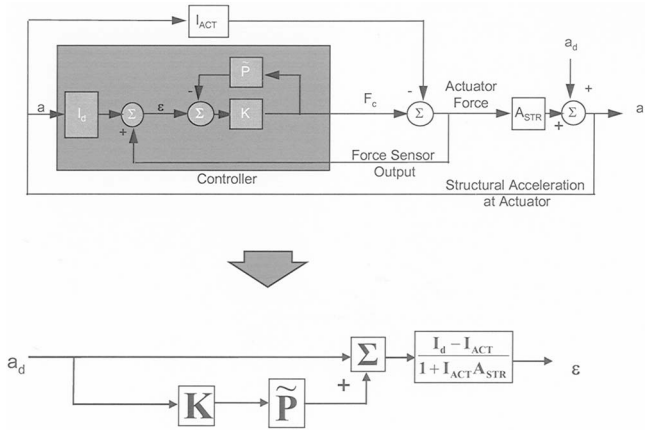


FIG. 7. Block diagram for implementing the impedance controller; A_{STR} =panel accelerance at the actuator attachment point; I_d=desired dynamic mass (inverse of accelerance); I_{ACT}=dynamic mass of the actuator as seen by the panel; K=control filter; and $\tilde{\mathbf{P}}$=plant transfer function.

$$\varepsilon = F_{\mathrm{ACT}} + I_d a_{\mathrm{ACT}} \tag{15}$$

where F_{ACT} is the force measured between the actuator and the panel, a_{ACT} is the acceleration on the panel at the actuator attachment point, and I_d is the desired dynamic mass.

In this approach the control filter coefficients are obtained using a linear least squares procedure similar to that described previously that seeks to minimize the mean square of the residual given in Eq. (15). The estimation is carried out as if the estimate of the plant transfer function $\tilde{\boldsymbol{P}}$ were a perfect match to the actual plant, $\boldsymbol{P}$. From Fig. 7 the residual can be expressed in terms of the plant, the control filter and the excitation spectrum as

$$|\varepsilon^2| = S_a(\omega) \left| \frac{I_d - I_{\mathrm{ACT}}}{1 + I_{\mathrm{ACT}} A_{\mathrm{STR}}} \right|^2 \{1 + \boldsymbol{K}\tilde{\boldsymbol{P}} + \boldsymbol{K}^*\tilde{\boldsymbol{P}}^* + \boldsymbol{K}\boldsymbol{K}^*|\tilde{\boldsymbol{P}}^2|\}. \tag{16}$$

In order to be able to control out of band amplification of the residual and to help ensure stability it is often useful to include control force weighting in the residual. The total residual then becomes

$$|\varepsilon_{\mathrm{total}}^2| = \boldsymbol{S}_a(\omega) \left| \frac{\boldsymbol{I}_d - \boldsymbol{I}_{\mathrm{ACT}}}{1 + \boldsymbol{I}_{\mathrm{ACT}} \boldsymbol{A}_{\mathrm{STR}}} \right|^2 \{1 + \boldsymbol{K}\tilde{\boldsymbol{P}} + \boldsymbol{K}^*\tilde{\boldsymbol{P}}^* + \boldsymbol{K}\boldsymbol{K}^*|\tilde{\boldsymbol{P}}^2|(1 + \beta(\omega))\} \tag{17}$$

where $\beta(\omega)$ is an arbitrary, positive, real function of frequency that when large will tend to suppress the control effort. An expression for an N tap FIR filter is now substituted for the control filter, K, in Eq. (17).

$$\boldsymbol{K}(\omega) = \sum_{n=1}^{N} \boldsymbol{K}_n \boldsymbol{e}^{-j\omega(n-1)\Delta T} \tag{18}$$

where $1/\Delta T$ is the sampling frequency and the K_n are the filter coefficients. Equation (17) is now integrated in frequency over the frequency band of interest, the derivative is taken with respect to each K_n and each result is set equal to zero. From that set of linear equations in the unknown filter

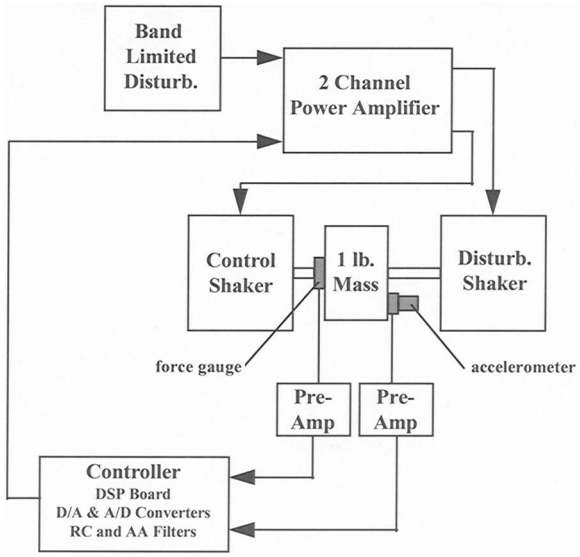


FIG. 8. Laboratory apparatus for implementing desired impedance. The two shakers are Ling 203, 6 lb. shakers. The force gauge is a PCB 208A02 (50.9 mv/lb); The accelerometer is a Wilcoxen 726T (95.5 mv/g) accelerometer;

coefficients K_n, the optimum FIR filter coefficients can be determined

To apply the approach in Fig. 7 the apparatus illustrated schematically in Fig. 8 was setup in the laboratory. In this setup the structure is a 1 lb mass driven by a disturbance shaker with band limited white noise. The impedance is implemented with a control shaker. The desired impedance was arbitrarily chosen to be a mass spring damper system with a 90 g mass, a loss factor of 0.5, and a natural frequency of 125 Hz. The system had a sampling rate of 2 kHz, the antialiasing and reconstruction filters were third-order butterworths with a cutoff frequency of 700 Hz, the plant model was 700 taps, the desired dynamic mass filter was 50 taps, the control filter was 200 taps for direct estimation and 700 taps for the filtered-X implementation, and the control band was limited to 30–200 Hz.

The dynamic masses implemented by the shaker are shown in Fig. 9 and compared to the desired result. The

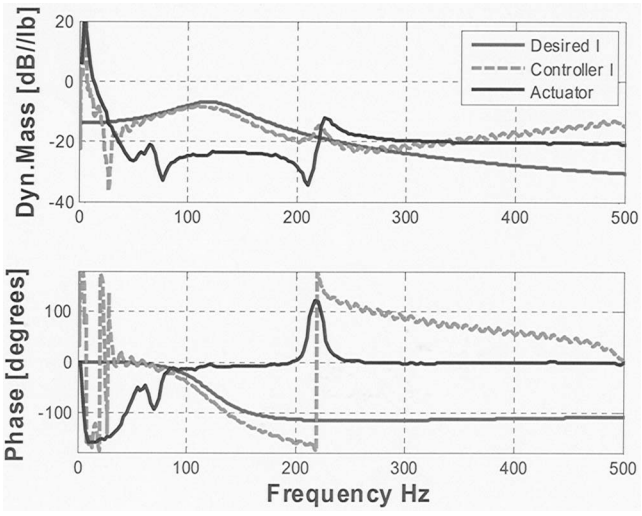


FIG. 9. Comparison of the desired dynamic mass and the dynamic masses achieved by the control system.

experimental dynamic mass was determined by measuring the ratio of force gauge output to accelerometer output when the mass was excited by the disturbance shaker. Also shown in Fig. 9 is the measured dynamic mass of the shaker. That result is essentially the uncontrolled dynamic mass.

Comparison of the desired dynamic mass with that achieved shows that the controller was successful in achieving the desired dynamic mass over the entire control band, nearly a decade in frequency. There is some phase error at high frequency that could be reduced by increasing the high frequency limit of the control force weighting, but all in all the result is very promising.

It is important to note that the results in Fig. 9 are somewhat fortuitous. As will be shown later, the fact that the structural mass [1 lb m (0.45 kg m), 0 dB in Fig. 9] is much greater than both the desired and actuator dynamic masses makes the plant approximately unity. Such a simple plant makes controller implementation easier. In addition, the fact that the actuator dynamic mass is much less than the desired dynamic mass in the control band also makes the implementation of the dynamic mass easier. In the next section the effect of more complex, realistic structural accelerances will be examined analytically using measured data along with the measured dynamic mass of a candidate actuator.

VI. APPLICATION TO A TEST PANEL

In this section the implementation of a desired dynamic mass is examined using measured accelerance data on a test panel and measured accelerance data on a candidate actuator. However, before proceeding to that calculation the optimum dynamic mass is estimated following the approach outlined in Sec. III, using the analytical model of panel radiation with 20 panel modes driven by a TBL as described in Sec. II. Figure 10(a) shows the estimated radiated power from the test panel with no dynamic mass and with the optimum noncausal dynamic mass attached at the location shown in Fig. 10(b). Figure 10(b) also shows the geometry of the test panel and the test conditions. A second-order oscillator approximation to the optimum dynamic mass (desired dynamic mass) was determined by trial and error. The oscillator parameters are a mass of 104 g; stiffness of 8×10^4 N/m; and damping coefficient of 520 N s/m, and the radiated power with that dynamic mass attached is also shown in Fig. 10(a) along with the parameters defining the oscillator. Figure 10(a) shows that both the optimum and simplified dynamic masses provide significant reductions in radiated power up to just short of 600 Hz. Above that frequency the performance degrades because the panel resonant frequencies are too close together. Recall from Sec. IV that good reductions in radiated power with a single actuator require lightly damped well separated panel resonant frequencies.

Up to this point the impact of the actuator dynamics on the performance of the impedance controller has not been considered. In the following the procedure outlined in Sec. V will be used to determine a causal filter K (see Fig. 7) that provides the best match to the desired dynamic mass. Because no measurements are available for the test panel relating radiated sound power to TBL and actuator excitation, the

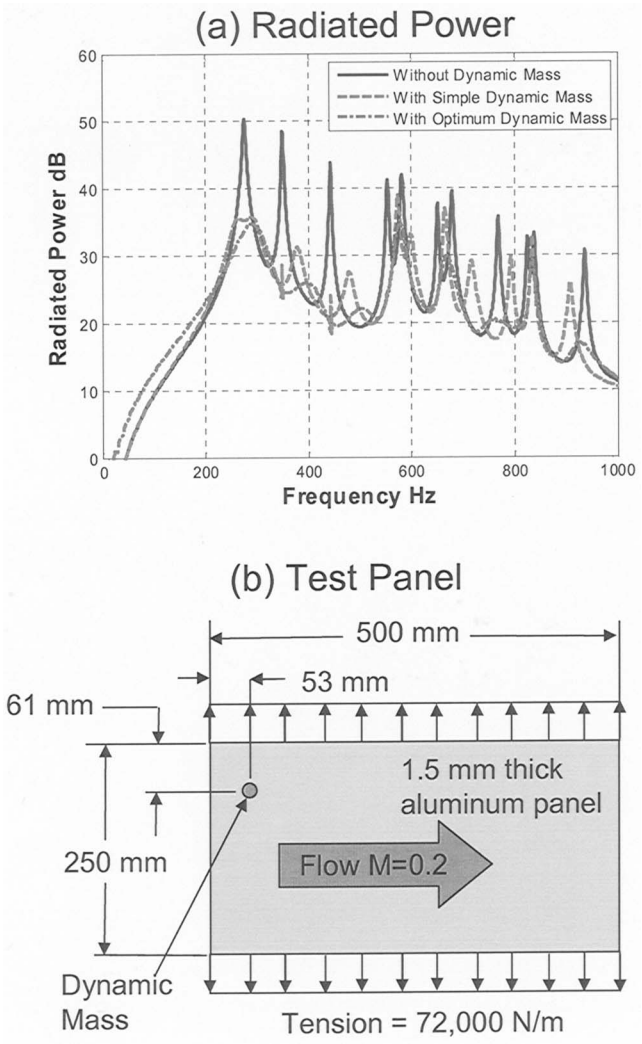


FIG. 10. Effect of a dynamic mass attached to the test panel (a) radiated power with no dynamic mass, with the optimum dynamic mass, and with a second-order oscillator approximating the dynamic mass and (b) geometry and in-plane tension of the test panel, location of the control actuator, and flow conditions.

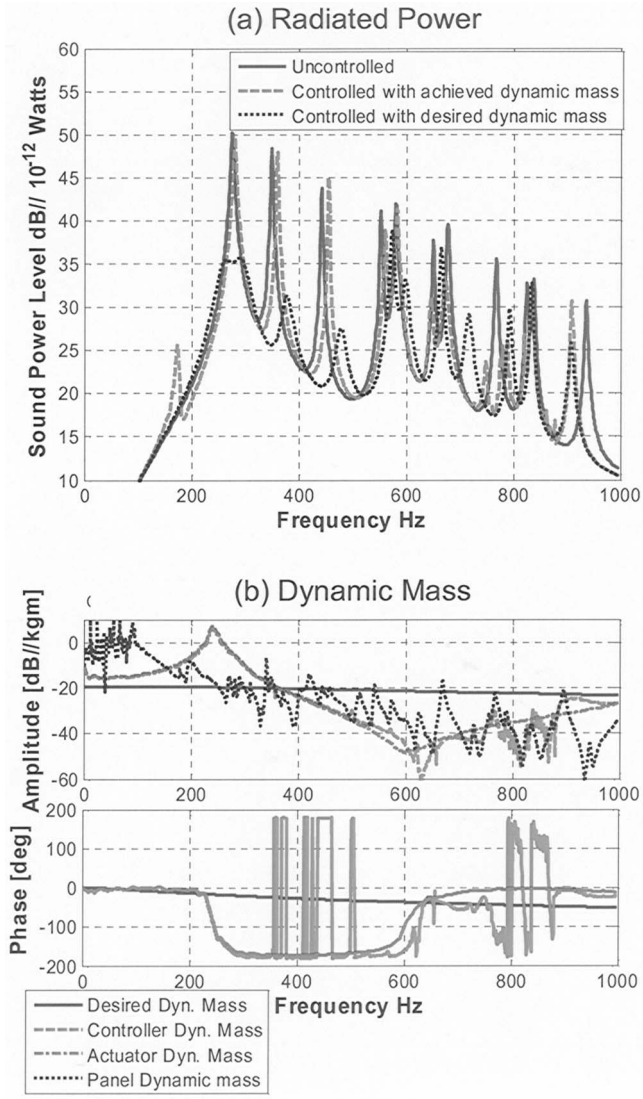


FIG. 11. Estimated performance using PCB 712A01 piezoelectric shaker on the test panel (a) reduction in radiated power and (b) comparison of dynamic masses.

analytical model of Sec. II will be used to compute the effect on radiated sound power of an actuator tuned to the desired dynamic mass.

Carrying out the procedure for determining the causal FIR filter K in Sec. V requires calculation of the plant transfer function. Equation (14) shows that the plant transfer function requires the test panel accelerance A_{STR}, the actuator dynamic mass I_{ACT}, and the desired dynamic mass, I_d. A_{STR} is available from accelerance measurements carried out by the National Aeronautics and Space Administration Langley Research Center (NASA LaRC) on a simulated aircraft panel in the NASA-LaRC Low Speed Acoustic Wind Tunnel (LSAWT). I_{ACT} is available from accelerance measurements also carried out by NASA-LaRC on a miniature commercially available actuator as might be used to implement the desired dynamic mass. I_d, the desired dynamic mass (second-order oscillator approximating the optimum dynamic mass), was obtained as indicated earlier.

The simulated aircraft test panel, on which the accelerance measurements were taken, and the operating conditions that will be used in the calculations to follow are shown in Fig. 10(b). The actuator, on which accelerance data were acquired, was a PCB Model 712A01 with a 100 g seismic mass. To estimate K the filter was constrained to be a FIR filter with 10 taps and the plant $\tilde{P}$ was also constrained to be a FIR filter but with 300 taps. The sampling rate was set to 2 kHz.

The change in radiated power from the panel when the actuator was activated is shown in Fig. 11(a). Also shown for comparison is the radiated power when the second-order oscillator approximation to the optimum dynamic mass is attached to the panel.

The dynamic masses are compared in Fig. 11(b). It is clear from Fig. 11(a) that the addition of the actuator tuned to the desired dynamic mass has not reduced the radiated power. It was expected that the power would have been reduced to the dotted curve in Fig. 11(a), the radiated power where the approximation to the optimum dynamic mass has been implemented on the panel. From Fig. 11(b) it is clear why the power was not reduced. The controller has done a very poor job of matching the actuator dynamic mass (the

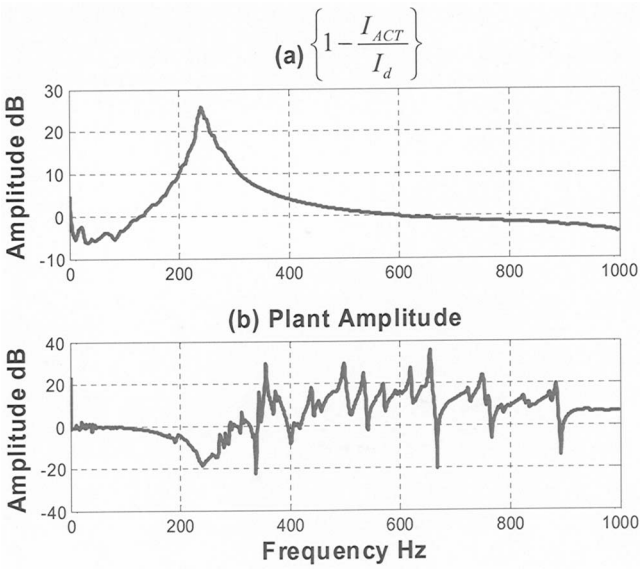


FIG. 12. Amplitude of the factors multiplying the controller residual in Eq. (19) to yield the dynamic mass residual.

dashed curve) to the desired dynamic mass (the solid curve). The question is why and what can be done about it?

In the Appendix the following relationship is derived between the controller residual and the dynamic mass residual:

$$\varepsilon_I = \frac{I_d - \tilde{I}_d}{I_d} = \tilde{\varepsilon}\left\{1 - \frac{I_{\mathrm{ACT}}}{I_d}\right\}\tilde{P} = \tilde{\varepsilon}\left\{1 - \frac{I_{\mathrm{ACT}}}{I_d}\right\} \times \left\{\frac{I_{\mathrm{STR}} + I_d}{I_{\mathrm{STR}} + I_{\mathrm{ACT}}}\right\}, \tag{19}$$

where ε_I is the dynamic mass residual, I_d is the desired dynamic mass, $\tilde{I}_d$ is the dynamic mass achieved by the controller, and $\tilde{\varepsilon}$ is the normalized controller residual as defined in the Appendix. The normalized controller residual is the residual that the controller seeks to minimize, $\varepsilon = \boldsymbol{F} + \boldsymbol{I}_d \boldsymbol{a}$, normalized to remove the actuator force F and the panel drive point acceleration, a, from the equation. Equation (19) indicates that the dynamic mass residual and the controller residual are not the same. Rather, the dynamic mass residual is the controller residual multiplied by two factors. If either of these factors is large, the control system may not provide a good match to the desired dynamic mass even if the controller residual is small.

It is now clear why the controller did such a poor job of providing the desired impedance. In Fig. 12 the two factors mentioned earlier are plotted. At low frequency $\{1 - I_{\mathrm{ACT}}/I_d\}$ is large, as shown in Fig. 12(a), because the actuator dynamic mass is large. At high frequency the plant transfer function is large, as shown in Fig. 12(b), because I_d is much larger than both I_{ACT} and I_{STR}. This is especially true for I_{STR} at the panel resonances, where as luck would have it, we are most interested in achieving good control. As a result throughout the frequency range of interest the dynamic mass residual is much greater than the controller residual making it difficult for the control system to match the achieved dynamic mass to the desired dynamic mass.

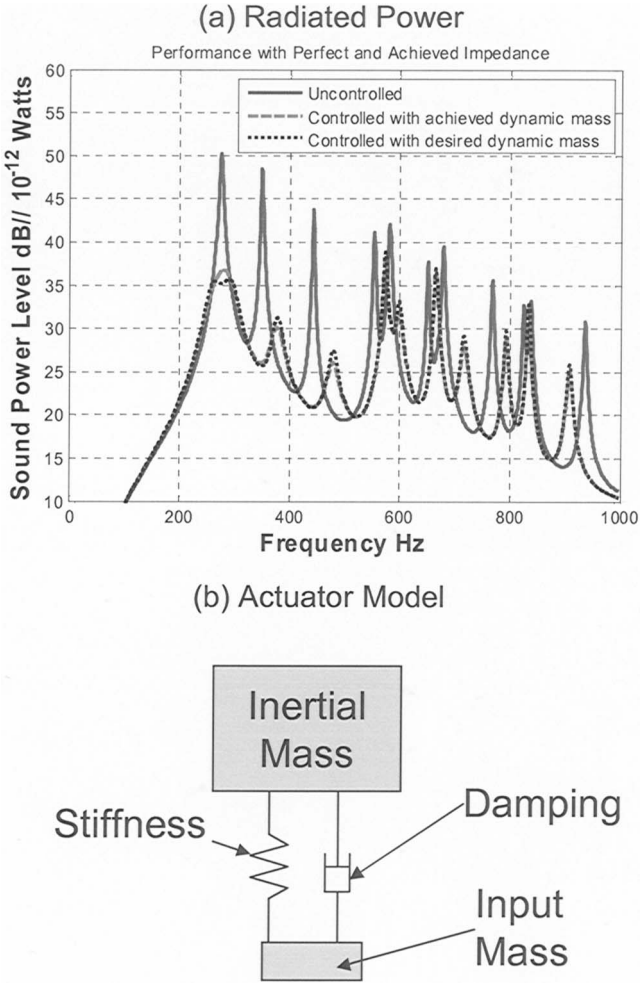


FIG. 13. Estimated performance using a shaker on the test panel tuned to be approximately the desired impedance (a) reduction in radiated power and (b) actuator model.

Under ideal circumstances to obtain good reductions in radiated power the designer would make I_d and I_{ACT} small compared to I_{STR}, which would make the $\tilde{P}$ close to unity. If, in addition, I_d is comparable in magnitude to I_{ACT}, the first factor $\{1 - I_{\mathrm{ACT}}/I_d\}$ would be small. Unfortunately the only term in Eq. (19) over which the designer has control is I_{ACT}. I_{STR} is set by the structure to be controlled, and if that structure is a lightly damped panel, it will be difficult to make I_{ACT} small compared to I_{STR} near any panel resonances. I_d is set by the physics of the problem and its amplitude will be of the same order of magnitude as I_{STR}.

The one strategy that seems to be viable is to make $I_{\mathrm{ACT}} \approx I_d$. The first term in brackets Eq. (19) then becomes small and has the added benefit that $\tilde{P}$ will become close to unity if $I_{\mathrm{STR}} < I_d$ and I_{ACT}. In Fig. 13(a) that strategy is explored. There, the radiated power is compared for the panel of Fig. 10(b) with no dynamic mass, with the desired dynamic mass (second-order oscillator defined previously) and with the dynamic mass achieved by the controller. For the performance calculation with the achieved dynamic mass the actuator is modeled as two masses separated by a spring and damper as shown in Fig. 13(b). The two masses are the inertial mass and the input mass. The input mass is the mass that is attached directly to the panel. The inertial mass is the

mass that the actuator pushes against to generate force. In most actuators the input mass is small so as to minimize the force drop across it when driving a structure with small impedance. Nevertheless it can have a significant effect on the actuator dynamic mass. Here the actuator inertial mass is made equal to the mass in the desired dynamic mass, the actuator stiffness is made 25% of the stiffness of the desired dynamic mass and the actuator damping coefficient is made to be 80% of that in the desired dynamic mass. The input mass is set to be 5 g, about the mass of a U.S. 5 cent piece. For estimating the control filter coefficients the control filter was constrained to be a FIR filter with 10 taps, the plant $\tilde{P}$ was also constrained to be a FIR filter but with 400 taps, and the sampling rate was set to 2 kHz. Figure 13(a) shows that the radiated powers from the panel with the achieved dynamic mass and with the desired dynamic mass are comparable, i.e., at the lowest resonance frequency of the panel the radiated power for the achieved dynamic mass is about 1 dB higher than that for the desired dynamic mass. In addition the reductions from the untreated panel are significant up to just below 600 Hz, where the closeness of the resonant frequencies makes significant reductions in radiated power with a single actuator difficult.

It is of interest to examine for this case how different the actuator dynamic mass and the desired dynamic mass can be and still obtain good reductions in radiated power from the panel. If the parameters of the actuator are significantly different from those of the desired dynamic mass, the controller will have to work much harder in adjusting the dynamic mass of the actuator to achieve the desired dynamic mass, potentially resulting in performance degradation. For example, if the input mass is increased from 5 to 25 g the radiated power increases by approximately 5 dB at the lowest panel resonance relative to the radiated power or the desired dynamic mass. Increasing the stiffness in the actuator to 4 times the stiffness in the desired dynamic mass increases the radiated power at the lowest panel resonance by 4 dB. Decreasing the stiffness has little effect on the radiated power. Increasing the damping coefficient in the actuator to 1.4 times the damping coefficient in the desired dynamic mass or decreasing it by 50% results in about a 5 dB increase in radiated power.

These results have demonstrated that by proper design of the actuator it is possible to achieve significant reductions in radiated power from a TBL excited panel near the panel resonances over a fairly broad range of frequencies by designing the dynamic mass of the actuator to have a dynamic mass that is similar to but not the same as the desired dynamic mass. Although the actuator stiffness can be significantly different from the stiffness in the desired mass without affecting the reduction in radiated power, the performance is more sensitive to the damping coefficient. The performance is not terribly sensitive to the actuator input mass. Nevertheless it should be as small as practical. These ranges of actuator parameters suggest that the controller has the potential to be able to adjust the actuator characteristics to deal with changes in operating conditions and panel to panel variations in an aircraft fuselage and provide significant reductions in radiated power near panel resonances over a fairly broad range of frequencies.

VII. CONCLUSIONS

This paper has presented the concept of adding a single active "tunable" dynamic mass to an aircraft panel to reduce the TBL generated sound power radiated into the aircraft interior. An approach to estimating the optimum dynamic mass was outlined, and it was shown that a simple second-order oscillator approximation to that dynamic mass can provide comparable performance. The generation of a desired dynamic mass using an active control system was demonstrated in the laboratory using a mass as the structure and two shakers, one for the disturbance and one, equipped with a force gauge and accelerometer, for the active impedance. Good agreement between the desired dynamic mass and that achieved by the control system was shown. When the performance of the concept was estimated analytically using the measured accelerance on a test panel and the measured accelerance of a commercially available miniature piezoelectric actuator, the system did not effectively reduce the power radiated when the panel was excited by a TBL. The source of the problem was found to be that the dynamic characteristics of the actuator placed such severe demands on the control system, that it could not control the actuator to effectively match the desired dynamic mass. By designing the actuator to have a dynamic mass that was similar to, but not the same as, the desired dynamic mass it was shown that good performance could be restored. Finally the degree of adjustability of the dynamic mass of the actuator by the control system suggests that the controller has the potential to be able to adjust the characteristics of a properly designed actuator to deal with changes in operating conditions and panel to panel variations in an aircraft fuselage and provide significant reductions in radiated power near panel resonances over a fairly broad range of frequencies.

ACKNOWLEDGMENTS

This work was supported in part by the National Aeronautics and Space Administration Langley Research Center (NASA-LaRC) under Contract No. NAS 120101. That support is gratefully acknowledged. The authors also wish to thank Dr. Gary Gibbs of NASA-LaRC who provided us with the panel and shaker accelerance data.

APPENDIX: DERIVATION OF THE RELATIONSHIP BETWEEN DYNAMIC MASS RESIDUAL AND THE CONTROLLER RESIDUAL

From Fig. 7(a) it is possible to derive the actuator force F, the drive point acceleration, a, and the residual ε in terms of the disturbance acceleration a_d:

$$\frac{F}{a_d} = \frac{KI_d - I_{\mathrm{ACT}} - K\tilde{P}I_{\mathrm{ACT}}}{(1 + I_{\mathrm{ACT}}A_{\mathrm{STR}})\{1 + K(\tilde{P} - P)\}}, \tag{A1}$$

$$\frac{a}{a_d} = \frac{1 + K\tilde{P} - K}{(1 + I_{ACT}A_{STR})\{1 + K(\tilde{P} - P)\}}, \tag{A2}$$

$$\frac{\varepsilon}{a_d} = \frac{(I_d - I_{ACT})(1 + K\tilde{P})}{(1 + I_{ACT}A_{STR})\{1 + K(\tilde{P} - P)\}}. \tag{A3}$$

From the previous equations, the dynamic mass achieved by the controller, $\tilde{I}_d$, can be obtained from the ratio of force to acceleration:

$$\tilde{I}_d = \frac{F}{a} = \frac{KI_d - I_{ACT} - K\tilde{P}I_{ACT}}{1 + K\tilde{P} - K}. \tag{A4}$$

It is interesting to compare the error in the achieved dynamic mass with the controller residual. The controller residual is defined to be

$$\varepsilon = F + I_d a. \tag{A5}$$

If $\tilde{I}_d$ is the dynamic mass achieved by the controller the force F becomes $-\tilde{I}_d a$ and the residual can be written as

$$\varepsilon = (I_d - \tilde{I}_d)a.$$

Substituting for a in this equation by using Eq. (A2) yields

$$\varepsilon = (I_d - \tilde{I}_d)\frac{1 + K\tilde{P} - K}{(1 + I_{ACT}A_{STR})\{1 + K(\tilde{P} - P)\}}a_d. \tag{A6}$$

Equation (A6) is the residual with the controller operating. It is helpful to normalize this result by the uncontrolled residual, which from Eq. (A3) is given by

$$\varepsilon = \frac{I_d - I_{Act}}{1 + I_{ACT}A_{STR}}a_d. \tag{A7}$$

Dividing Eq. (A6) by Eq. (A7) and dividing top and bottom by I_d yields the normalized residual, $\tilde{\varepsilon}$:

$$\tilde{\varepsilon} = \frac{\left(1 - \frac{\tilde{I}_d}{I_d}\right)}{\left(1 - \frac{I_{ACT}}{I_d}\right)}\frac{1 + K\tilde{P} - K}{\{1 + K(\tilde{P} - P)\}} \tag{A8}$$

Examination of Fig. 7(a) shows that in order for the controller to perform well it is necessary that $K \approx -P^{-1}$. The extent to which this is possible depends on whether the inverse of P is causal or nearly causal and on the accuracy of the algorithm that computes K. If $K \approx -P^{-1}$ and $\tilde{P}$ is a good match to P, then, and the Eq. (A8) simplifies to

$$\tilde{\varepsilon} = \frac{\left(1 - \frac{\tilde{I}_d}{I_d}\right)}{\left(1 - \frac{I_{ACT}}{I_d}\right)}\tilde{P}^{-1}.$$

Rearranging terms yields a simplified relationship between the normalized controller residual $\tilde{\varepsilon}$ and the dynamic mass residual, $\varepsilon_I = (I_d - \tilde{I}_d)/I_d$,

$$\varepsilon_I = \frac{I_d - \tilde{I}_d}{I_d} = \tilde{\varepsilon}\left\{1 - \frac{I_{ACT}}{I_d}\right\}\tilde{P} = \tilde{\varepsilon}\left\{1 - \frac{I_{ACT}}{I_d}\right\} \times \left\{\frac{I_{STR} + I_d}{I_{STR} + I_{ACT}}\right\} \tag{A9}$$

[1] D. Guicking, J. Melcher, and R. Wimmel, "Active impedance control in mechanical structures," Acustica **69**, 39–52 (1989).
[2] R. J. Bobber, "Active load impedance," J. Acoust. Soc. Am. **34**, 282–288 (1962).
[3] T. Samejima "A state feedback electro-acoustic transducer for active control of acoustic impedance," J. Acoust. Soc. Am. **113**, 1483-1491 (2003) (Abstract).
[4] M.-A. Galland, O. Lacour, and D. Thenail, "Active control of enclosed sound fields via impedance changes (A)," J. Acoust. Soc. Am. **105**, 1244 (1999) (Abstract).
[5] D. Guicking and E. Lorenz, "An active sound absorber with porous plate," J. Vibr. Acoust. **106**, 389–392 (1984).
[6] D. Thenail, M. Galland, M. Sunyach, and M. Sunback, "Active enhancement of the absorbent properties of a porous material," Smart Mater. Struct. **3**, 18–25 (1994).
[7] J. J. Finneran, and M. C. Hastings, "Active impedance control within a cylindrical waveguide for generation of low-frequency, underwater plane traveling waves," J. Acoust. Soc. Am. **105**, 3035–3043 (1999).
[8] C. R. Fuller, M. J. Bronzel, C. H. Gentry, and D. E. Whittington, "Control of sound radiation reflection with adaptive foams," Proceedings of NoiseCon 1994, Fort Lauderdale FL, pp. 429–436.
[9] S. Beyenne and R. Burdisso, "A new hybrid passive active noise absorption system," J. Acoust. Soc. Am. **106**, 2646–2652 (1984).
[10] J. Yuan, "Causal impedance matching for broad band hybrid noise absorption," J. Acoust. Soc. Am. **113**, 3226–3232 (2003).
[11] J.-M. F. Bras, M. E. Johnson, and C. R. Fuller, "Active control of mechanical impedance for characterizing actuator performance(A)," J. Acoust. Soc. Am. **104**, 1851 (1998) (Abstract).
[12] L. Benassi and S. J. Elliott, "Global control of a vibrating plate using a feedback-controlled inertial actuator," J. Sound Vib. **28**, 69–90 (2005).
[13] C. Deffayet and P. A. Nelson, "Active control of low-frequency harmonic sound radiated by a finite panel," J. Acoust. Soc. Am. **84**, 2192–2199 (1988).
[14] S. D. Snyder, P. Jie, and C. H. Hansen, "A general analysis of the active control of sound radiation from a baffled panel," J. Acoust. Soc. Am. **86**, S32 (1989).
[15] J. Pan, S. D. Snyder, C. H. Hansen, and C. R. Fuller, "Active control of far-field sound radiated by a rectangular panel—A general analysis," J. Acoust. Soc. Am. **91**, 2056–2066 (1992).
[16] L. Benassi and S. J. Elliott, "The equivalent impedance of power-minimizing vibration controllers on plates," J. Sound Vib. **283**, 47–67 (2005).
[17] J. Pan and C. Bao, "Analytical study of different approaches for active control of sound transmission through double walls," J. Acoust. Soc. Am. **103**, 1916–1922 (1998).
[18] B.-T. Wang, C. R. Fuller, and E. K. Dimitriadis, "Active control of noise transmission through rectangular plates using multiple piezoelectric or point force actuators," J. Acoust. Soc. Am. **90**, 2820–2830 (1991).
[19] J. Pan, C. H. Hansen, and D. A. Bies, "Active control of noise transmission through a panel into a cavity: I. Analytical study," J. Acoust. Soc. Am. **87**, 2098–2108 (1990).
[20] A. R. D. Curtis and P. J. Remington, "Simulation model for the control of sound radiation from an aircraft panel excited by a turbulent boundary layer using impedance control," BBN Technical Memo 1191, BBN Technologies, 1997.
[21] G. M. Corcos, "The resolution of turbulent wall pressures at the wall of a boundary layer," J. Fluid Mech. **18**, 59–70 (1964).
[22] B. M. Efimtsov, "Characteristics of the field of turbulent wall pressure fluctuations at large Reynold's numbers," Sov. Phys. Acoust. **28**, 289–292 (1982).
[23] W. R. Graham, "Boundary layer induced noise in aircraft," AIAA paper No. DGLR/AIAA 92–02–015 1992, pp. 87–96.
[24] E. Morari and E. Zafiriou, *Robust Process Control* (Prentice Hall, Englewood Cliffs, NJ, 1989).

Assessment of asphalt concrete acoustic performance in urban streets

S. E. Paje,[a] M. Bueno, F. Terán, U. Viñuela, and J. Luong
Laboratory of Acoustics Applied to Civil Engineering (LA^2IC), Department of Applied Physics-E.T.S.I. Caminos and E.S. Informática, Universidad de Castilla-La Mancha (UCLM), 13071 Ciudad Real (Spain)

(Received 14 May 2007; revised 4 December 2007; accepted 4 December 2007)

Geo-referenced close proximity rolling noise and sound absorption measurements are used for acoustical characterization of asphalt concrete surfaces in an urban environment. A close proximity noise map of streets with low speed limits is presented for a reference speed of 50 km/h. Different pavements and pavement conditions, common in urban streets, are analyzed: dense and semidense asphalt concrete, with Spanish denomination D-8 and S-12, respectively, and on the other hand, dense pavement at the end of its service life (D-8*). From the acoustics point of view, the most favorable surface, by more than 4 dB(A) compared with the S-12 mix, is the smoothest surface, i.e., the D-8 mix, even though it presents a minor absorption coefficient in normal incidence. Noise levels from dense surfaces (D-8) increase significantly over time, principally due to the appearance of surface defects such as cracks and ruts. Longitudinal variability of the close proximity tire/pavement noise emission and surface homogeneity are also analyzed.
 [DOI: 10.1121/1.2828068]

PACS number(s): 43.50.Lj, 43.50.Ed, 43.50.Rq, 43.55.Ev [KA] Pages: 1439–1445

I. INTRODUCTION

In recent years, environmental noise has become a serious issue for civil infrastructure and environmental administration due to public concern over the subject of noise pollution. The most significant deterioration of environmental acoustics conditions comes from road traffic transportation, particularly in built-up areas. In addition, the volume of road traffic has increased significantly in the last few decades and at present gives a contribution which in some cases can reach 50% of the total noise pollution.[1] In order to prevent or reduce harmful effects of traffic noise, different noise abatement measures are required to achieve improvements in the environmental situation. Such measures comprise items like the renewal of road surfaces with different types of noise reducing pavements or the replacement of rough pavements by smoother layers.[2]

The predominant noise source is the combination of the tire/pavement interaction and the propulsion systems (engine, exhaust, transmission, intake, etc.) of the vehicles. Generally, tire/pavement interaction is the principal source of noise for speeds above 40 km/h in the case of most modern cars.[3] For these reasons, a great effort is currently being made in the classification of road on the basis of their acoustic performances and in evaluation of new pavements with noise control features.[4–10] However, studies to evaluate close proximity noise-measuring procedures applied to real cases are still required, which would allow administrators to reduce the noise impact on people's quality of life from pavement surfaces.

In this study, procedures based on close proximity method for measuring tire/pavement noise[10–13] of three in-service pavement types are used. In addition, measurements of the acoustic absorption of core samples with an impedance tube are carried out. Test vehicle and monitoring techniques based on continuous tire/pavement close proximity measurements developed in LA^2IC are employed. The A-weighted sound levels are geo referenced, thus, sound pressure levels are suitable for their inclusion in a geographic information system.

The overall aim of this work is to contribute, through geo-acoustic characterization of urban pavement surfaces, to the control of traffic noise via the assessment of pavement surfaces in roads with low speed limits.

II. EXPERIMENTAL PROCEDURE

A. Sound absorption measurements

To evaluate the acoustic characteristics of core samples a 4206 Bruel&Kjaer (B&K) impedance tube was employed. The impedance tube consists of both 100 and 29 mm inner diameter tubes (i.e., two measurement settings) with a loudspeaker mounted at one end. The B&K impedance tube is suitable for both low frequency and high frequency measurements over the extended frequency range (50 Hz–1.6 kHz and 500 Hz–6.4 kHz), by the combination of measurements with the different diameters. The materials tested were core samples that were taken from actual street pavements at different locations in Ciudad Real (Spain). These were covered laterally with a thin film of Teflon to eliminate the air gap between the core samples and the tube. Broadband, stationary random or pseudo-random sound waves, which propagated as plane waves, were generated from the noise source of a B&K multi-analyzer Pulse system type 3560 B-T06, with a power amplifier type B&K 2716C. Two 1/4 in. condenser microphones type 4187, placed along the tube at fixed locations were used to simultaneously measure the sound

[a]Author to whom correspondence should be addressed. Electronic mail: santiago.exposito@uclm.es

FIG. 1. (Color online) Microphone setup used for the test measurements, showing their positions in relation to the reference tire Pirelli P6000 205/55 R16.

wave pressure on the standing-wave interference pattern produced by the superposition of forward and backward traveling waves inside the tube. The Pulse multi-analyzer was used to obtain the normal-incidence acoustic properties by calculating the two-microphone transfer function between the input signal of the loudspeaker and the microphone output at each of the two measurement positions located near the sample. The Pulse software program Material Testing type 7758 was used for calibration, correction for microphone mismatch, measurements and display of the real and imaginary components of the acoustic properties with frequency.

To reduce the effect of sample mounting errors and material variability, two different core samples and four different measurements in each case were performed; the results of the four individual tests were then averaged to give the results presented in this paper. Sound absorption was determined over the extended frequency range with a frequency resolution of 1 Hz and with 100 spectrum averages at the microphone positions.

B. Geo-referenced close proximity noise emission measurements

The A-weighted sound pressure levels emitted by the rolling of a reference tire, Pirelli P6000 205/55 R16, were measured by three BSWA MP201 1/2 in. microphones located close to the tire, which were mounted on a standard passenger vehicle. The measurement setup diagram is shown in Fig. 1. A portable NI Compact Rio control and acquisition system with a four-channel module and a cRio Mobile Module (GPS), for global position determination, were used to constantly carry out geo-referenced acoustical measurements. A digital tachometer was used during the test to continuously measure the vehicle speed for speed correction. The reference speed was 50 km/h, since the study was carried out in an urban environment. Before the test measurements, the reference tire was warmed up by driving for more than 20 min and the sensitivity of the whole acoustic measurement setup was checked with an acoustic calibrator 4231 B&K.

The tire selected in this research was at the end of its useful service life; it has previously run 40 000 km. This reference tire was selected with the purpose of representing one tire actually used on standard passenger automobiles in Spain; vehicles which dominate the urban traffic composition

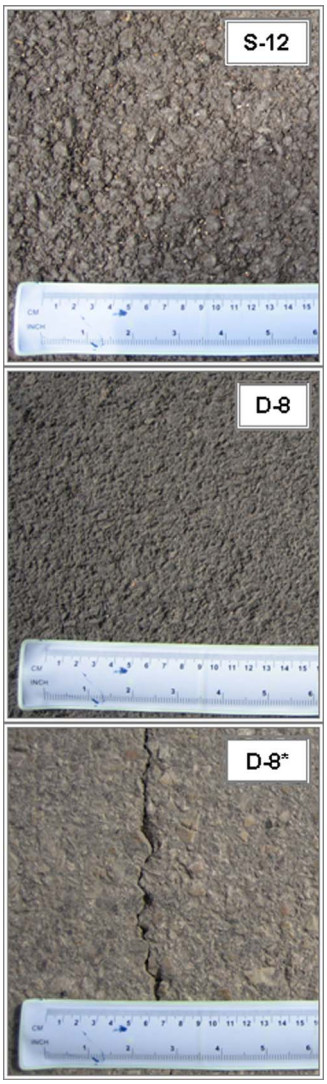


FIG. 2. (Color online) Representative photographs of the tested surfaces at Ciudad Real, showing the texture of the fine, the coarse, and the worn condition asphalt surfaces, D-8, S-12 and D-8* asphalt mixes, respectively.

(see Fig. 1). The repeatability of the measurements had previously been evaluated, to ensuring that the same close proximity tire/pavement noise emission was obtained in successive measurements with identical conditions.

C. Pavement surfaces and location

Three different surfaces of streets where 50 km/h represents the typical vehicle speed were selected as the test tracks. The surface selection—dense asphalt concrete (or dense graded) and semidense asphalt concrete (or semigap graded) with Spanish denomination D-8 and S-12, respectively—was made bearing in mind one of the more representative mixtures that can typically be found on streets and urban roads in Spanish cities. A dense surface with 2–6% air void content and a semidense asphalt concrete, which is a type of asphalt concrete that is between a dense and a porous asphalt concrete, with an air void content of 6–12%, were studied. Photographs of the surfaces investigated are shown in Fig. 2. The pavement surface S-12 had texture characteristics rougher than surface D-8, as is shown

TABLE I. Properties of dense (D) and semidense (S) pavement types used in this study.

Section name	Type of mix (Spanish denomination)	Thickness of core d(mm)	Year	A-weighted sound level Close proximity L_{CPtr} dB(A) 50 km/h-Pirelli P6000
SO	D-8	68	2006	86
AD_1	S-12	70	2005	90
AD_2	D-8[a]	92	1994	94

[a]Quite worn condition.

visually in Fig. 2. The surface D-8 was laid one month before the measurements for this research. It presented a maximum aggregate size of 10 mm. Other measurements were taken over a D-8* surface, which was laid 12 years before our study, and provided surface defects such as cracks and ruts (see Fig. 2). The surface S-12 was laid ten months before the measurement for this study and presented a maximum aggregate size of 20 mm. The average thicknesses of the surface layers and other properties of the materials used are listed in Table I. Other specifications for these surfaces are given in the Official State Bulletin *B.O.E.*[14]

III. FIELD MEASUREMENTS AND DISCUSSION

A. Characterization of sound absorption for normal incidence

Sample cores were extracted from the streets by removing surface material in two locations, for each of the in-service pavements studied. All the cores were then subjected to a series of absorption measurements using the two-microphone impedance tube method. The samples were placed in the impedance tube and the complex acoustic transfer function was then measured. It is defined as

$$H_{12} = \frac{p_2 p_1^*}{p_1 p_1^*} = |H_{12}| e^{j\phi} = H_r + jH_i, \tag{1}$$

where p_1 and p_2 are the Fourier transforms of the temporal acoustic pressure at the two microphone positions:

$$p_1 = \hat{p}_I e^{jk_0 x_1} + \hat{p}_R e^{-jk_0 x_1}$$

$$p_2 = \hat{p}_I e^{jk_0 x_2} + \hat{p}_R e^{-jk_0 x_2}. \tag{2}$$

Subscripts I and R denote incident and reflected, respectively. The terms $\hat{p}_R$ and $\hat{p}_I$ are the magnitudes of p_R and p_I at the surface of the sample ($x=0$). The distances of the two microphone positions from the surface of the sample are x_1 and x_2. As $\hat{p}_R = r\hat{p}_I$, then

$$H_{12} = \frac{e^{jk_0 x_2} + re^{-jk_0 x_2}}{e^{jk_0 x_1} + re^{-jk_0 x_1}}. \tag{3}$$

Transposing the above equation yields r, the sound reflection factor at $x=0$. The sound absorption coefficient, α, is then obtained in terms of r by

$$\alpha = 1 - |r|^2 = 1 - r_r^2 - r_i^2. \tag{4}$$

Figure 3 shows the normal incidence sound absorption spectra in the extended frequency range, for the pavements studied. The averaged normal-incidence sound absorption coefficients do not exhibit significant absorption peaks. In all asphalt concrete samples, the coefficient shows maximum values below 0.3. However, as Sandberg[11,15] has pointed out, an average sound absorption coefficient in the range 0.1–0.2 could affect the noise from porous road. As is shown in Fig. 3, the absorption coefficient from the coarse aggregate sample S-12 is slightly higher than that for the fine aggregate sample D-8, but in both cases, D-8 and S-12 surfaces, there are no absorption peaks coinciding with the typical frequency spectra generated by traffic noise, which is peaked at 600 Hz for low speed roads and 1 kHz for high-speed roads.[16] The standardized traffic noise spectrum, EN 1793-3 or ISO 717-1, represents a mix of light and heavy vehicle traffic traveling at speeds of about 50 km/h. In contrast to the sound absorption coefficient of the dense (D-8) and semidense (S-12) surfaces, which are below 0.3 over the extended frequency range, porous road surfaces can be designed to present an absorption coefficient peaked, for example, at around 1 kHz (see, for comparison, a sound absorption spectrum for a PA-12 porous surface, Spanish denomination, in Fig. 3). The sample D-8*, which presented surface defects and was polished, does not show an absorption peak. It is to be expected that such a surface condition will generate high sound reflection and will induce vibrations on the tire, increasing the close proximity noise emissions.

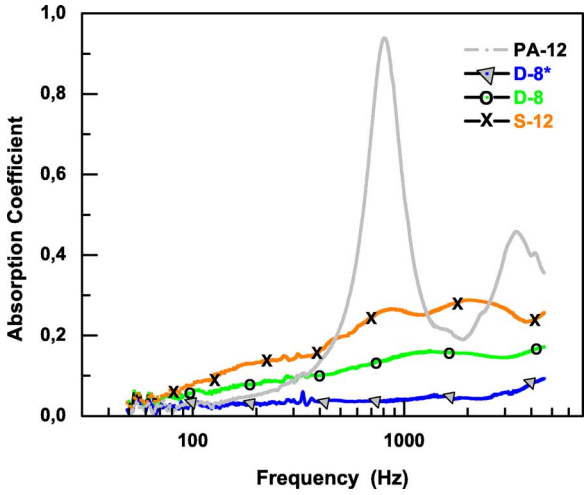


FIG. 3. (Color online) Comparison of normal incidence acoustic absorption spectra of core samples taken from asphalt mixes D-8, D-8*, S-12 and PA-12. The porous surface PA-12 has an air void content of 20%, and presents a maximum aggregate size of 18 mm. The thickness of the cores is around 50 mm.

The observed differences among the dense and semidense asphalt mixes in the sound absorption behavior may be attributed, mainly, to the effect of air void content (also known as pavement surface porosity) which takes values of 2–6% for dense surfaces and 6–12% for semidense surfaces. Other physical parameters, however, should be considered in sound propagation in materials, such as pavement thickness, air flow resistance, and tortuosity (see, for example, von Meier[16] and Berengier[17] *et al.*). The maximum value of the absorption coefficient is proportional to the air void content and, in general, responsible for the increase in the overall noise absorption.[11,16,18]

These absorption measurements on the dense D-8 surface, the D-8* surface in quite worn conditions, and the semidense S-12 asphalt concrete suggest that traffic noise from an S-12 coarse mix aggregate could be the lower, given their higher sound absorption value. However, the maximum size of the S-12 aggregate mix (20 mm), versus that for the D-8 surface (10 mm), may result in more tire tread impact noise and, thus, higher overall close proximity noise emitted to the surroundings. For this reason, the above absorption measurements are still preliminary for assessing the acoustic properties of these pavements and, for example, noise from close proximity tire/pavement interaction is necessary for a more complete assessment of these surface layers. Both properties, acoustical absorption and close proximity noise, are expected to be of great utility for the understanding and optimizing of asphalt concrete acoustic performance.

B. Global noise fluctuation profile

Figure 4 shows the fluctuation of the close proximity noise levels from arithmetically averaged microphone 1 and microphone 3, continuously measured on the D-8 and D-8* dense asphalt concrete, and S-12 semidense asphalt concrete, lengthways on a test segment of the streets. The variations of the level of the central microphone, M2, showed a similar behavior. The vehicle speed, around 50 km/h, was monitored by the digital tachometer and by the GPS system during the test.

The results of the measurements in Fig. 4(a) indicate that D-8 dense asphalt, the smoothest surface, provides continuous tire/pavement close proximity noise levels, L_{CPtr}, of around 86 dB(A). Higher noise levels [90 dB(A)] were measured over S-12 semidense asphalt concrete, at the same current speed of around 50 km/h. Over the worn D-8* dense asphalt, laid 12 years before our study, the measured rolling level for the same current speed was around 94 dB(A), some 8 dB(A) higher than the noise level measured over the same type of mix, a D-8 dense asphalt, but laid one month before these acoustical measurements. The above L_{CPtr} levels [86 dB(A), 90 dB(A) and 94 dB(A)] were averaged values over a distance of 100 m.

The acoustic behavior of the three surfaces was corrected for speed deviations.[11] Thus, Fig. 5 shows the averaged (Mic 1 and Mic 3) close proximity noise fluctuations, at 50 km/h nominal reference speed, after speed correction by equation

$$L_{corr}(t) = L_{meas}(t) - B \log_{10}\left(\frac{v(t)}{v_{ref}}\right) \tag{5}$$

as a function of the distance covered. The acoustical monitoring test developed on these surfaces indicated that the highest tire/pavement close proximity noise levels were found for the D-8* 12 year old surface, a polished surface with surface defects such as cracks and ruts. The fluctuation of L_{CPtr} noise levels of the different pavement surfaces reflects different degrees of longitudinal surface homogeneity.

The speed constant B was obtained from speed-noise measurements by means of a regression, as shown in the insets of Figs. 4(a)–4(c). The coefficients A and B obtained

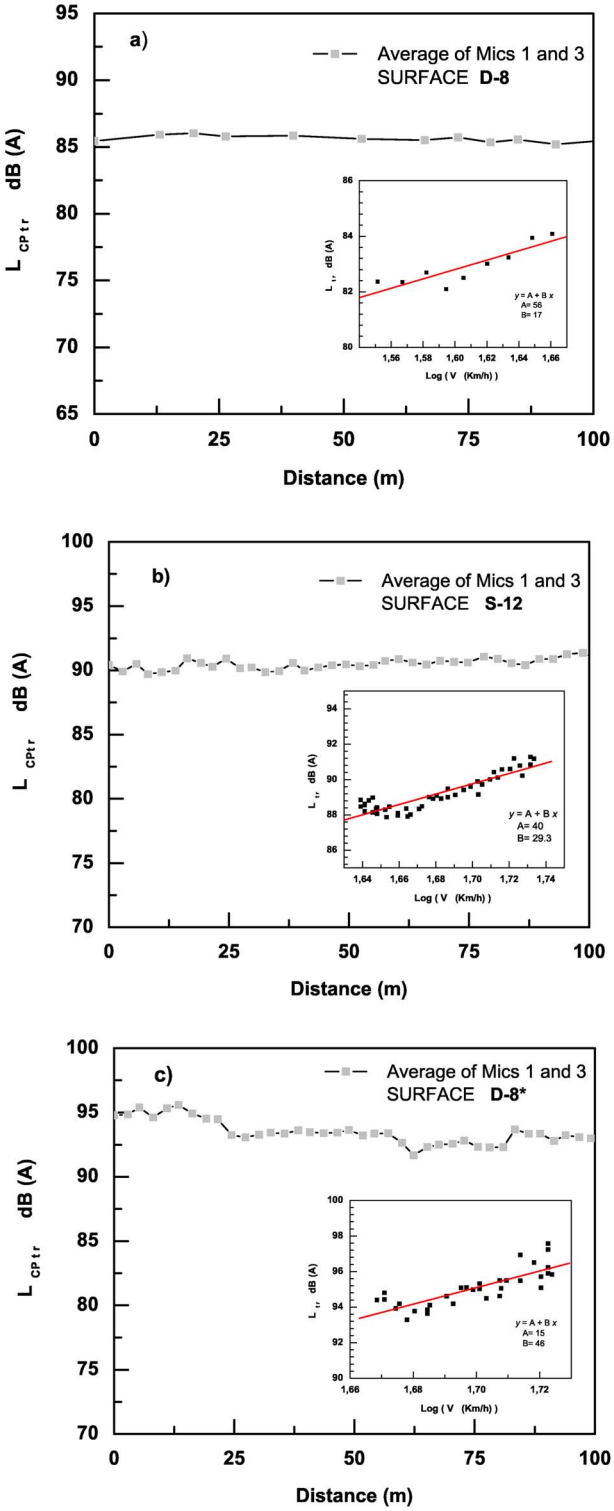


FIG. 4. (Color online) Variations with distance of the arithmetically averaged noise levels from microphones 1 and 3 for a reference speed of 50 km/h. (a) Rolling on a D-8 asphalt concrete layer, (b) rolling on an S-12 asphalt concrete layer and (c) rolling on D-8* asphalt concrete layer. The insets show the noise level L_{CPtr} versus the logarithmic vehicle speed, with their corresponding linear regressions. The relationship $L=A+B\log_{10}(V(\text{km/h}))$ is used, with the coefficients B and A shown in the inset.

from the fitting $L_{CPtr}=A+B\,log(V)$ are also shown in Fig. 4. The parameter V is the speed in km/h and A and B are coefficients. In Fig. 6, the coefficients A and B are given for each asphalt concrete surface or surface condition studied, for the

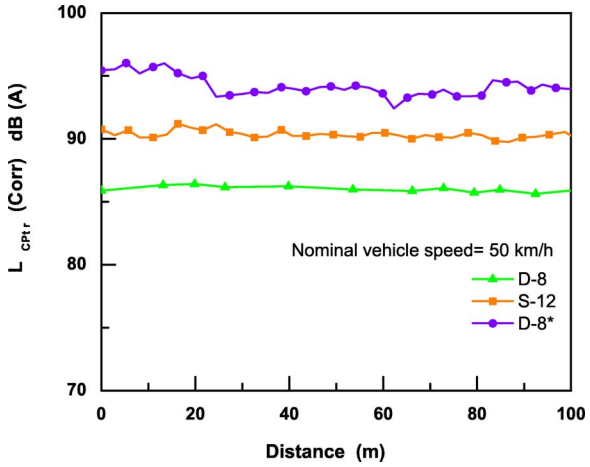


FIG. 5. (Color online) Comparison of the fluctuations with distance of the arithmetically averaged noise level from microphones 1 and 3 for a reference speed of 50 km/h, after speed correction using Eq. (5).

same reference tire. The pavement surfaces are identified in this figure by different symbols. The analysis of the results by a linear regression gives the relationship between the coordinates A and B shown in this figure. Similar values were obtained by other authors using the trailer method for truck tires on a GRB-S surface.[11] This drum surface is a replica of dense asphalt concrete and is considered the most common surface type in urban and suburban areas in most countries around the world (see Ref. 11). In view of these results, the asphalt concrete D-8 surface, or surfaces with similar physical properties, is more suitable for a low-speed streets, due to its low coefficient A and high coefficient B. Following the same reasoning, an S-12 surface should be chosen for streets with higher speed limits,[3,11] because its acoustical behavior improves with speed, due to its lower coefficient B. Other asphalt surfaces, with deviations from the linear regression toward the upper left part of the figure, imply that the surfaces will be noisier at low speeds. The further to the right the surface is located, the higher its slope, suggesting that these surfaces are less favorable at higher speeds and more favorable at lower speeds, provided their coefficient A is low.

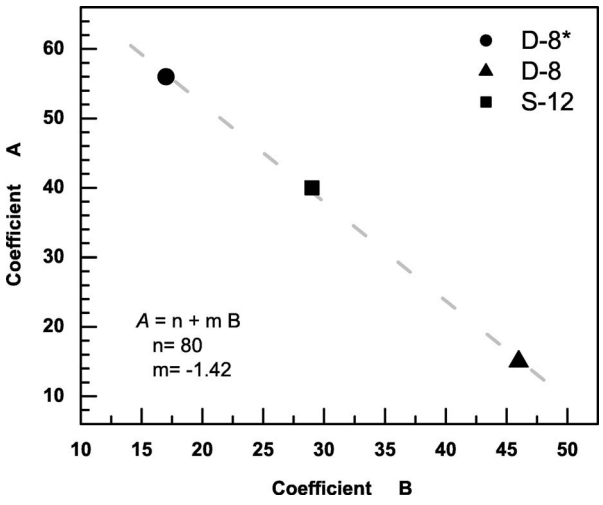


FIG. 6. Relation between coefficients A and B from the L_{CPtr}/speed regressions obtained for the D-8 (▲), S-12 (■) and D-8* (●) surfaces.

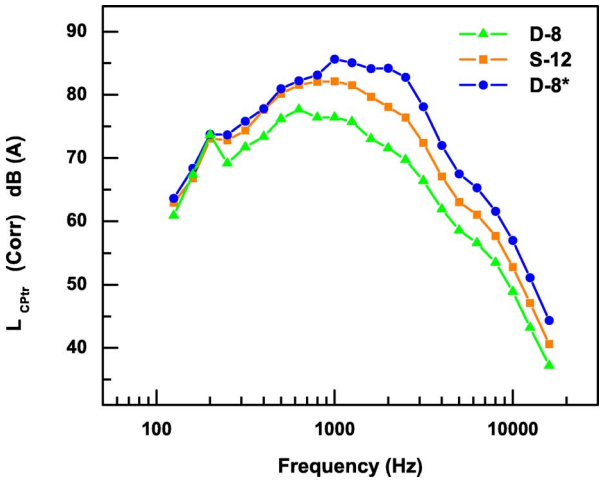


FIG. 7. (Color online) Comparison of averaged close proximity noise spectra at 50 km/h for the three different surfaces studied, over the 100 m tracks.

C. Spectral analysis

Close proximity noise emissions from tire/pavement interaction, rolling at 50 km/h as reference speed, were analyzed in third octave bands between 150 Hz and 16 kHz. Figure 7 shows the average of the spectra from the rolling over 100 m, for the three surfaces studied. Each third octave band level derives from the average of the values recorded from microphones 1 and 3 after the speed correction [Eq. (5)]. Significant spectral differences between the pavement surfaces appear above 600 Hz. The spectrum with maximum amplitude corresponds to the 12 year old D-8* surface. A shift to lower frequencies of the maximum of the spectrum is observed for quiet surfaces, i.e., surfaces with lower L_{CPtr} noise levels. The dense surface D-8 shows a peak near 600 Hz. These results reflect the different spectral behavior, on the one hand between dense and semidense surfaces, with different maximum stone size and air void content, and on the other, between the same type of surfaces but in different wear conditions. This may be taken as the representative frequency spectrum of asphalt concrete surfaces over a test distance of 100 m. It should be noted that it is not uncommon that open surfaces (S-12), with slightly greater levels of acoustic absorption, display higher L_{CPtr} levels than smooth-textured dense surfaces (D-8) (see Fig. 3). This is because in the tire/pavement noise close proximity measurements, for the surfaces compared in this research, and the reference tire used, the main noise generation mechanism seems to be the tire vibrations induced by the pavement surface texture and/or defects.

D. Variability of close proximity sound emission

Another parameter which should be considered in acoustical assessments of pavement surfaces is the variability. In this research, the variability of the close proximity tire/pavement sound emission was studied along a test section of the three different surfaces. The A-weighted sound level was measured for the reference tire used in this research. The variation of the noise level along the tested surface, after speed correction for close proximity sound levels, could be

considered to be an indication of longitudinal surface homogeneity, or the state of maintenance of these urban surfaces.[10] The averaged sound levels over the tested surfaces are 86 dB(A) for D-8, 94 dB(A) for D-8* and 90 dB(A) for S-12 asphalt concrete surfaces, and their standard deviation around these means are 0.23 dB(A), 0.81 dB(A) and 0.32 dB(A), respectively. These standard deviations may be interpreted as the A-weighted close proximity tire-pavement sound emission variability of the surfaces, over the 100 m test sections representative of these pavements. On the other hand, for the above mentioned reference vehicle speed and tire, the total range of noise level variations within the selected sections is around 0.78 dB(A), 3.6 dB(A) and 1.45 dB(A) for surfaces D-8, D-8* and S-12, respectively. These results are consistent with the above discussion, which considers the principal noise generation mechanism to be the tire vibrations induced by the pavement surface texture and/or defects.

IV. CONCLUSIONS

Acoustical properties of different pavement surfaces of urban streets have been reported. Geo-referenced close proximity noise measurements and sound absorption from core samples were performed. Geo-referenced sound levels for each passage of the test vehicle, sound spectra averaged over a distance of 100 m, and sound absorption coefficients, are analyzed to assess the acoustical performance of dense asphalt and semidense asphalt with Spanish denomination D-8 and S-12. Close proximity noise levels from the tire/pavement interaction, L_{CPtr}, of 86 dB(A), 90 dB(A) and 94 dB(A) for D-8, S-12 and D-8* asphalt mixes, respectively, are found at a reference speed of 50 km/h. The sections tested are analyzed to provide the variability. The D-8 dense asphalt concrete surface, 0/10 grading, presents the lower variability, 0.23 dB(A), and the highest variability is found in the D-8*, the older surface, which is at the end of its service life. This surface condition shows a variability of 0.81 dB(A). The semidense asphalt concrete surface S-12 (grading 0/20) presents an intermediate value of 0.32 dB(A). From an acoustical point of view, the D-8 mix provides the most homogeneous surface of all of the tested pavements, possibly due to its smallest maximum size aggregate. In contrast, the S-12 mix has a high variability because of its open-graded mix type of texture. The frequency spectrum analysis over the test section shows noticeable differences between the pavements.

A first step toward a close proximity noise map of streets has been presented in this study for the first time, as can be seen in Fig. 8. From this, the most favorable surface, by more than 4 dB(A), is concluded to be the finest surface, D-8, rather than the coarse asphalt surface, even though the latter presents a slightly higher absorption coefficient in normal incidence. A plan of the test track in Ciudad Real (Spain), including the location of the test streets, is shown in Fig. 8(a).

The relation between the L_{CPtr} level and the logarithm of speed can be described by a linear relationship, with slopes for regressions ranging from 17 (D-8) to 46 (D-8*). This

FIG. 8. (a) Geo-referenced close proximity noise map of streets with different pavement surfaces and pavement conditions in the city of Ciudad Real (Spain). The tire/pavement interaction is measured at 50 km/h. (b) Photographs of the three test tracks.

relationship is found to depend heavily on the type of pavement and pavement condition, and could serve local administrations when selecting the most appropriate type of pavement as a function of the speed of the street and in pavement rehabilitation works.

The influence of the surface on the close proximity tire/pavement noise emission was measured using a self-powered vehicle, without enclosures. For this reason, the measurements of asphalt concrete surfaces were developed without traffic, thus, ensuring that sound results were not disturbed.

Finally, in this study we have endeavored to quantify close proximity noise emissions and to demonstrate procedures for classifying the influence of asphalt concrete surfaces in city streets with low speeds limits.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science, MEC, through Grant No. BIA2006-04783 and by Grant Nos. PCI05-025 and PAC07-0098-7433 of Junta de

Comunidades de Castilla-La Mancha. We thank *Ayuntamiento de Ciudad Real* and *Laboratorio de Obras Públicas-CR (JCCM)* for their valuable assistance. One of the authors, J.L., from Université Pierre et Marie Curie, wishes to thank the University of Castilla-La Mancha for the warm hospitality.

[1]See, for example: A. Rust and J. Affenzeller, "CALM—An EU Network for Strategic Planning of Future Noise Research," (Euronoise, Naples 2003). This is available online at: http://www.calm-network.com/Euronoise2003-Paper.pdf (last viewed online 25 November, 2007).

[2]SMILE workshop "Guidelines for Road Traffic Noise Abatement" Berlin, Germany, 2003. A paper copy is available from French Agency for the Environment and Energy Management, Center de Sophia Antipolis, 500 route des Lucioles, FR-06560 Valbonne. It is also available online at the site: http://www.smile-europe.org/PDF/guidelines_noise_en.pdf (last viewed online 25 November, 2007).

[3]U. Sandberg, "Tyre/road noise—Myths and realities," *Proceedings of the 2001 International Congress and Exhibition on Noise Control Engineering* (the Hague, the Netherlands, 2001).

[4]The interest of this subject is reflected in the number of workshops, see, for example: Jornada Técnica CEDEX "Incidencia de la capa de rodadura en la reducción del ruido del tráfico en carreteras" (Technical workshop CEDEX "Influence of the road surface on traffic noise reduction") (Madrid, Spain, 2006). http://www.cedex.es/castellano/publicador/formacion/2006/JT_Ruido_Rodadura.pdf (last viewed online 25 November, 2007).

[5]U. Sandberg, "Road traffic noise—the influence of the road surface and its characterization," Appl. Acoust. **21**, 97–118 (1987).

[6]T. Bennert, D. Hanson, A. Maher, and N. Vitillo, "Influence of pavement surface type on tire/pavement generated noise," J. Test. Eval. **33**, 94–100 (2005).

[7]T. Fujiwara, S. Meiarashi, Y. Namikawa, and M. Hasebe, "Reduction of equivalent continuous A-weighted sound pressure levels by porous elastic road surfaces," Appl. Acoust. **66**, 766–778 (2005).

[8]A. Fortier Smit, "Sound pressure and intensity evaluations of low noise pavement structures with open graded asphalt mixtures," *Proceedings of the Inter-noise 2006 Congress* (Honolulu 2006).

[9]J. A. Reyff and P. Donavan, "Studies of the effect of aging of quiet pavements on tire/pavement noise (A)," J. Acoust. Soc. Am. **118**, 1915 (2005), abstract.

[10]S. E. Paje, M. Bueno, F. Terán, and U. Viñuela, "Monitoring road surfaces by close proximity noise of the tire/road interaction," J. Acoust. Soc. Am. **122**, 2636–2641 (2007).

[11]U. Sandberg and J. A. Ejsmont, *Tyre/Road Noise Reference Book* (Informex, Harg, SE-59040 Kisa, Sweden, 2002).

[12]J. A. Ejsmont and P. Mioduszewski, "Certification of vehicles used for close proximity measurements of tire/pavement noise," *Proceedings of the Inter-noise 2006 Congress* (Honolulu, 2006).

[13]W. T. Hung, W. G. Wong, C. F. Ng, and M. B. Kwok, "The construction of a close proximity trailer to measure road-tire noise in Hong Kong," *Proceedings of the Inter-noise 2006 Congress* (Honolulu, 2006).

[14]B.O.E. *n^o 83, page 14446,8 (2004) orden FOM/891/2004*. The official Spanish state bulletin is also available, in Spanish, on line at: http://www.boe.es/g/eng/index.php (last viewed online 25 November, 2007).

[15]U. Sandberg, "Low noise road surfaces—A state-of-the art review," *Proceedings of the Euro-symposium, The Mitigation of Traffic Noise in Urban Areas* (LCPC, Nantes, France, 1992).

[16]A. von Meier, "Thin porous surface layers—design principles and results obtained," *Proceedings of Eurosymposium, The mitigation of traffic noise in urban areas LCPC* (1992).

[17]M. C. Berengier, M. R. Stinson, G. A. Daigle, and J. F. Hamet, "Porous road pavements: Acoustical characterization and propagation effects," J. Acoust. Soc. Am. **101**, 155–162 (1997).

[18]A. von Meier, G. J. van Blokland, and J. C. P. Heerkens, "Noise optimized road surfaces and further improvements by tyre choice," *Proceedings of INTROC* (Gothenburg, Sweden, 1990).

The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places?

Robert Pheasant, Kirill Horoshenkov,[a] and Greg Watts
School of Engineering, Design and Technology, University of Bradford, Bradford, West Yorkshire, BD7 1DP, United Kingdom

Brendan Barrett
Department of Optometry, School of Life Sciences, University of Bradford, Bradford, West Yorkshire, BD7 1DP, United Kingdom

(Received 11 September 2007; revised 5 December 2007; accepted 6 December 2007)

Prior to this work no structured mechanism existed in the UK to evaluate the tranquillity of open spaces with respect to the characteristics of both acoustic and visual stimuli. This is largely due to the fact that within the context of "tranquil" environments, little is known about the interaction of the audio-visual modalities and how they combine to lead to the perception of tranquillity. This paper presents the findings of a study in which visual and acoustic data, captured from 11 English rural and urban landscapes, were used by 44 volunteers to make subjective assessments of both their perceived tranquillity of a location, and the loudness of five generic soundscape components. The results were then analyzed alongside objective measurements taken in the laboratory. It was found that the maximum sound pressure level (L_{Amax}) and the percentage of natural features present at a location were the key factors influencing tranquillity. Engineering formulas for the tranquillity as a function of the noise level and proportion of the natural features are proposed.
 [DOI: 10.1121/1.2831735]

PACS number(s): 43.50.Qp, 43.50.Sr, 43.50.Rq, 43.66.Lj [KA] Pages: 1446–1457

I. INTRODUCTION

It is something of a paradox that in an age where technological innovations such as mobile phones, personal computers and multi-media entertainment systems, which were designed to ease the stresses and strains of everyday living, have in the main failed to deliver an increase in rest and relaxation time. Instead the majority of modern electronic devices and systems of transportation have served to accelerate time and compress space,[1] thereby prolonging periods of cognitive intensity and making less distinguishable the division between our professional and private lives. In order to cope with the excessively high levels of mental stimulation that are characteristic of living and functioning in a globalized world it is imperative that we allow ourselves restorative periods of cognitive quiet (tranquillity), in which to regain our composure.[2]

Tranquillity, however, is a very personal construct, which requires people to draw upon a complex array of sensory, experiential and emotional inputs, in order to bring about what is ultimately a state of being, that has no reality outside of their interpretations and perceptions.[3] Conversely the concept of "tranquil space," i.e., space that can facilitate a state of tranquillity, is much more tangible as it suggests a reconciliation between mental space and the physical and social spaces in which we all live and work.[4] In order to accommodate the need for cognitive restoration, areas within the natural and built environment that comprise the ingredients essential in facilitating a state of tranquillity are therefore required. For the purposes of this project, the extent to which a place is considered to be tranquil is defined by how much individuals think a particular setting is a quiet, peaceful and attractive place to be, i.e., a place to get away from "everyday life."[5]

The fundamental definition of a tranquil place is rather complex. Within the concept of Attention Restoration Theory (ART), Kaplan suggested that restorative environments need to provide an element of "soft fascination" if they are to be effective.[6] This essential ingredient is deemed to occur when there is enough interest in the surroundings to hold attention, but not so much that the ability to reflect and relax becomes impeded. In essence soft fascination provides a pleasing level of sensory input that involves no cognitive effort other than removing oneself from an overcrowded mental space. In contrast, "directed attention," requires a significant amount of cognitive effort and it is known that prolonged periods of sustained mental activity can lead to directed attention fatigue. This condition has the potential to impact on performance and bring about negative emotions, irritability and decreased sensitivity to interpersonal cues.[7]

In developing ART Kaplan listed four main factors of restorative environments. Two of these factors are seeking cognitive quiet fostered by soft fascination and psychologically distancing oneself from directional attention. The other two are clearing the head, and reflecting on life's priorities, possibilities and goals.

[a]Author to whom correspondence should be addressed. Electronic mail: k.horoshenkov@bradford.ac.uk.

Significant contributions in understanding the essential physical and psychological qualities of tranquil environments have been made by Herzog *et al.*[5,6] and these can be applied globally when characterizing restorative space. However, the types of environments that serve as preferential tranquil space are not globally uniform, due to the vast geographic differences that exist between the USA and other developed countries including the United Kingdom, where, for example, deserts and large scale forests do not exist. In common with many advanced economies in Europe and in the Far East the United Kingdom has lost vast tracts of green space both within the inner city and along the urban–rural fringe due to development activity. Despite still having around 27,000 public parks and gardens[8] the number of tranquil spaces in the UK is becoming seriously compromised and has prompted systematic research into tranquillity mapping.[3,9] This work has been conducted along with attempts to define and characterize "quiet areas" in response to the European Directive on the Assessment and Management of Environmental Noise (END).[10] Although defining quiet areas in accordance with the END using purely acoustical measures is an important step in protecting tranquil spaces, there is a need to go further and integrate both aural and visual factors into an overall descriptor that will be sufficiently precise and practical.

There have been several attempts to study the cross-modal interaction[11] and the link between the perceived degree of noise annoyance and specific visual settings.[12–14] However, the relevant importance of visual and auditory stimuli for the construction of a tranquil space is not yet fully understood. The purpose of this work is to investigate how the auditory and visual modalities influence the construction of tranquil space in a real world setting by adopting a two-stage experiment strategy. Stage 1 involved a tranquillity ranking exercise using purely visual stimuli presented as still images, and Stage 2 was an experiment conducted in an anechoic chamber using video and audio data to assess both the perceived tranquillity and the soundscape quality of 11 English landscapes.

The paper is organized as follows. Sections II and III describe the two-stage experimental methodology and Sec. IV presents the results of the objective analysis and subjective experiments. The results showed that the A-weighted maximum sound pressure level, L_{Amax}, and the percentage of natural features at any given location, are the key factors most closely associated with how tranquil an environment is perceived to be. A range of other noise indices were analyzed in the laboratory and in the combined audio-visual condition L_{Amax} gave the greatest correlation ($R^2=0.85$), against the mean tranquillity rating of each location compared to L_{Aeq} and L_{A10} for which the coefficient of determination had values of 0.63, and 0.58, respectively. Section V discusses the key findings of the research and proposes a simple engineering formula for tranquillity while conclusions are drawn in Sec. VI.

II. METHODOLOGY (STAGE 1)

A database of landscape images was created by taking 360 photographs from across England in the summer of 2005, using a Canon EOS 50E SLR camera loaded with 200 ASA film. The chosen angle of view was generally suitable for taking typical landscape views, i.e., telephoto shots were avoided and the photographs were taken from a position of rest. One hundred images were randomly selected from this database to cover a representative range of English landscapes. These landscapes were identifiable but not essentially familiar to the subjects taking part in the experiment. The selected photographs were composed of 20 color images from each of the following five generic landscape classifications: (i) mountainous and wilderness; (ii) coastal; (iii) parks and gardens; (iv) rural; and (v) urban environments, and the images to be included from the initial data set identified by the use of random number tables. These images were presented to 105 subjects (47 males, and 58 females, average age 32 years) in the form of 15 cm $\times$ 10 cm photographs and ranked in terms of the perception of tranquillity that they generated in the participants. Rankings were based on a scale of 0–100 with 1 being most tranquil. In this exercise the photographs were placed on a large table at randomized positions and the subjects were asked to arrange them in order of preference in terms of the tranquillity they perceived at each location. In each case it was left to the individual participant to decide upon the value judgments they drew upon and the sorting process they adopted while placing the photographs in rank order. On average it took individuals approximately 30 min to complete the task.

A. Photographic ranking exercise

The primary aim of the ranking exercise was to develop a systematic approach to identifying landscape types which have the potential to facilitate tranquillity, in order that they could be filmed and used as representative samples during the Stage 2 subjective assessment experiments. The secondary aim was to use the photographs to calculate the percentage of natural and anthropocentric space occupied in each image, in order to determine if a correlation existed between this and the image's ranked position in terms of its perceived tranquillity. In this project anthropocentric influence was defined as the amount of space occupied by people, man-made structures and objects including machines, monuments and buildings, but not including dry stone walls, as for many these are considered an intrinsic part of the English landscape. Everything else was considered a natural feature even if it had been fashioned by the hand of man.

The results of the ranking exercise for the entire sample were statistically tested by calculating the degree of agreement between the subjects using Kendall's Coefficient of Concordance[15] and a score of 0.47 was achieved. This correlation is statistically significant at the 0.1% level and clearly indicates a good measure of agreement between the participants, and suggests that the subjects were using similar criteria in assessing tranquillity.

B. Objective measures

In addition to the ranking exercise, objective measures of the scenes contained within the photographs were also taken. The percentage composition of each photograph was

TABLE I. Location designators and geographic descriptors of venues used in the Stage 2 experiments.

Location designator	Place name	Generic description
BSI	Bosigran Cliffs—Cornwall	Sea Cliffs
GS	Grasmere Lake—Lake District National Park	Lake
BCG	Baildon Community Garden—West Yorkshire	Community Garden
CR	Carrick Roads—Cornwall	Approaches to Falmouth Harbour
BB	Baildon Bank—West Yorkshire	Disused Quarry
SC	Sennen Cove—Cornwall	Coast
CC	Carland Cross—Cornwall	Wind Farm
CH	Chatsworth House—Derbyshire	Stately Home
HP	Hawksworth Pond—West Yorkshire	Rural Pond
OM	Otley Market—West Yorkshire	Market Town
BS	Building Site—West Yorkshire	Construction Site

determined by overlaying a 10×10 grid and estimating the percentage area occupied by natural features (including agricultural land but excluding sky), and anthropocentric influence (including the percentage of space occupied by people and the percentage of space occupied by man-made structures or machines). The length of the vista on each image was calculated using "Memory-Map OS-5" mapping software, while the number of people present in any scene was simply counted.

III. METHODOLOGY (STAGE 2)

Stage 2 of the experiment strategy involved capturing audio and video footage from those locations ranked at ten percentile intervals during the photographic ranking exercise, with the exception of the building site which was ranked as the least tranquil location. This site had significantly altered since the still images were taken therefore a location comparable to the original site was used. These ten images, along with the location assessed as being "most tranquil," gave 11 contrasting environments for use in the subjective assessment exercise (see Table I and Fig. 1).

The data were captured using a Canon XM 2 camcorder which was positioned on top of a mannequin's head with a binaural audio feed provided by two Bruel and Kjær (B&K) type 4188 condenser microphones. These were positioned

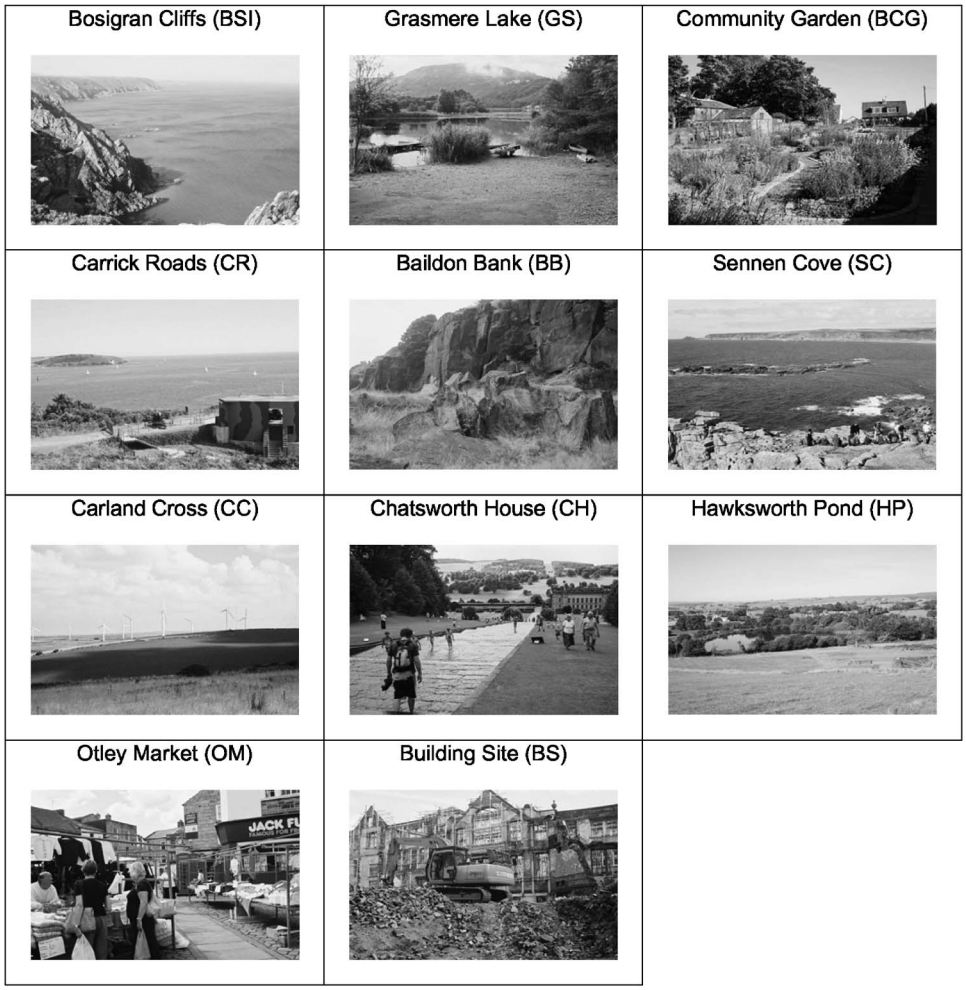


FIG. 1. Eleven locations used in the Stage 2 experiments.

FIG. 2. Recording equipment used to capture the audio-video data.

inside the ear cavities and covered by ear defender style wind shields (see Fig. 2). Where possible the video footage was taken from exactly the same position as its corresponding still image and the camera was swept from left to right over a 32 s period. This included an 8 s sweep from the left of the vista to the center, followed by a 16 s stationary view of the landscape before sweeping right for the final 8 s. Every attempt was made to include as much acoustic context as possible within the footage and the associated WAV files were left unedited in terms of the recorded audio information. Adobe Premier 6.5 software was used to capture video and audio data recorded on a DV tape and both audio channels were calibrated using a standard B&K Type 4230 calibrator, prior to and after the measurement. The recording levels on the audio channels in the camcorder were manually adjusted to match the audio channel sensitivity between individual recordings taken on separate dates.

The optimal video clip duration was determined during a pilot study in which 12 subjects were presented with combined video and audio data over six escalating exposure times (2, 4, 8, 16, 32, and 64 s). They were each asked to make an assessment of the perceived tranquillity of each clip using a scale of 0–10, with 0 being least tranquil and 10 being most tranquil. The results were averaged and graphed and a time scale of 32 s identified as being the point at which the subjects' assessments of tranquillity remained constant.

The Stage 2 experiments were conducted in an anechoic chamber equipped with a Pioneer (PDP—506XDE) plasma screen, a Samsung DVD—R125 player and two sets of Technics RP-295 stereo headphones. Prior to commencing the experiment the headphones were calibrated against a 94 dB 1 kHz pure tone prerecorded on a DVD via the two camcorder audio channels, using the sensitivity settings identical to that used for normal in-situ recordings. The binaural head with microphones connected to a Bruel and Kjaer sound level meter Type 2231 was used in this calibration exercise. The maximum and minimum sound pressure levels that the subjects would be exposed to were determined exactly in this way.

A. Subjective assessment of tranquillity

The aims of the subjective assessments were for volunteers to determine how tranquil they found the 11 locations to be and how loud they estimated five generic soundscape components to be when presented with both single (i.e., visual only or audio only) and bi-modal (audio-visual) stimuli. Prior to the experiment the subjects were told that for the purpose of this research a tranquil environment was one that they considered a quiet, peaceful and attractive place to be, i.e., a place to get away from "everyday life"[5] and were told to draw upon whatever value judgments they deemed necessary when making their assessment. They were also informed that for the purpose of the exercise the environments that they saw could be considered as steady state.

The volunteers were provided with subjective response sheets on which to record both their loudness and tranquillity assessments and were asked to score on a scale of 0–10 (0 =not at all and 10=very much), how sensitive they considered themselves to be to noise. The participants were also asked to declare whether they were aurally or visually impaired and, following the experiment, they were asked to complete a questionnaire on how strongly certain factors had influenced their assessments of tranquillity.

Using headphones and the plasma screen, 44 subjects (20 male and 24 female, average age 35 ± 14.1 years), were presented sequentially with audio only, video only and combined audio-video data streams which they scored on a scale of 0–10 with 0 representing the least tranquil and 10 the most tranquil. The scoring was completed by circling the number in a table corresponding to the value of the tranquillity perceived by the subject. The subjects were seated 2 m from the center of the plasma screen and the data were presented to each subject four times per experimental condition in a balanced design to reduce order effects. Each audio/video clip lasted for 32 s. Using headphones and sitting in pairs, the first group of 22 subjects were presented with the stimuli in the following order: (i) audio only, (ii) video only and (iii) combined audio/video data. In the case of the second group of subjects the order was changed to: (i) video only, (ii) audio only and (iii) combined audio/video data. Analysis of the results showed no statistically significant difference in tranquillity assessments between the two groups. Some differences in tranquillity assessments were identified due to

TABLE II. Five generic soundscape components used in the subjective loudness evaluation.

Sound source	Definition
Human (H)	Sounds made by people including musical instruments and bells
Mechanical (M)	Sounds emitting from anything man-made, excluding musical instruments and water features
Biological (B)	Sounds made by living organisms excluding human beings, e.g., farm animals, bird song, humming bees
Weather (WX)	Sounds made by the wind, e.g., wind in trees, telegraph wires and thunder/lightening
Water (WA)	Sounds made by water, e.g., rapids, waves, rain, fountains etc.

gender, but those attributable to age and ethnicity were not examined. These possible effects will be investigated in a further stage of the study.

In the video only condition women rated seven out of the 11 locations higher in terms of perceived tranquillity than the men, although on average the degree of difference between the two groups was <0.5 on the tranquillity rating scale. This contrasts with the audio only condition where the men scored all 11 of the locations higher than the women by approximately 0.8 of one tranquillity point.

When assessing audio only data the locations that were given higher than average tranquillity ratings by the male subjects were Grasmere, Carrick Roads, Baildon Bank, Hawksworth Pond, Carland Cross, and Baildon Community Garden. Each of these locations contained either mechanical or human noise; however, in the case of GS, BB, HP and CC, the levels of mechanical noise were significantly higher than the majority of other locations.

In the combined audio-visual condition the gender difference in tranquillity ratings was <0.4 with the largest discrepancies in this condition being Carland Cross wind farm and Hawksworth Pond. For these two environments the men rated them one tranquillity point higher than the women.

Analysis of the results suggests that women rely more on visual stimuli when characterizing landscapes than men do and that men accept higher levels of mechanical and human generated noise than women when constructing tranquil space. This raises some interesting points for landscape design and management and is the subject of ongoing research.

B. Subjective assessment of loudness

At the start and finish of the experiment the subjects were asked to make an assessment of the loudness of a 1 kHz tone played to them via the headphones at levels corresponding with the loudest and lowest sound sources that they would encounter throughout the experiment. These data were used to determine the subjective loudness limits of the objective dynamic range covered by the recorded sounds, and scale accurately the bounds of the loudness assessment.

Before being presented with the last set of repeat data the subjects were told that in addition to assessing the tranquillity of each location they would also have to assess the loudness of each of the soundscapes based on the five generic soundscape descriptors (see Table II) listed on their response sheets. The loudness of each of these five descriptors was assessed using magnitude estimation which categorized the loudness as being either quiet, moderately quiet, moderately loud or loud. The subjects were also instructed to make an assessment of the perceived soundscape based on video only data, i.e., in the absence of any audio stimulus and in the presence of a silent video clip.

C. Analysis methods

The results of the subjective assessments were collated and analyzed using SPSS and TableCurve2D software to determine the statistical significance of the data and the relationships between several independent or predictor variables. In this analysis the mean and standard deviation for the tranquillity rating, subjective and objective loudness and the proportion of the natural features present were used, together with the objectively measured noise indices L_{Amax}, L_{Aeq}, L_{Amin}, L_{A10}, and L_{A90} for each of the 11 locations. The objective noise indices were calculated as the mean between the two channels of the binaural head over the 32 s recording interval using a specially developed MATLAB subroutine with the averaging interval of 10 ms, and the calibration tone recordings. The maximum loudness in sones was determined using the 1/3-octave band method detailed in (Ref. 16). A summary of numerical data obtained from the analysis is presented in Table III.

IV. RESULTS

In order to identify the correlations that existed between the subjective and objective variables involved in this research the subjects' mean assessments of tranquillity for each location and each experimental condition were contrasted against a range of independent variables utilizing the SPSS (version 14.0), Linear Regression (Stepwise) analysis tool. These included L_{Amax}, L_{Aeq}, L_{A90}, L_{A10}, the percentage of natural features present within each location, and the perceived loudness of the corresponding soundscape components.

A. Subjective and objective measures

Figure 3 shows the mean tranquillity rating of each location for each of the experimental conditions, plotted against L_{Aeq} and L_{Amax}. Figure 4 presents a summary of the standard deviation data for the tranquillity assessment. It is convenient to discuss the results of the 11 locations using

TABLE III. Summary of numerical data.

Location	Tranquillity Rating (Audio Only)	Tranquillity Rating (Video Only)	Tranquillity Rating (Audio-video)	Mean % Natural Features Present in 32 second recording	L_{Amax} dB	L_{Amax}-Sones	L_{Aeq} dB
BSI	7.65	8.73	8.34	99.99	57.92	7.65	51.96
GS	5.73	7.67	7.18	92.03	58.74	5.09	47.63
BCG	6.91	6.62	7.02	84.92	59.32	10.45	44.43
CR	5.41	5.34	5.51	70.39	59.37	3.93	43.84
BB	4.80	7.15	5.55	100	63.22	9.81	51.84
SC	4.65	7.57	6.69	98.15	62.33	11.34	58.03
CC	2.69	5.25	3.57	96.30	67.28	36.98	61.59
CH	1.61	2.79	1.64	75.13	79.39	26.03	78.35
HP	0.64	1.04	0.56	84.19	80.33	14.35	65.89
OM	2.43	6.67	3.64	0.00	77.80	21.62	66.60
BS	1.87	4.88	2.98	7.87	78.26	33.68	71.36

three groupings, i.e., those with A-weighted values <55 dB, those with values ranging from 55 to 75 dB and those above 75 dB. These have been annotated as Groups A, B & C, respectively. It can be seen that once presented with bi-modal data the subjective assessments of tranquillity move to a position that approximates a mid-way value between those shown for the audio and video only responses.

Group A comprises tranquillity data plotted against the equivalent continuous sound pressure level below 55 dB. A high proportion of natural features ($>70\%$) are characteristic of this group, within which a relatively poor correlation between the mean tranquillity ratings and L_{Aeq} is observed. This can be attributed to the relatively small range of the independent variable and sample size.

The landscapes which fall into Group C are characterized by relatively high maximum sound pressure levels (>75 dB) and variable proportions of natural features. It can be seen from Fig. 3 that the response of Group C to visual stimuli is greater than that of Group B in comparison to the audio only condition. However, this change is exaggerated by the fact that two of the locations, Chatsworth House and Hawksworth Pond, both comprise a high percentage of natural features, i.e., 75% and 84%, respectively. In the case of Hawksworth Pond the auditory context is not easily determined due to the hidden nature of the dominant noise source.

Of the 11 locations used in the experiment, Chatsworth House, which is considered something of a national treasure by the 650,000 visitors it attracts each year,[17] could also be viewed as an outlier. We note that the quoted A-weighted maximum level of 79 dB was recorded 1m from the water edge. Given the linear extent of the water channel one can argue that a 3 dB reduction to this level per doubling the distance could be expected if the recordings were to be taken further away from the running water. Sound pressure level measurements taken at 12 m away from the water channel, i.e., where the public like to sit and enjoy the view, suggest that the L_{Amax} can reduce to 61 dB. Therefore, the subjective tranquillity rating for this specific location would depend on the proximity to the water feature and may become comparable to that obtained in the cases of Sennen Cove and Baildon Bank.

This agrees with the idea that subjects like to spatially distance themselves from directed attention, preferring instead the soft fascination that remaining in acoustic range of the water feature provides. The aircraft noise which is the source of L_{Amax} at Hawksworth Pond, however, is omnipresent and therefore not escapable within the considered landscape. The other two locations, Otley Market and the Building Site, comprise a low proportion of natural features, 0% and 8%, respectively, and high levels of mechanical noise. The tranquillity rating of these sites is low and consistent for all three combinations of audio and video stimuli presented in the subjective assessment experiment (see Fig. 3).

Group B comprises the tranquillity data plotted against the maximum sound pressure level in the range of 55 -75 dB. The data in this group show a good correlation of 77% between the tranquillity rating and the L_{Amax} which is quantified in Sec. V. Within this group the Baildon Community Garden and Carrick Roads were rated slightly lower in the visual condition than they were in the audio only condition. In the case of the Community Garden (the location with the lowest standard deviation in all three experimental conditions), the dominant noise source was birdsong, which at 59 dB, and with an averaged loudness assessment of moderately quiet, the subjects evidently found pleasing. The urban nature of the location, which is situated on the site of an old school close to the village center accounts for its lower rating in the visual condition.

The dominant noise source at Carrick Roads came from people speaking; however, it was clear from what was being said that the people were involved in some sort of sightseeing activity. This provided the subjects with enough locational context on which to make their tranquillity assessments. Once presented with the visual context of the environment, which included a red brick World War 2 gun emplacement overlooking the sea, the perceived value of the

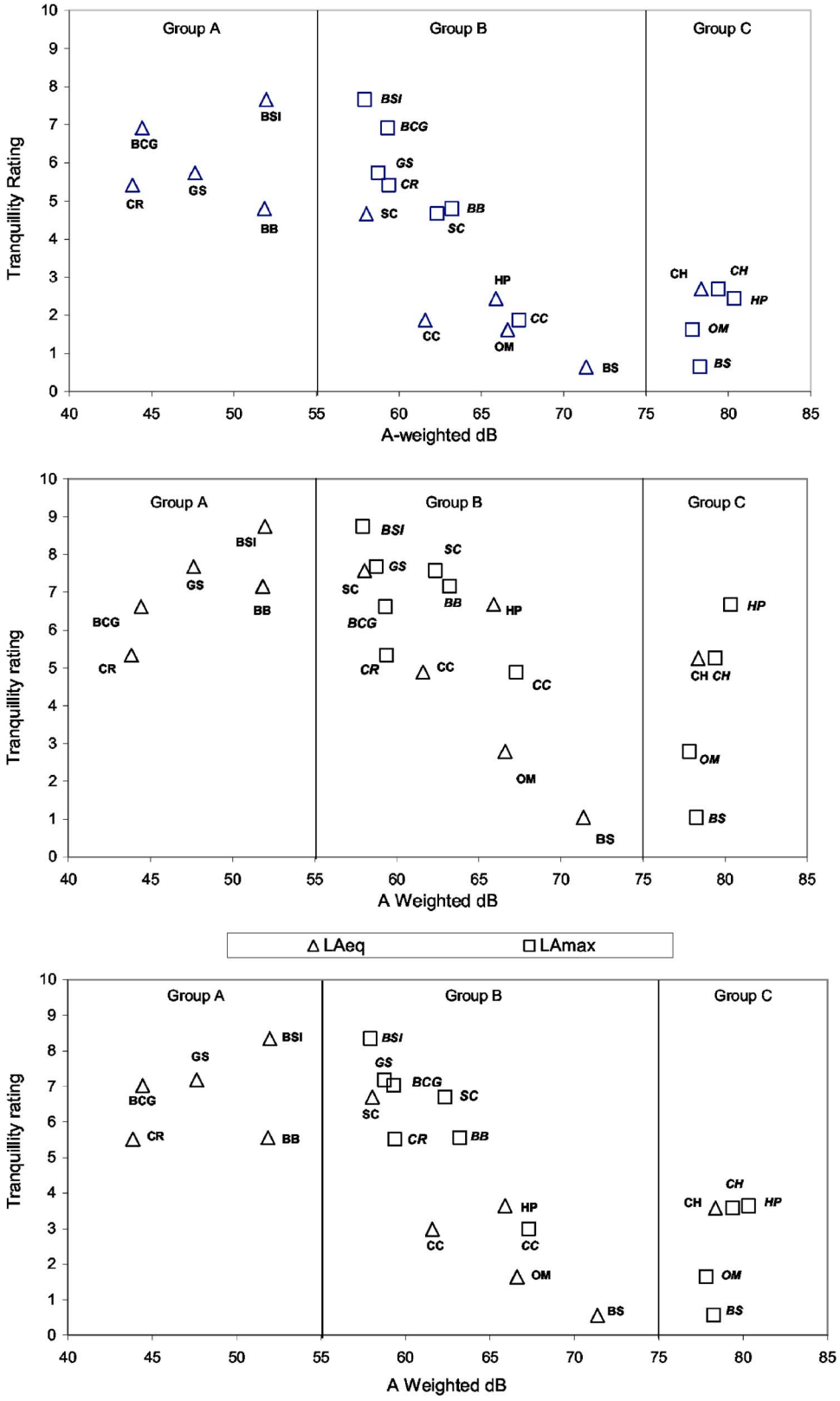


FIG. 3. (Color online) Mean tranquillity rating plotted against L_{Aeq} and L_{Amax} for the audio, visual and combined audio-visual experimental conditions.

location dropped slightly. This compares starkly with Carland Cross wind farm which more than doubled its audio only tranquillity rating once the visual data had been provided. However, as can be seen in Fig. 4, the standard deviation of the wind farm shows that there was a high degree of variance among the subjects, adding further weight to the contested status of wind turbines in general.

In each condition only those variables which could have noticeably influenced the subjects' assessment of tranquillity were included. For example, in the audio only condition variables that relied on visual information such as the percentage of natural features occupying each scene, or the number of people present, were not included. Likewise in the video only condition L_{Amax} and L_{Aeq} were omitted. However, the

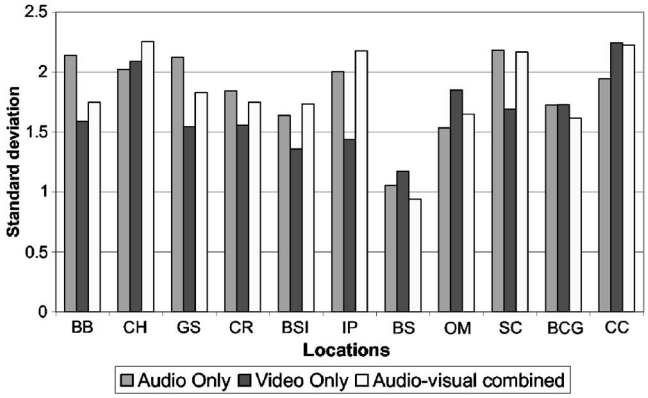


FIG. 4. Standard deviation of tranquillity assessments for each experimental condition.

perceived levels of loudness for the five generic soundscape categories were included as it was felt that these could reasonably influence how tranquil a particular environment was felt to be by providing an element of acoustic context.

When assessing the impact of environmental noise it is usual within the EU to use L_{den} which relies on the L_{Aeq} scale. However, the World Health Organization (WHO) advises in their Guidelines for Community Noise,[2] that measurements based solely on L_{Aeq} values do not fully characterize most noise environments and do not adequately assess the health impacts on human beings. Their guidance also states that it is important to measure the maximum noise levels and the number of noise events when deriving guideline values.

When proposing guideline values for specific environments in Ref. 2 the WHO states that the critical health effect of allowing the ratio of intrusive noise to exceed the background noise in parks and conservation areas, is a disruption of tranquillity. Given these factors it has been deemed appropriate to examine L_{Amax} in addition to L_{Aeq}. This will be achieved by first looking at the relationship between L_{Aeq} and the tranquillity rating before establishing how L_{Amax} influenced the subjective responses.

B. L_{Aeq} Groups A & B

Group A, which is widely distributed below 55 dB does not show any systematic dependency between L_{Aeq} and tranquillity, which is determined largely in response to visual stimuli alone. When analyzing both the audio only, and the combined audio-visual conditions, no meaningful correlation existed between tranquillity and any of the independent variables. In the video only condition the number of people present in the scene and the perceived loudness of mechanical noise were the two variables identified in the regression analysis. However, with an R^2 value of 0.20 this correlation was considered very weak.

In Group B a weak correlation existed between L_{Aeq} and the perceived loudness of biological noise in the audio only condition. In this group the video only condition showed the best result with the percentage of natural features being the single variable that influenced the tranquillity rating. Stepwise analysis of the combined audio-visual condition showed L_{Aeq} as being the sole influencing factor. The one-dimensional regression analysis of the data within this group suggests that the tranquillity rating (TR), depends strongly on the L_{Aeq} and can be closely approximated by the following linear equation:

$$TR = 29.7 - 0.41L_{Aeq} \quad (1)$$

for which the coefficient of determination $R^2=0.80$. This equation is valid for the range of the L_{Aeq} values being between 55 and 72.4 dB.

C. L_{Amax} Groups B & C

1. Audio only condition Group B

When responding to audio only data subjects made their tranquillity assessments of Group B based on L_{Amax} and the perceived loudness of human noise. Within this experiment subjects showed that they were well capable of applying auditory scene analysis, i.e., the ability to automatically partition incoming sounds with similar regularities,[18] to each of the 32 s data streams played to them. It was by applying this technique that they were able to generate, and subsequently categorize, a number of "auditory streams" that corresponded to a single sound source. This information then enabled them to make reasonably consistent assessments of the soundscape and not only identify which of the five generic components were present, but also allocate them a loudness score. The fact that the perceived level of loudness of human noise correlated so well suggests that when working on hearing alone individuals place significant value on information relating to other people. This may be linked to an instinctive survival response where understanding the context in which a potential threat exists is an imperative in activating or depressing the fight or flight mechanism. More work is required in the area of environmental acoustics to test out this hypothesis.

2. Audio only condition Group C

Once above 75 dB the results became ambiguous and no significant correlation existed between the variables in the audio only condition for Group C. This may be due to the combined affects of the small sample size of the group, and the potentially negative impact on tranquillity of variables such as the perceived level of human and mechanical noise, the number of people present, and the objectively measured L_{Amax}. Combined these factors make it difficult to draw any coherent conclusions and more data are required to determine under what conditions tranquil space could exist in either natural or built environments with a soundscape greater than 75 dB. Such work would build on the findings of Yang and Kang,[19] who together identified an L_{Aeq} of 76 dB as being the level at which people find the soundscape of urban spaces noisy.

3. Video only condition Group B

The regression analysis of the video only condition identified the perceived level of mechanical noise and the number of people present in a scene to be the key correlates. This shows that when attempting to gauge how tranquil an environment is in response to visual only data, individuals utilize

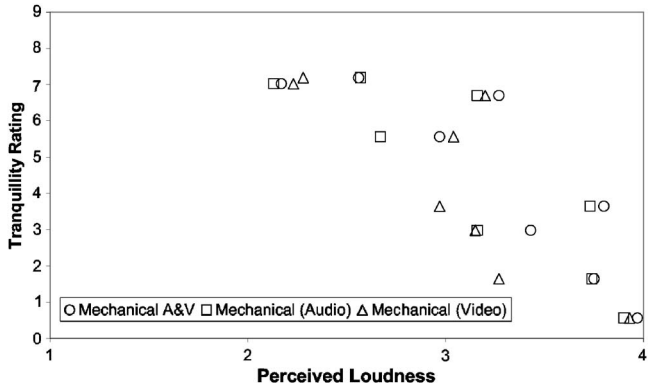


FIG. 5. Mean perceived loudness of mechanical noise plotted against tranquillity rating.

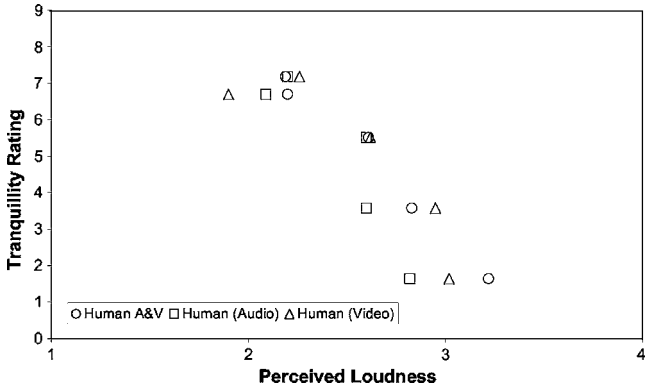


FIG. 6. Mean perceived loudness of human noise plotted against tranquillity rating.

the same hierarchy of influencing factors that they did in the audio only presentation. In this condition L_{Amax} is replaced by the perceived loudness of mechanical noise and the perceived loudness of human noise is replaced by the actual number of people present.

4. Video only condition Group C

When the locations that comprise Group C were shown to the subjects in a video only data stream the variables that correlated were: the perceived levels of mechanical, biological and human noise. All other factors were excluded from the models. This suggests that for environments above 75 dB estimating the acoustic context is more important when assessing tranquillity than the visual cues.

5. Combined audio visual condition Groups B and C

Once presented with bi-modal information that provided as much visual and acoustic context as was technically possible, the subjects were able to make a judgment based on near sensory certainty. The key variables which control the tranquillity rating of the locations falling within these groups were L_{Amax} and the percentage of natural features present in the scene. These variables are likely to predominate through a complex synthesis of auditory scene analysis, visual congruence provided by scene perception, mankind's deep affiliation with nature, which Wilson describes as biophilia,[20] and preference for a particular environment.[5] All of which allowed the subjects to take what began as undifferentiated space and endow it with a lesser or greater degree of value. A more detailed investigation of the experimental data reveals that the effect of the percentage of natural features on the perceived tranquillity of the scene is relatively small in comparison with that of the L_{Amax}.

Linear regression analysis of this data suggests the following approximation to the tranquillity rating for 55 dB $<$LAmax$<$75 dB (R^2=0.77)

$$TR = 34.0 - 0.46L_{Amax}, \quad 55\ \text{dB} < L_{Amax} < 75\ \text{dB}. \quad (2)$$

Above the level of 75 dB (Group C) the percentage of natural features present in the scene was the only variable that correlated with the measured tranquillity rating data.

6. Groups A, B & C combined

When analyzing all 11 locations together some of the subtleties described above become obscured and in the audio only scenario L_{Amax}, the perceived level of mechanical noise and the perceived level of biological noise become the key factors. Whereas in the video only condition the percentage of natural features was the only correlate. However, in the combined audio-visual condition, the key determinants for all of the locations were: L_{Amax}, the percentage of natural features and the perceived loudness of biological noise; which the subjects in these experiments never assessed as being greater than moderately quiet.

D. Loudness estimation

The results have shown how important the perceived loudness of individual soundscape components are when an individual is presented with both single and bi-modal stimuli. In order to facilitate the analysis of the loudness data the assessments made by the subjects were each allocated a numerical value; (1=quiet, 2=moderately quiet, 3=moderately loud, 4=loud), and weighted averages for each of the five soundscape components present at each location calculated. These were then plotted against the mean tranquillity rating for each of the 11 venues. The results are shown in Figs. 5–8, where it can be seen that as the perceived loudness of human and mechanical noise increases the tranquillity rating falls, and that as the perceived loud-

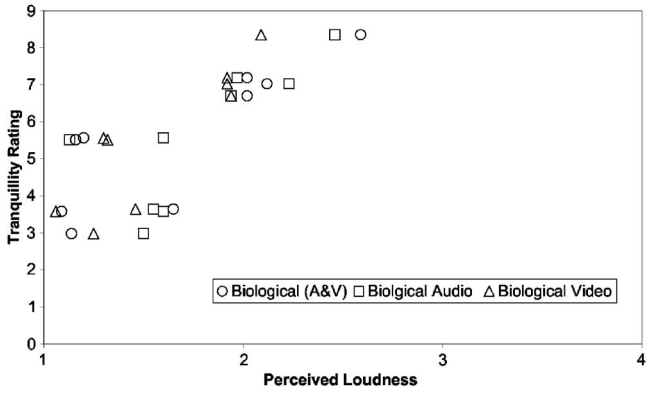


FIG. 7. Mean perceived loudness of biological noise plotted against tranquillity rating.

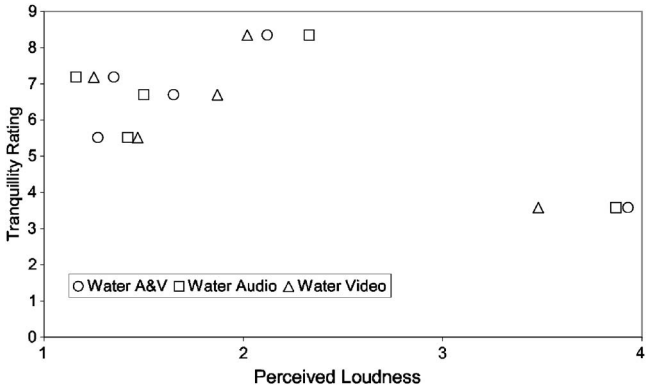


FIG. 8. Mean perceived loudness of water noise plotted against tranquillity rating.

ness of biological noise increases so does the tranquillity rating. This provides an indication of an interesting effect which if fully validated could provide a powerful design tool to those charged with landscape management and design. In the case of water noise the tranquillity rating rises until the loudness is considered to be moderately loud, beyond which point it drops significantly. This suggests an inverted U relationship similar to that reported by Brown[21] when exploring the effects of streamflow on recreation. No obvious trends were observed for the perceived loudness of noise made by the weather.

Figure 9 shows the tranquillity rating plotted against the loudest sound source per location in sones, which was calculated using 01 dB software for the audio and combined audio-visual experimental conditions. This data further confirms the importance of visual stimuli in providing auditory context to a sound environment.[22] It can be suggested that in the majority of cases the tranquillity rating increases when bi-modal data are presented to the subject.

V. DISCUSSION

A. Creating optimum tranquil space in an urban environment

The perceived levels of tranquillity in environments that combine a high percentage of natural features with low levels of mechanical or human noise, such as the sea cliffs at Bosigran, for example, are not contested, nor are environments such as construction sites which offer no opportunity for cognitive recovery. Urban "green space" however, is positioned somewhere between the two extremes. Yet with careful management of the visual and acoustic environments, engineered solutions to the problems associated with creating urban tranquil space can be achieved. An example is Baildon Community Garden with an average bi-modal tranquillity rating of 7.0 (photo BCG in Fig. 1), which was valued almost equally in both single and bi-modal sensory conditions. This location is also supported by a relatively low standard deviation of the mean tranquillity ratings (see Fig. 4). Here the important components identified across all the sites were L_{Amax} and the percentage of natural features present (excluding the sky). The following section attempts to incorporate these essential elements into a landscape management planning tool.

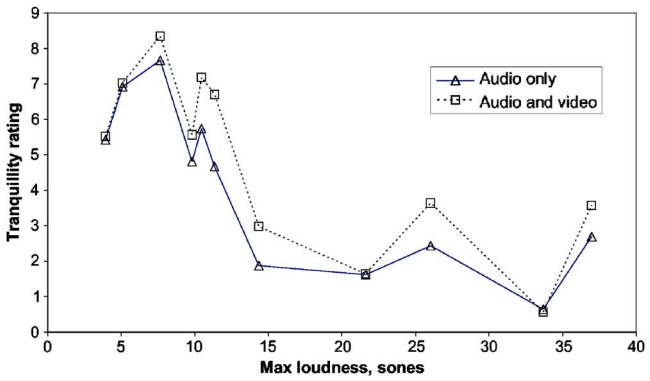


FIG. 9. (Color online) Maximum loudness in sones plotted against mean tranquillity ratings for the audio and combined audio-visual conditions.

B. Proposed landscape management planning tool

By using audio-visual tranquillity ratings across all 11 sites it has been possible to establish a reasonably robust linear expression that once fully validated could form the basis of a landscape management planning tool. The variables L_{Amax} and the percentage of natural features were the two independent variables found to explain the greatest degree of variance in the dependent variable and were statistically significant at the 0.1% level. The R^2 value of the analysis was 0.52 which is considered acceptable for this area of work. The following expression enables a tranquillity rating to be determined:

$$TR = 13.93 - 0.165L_{Amax} + 0.027NF,$$

$$(40\ dB < L_{Amax} < 85\ dB), \quad (3)$$

where NF is the proportion of natural features (%).

In the case when the L_{Aeq} needs to be used instead of L_{Amax}, Eq. (3) becomes

$$TR = 8.57 + 0.036NF - 0.11L_{Aeq}. \quad (4)$$

The coefficient of determination in this case was found to be 0.49.

The nature of function (3) is shown in Fig. 10 where TR is plotted against the two independent variables to form a three-dimensional graph. Equation (3) can be used in a variety of ways though care should be taken when extrapolating beyond the range of the data used in this study. For example, assuming a high tranquillity rating of 7.5 or above (i.e., the upper quartile) is sought in an urban space then Eq. (3) can be used to determine the range of variables that are likely to achieve this result. Figure 11 shows this relationship which is essentially a plan view of the top slice in Fig. 10.

Specifically, the relation suggests that an $L_{Amax} \leqslant 55$ dB is required with 100% natural features present. Lower noise limits are required as the degree of natural features falls. Note that an L_{Aeq} value of below 55 dB has been suggested for the establishment of Quiet Areas in agglomerations for England[8] based on an extensive review of the literature. This implies a somewhat higher level of L_{Amax} but is sufficiently close as to offer some support for the approach

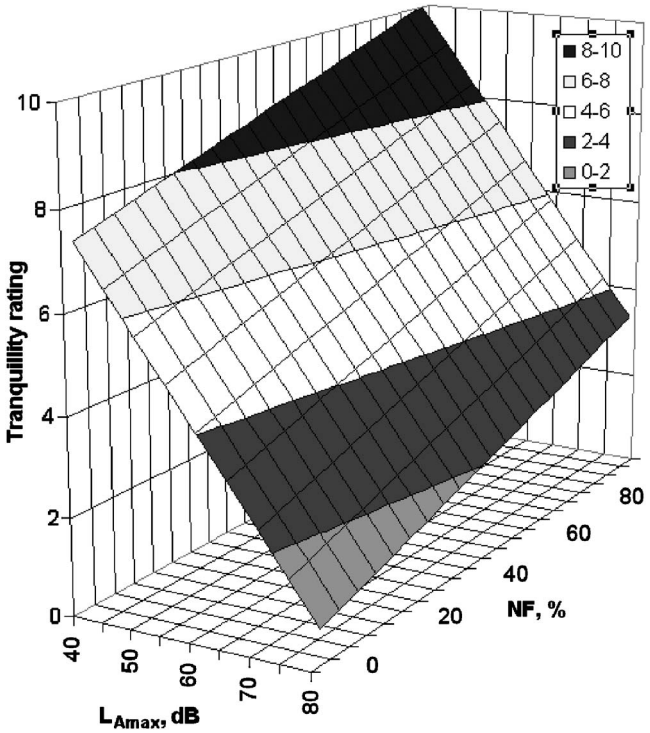


FIG. 10. Tranquillity bands plotted against L_{Amax} and percentage of natural features.

taken here. The noise level limit should be for man-made sounds and not natural sounds such as running or falling water or bird song. Clearly natural features have a part to play in improving the perception of urban spaces. Some recent support to this notion comes from a laboratory study at Chalmers University, where Johansson[23] showed in a series of audio-visual simulations that the adverse effects of low frequency ventilation noise could be positively influenced by the presence of vegetation.

VI. CONCLUSIONS

This research represents the first step in understanding the effects of audio-visual interaction on the perception of tranquillity when responding to single and bi-modal stimuli. The reliance of acoustic information, both actual and perceived, in the decision making process, can be interpreted as an important factor when we consider that six out of the eight correlations identified in the regression analysis included as an independent variable, either L_{Amax} or the perceived loudness of one or more of the generic soundscape categories.

This is a key factor in the search for cognitive quiet as individuals rely strongly on being able to contextualize the quality of the acoustic environment,[22] as is the context of anthropocentric activity when reacting to audio and visual stimuli alone. When using bi-modal data to assess tranquillity, the percentage of natural features present within a scene is also a significant influencing factor.

From the regression analysis it has been possible to propose approximate linear expressions for tranquillity using L_{Amax}, L_{Aeq} and the percentage of natural features within an environment as the coefficients. This research has indicated that given the right combination of natural features and soundscape components, tranquil space can be engineered and created within an urban environment. In addition, the results of the subjective loudness analysis suggest that as the perceived loudness of human and mechanical noise increases the tranquillity rating falls, and that as the perceived loudness of biological noise increases so does the tranquillity rating.

It has been shown that in order to achieve a high level of tranquillity above 7.5, the percentage of natural features present should be close to 100% and man-made noise sources should be characterized by an $L_{Amax} < 55$ dB or $L_{Aeq} < 42$ dB. However, more work is required to acquire more visual and audio data, especially in relation to environments for which high and low noise levels are characteristic and the proportion of natural features is diverse. These data would allow a more accurate planning tool to be developed and validated.

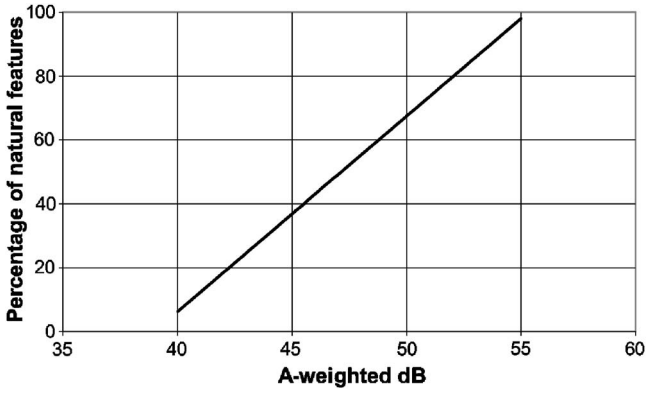


FIG. 11. Linear plot showing the values of L_{Amax} and percentage of natural features required to achieve a tranquillity rating of 7.5 or greater.

[1]D. Harvey, *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change* (Blackwell, Cambridge, MA, 1990).
[2]www.who.int/docstore/peh/noise/commnoise4.htm "accessed 15/08/2007"
[3]R. MacFarlane, C. Haggett, D. Fuller, H. Dunsford, and B. Carlisle, "Tranquillity mapping: +developing a robust methodology for planning support," Report to the Campaign to Protect Rural England, Countryside Agency, North East Assembly, Northumberland Strategic Partnership.
[4]H. Lefebvre, *The Production of Space*, translated by D. Nicholson-Smith (Blackwell, Oxford, 1991).
[5]T. R. Herzog and G. J. Barnes, "Tranquillity and preference revisited," J. Environ. Psychol. **19**, 171–181 (1999).
[6]S. Kaplan and R. Kaplan, *The Experience of Nature: A Psychological Perspective* (Cambridge University Press, New York, 1989).
[7]T. Hartig, F. Kaiser, and P. Bowler, "Further development of a measure of perceived environmental restorativeness," Institute of Housing Research, Working paper No. 5 (1997).
[8]G. Watts, P. Morgan, and A. Abbott, "A proposal for identifying Quiet Areas in accordance with the Environmental Noise Directive (END)," *Proceedings of IOA Autumn Conference* Vol. **8**, pt 7 (2006).
[9]www.cpre.org.uk/library/results/tranquillity "accessed 15/08/2007"
[10]Commission of the European Communities, "Directive 2002/49/EC of the European Parliament and of the Council relating to the assessment and management of environmental noise. L189/12," European Commission, Brussels, Belgium (2002).
[11]J. Heron, D. Whitaker, and P. McGraw, "Sensory uncertainty governs the extent of audio-visual interaction," Vision Res. **44**, 3119–3130 (2004).
[12]S. Viollon, C. Lavandier, and C. Drake, "Influence of visual settings on sound ratings in an urban environment," Appl. Acoust. **63**, 493–511 (2002).
[13]Northumberland National Park Authority and Durham County Council,

Centre for Environmental & Spatial Analysis, Northumbria University. (2004).
[14]G. Watts, L. Chinn, and N. Godfrey, "The effects of vegetation on the perception of traffic noise," Appl. Acoust. **56**, 39–56 (1999).
[15]S. Siegel and N. J. Castellan, "Nonparametric statistics for behavioural sciences," Second Edition, McGraw-Hill, New York, p. 263 (1988).
[16]International Standards Organization. ISO 532 Acoustics—Method for calculating loudness level, 1975.
[17]www.chatsworth.org/learning/theestate.htm "accessed 15/08/2007"
[18]J. G. Neuhoff, *Ecological Psychoacoustics* (Elsevier-Academic, New York, 2004), pp. 15–44.
[19]W. Yang and J. Kang, "Acoustic comfort in urban open public spaces," Appl. Acoust. **66**, 211–229 (2005).
[20]E. O. Wilson, *Biophilia* (Harvard University Press, Cambridge, MA, 1984), p. 22.
[21]T. C. Brown, J. G. Taylor, and B. Shelby, "Assessing the direct effects of streamflow on recreation: A literature review," Water Resour. Bull. **27**, 979–989 (1992).
[22]D. Botteldooren and B. De Coensel, "Quality labels for the quiet rural soundscapes," *Proceedings of Internoise 2006*, Honolulu, Hawaii, December 2006.
[23]D. Johansson, "Subjective response to audio-visual environments," Masters Thesis **25**, Department of Civil and Environmental Engineering, Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden (2005).

Evaluating airborne sound insulation in terms of speech intelligibility

H. K. Park, J. S. Bradley,[a] and B. N. Gover
National Research Council, Montreal Rd., Ottawa, Canada, K1A 0R6

(Received 13 August 2007; revised 12 November 2007; accepted 6 December 2007)

This paper reports on an evaluation of ratings of the sound insulation of simulated walls in terms of the intelligibility of speech transmitted through the walls. Subjects listened to speech modified to simulate transmission through 20 different walls with a wide range of sound insulation ratings, with constant ambient noise. The subjects' mean speech intelligibility scores were compared with various physical measures to test the success of the measures as sound insulation ratings. The standard Sound Transmission Class (STC) and Weighted Sound Reduction Index ratings were only moderately successful predictors of intelligibility scores, and eliminating the 8 dB rule from STC led to very modest improvements. Various previously established speech intelligibility measures (e.g., Articulation Index or Speech Intelligibility Index) and measures derived from them, such as the Articulation Class, were all relatively strongly related to speech intelligibility scores. In general, measures that involved arithmetic averages or summations of decibel values over frequency bands important for speech were most strongly related to intelligibility scores. The two most accurate predictors of the intelligibility of transmitted speech were an arithmetic average transmission loss over the frequencies from 200 to 2.5 kHz and the addition of a new spectrum weighting term to R_w that included frequencies from 400 to 2.5 kHz.
[DOI: 10.1121/1.2831736]

PACS number(s): 43.55.Rg [NX] Pages: 1458–1471

I. INTRODUCTION

In many situations it is important to accurately rate the sound insulation of walls to the transmission of airborne sounds. These situations could include common walls between two apartments or two row houses, walls between classrooms, or walls between offices and meeting rooms. An "accurate" rating of the sound insulation implies that values of a physical rating measures are strongly correlated with some mean subjective ratings of a panel of listeners. For example, subjects could rate the loudness or annoyance of various types of transmitted sounds. In this paper sound insulation ratings are related to the speech intelligibility scores of subjects listening to transmitted speech sounds. The physical measures for rating sound insulation that were considered included: several standard sound insulation ratings, measures intended to rate the intelligibility of speech, as well as a number of variations of these measures and other related quantities.

Airborne sound transmission through partitions separating dwellings and other spaces is measured over a range of frequencies in standardized tests. In North America the American Society for Testing and Materials (ASTM) E90[1] procedure is used in the laboratory and the ASTM E336[2] procedure is used in field situations. In most other countries the ISO 140 procedures[3] are usually followed to measure airborne sound transmission through walls and floors. These procedures are all very similar and include single number ratings to reduce the results of measurements at a number of frequencies to a single numerical value. The STC (*Sound Transmission Class*) from the ASTM E413 standard[4] and the R_w (*Weighted Sound Reduction Index*) from the ISO 717-1 standard[5] are quite similar in their derivation and are widely used to specify the required airborne sound insulation of common walls in various situations such as between apartments and row houses.

The major differences between STC and R_w are slightly different ranges of included test frequencies, and the "8 dB rule" that is included in the STC rating. The STC rating is derived from sound transmission loss values from 125 Hz to 4 kHz, while the ISO R_w uses the measured sound reduction values from 100 to 3.15 kHz. The 8 dB rule used in calculating STC values limits the maximum allowed deviation below the rating contour at one test frequency to be no more than 8 dB. Although there has been much discussion about the merits of the 8 dB rule,[6,7] it is currently a part of the ASTM standard. The ISO procedure has no such limitation but for both the ASTM and ISO approaches the cumulative deviation below the rating contour summed over all 16 1/3-octave bands cannot exceed 32 dB.

The ISO 717 procedure also includes possible spectrum adaptation terms that are intended to make the ISO procedure more appropriate for rating airborne sound insulation for specific types of sounds.[8] The C-type correction makes the result more representative of an A-weighted sound insulation rating to a pink noise source. The C_{tr}-type correction is intended to provide results that are more representative of adverse responses to road traffic noise sources.

It is important that ratings accurately indicate the perceived rank ordering of the sound insulation of different partitions so that they can be designed to maximize occupant

[a]Author to whom correspondence should be addressed. Electronic mail: john.bradley@nrc-cnrc.gc.ca

satisfaction.[9–13] Vian *et al.*[14] conducted subjective listening tests in the laboratory to assess the adequacy of a French rating method for sound isolation in buildings based on an A-weighted level difference and using a pink noise source signal. They concluded that annoyance responses were most strongly correlated with A-weighted level differences limited to include information only in the 1/3-octave bands from 125 to 4k Hz. Annoyance responses were less strongly correlated with A-weighted level differences for a broader range of included frequencies and not significantly related to the un-weighted sound pressure levels of transmitted music sounds. Annoyance responses were also related to differences among music samples and to the slope of the simulated transmission loss versus frequency curves. The slope between 100 Hz and 1 kHz was said to be the most important portion of the sound insulation curve. Music samples that included words in the language of the listeners (French) were more annoying than those they did not understand (English). The tests included listening to the 144 combinations of 12 different pieces of music and 12 simulated transmission loss versus frequency curves. The simulated sound transmission loss curves were constructed to systematically vary key features such as the slope and the location of coincidence dips and did not exactly model the characteristics of real walls.

Tachibana *et al.*[15] performed an experiment to establish a rating method for airborne sound insulation in buildings using loudness evaluations. They included 11 models to represent sound insulation characteristics and three models for the noises incident on the walls, which were electronically synthesized by filtering the output of a random noise generator. The various test sounds were modified to simulate transmission through the 11 different wall models and were presented to subjects from a loudspeaker in an anechoic room. Using the method of adjustment, subjects determined when reference sounds were equally loud to the sounds transmitted through the simulated walls. Although Steven's Perceived Level was well correlated with the loudness judgments, the arithmetic mean of the transmitted sound levels from 63 to 4k Hz was judged to be the most successful predictor of the loudness responses.

Recent work by Gover and Bradley,[16,17] evaluated a number of measures for rating the speech security of meeting rooms, where speech security indicates very high levels of speech privacy. Although the Articulation Index (AI) and Speech Intelligibility Index (SII) were generally good predictors of the intelligibility of transmitted speech, they were less successful at very low levels of intelligibility because both measures approach gradually to 0 for such situations. Frequency-weighted signal-to-noise ratios (SNRs) (in decibels) were found to be more widely applicable and were also used to assess conditions where speech was just audible but not intelligible.

A frequency-weighted average signal-to-noise ratio with the same frequency weightings as the Speech Intelligibility Index, referred to as SNR_{sii22}, was found to be best correlated with intelligibility scores and subjective evaluations of the threshold of intelligibility. However, a frequency-weighted average signal-to-noise ratio with uniform frequency weightings, SNR_{uni32}, was better correlated with perceptions of the threshold of audibility of the speech. The SNR_{uni32} measure was judged to be the best measure for predicting both the audibility and the intelligibility of speech transmitted through several simulated walls. In these measures the 22 and 32 refer to the level below which the 1/3 octave band signal-to-noise ratios are clipped to avoid contributions lower than −22 dB or −32 dB to the SNR_{sii22} and SNR_{uni32} values, respectively.[16]

To be successful, sound insulation ratings should correlate strongly with subjective evaluations of the acoustical conditions that result when unwanted sounds are transmitted through various partitions. Although this may involve many different types of unwanted sounds, speech is one of the more frequent types. Even for speech sounds, there are many different ways to evaluate the adverse effects of unwanted speech sounds from adjacent spaces. For example, one could rate how annoying or disturbing they are. Where high speech privacy is required it is important that the transmitted speech be not very intelligible. If we assume that the less intelligible the transmitted speech is the less disturbing it will be, then assessing the intelligibility of transmitted speech sounds can be a more precise subjective rating of sound insulation for speech sounds.

This current work therefore examines the merits of a number of measures of airborne sound insulation in terms of the intelligibility of transmitted speech sounds. Subjects rated the intelligibility of speech played through 20 different simulated walls chosen to cover a wide range of different sound insulations and based on the measured characteristics of real walls. Further studies are planned to consider sound insulation rating in terms of the annoyance of transmitted speech and other sounds. However, the current initial investigation is limited to ratings of the intelligibility of speech because there are well-established procedures to obtain quite accurate speech intelligibility scores and because these intelligibility ratings are expected to be strongly related to other subjective ratings of airborne sound insulation.

II. EXPERIMENTAL METHOD

A. Test facility

All tests were conducted in a sound isolated acoustically "dead" room. This had dimensions: 9.2 m long by 4.7 m wide by 3.6 m high. The interior walls were covered with 0.1-m-thick sound absorbing foam. The floor was carpeted and the ceiling consisted of a suspended T-bar system with 1-in.-thick glass fiber ceiling tiles. The structure is a massive concrete construction that was vibrationaly isolated from the rest of the building. The measured background noise level in the room with no experimental sounds was about 12 dBA.

The test speech was played over loudspeakers positioned 2 m in front of the subject. The simulated background noises were played over another set of loudspeakers positioned above the T-bar ceiling, directly above the subject.

The speech sounds were modified to simulate propagation through various walls using programmable digital audio equalizers that were part of two Yamaha Digital Mixing Engines (DME32). These are very flexible signal processing boxes, able to perform the functions of many interconnected

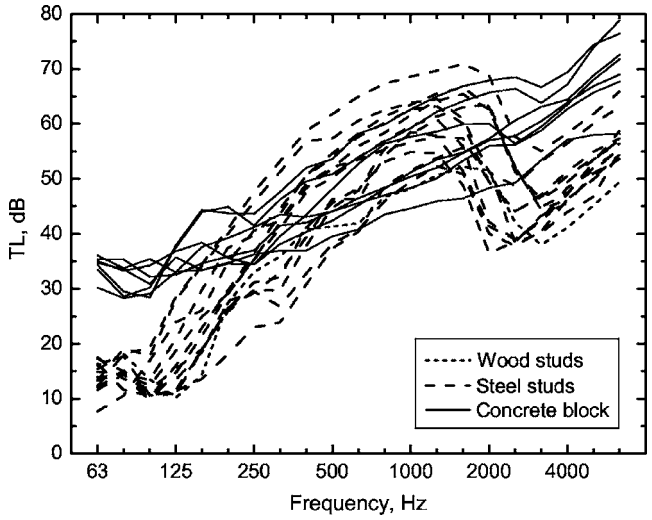


FIG. 1. Sound transmission loss (TL) vs 1/3-octave band frequency for the 20 walls simulated in the listening tests illustrating the wide range of included characteristics. Walls containing wood studs, steel studs, and concrete blocks are separately identified by different line types.

devices and can easily be controlled by computer using a MIDI interface. One component in each DME32 was used to equalize the playback path through the power amplifiers and loudspeakers to be flat at the position of the listener's head (± 1 dB from 60 to 12,000 Hz).

B. Speech and noise sounds

The speech tests used the Harvard sentences.[18] These are phonetically balanced English sentences with content that is of low predictability. This is important to minimize the effects of guessing. The sentences were all recorded by the same clear speaking male talker and were all played, at the same fixed source level, through 1/3-octave band equalizers set to represent the transmission loss characteristics of 20 different walls. Thus, the variations in the speech sounds that subjects heard were due only to the different simulated transmission characteristics.

The simulated ambient noise had a −5 dB/octave spectrum shape and was intended to approximate typical indoor ventilation noise. The overall level of the noise was 35 dBA and was the same for all simulated walls and test sentences. This level is commonly found at quieter times in homes[19] and meeting rooms[17] and is the recommended maximum ambient noise level for school classrooms.[20]

C. Simulated wall transmission characteristics

The sound transmission characteristics of 20 different real walls were selected from a large database of test results. A large number of measured wall characteristics were first assembled that included a broad range of STC ratings. From these, 20 wall characteristics with STC ratings evenly distributed from STC 34 to STC 58 were selected. For the chosen fixed speech playback level, these were expected to lead to intelligibility scores varying between 0% and 100%.

The sound transmission loss versus frequency curves for the 20 selected walls are shown in Fig. 1. The shapes and overall levels of the transmission loss values varied considerably and the data represent a broad range of real walls. The walls containing wood studs, steel studs and concrete blocks are separately identified in this figure and are seen to have quite different characteristics.

Table I provides a summary of the wall constructions and their STC and R_w ratings. The walls are common constructions in North America. The descriptor codes are explained in Table II. For example, wall number 17 is described as, G16_GFB90_SS90_2G16, which indicates the various layers of the construction from one side to the other.

TABLE I. Summary of simulated wall constructions with their STC and R_w ratings.

No.	Descriptor	STC rating	R_w rating
1	G13_GFB90_WS89_G13	34	37
2	G13_SS65_G13	34	33
3	G16_SS65_G16	35	37
4	G16_SS90_G16	36	37
5	G16_SS90_G16	37	36
6	G13_GFB90_SS90_G13	39	41
7	G16_SS40_AIR10_SS40_G16	39	38
8	G13_GFB90_SS90_G13	40	42
9	G13_GFB65_SS65_G13	43	43
10	BLK90	44	44
11	G16_MFB40_SS90_G16	45	45
12	BLK140	47	47
13	G16_GFB90_SS90_G16	47	45
14	BLK190_PAI	48	48
15	G16_BLK190_G16	49	50
16	BLK190	50	50
17	G16_GFB90_SS90_2G16	52	50
18	2G13_GFB90_SS90_2G13	53	52
19	PAI_BLK140_WFUR40_GFB38_G13	56	55
20	PAI_BLK140_GFB38_WFUR40_G13	58	56

TABLE II. Explanation of symbols used to describe the simulated wall constructions in Table I.

Descriptor	Explanation
AIR	Air space
BLK	Concrete block
G	Gypsum board
GFB	Glass fiber batt
MFB	Mineral fiber batt
PAI	Paint
SS	Steel stud
WFUR	Wood furring
WS	Wood stud

In this case the construction includes: 16 mm gypsum board (G16), 90 mm glass fiber batts (GFB90), 90 mm steel studs (SS90), and two layers of 16 mm gypsum board (2G16).

D. Subjects and listening tests

Subjects were all employees of the National Research Council (NRC) who volunteered to do the test. They were not paid or rewarded in any way and the experimental protocol was approved by the NRC Research Ethics Board (protocol 2006-27). A total of 15 subjects completed the test. As will be described later, this was found to be an adequate number of subjects because the standard errors of the mean speech intelligibility scores were quite small.

All subjects were first given a hearing sensitivity test. Their pure tone average hearing levels varied from −1.7 dB to 9.2 dB (averaged for both ears over the test frequencies 500, 1k and 2k Hz). By comparing the average hearing levels of the subjects with the expected distribution of a population of normal hearing listeners from ISO 7029 (1984),[21] the subjects' average result approximately matched the 50th percentile of the ISO results.

To familiarize subjects with the types of sounds they would hear, they first listened to ten different test sentences played through ten different simulated walls that varied from very low to very high STC rating. They were told that the practice sentences were representative of the full range of conditions that they would hear in the complete test. In the complete test, listeners heard five different Harvard sentences through each of the 20 simulated walls for a total of 100 sentences. The order of the sentences and of the walls was randomized so that subjects heard conditions in one of three different randomized orders of the tests.

After hearing each test sentence, subjects repeated back the words that they had understood, using a microphone, to the experimenter who was located outside the test room. The percentage of the words in each sentence that they had understood was their speech intelligibility score.

Only the simulated transmission characteristics of the walls were varied. The effective speech source level and the ambient noise level at the listener's position remained constant throughout the tests.

E. Analyses

The results were analyzed in terms of the average intelligibility scores for all listeners and all test sentences for each wall. That is, each average speech intelligibility score was an average of scores for 5 sentences ×15 subjects or 75 different test scores.

Most of the analyses consisted of plots of the mean speech intelligibility scores versus a sound insulation rating measure such as STC. To test the strength of the correlation between the intelligibility scores and the sound insulation ratings, best-fit regression lines were fitted to the results using a Boltzmann equation and the related R^2 values (coefficient of determination) were calculated. The Boltzmann equation is given by the following

$$y = \frac{A_1 - A_2}{1 + e^{(x-x_0)/dx}} + A_2, \tag{1}$$

where

A_1 = y value for $x=-\infty$ (0% or 100%)
A_2 = y value for $x=+\infty$ (0% or 100%)
x_0 = x value of mean y value, that is, the x value when $y=50\%$ in our case
dx = relates to the slope of the mid-part of the regression line.

The Boltzmann equation represents a sigmoidal shaped curve that approaches asymptotically to 0% and 100% intelligibility for the extreme values of the sound insulation measure (i.e., the A_1 and A_2 values were set to either 0% or 100%). As many of the later results in this paper indicate, the Boltzmann equation is a good fit to the relationships between speech intelligibility scores and the various sound insulation rating measures.

All of the results presented were statistically significant. Since there were always 20 data points and the same format of regression equation, the significance is simply related to the R^2 value. Any R^2 value $\geqslant 0.193$ is statistically significant ($p<0.05$) and any R^2 value $\geqslant 0.317$ is statistically significant ($p<0.01$).

III. EVALUATIONS OF EXISTING MEASURES

In this section the results of evaluations of the relationships between mean speech intelligibility scores and various existing measures that are either described in current standards or have been suggested in previous studies. These include: standard ratings of airborne sound insulation, speech intelligibility measures, arithmetic average transmission loss values and loudness related quantities. Each of the following three subsections presents the results for one group of existing measures.

A. Evaluation of standard sound insulation measures

The standard sound insulation ratings STC and R_w are first evaluated. Figure 2 plots both the STC ratings of the simulated walls and the mean intelligibility scores (with associated standard errors) versus the wall number (as listed in Table I). It is seen that the standard errors of the mean speech

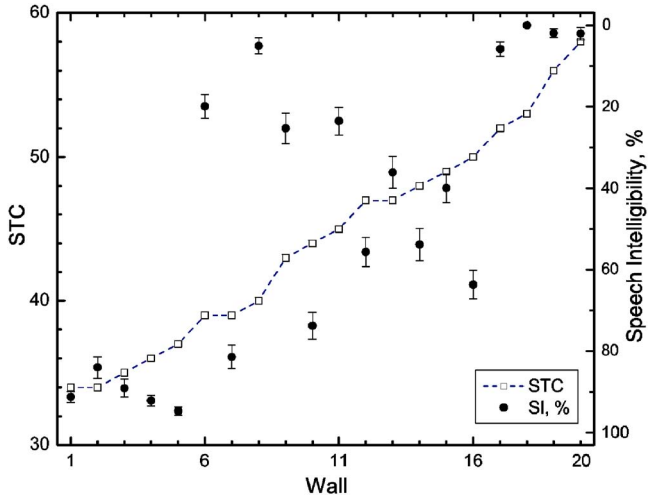


FIG. 2. (Color online) Mean speech intelligibility (SI) scores (right hand axis) with error bars indicating ±1 standard error and Sound Transmission Class (STC) values (left hand axis) plotted vs wall number (given in Table I).

intelligibility scores are small enough to describe the differences between the scores for the different walls. The walls are in order of increasing STC rating and the plot illustrates the even distribution of the included STC values. However, it is clear that this order does not correspond well to the decreasing speech intelligibility scores.

When these mean intelligibility scores are plotted versus the STC values of the simulated walls [see Fig. 3(a)] and a Boltzmann type regression line is fitted to the data, the results are statistically significant but with a modest R^2 value (R^2=0.510). STC is therefore not a very good predictor of the intelligibility of the transmitted speech.

Figure 3(b) plots mean intelligibility scores versus values of the standard R_w rating for the simulated walls. The scatter is similar to that for the STC ratings and again there is a significant but weak relationship (R^2=0.542). This is not surprising because of the similarity of the STC and R_w procedures. Since neither STC nor R_w is strongly related to the intelligibility scores, in their currently standardized form, neither is a very good predictor of the intelligibility of the transmitted speech.

B. Evaluation of signal-to-noise ratio speech intelligibility type measures

Various measures have been developed to rate the expected intelligibility of speech based on signal-to-noise ratios with frequency weightings to account for the relative importance of the signal-to-noise ratio in each frequency band to the intelligibility of the speech. Several of this type of measure are evaluated in this section to determine how well they are related to the intelligibility of speech transmitted through walls.

Figure 4(a) plots the mean intelligibility scores versus values of the Articulation Index (AI)[22] The results in Fig. 4(a) have higher R^2 values than those in the previous section indicating that the AI is a better predictor of the intelligibility of the transmitted speech in spite of the large spectral modification resulting from transmission through the simulated

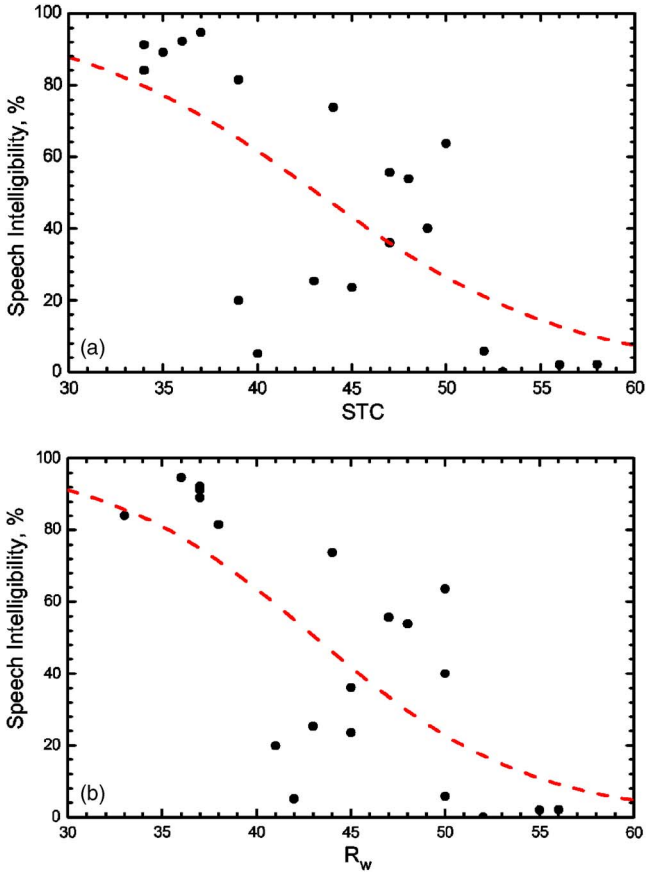


FIG. 3. (Color online) Speech intelligibility scores vs (a) STC ratings (R^2=0.510) and (b) R_w ratings (R^2=0.542) of the simulated walls.

walls. Of course AI is the result of considerable research effort over many years[23] to develop a measure for predicting expected speech intelligibility.

Figure 4(b) plots the mean intelligibility scores versus the Speech Intelligibility Index (SII).[24] These results are very similar to those for AI values in the previous figure because the two measures are so similar. The R^2 values when SII is the predictor in Fig. 4(b) are very slightly higher than those for AI values in Fig. 4(a). Both AI and SII are seen to be relatively successful predictors of these intelligibility scores as would be expected.

Because AI and SII values approach gradually to zero for low signal-to-noise ratios, they are not as useful for cases of very low speech intelligibility. To solve this problem Gover and Bradley[16] proposed a simple frequency-weighted signal-to-noise ratio in which the frequency weightings were those of the SII measure. Figure 5(a) shows a plot of mean intelligibility scores versus values of this SII-weighted signal-to-noise ratio (SNR_{sii22}). The relationship is highly significant (R^2=0.913) and agrees well with the previously published regression line,[16] which was based on the characteristics of only four different walls. Very similar results were obtained when using the frequency weightings in the AI measure (R^2=0.896), because the two measures (SNR_{ai} and SNR_{sii22}) are so similar.

Gover and Bradley[16] also related estimates of the onset or threshold of intelligibility to SNR_{sii22} values. If subject's intelligibility scores indicted they had understood at least one

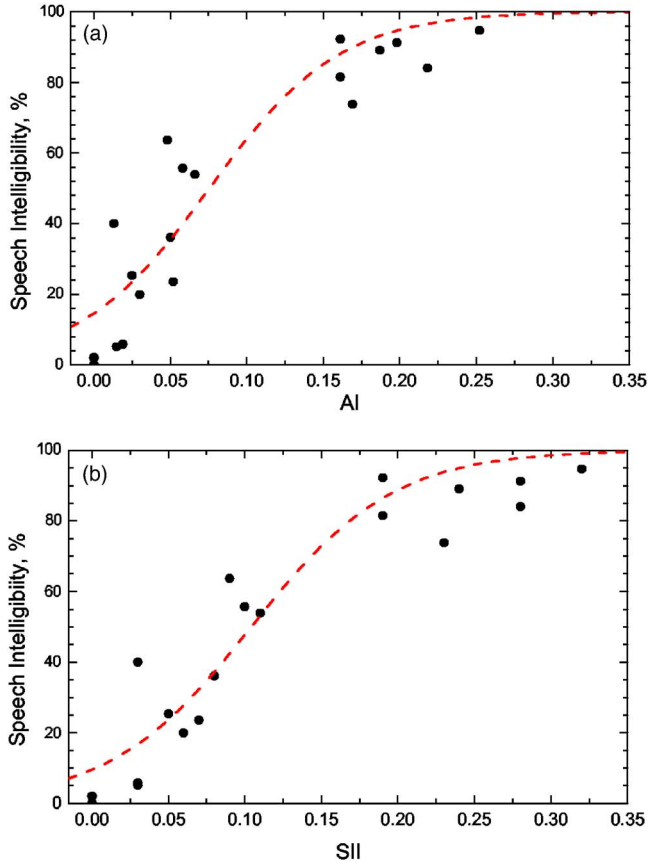


FIG. 4. (Color online) (a) Speech intelligibility scores vs AI ratings of the test conditions (R^2=0.864), and (b) speech intelligibility scores vs SII ratings of the test conditions (R^2=0.899).

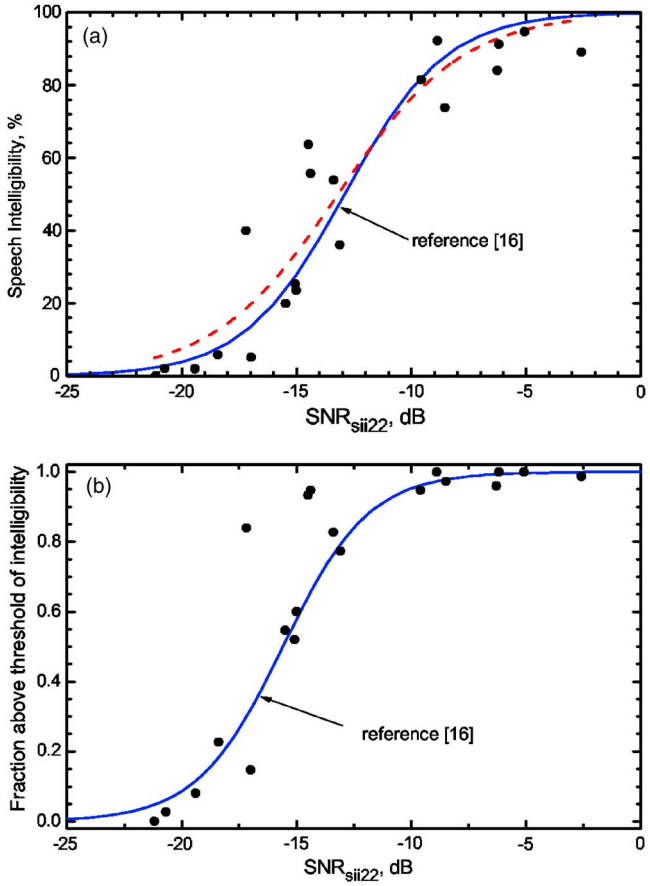


FIG. 5. (Color online) (a) Speech intelligibility scores vs SNR_{sii22} ratings of the test conditions (R^2=0.913). Solid line from previous experiment Ref. 16 and (b) fraction of responses above the threshold of intelligibility versus SNR_{sii22} ratings of the test conditions (R^2=0.822). The solid line is from a previous experiment Ref. 16 and is essentially the same as the best fit to the new data.

word, they were considered to be above the threshold of intelligibility. For each wall, the fraction of the responses indicating that at least one word was understood were determined. When plotted versus SNR_{sii22} values, the point at which 50% of the subjects could understand at least one word was defined as the threshold of intelligibility. The intelligibility scores in the current tests were examined to obtain similar estimates of the onset of intelligibility. The fraction of the responses indicating at least one word was understood is plotted versus SNR_{sii22} values in Fig. 5(b). The solid line is the best fit Boltzmann equation from the previous study and is seen to be a good fit to the data from the current tests.

The previous work found that although the SNR_{sii22} measure was a good predictor of speech intelligibility scores and of the threshold of intelligibility results, a uniform weighted signal-to-noise ratio, SNR_{uni32} was a better compromise as a predictor of both intelligibility and audibility judgments. The intelligibility scores from the current study are plotted versus SNR_{uni32} values in Fig. 6(a). In this figure the dashed line is a best fit to the current data and the solid line is the best fit line from the previous study.[16]

The A-weighted level difference is sometimes proposed as a simple measure of the intelligibility of speech. Mean speech intelligibility scores are plotted versus the difference between the A-weighted level of the transmitted speech and the A-weighted noise level, SNR(A), in Fig. 6(b). There is clearly not a good relationship with speech intelligibility scores. SNR(A) values are not accurate indicators of the intelligibility of speech in noise and should not be used for predicting expected speech intelligibility.

C. Evaluation of loudness and frequency weighted TL summations

In this section several measures that average transmission loss (TL) values over various frequency ranges are considered. In Fig. 7(a), the mean intelligibility scores are plotted versus values of the Articulation Class (AC) divided by 10 for the simulated walls. (In this plot the AC values were divided by 10 to help comparisons by giving similar magnitude values to the other quantities). AC values are a frequency-weighted summation of transmission loss values over the 1/3 octave bands from 200 to 5000 Hz.[25] The frequency weightings are those from the AI measure. Although the R^2 values are quite high, intelligibility scores are a little less well correlated with AC values than with the various signal-to-noise ratio type speech intelligibility measures in the previous section. Of course, if speech source levels and ambient noise levels were also varied in this experiment, AC values would be a less successful predictor of speech intelligibility scores than signal-to-noise ratio type measures.

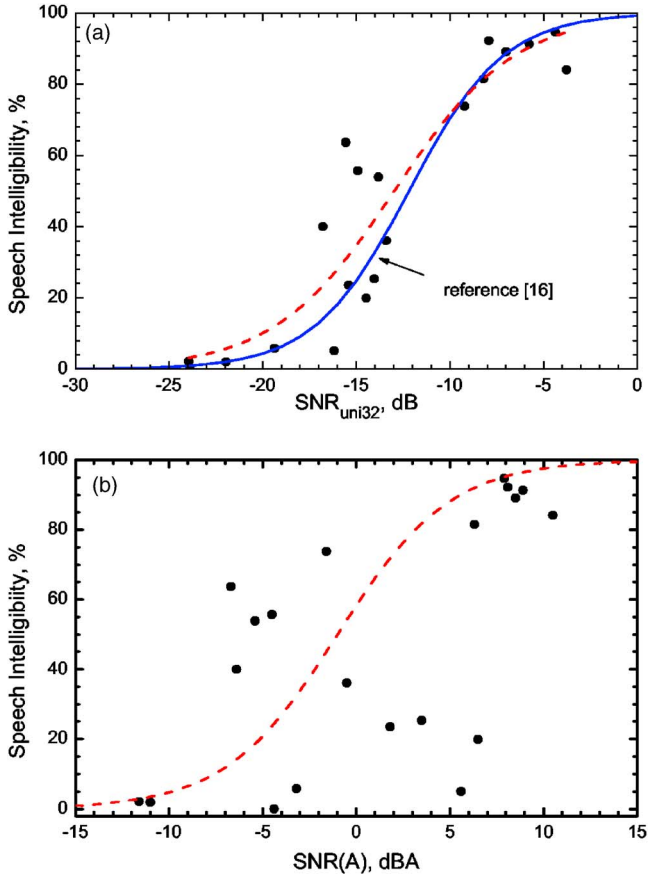


FIG. 6. (Color online) (a) Speech intelligibility scores vs SNR_{uni32} ratings of the test conditions (R^2=0.853). Solid line from previous experiment Ref. 16 and (b) mean speech intelligibility scores vs SNR(A) ratings of the test conditions (R^2=0.259).

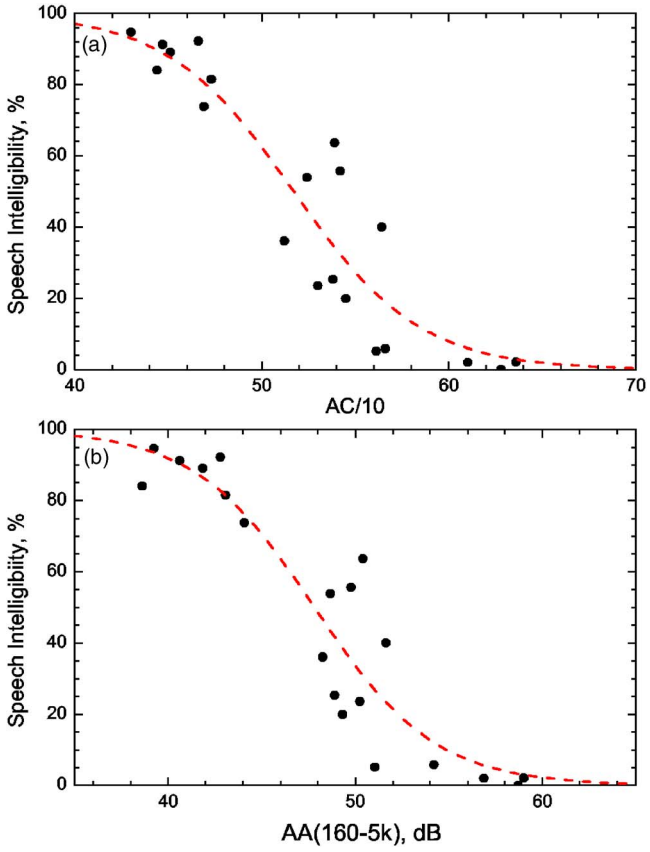


FIG. 7. (Color online) (a) Speech intelligibility scores vs the AC/10 ratings of the simulated walls (R^2=0.856), and (b) speech intelligibility scores vs AA(160-5k) ratings of the simulated walls (R^2=0.853).

However, the intent here was to evaluate the various measures as sound insulation ratings of the walls.

The success of the uniform weighted signal-to-noise ratio, SNR_{uni32}, in previous speech security studies,[16,17] suggests that a simple arithmetic average of transmission loss value over speech frequencies should be a reasonably good predictor of the intelligibility of transmitted speech. Values of the Arithmetic Average transmission loss [AA(160-5k)] were calculated over the frequencies from 160 Hz to 5 kHz. Mean intelligibility scores are plotted versus these AA(160-5k) values in Fig. 7(b). The R^2 values indicate that this measure is a better predictor of intelligibility scores than the standard STC and R_w ratings but not quite as good as the better speech intelligibility type measures in the previous section. However, this will be seen to be strongly influenced by the frequency bands that are included in calculating the average transmission loss value.

Many broadband acoustical measures are derived by adding or averaging the energy over a number of frequency bands to get a single broadband measure. For example, an A-weighted sound level can be obtained by A weighting the levels in each 1/3-octave band and adding the sound energies of each weighted 1/3-octave band level to obtain the broadband A-weighted level. This leads to different results than the arithmetic averaging of decibel measures, which is a part of many speech intelligibility measures described in the previous section, such as AI and SII. Several quantities were investigated that involve energy averaging or energy summation of the information from various frequency bands. For example, when speech intelligibility scores were plotted versus values of the energy average of the transmission loss values from 160 to 5000 Hz, the result was less successful (R^2=0.446) than were STC values and for most of the previous results.

STA is an A-weighted sound transmission loss measure.[26] It is calculated by A-weighting and summing the 1/3-octave band transmission loss values (from 100 to 4k Hz) on an energy basis as shown in the following equation:

$$\mathrm{STA} = -10 \cdot \log\left\{\frac{1}{17}\sum_{b=100}^{4k} 10^{(-TL_b + Awt_b)/10}\right\}, \text{ dB}, \qquad (2)$$

where, Awt_b is the A-weighting attenuation in band "b" and TL_b is the sound transmission loss in band "b." Although this had been a successful predictor of survey response ratings of the sound insulation of common walls between apartments and row houses, it was only weakly correlated with speech intelligibility scores (R^2=0.361).

While A weighting can be thought of as a simple loudness measure, the Zwicker loudness level[27] procedure is intended to give ratings more closely related with loudness

judgments. The Zwicker loudness (sones) of the transmitted speech sounds were only weakly related with the speech intelligibility scores ($R^2=0.305$).

Although some measures such as AC and AA(160-5k) are somewhat more strongly correlated with intelligibility scores than the standard STC and R_w measures, they are not as well correlated with intelligibility scores as the speech intelligibility type measures. This is, perhaps, reassuring in that the speech intelligibility measures in the previous section are based on extensive research to understand how we respond to speech in noise.

IV. EVALUATIONS OF VARIATIONS OF EXISTING MEASURES

This section presents the results of several variations of existing measures and explorations of the most important frequencies to be included in various quantities. The first section below explores the more general question of the importance of the frequencies included in rating measures. These results are used to guide the investigations in the following subsections in which systematic variations in a number of the existing sound insulation ratings are explored.

A. Effects of included frequencies

The correlations of intelligibility scores with various possible sound insulation rating measures are expected to be influenced by the frequencies included in their calculation. For example, if very low frequency transmission loss information is included, this is likely to reduce the correlations, since the added information does not contain information related to the intelligibility of the transmitted speech. The AI measure includes information in the 1/3-octave bands from 200 Hz to 5 kHz, while the SII measure includes values from the 160 Hz to the 8 kHz 1/3-octave bands. Neither includes low frequency information unrelated to speech intelligibility. However, a different range of frequencies may be more important for predicting speech intelligibility in this study because the transmission through the various walls modifies the spectral characteristics of the speech by severely attenuating the higher frequencies.

Figure 8 illustrates the more important frequency bands in this study by showing the results of simple correlations of the intelligibility scores with the transmission loss values in each 1/3-octave band. It is seen that including the 1/3-octave bands from 160 Hz to 2.5 kHz yields the highest correlations and hence these bands are most important for predicting intelligibility scores. Lower frequency bands are less important because there is less speech energy in human voices below 160 Hz. The higher frequency bands above 2.5 kHz are less important because little speech energy in this range is effectively transmitted through most walls. This is different than the range of included frequencies in the AI and SII measures that were developed to consider natural unmodified speech. The severe filtering that results from sound transmission through typical walls removes the significance of these higher frequencies.

Further analyses were carried out to determine the optimum range of frequencies to include when predicting the

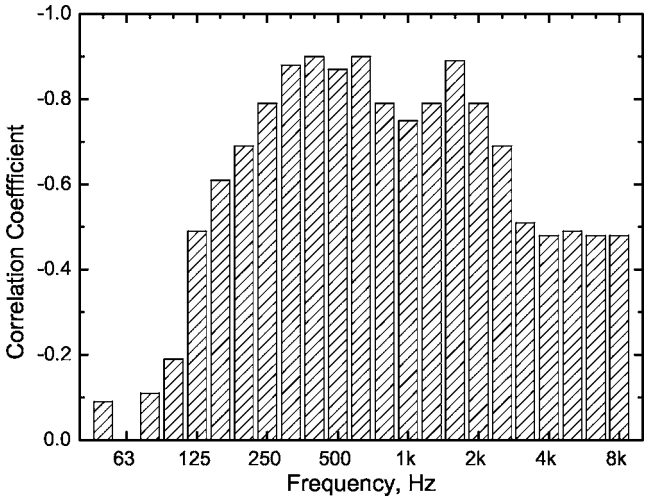


FIG. 8. Results of correlations between mean intelligibility scores and individual 1/3-octave band sound transmission loss values.

intelligibility of transmitted speech. These analyses involved recalculation of an Arithmetic Average (AA) transmission loss measure for varied lowest and highest included frequency bands. The previous AA results in Fig. 7(b) were calculated by arithmetic averaging of the sound transmission loss values for the frequency bands from 160 to 5k Hz inclusive. In these new analyses the lowest included frequency was systematically increased from 63 to 2000 Hz. At the same time, the highest included 1/3-octave band was varied from 200 to 6300 Hz. This resulted in a matrix of correlation coefficients that are shown in Fig. 9. This contour plot shows the combinations of highest and lowest included 1/3-octave

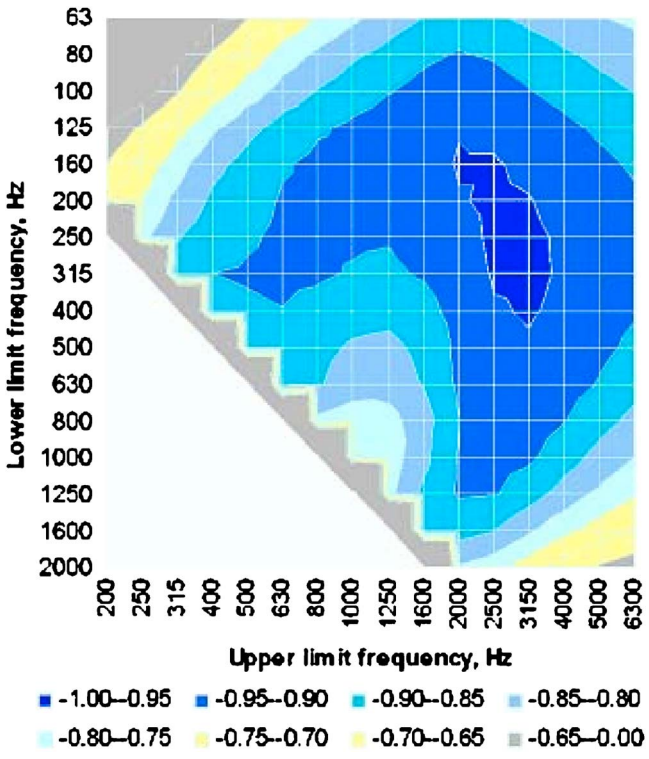


FIG. 9. (Color online) Correlation coefficients from correlating mean speech intelligibility scores with arithmetic average sound transmission loss values for varied highest and lowest included frequencies.

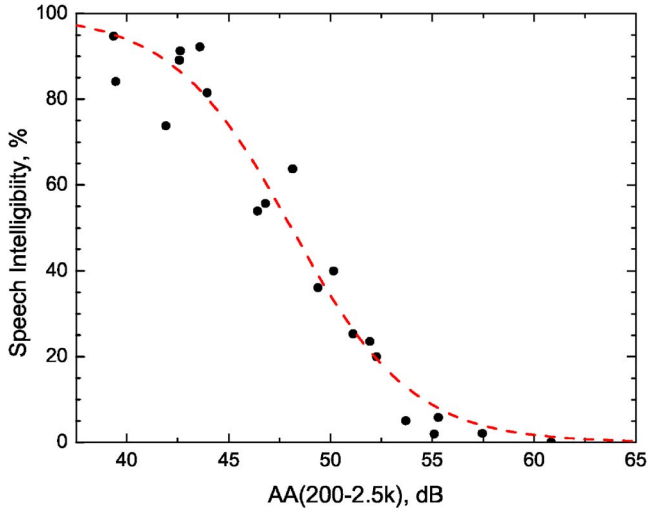


FIG. 10. (Color online) Speech intelligibility scores vs AA(200-2.5k) ratings of the simulated walls (R^2=0.959).

band frequencies, that lead to the highest correlations with intelligibility scores. The darkest shaded area towards the upper right of the graph shows the highest magnitude correlation coefficients with values of −0.95 to −0.96.

The highest magnitude correlation coefficient is obtained when the lowest included frequency is increased to about 200 Hz and the highest included frequency is reduced to about 2.5 kHz. Some other adjacent combinations can result in essentially the same correlation coefficient.

The success of the Arithmetic Average transmission loss with an included frequency range of from 200 Hz to 2.5 kHz is illustrated in Fig. 10, which plots speech intelligibility scores versus this measure. The data points follow the best-fit regression line quite closely and a quite high R^2 value was obtained (R^2=0.959). It is clear that by eliminating the irrelevant information at lower and higher frequencies, better correlations can be achieved with intelligibility scores.

A similar type of systematic analysis of the importance of the included frequencies was also tried for an energy average type measure. The correlation coefficients were largest in magnitude for a restricted range of intermediate frequencies. This limited range of frequencies would probably only be useful for speech and the results indicated weaker relationships than those for the arithmetic averages of transmission loss values.

B. Variations of standard measures

1. Variations of maximum and total deficiency rules

This section describes the results of further analyses to investigate the causes of the limitations of the STC measure and how it might be improved. One of the key differences between STC and the R_w is the inclusion of the 8 dB rule in the ASTM measure. The 8 dB rule limits the maximum deviation that is allowed below the rating contour and can lead to some peculiar results. For example, the transmission loss values of the two walls illustrated in Fig. 11(a) look quite different, but have the same STC 39 rating. This is because the 8 dB rule limits the STC rating. From Table I, the R_w ratings for these walls are seen to be 38 and 41.

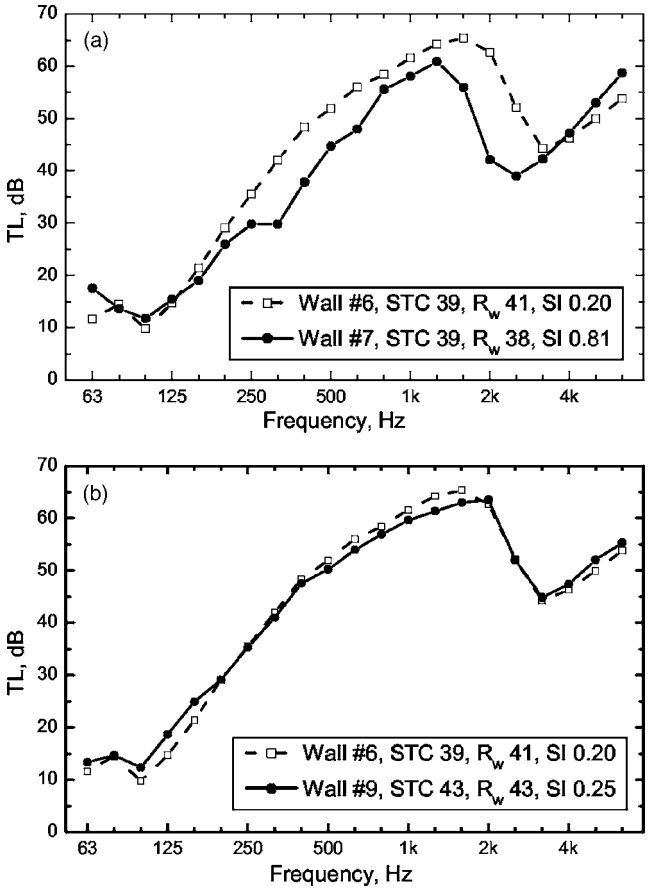


FIG. 11. (a) Comparison of the sound transmission loss vs frequency characteristics of two walls having the same STC 39 rating, (b) comparison of two very similar sound transmission loss vs frequency characteristics of two walls having different STC ratings (STC 39 and STC 43). In each case the legend gives the STC, R_w and the mean SI score for each wall. (See Table I for description of the constructions of walls Nos. 6, 7, and 9).

On the other hand, Fig. 11(b) illustrates the sound transmission loss values for two walls that seem to have very similar characteristics but have different STC values. Again this is a result of the 8 dB rule and the R_w ratings (that do not include an 8 dB rule) from Table I are more similar than the STC ratings for these walls.

In both cases the confusing results are due to the application of the 8 dB rule. To test the appropriateness of the 8 dB rule, modified STC values were calculated without an 8 dB rule. Speech intelligibility scores are plotted versus this modified STC value in Fig. 12. The modified STC measure without the 8 dB rule is significantly related to speech intelligibility scores. The results for the standard STC measure showed an R^2 value of 0.510 for the Boltzmann equation fit. Therefore the STC measure is a better predictor of the intelligibility of transmitted speech when the 8 dB rule is removed.

It is possible that some variation in the magnitude of the 8 dB rule might improve it. Therefore STC values were recalculated with an 8 dB type rule but with the magnitude of the allowed maximum deficiency varied from 1 to 16 dB as well as the no 8 dB rule case. Each set of modified STC values was correlated with the mean speech intelligibility scores and the resulting correlation coefficients are plotted in Fig. 13(a).

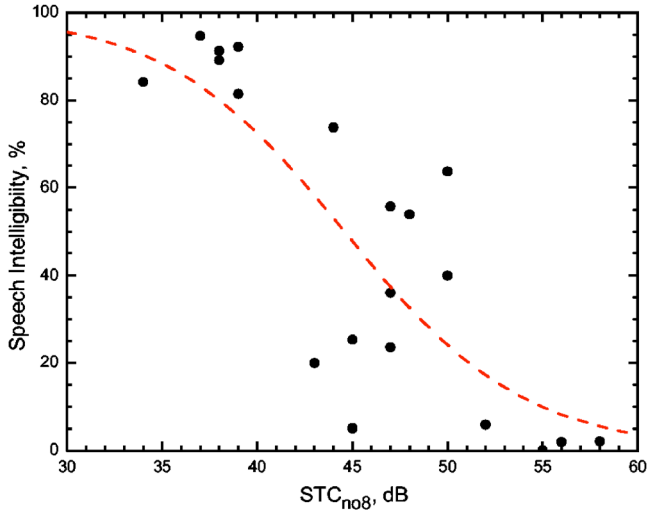


FIG. 12. (Color online) Speech intelligibility scores vs modified STC ratings without an 8 dB rule for the simulated walls (R^2=0.661).

The results in Fig. 13(a) show that the modified STC measure correlates best with speech intelligibility scores when there is no maximum deficiency limit (i.e., no 8 dB rule).

The STC procedure also includes a limit on the total deficiency summed over all frequencies from 125 Hz to 4 kHz. Further analyses were carried out to determine whether varying the 32 dB limit could improve the STC measure. Figure 13(b) shows correlation coefficients between speech intelligibility scores and the modified STC values for varied maximum allowed deficiency from 0 to 60 dB and with the 8 dB rule included and unmodified. Figure 13(b) also shows correlation coefficients between speech intelligibility scores and modified STC values with the maximum acceptable deficiency varied from 0 to 60 dB and without the 8 dB rule.

When the 8 dB rule is included, Fig. 13(b) indicates that a little smaller maximum deficiency would lead to slightly improved correlations with speech intelligibility score. However, when the 8 dB rule was excluded, the results in Fig. 13(b) show that when the total deficiency is equal to or greater than about 30, further increases have very little effect on the correlations with intelligibility scores. The standard 32 dB total deficiency value seems to be safely into the range where it works well and is insensitive to changes for the situation without an 8 dB rule. This result suggests that one should not contemplate changing the total deficiency limit in either the STC or R_w measures, because the currently used value of 32 dB seems to be acceptable for predicting the intelligibility of transmitted speech.

2. Variations of spectrum adaptation terms

The ISO 140 procedure includes the option of adding spectrum adaptation terms, to better predict the expected responses for specific types of sounds. The C-type spectrum weighting is intended to provide results similar to an A-weighted sound transmission loss to a pink noise source.

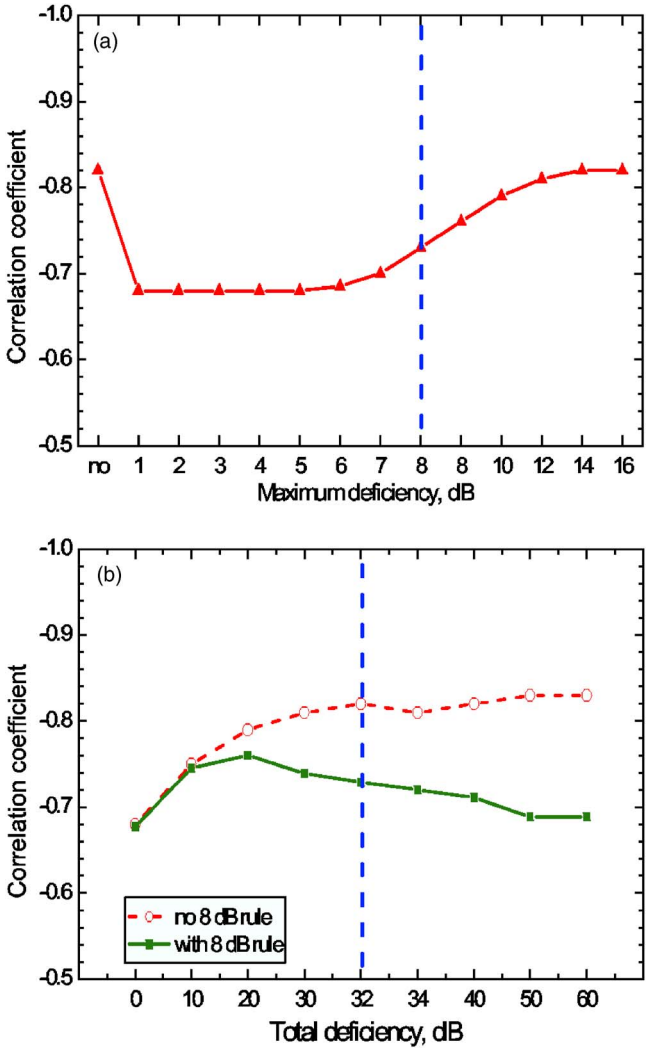


FIG. 13. (Color online) (a) Correlation coefficients between speech intelligibility scores and a modified STC value for which the maximum acceptable deficiency was varied from 1 to 16 dB including a "no" maximum deficiency limitation case, and (b) correlation coefficients between speech intelligibility scores and modified STC values for which the total acceptable deficiency was varied from 0 to 60 dB with and without an 8 dB rule. Vertical dashed lines indicate the standard values of these criteria.

The C_{tr}-type weighting is intended to provide sound insulation ratings that better relate to responses to road traffic noise transmitted through various constructions.

When either of these spectrum weightings was added to the R_w values, the correlations with speech intelligibility scores were greatly reduced. This is not surprising because neither is intended to improve the prediction of sound insulation ratings for speech sounds. The results for the R_w+C values led to very similar weak correlations with speech intelligibility scores (R^2=0.359) as those for the A-weighted transmission loss measure STA (R^2=0.361). This may be because the C-type spectrum correction is intended to produce results equivalent to an A-weighted transmission loss measure. The C_{tr}-type spectrum correction is intended to give better predictions for road traffic noise, which is spectrally quite different than speech and explains why it leads to much reduced correlations (R^2=0.205).

New spectrum weightings were investigated to deter-

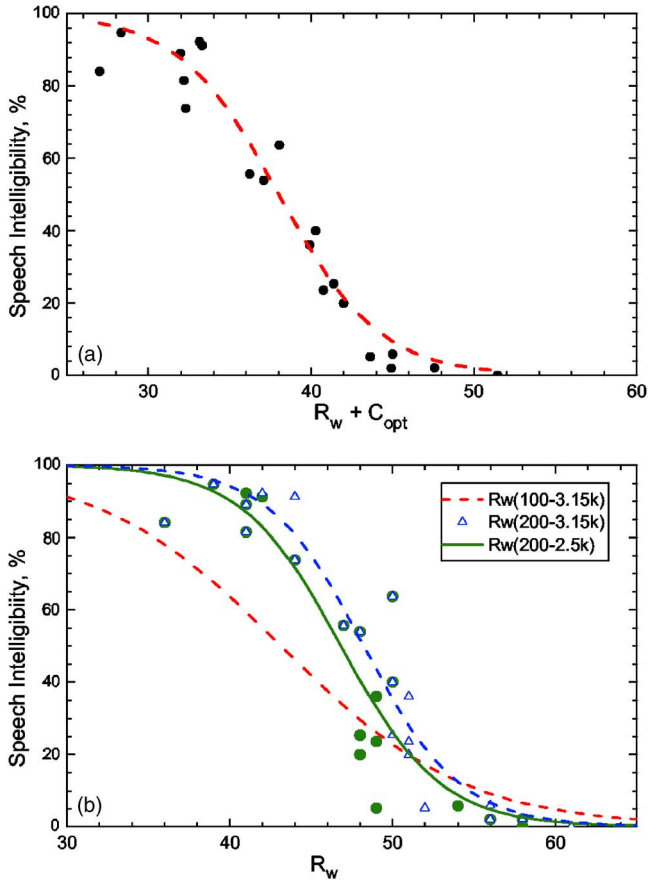


FIG. 14. (Color online) (a) Speech intelligibility scores vs R_w ratings with an added C_{opt}-type spectrum weighting correction for the simulated walls (R^2=0.957) and (b) speech intelligibility scores versus modified R_w ratings with varied included frequencies. (100–3.15 kHz, R^2=0.542; 200–3.15 kHz, R^2=0.842; 200–2.5 kHz, R^2=0.922).

mine if they would lead to better predictors of the intelligibility of the transmitted speech. This was first done by systematically varying the highest and lowest frequency limits of the included 1/3-octave bands. This is similar to what was done in Sec. IV A to develop an improved arithmetic average measure. However, here the reduced frequency range was used to create an ISO type spectrum weighting correction. For this simple form of spectrum weighting, the best result was obtained by limiting the included frequencies to the range from 400 Hz to 2.5 kHz. These frequencies were given a 0 dB weighting and all other frequencies were given a −50 dB weighting. When mean intelligibility scores were plotted versus the combination of $R_w+C_{400-2.5k}$ there was quite a good fit with an associated R^2 value of 0.941.

This quite simple spectrum weighting correction was further improved by adding additional frequency bands with small amounts of attenuation. This led to further small improvements in the relationship with speech intelligibility scores. Only two more 1/3-octave bands were included. The 315 Hz band was included with a weighting of −8 dB and the 3.15 kHz band was included with a weighting of −3 dB. When mean speech intelligibility scores are plotted versus R_w values with this new C_{opt} spectrum weighting term [see Fig. 14(a)], the related R^2 value for the Boltzmann fit was 0.957.

V. VARIATIONS OF INCLUDED FREQUENCIES AND RATING CONTOUR

The STC and R_w measures are also a little different in the range of frequencies that are included in each. Results in Sec. IV A, where the range of included frequencies was investigated, indicated that eliminating irrelevant frequencies could improve the correlations with speech intelligibility scores. For an arithmetic average type transmission loss measure, a frequency range of 200 Hz to 2.5 kHz was found to provide maximum correlations with speech intelligibility scores. Attempts were made to create improved STC and R_w values by limiting the frequencies included in their calculation. Because it has already been shown that removing the 8 dB rule from the STC calculation improves STC as a predictor of speech intelligibility and the inclusion of the 4 kHz 1/3-octave band is easily shown to degrade predictions of these speech intelligibility scores, only the results of variations of R_w values are included.

Variations of the R_w rating were considered in an attempt to find measures that would better predict the intelligibility of transmitted speech. Figure 14(b) compares results for three different ranges of included frequencies. In this figure, speech intelligibility scores are plotted versus the standard R_w (100–3.15 kHz) and two modified versions of R_w for included frequency ranges of 200–3.15 kHz, and 200–2.5 kHz.

Although values of the standard version of R_w (100–3.15 kHz) are not so strongly related to speech intelligibility scores (R^2=0.542), the two variations with a more limited range of included frequencies are much improved predictors of speech intelligibility scores. When the included frequencies were limited to 200–3.15 kHz, R^2 was 0.842, and for a 200 to 2.5 kHz range R^2 was 0.9222. Again limiting the included frequencies to those most relevant for transmitted speech provides improved correlations with intelligibility scores.

As a final variation of the R_w measure, sound insulation ratings were calculated following the R_w procedure except that a different shape of rating contour was used. This rating contour was the same as that of the C_{opt} spectrum adaptation term. These R_{wopt} values also led to improved relationships with speech intelligibility scores (R^2=0.951). However, further variations were not considered because the new spectrum adaptation terms were thought to be more accurate and also practically more acceptable.

VI. SUMMARY AND DISCUSSION OF RESULTS

A large number of variations of sound insulation rating measures have been tested as predictors of the intelligibility of speech transmitted through 20 different simulations of real walls. These results are summarized in Fig. 15. This figure plots the R^2 values for Boltzmann equation best-fit regression lines to the data. Table III includes the same values as well as the range of frequencies included in each measure and a short description of each measure.

While using the standard STC and R_w measures led to only moderate R^2 values between 0.5 and 0.6, many other forms of sound insulation rating measures led to much higher

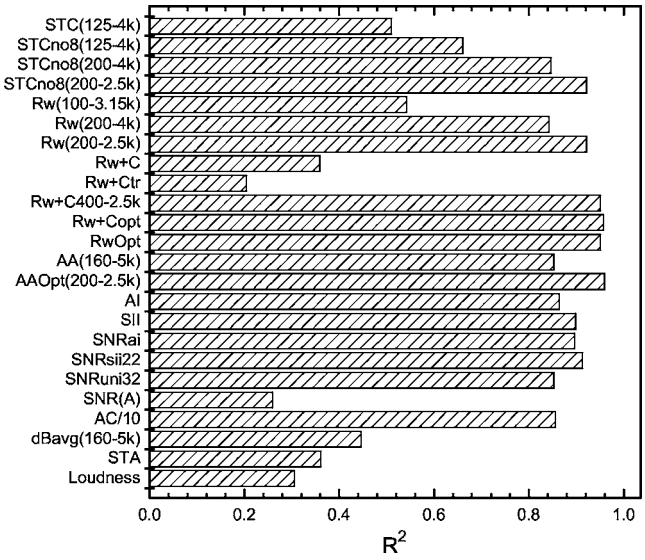


FIG. 15. Summary of R^2 values for Boltzmann equation best-fit regression lines to plots of mean speech intelligibility scores vs the various sound insulation ratings. The two dashed vertical lines indicate the R^2 values corresponding to the significance levels $p<0.05$ ($R^2<0.193$) and $p<0.01$ ($R^2<0.317$).

R^2 values. The highest R^2 values are above 0.9 and indicate very strong relationships between the speech intelligibility scores and the particular sound insulation measures.

The standard measures STC and R_w were not seen to be good predictors of the intelligibility of speech transmitted through the 20 simulated walls (R^2 values were 0.510 and 0.542, respectively). Removing the 8 dB rule from the STC calculation did improve it, but the R^2 values only increased a small amount (0.510–0.661).

Adding the C-type or C_{tr}-type spectrum adaptation terms to the R_w measure did not improve the correlation with the intelligibility of the transmitted speech. However, creating a new spectrum adaptation term for speech did considerably increase R^2 values. (See results for $R_w+C_{400\text{-}2.5k}$ and R_w+C_{opt} in Table III and Fig. 15).

A measure based on energy summation of the information across speech frequencies (160–5k kHz) was less well correlated with intelligibility scores ($R^2=0.446$) than the standard STC and R_w measures.

Speech intelligibility type measures that are based on arithmetic averages or summations of frequency-weighted signal-to-noise ratios were generally better predictors of intelligibility scores than the standard sound insulation measures. This method of combining the information across frequencies is based on the concepts that led to the Articulation Index (AI) measure. A simple uniform-weighted arithmetic average over the frequencies from 200 to 2.5k kHz, AA(200-2.5k), was a very successful predictor of intelligibility scores (R^2 value 0.959). However, a number of other

TABLE III. Summary of R^2 values for Boltzmann equation fits to plots of speech intelligibility scores vs various sound insulation ratings as well as the lowest (F_1) and highest (F_2) included frequencies.

Symbol	F_1, Hz	F_2, Hz	R^2, Boltzmann	Description
STC	125	4k	0.510	Standard STC
STC_{no8}	125	4k	0.661	STC without an 8 dB rule
STC_{no8}	200	4k	0.846	Modified STC without an 8 dB rule
STC_{no8}	200	2.5k	0.922	Modified STC without an 8 dB rule
R_w	100	3.15k	0.542	Standard R_w
R_w	200	3.15k	0.842	Modified R_w
R_w	200	2.5k	0.922	Modified R_w
R_w+C	100	3.15k	0.359	R_w with C-type spectrum correction
R_w+C_{tr}	100	3.15k	0.205	R_w with C_{tr}-type spectrum correction
$R_w+C_{400\text{–}2.5k}$	400	2.5k	0.941	R_w with C_{sps}-type spectrum correction
R_w+C_{opt}	315	3.15k	0.957	R_w using simplified speech shape rating contour
R_{wOpt}	100	3.15k	0.951	R_w using simplified speech shape rating contour
AA(160-5k)	160	5k	0.853	Arithmetic average transmission loss
AA_{opt}(200-2.5k)	200	2.5k	0.959	Arithmetic average transmission loss
AI	200	5k	0.864	Articulation Index
SII	160	8k	0.899	Speech Intelligibility Index
SNR_{ai}	200	5k	0.896	AI-weighted signal-to-noise ratio
SNR_{sii22}	160	8k	0.913	SII-weighted signal-to-noise ratio
SNR_{uni32}	160	5k	0.853	Uniform weighted signal-to-noise ratio
SNR(A)	50	8k	0.259	A-weighted speech–noise level difference
AC/10	200	5k	0.856	Articulation Class
dBavg(160-5k)	160	5k	0.466	Energy average of transmission loss values
STA	100	5k	0.361	A-weighted transmission loss
Loudness	25	12.5k	0.305	Zwicker loudness, Sones

related measures (AI, SII, SNR_{ai}, SNR_{sii22}, SNR_{uni32}, and AC) were all quite strongly related to intelligibility scores (R^2 values 0.853–0.913) in this experiment where speech source levels and ambient noise levels were not varied. Of course, in more realistic situations where the ambient noise levels could also vary, the AC rating would be less successful because it is not a signal-to-noise type measure and does not account for the effects of variations in ambient noise levels. However, AC values are seen to be a good single-number rating of the attenuation of the walls when predicting the intelligibility of transmitted speech. A simple A-weighted speech-noise level difference was only weakly related to the intelligibility scores and hence is not a good indicator of the intelligibility of transmitted speech.

The success of the speech intelligibility related measures is presumably related to the arithmetic average or summation process that assumes that each frequency band contributes independently to the intelligibility of the speech. Their success is also related to the various frequency weightings that minimize the influence of frequency components not important for speech intelligibility. Both factors are thought to be important for accurate predictions of the intelligibility of the transmitted speech. Sound insulation ratings that do not include these factors are found to be less successful predictors of the intelligibility of the transmitted speech.

The question of the most appropriate frequency range was considered and indicated that speech levels in the 1/3-octave bands from 200 Hz to 2.5 or 3.15 kHz are the most important. Several approaches to limiting the sound insulation rating to the speech frequency range were tried. Although all led to increased R^2 values, some approaches were more effective than others. A simple arithmetic average transmission loss over the frequencies from 200 to 2.5 kHz was the best approach with an R^2 of 0.959 for the fit with speech intelligibility scores. This was most successful because it incorporated both the arithmetic averaging of TL values in decibels and limited included sounds to those frequencies most important for speech.

Creating a new spectrum adaptation correction for the R_w measure was also very successful. By using a new spectrum adaptation term, C_{opt}, that approximated a band pass filter for speech frequencies, an R^2 value of 0.957 was achieved. This is an appealing approach to achieving an improved sound insulation rating for speech because it builds on the accepted ISO standard approach and is nearly as accurate as the simple arithmetic average over the speech frequencies.

The success of the arithmetic average measures would appear to support the results of Tachibana *et al.*[15] who found a simple arithmetic average transmission loss to be a good predictor of subjective ratings of sound insulation. However, their subjective ratings were in terms of loudness ratings and not in terms of speech intelligibility.

These new results do agree with recent studies of the speech security of meeting rooms[16,17] that found frequency weighted signal-to-noise ratio type measures to be good predictors of the intelligibility of transmitted speech. In this earlier work, a signal-to-noise ratio with the same frequency weightings as the SII measure (SNR_{sii22}), was found to be the best predictor of intelligibility scores and a uniform weighted signal-to-noise ratio (SNR_{uni32}) was found to be the best predictor of the threshold of audibility of the transmitted speech. The previous study found the SNR_{uni32} measure was a good predictor of both the audibility and the intelligibility of speech. In the current results, the SNR_{sii22} measure is again seen to be a good predictor of intelligibility scores ($R^2 = 0.913$). However, the new results in the current study suggest that an even more restrictive frequency range than the 160–5 kHz range previously used with the SII weighting, such as those in the AA(200–2.5k) and $Rw+C_{opt}$ ratings, can create even more successful predictors of the intelligibility of transmitted speech.

These new results offer some hope that new sound insulation ratings can be developed that would be more accurate indicators of subjective ratings of transmitted sounds. However, further experiments are first required to investigate subjective ratings of other responses and for other types of sounds such as music and various noises.

VII. CONCLUSIONS

The STC and R_w ratings are not very accurate predictors of the intelligibility of speech transmitted through walls. Removing the 8 dB rule from the calculation of STC values only led to modest improvements in the prediction accuracy of intelligibility scores.

Energy summation of transmission loss values or signal-to-noise ratios over various frequency bands did not lead to the most accurate predictors of intelligibility scores. A-weighted speech-noise level differences were particularly inaccurate indicators of intelligibility scores.

The best results were found for measures that included arithmetic averages or summations of decibel values over a range of frequency bands limited to those most important for the intelligibility of speech. This included most measures that are related to or derived from previously accepted speech intelligibility measures.

The most accurate prediction of speech intelligibility scores was found for an Arithmetic Average transmission loss value AA(200-2.5k). A new spectrum adaptation term, when combined with R_w values, was essentially equally successful (R_w+C_{opt}).

Although these results indicate some very successful sound insulation rating measures, they have only been tested for speech sounds and specifically for the intelligibility of speech transmitted through walls. Other types of sounds and other types of subjective ratings must also be considered in future studies to verify the more general applicability of the measures that were successful in this work.

ACKNOWLEDGMENTS

The first author's contribution to this work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) and The Regional Research Universities Program/Biohousing Research Institute.

[1]ASTM E90—92, "Standard Test Method for Laboratory Measurement of Airborne Sound Transmission loss of Building Partitions and Elements" (Am. Soc. for Testing and Materials, Philadelphia, 1992).

[2]ASTM E336-97, "Standard Test Method for Measurement of Airborne Sound Insulation in Buildings" (ASTM International, West Conshohocken, PA, 1997).
[3]ISO 140 "Acoustics—Measurement of sound insulation in buildings and of building elements"—Part 3, "Laboratory measurement of airborne sound insulation of building elements" (2004), Part 4, "Field measurements of airborne sound insulation between rooms" (International Organization for Standardization, Geneva, 1998).
[4]ASTM E413, "Classification for Rating Sound Insulation" (Am. Soc. for Testing and Materials, Philadelphia).
[5]ISO—717-1, "Acoustics—Rating of Sound Insulation in Buildings and of Building Elements—Part 1: Airborne Sound Insulation" (International Organization for Standardization, Geneva, 1996).
[6]D. M. Clark, "Subjective study of the Sound-Transmission Class system for rating building partitions," J. Acoust. Soc. Am. **67**(3), 676–682 (1970).
[7]L. W. Sepmeyer, "Study of the sound transmission class system for rating building partitions-Another view," J. Acoust. Soc. Am. **80**(5), 1404–1407 (1986).
[8]P. Wright and L. Fothergill, "The spectrum adaptation terms in BS EN ISO 717-1:1997," Acoustics Bulletin **23**(6), 5–8 (1998).
[9]C. Grimwood, "Complaints about poor sound insulation between dwellings in England and Wales," Appl. Acoust. **52**(3/4), 211–223 (1997).
[10]F. J. Langdon, I. B. Buller, and W. E. Scholes, "Noise from neighbours and the sound insulation of party walls in houses," J. Sound Vib. **79**(2), 205–228 (1981).
[11]F. J. Langdon, I. B. Buller, and W. W. Scholes, "Noise from neighbours and the sound insulation of party floors and walls in flats," J. Sound Vib. **88**(2), 243–270 (1983).
[12]J. H. Rindel, "The relationship between sound insulation and acoustic quality in dwellings," Proceedings Internoise **98**, 819–824 (1998).
[13]J. S. Bradley, "Grouped subjective ratings of airborne sound insulation," J. Acoust. Soc. Am. **105**(2), 1175 (1999).
[14]J.-P. Vian, W. F. Danner, and J. W. Bauer, "Assessment of significant acoustical parameters for rating sound insulation of party walls," J. Acoust. Soc. Am. **73**(4), 1236–1243 (1983).
[15]H. Tachibana, Y. Hamado, and F. Sato, "Loudness evaluation of sounds transmitted through walls—Basic experiments with artificial sounds," J. Sound Vib. **127**(3), 499–506 (1988).
[16]B. N. Gover and J. S. Bradley, "Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms," J. Acoust. Soc. Am. **116**(6), 3480–3490 (2004).
[17]J. S. Bradley and B. N. Gover, "Designing and assessing the architectural speech security of meeting rooms and offices," IRC Research Report No. RR-187, July (2006).
[18]IEEE "Recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246 (1969).
[19]J. S. Bradley, "Acoustical measurements in some Canadian homes," Can. Acoust. **14**(4), 19–21, 24 (1986).
[20]American National Standards Institute (ANSI) Standard S12.60. "Acoustical performance criteria, design requirements, and guidelines for schools" (American National Standards Institute, New York, 2002).
[21]ISO 7029-1984 (E), "Acoustics—Threshold of hearing by air conduction as a function of age and sex for otologically normal persons" (International Organization for Standardization, Geneva, 1984).
[22]ANSI S3.5-1969, American National Standard, "Methods for the calculation of the articulation index" (Acoustical Society of America, Melville, 1969).
[23]N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**(1), 90–119 (1947).
[24]ANSI S3.5-1997, American National Standard, "Methods for calculation of the speech intelligibility index" (Acoustical Society of America, Melville, 1997).
[25]ASTM E1110-01, "Standard classification for determination of articulation class," (ASTM International, West Conshohocken, PA).
[26]J. S. Bradley, "Subjective rating of party walls," Can. Acoust. **11**(4), 37–45 (1983).
[27]E. Zwicker, H. Fastl, and C. Dallmayr, "BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B," Acustica **55**, 63–67 (1984).

Sound field separation technique based on equivalent source method and its application in nearfield acoustic holography

Chuan-Xing Bi,[a] Xin-Zhao Chen, and Jian Chen
Institute of Sound and Vibration Research, Hefei University of Technology, Hefei 230009, People's Republic of China

(Received 12 June 2007; revised 11 December 2007; accepted 4 January 2008)

A technique for separating sound fields using two closely spaced parallel measurement surfaces and based on equivalent source method is proposed. The method can separate wave components crossing two measurement surfaces in opposite directions, which makes nearfield acoustic holography (NAH) applications in a field where there exist sources on the two sides of the hologram surface, in a reverberant field or in a scattered field, possible. The method is flexible in applications, simple in computation, and very easy to implement. The measurement surfaces can be arbitrarily shaped, and they are not restricted to be regular as in the traditional field separation technique. And, because the method performs field separation calculations directly in the spatial domain—not in the wave number domain—it avoids the errors and limitations (the window effects, etc.) associated with the traditional field separation technique based on the spatial Fourier transform method. In the paper, a theoretical description is first given, and the performance of the proposed field separation technique and its application in NAH are then evaluated through experiments.
 [DOI: 10.1121/1.2837489]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.20.Rz [EGW] Pages: 1472–1478

I. INTRODUCTION

The attempt to separate sound fields by two-surface method first appeared in 1956.[1] In Pachner's paper, he described a technique for determining the traveling and standing components of an arbitrary scalar wave field from the measured instantaneous values of the field on the surface of two spheres surrounding the emitter. The method was directed toward determining the directivity patterns of transducers. In 1980, Weinreich and Anold[2] proposed a technique for separating outgoing waves from the source and incoming waves from reflections by making the measurement on two closely spaced parallel surfaces. On the basis of nearfield acoustic holography (NAH) technique by Williams and Maynard[3] and the method by Weinreich and Anold,[2] Frisk *et al.*[4] derived a field separation technique based on the spatial Fourier transform (SFT) method for measuring the reflection coefficient in the field of underwater acoustics. The method was studied and further developed by several researchers. Tamura[5] extended the method to measure the plane-wave reflection coefficient at oblique angles of incidence, and the error due to the finite measurement area was discussed in detail. Hu and Bolton[6] and Tamura[7] validated the method by experimental studies. Cheng *et al.*[8,9] extended the method to separate the scattered field of a complex shape and developed a technique to separate the incident and scattered fields in cylindrical coordinates. Recently, Yu *et al.*[10] successfully used the field separation technique to remove noise sources for NAH applications.

The major advantage of the field separation technique based on the SFT method is that the usage of the SFT method makes it easy and simple for applications. However, the method has its limits. The measurement surface must be regular, such as planar, cylindrical or spherical. The accuracy is limited because of the errors and limitations (the window effects etc.) associated with the SFT method. Recently, statistically optimized nearfield acoustic holography using an array of pressure-velocity probes[11] was used to avoid the errors and limitations associated with the SFT method, and boundary element method using a double layer[12] was proposed to overcome the restriction on the shapes of measurement surfaces.

The equivalent source method (ESM),[13,14] also named wave superposition method,[15,16] source simulation method,[17] etc. was proposed by Koopmann for solving the acoustic radiation problem of closed sources. It has been proved to be an effective method for acoustic radiation, scattering and NAH problems by several researchers.[13–19] In the method, the sound field is approximated by a number of equivalent sources placed inside a vibrating structure, and the strengths of the equivalent sources are determined by matching the boundary conditions on the source or on the hologram surface. The ESM may also be applied to flat or curved surface. Recently, Sarkissian successfully used the method to extend the measurement surface tangentially outward[20] and realize the patch NAH which was measured on a curved surface near a radiating structure.[21].

The purpose of this paper is to use the ESM to realize the field separation. In the method, measurements are performed on two closely spaced parallel surfaces with arbitrary shapes, relationships between the pressures on the two measurement surfaces and the sources on the two sides of the measurement surfaces are established by ESM, and field separation calculations are performed by solving the combined equations directly in the spatial domain. In the paper, a

[a]Author to whom correspondence should be addressed. Electronic mail: cxbi@hfut.edu.cn

theoretical description of the ability of the method to separate wave components crossing two closely spaced parallel measurement surfaces in opposite directions is first given. The performance of the proposed method and its application in NAH are then evaluated through an experiment.

II. EQUIVALENT SOURCE METHOD

The basic idea of the ESM is to replace the actual acoustic field by the superposition of fields produced by a number of equivalent sources placed inside a vibrating structure.[13–17] Here, the source strengths of the equivalent sources are determined from the boundary conditions of the vibrating structure, i.e., surface pressure or normal surface velocity. Furthermore, the field on and near a measurement surface with flat or curved shape in an acoustic field can also be approximated by the superposition of fields produced by a number of equivalent sources.[20] Here, the equivalent sources are placed on a fictitious surface near the measurement surface, and these source strengths are determined by applying boundary conditions on the measurement surface.

Figure 1 contains the geometry, showing a source radiating a harmonic field, a measurement surface S and a fictitious surface S^* on which the equivalent sources are distributed. Here, the measurement surface S and the fictitious surface S^* can be a closed, flat or curved surface. The acoustic quantities at a field point r which is located on or near the measurement surface S is approximated by the superposition of equivalent sources placed on the fictitious surface S^*. Suppose there are M measurement points on the measurement surface S and N equivalent sources on the fictitious surface S^*. The pressure and particle velocity produced by the ith equivalent source at the field point r is $p_i^*(r)$ and $v_i^*(r)$, respectively. Then, the actual pressure and particle velocity at the field point r can be approximated as

$$p(r) = \sum_{i=1}^{N} w_i p_i^*(r), \tag{1}$$

$$v(r) = \sum_{i=1}^{N} w_i v_i^*(r), \tag{2}$$

where w_i is the source strength of the ith equivalent source. By Eq. (1), the pressure at the M measurement points can be expressed by

$$p_S(r_j) = \sum_{i=1}^{N} w_i p_{\mathrm{Si}}^*(r_j), \quad j = 1,2,\ldots,M. \tag{3}$$

Equation (3) can be rewritten in matrix form as

$$\mathbf{p}_S = \mathbf{p}_S^* \mathbf{W}, \tag{4}$$

where

$$\mathbf{p}_S^* = \begin{bmatrix} p_{S1}^*(r_1) & p_{S2}^*(r_1) & \cdots & p_{SN}^*(r_1) \\ p_{S1}^*(r_2) & p_{S2}^*(r_2) & \cdots & p_{SN}^*(r_2) \\ \vdots & \vdots & \vdots & \vdots \\ p_{S1}^*(r_M) & p_{S2}^*(r_M) & \cdots & p_{SN}^*(r_M) \end{bmatrix}, \tag{5}$$

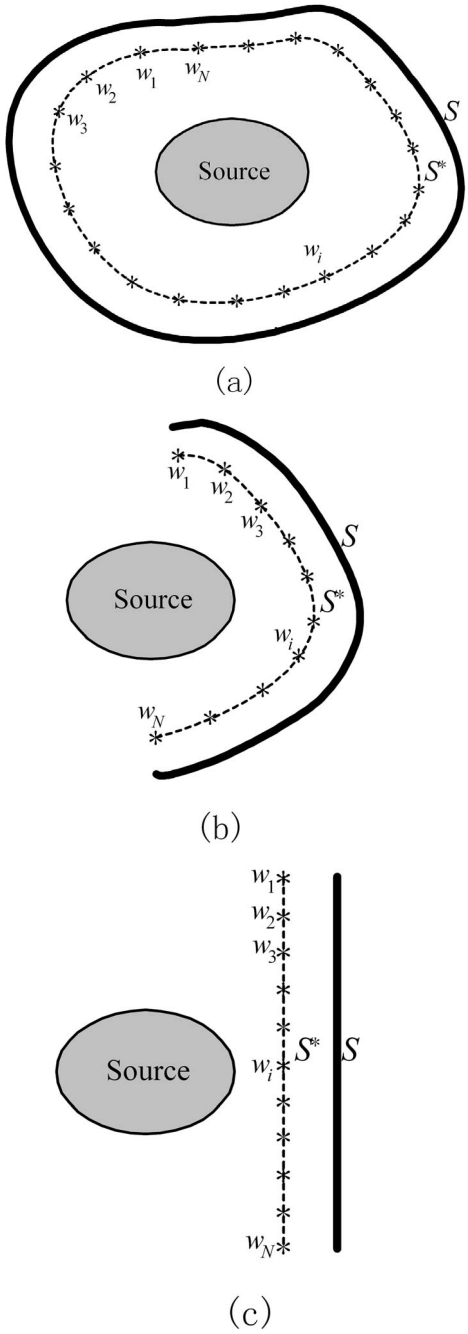


FIG. 1. Geometry of a measurement surface S and a fictitious surface S^* on which the equivalent sources are distributed: (a) Closed surface; (b) curved surface; (c) flat surface.

$$\mathbf{W} = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix}^T, \tag{6}$$

where $\mathbf{p}_S$ is the pressure column vector of the M measurement points, $\mathbf{W}$ is source strength column vector of the N equivalent sources, $\mathbf{p}_S^*$ is the transfer matrix from the N equivalent sources to the M measurement points, and the superscript "T" denotes the matrix transpose.

For Eq. (4), if $M \geqslant N$, the source strength vector $\mathbf{W}$ can be uniquely determined by utilizing the generalized inverse transform as

$$\mathbf{W} = (\mathbf{p}_S^*)^+ \mathbf{p}_S, \tag{7}$$

where the superscript "+" denotes the pseudo-inverse. Because the inverse process in Eq. (7) is generally very sensitive to measurement errors, when the ESM is used to realize the reconstruction of sound field, the regularization should be applied.[20,21] In the following separation technique, because no back propagation is needed, the regularization is also not needed.

By Eqs. (1), (2), and (7), the pressure and particle velocity at any field point r can be predicted, respectively.

III. SOUND FIELD SEPARATION TECHNIQUE BASED ON ESM

In order to identify a target source in a field where there exist other sources, reflecting interfaces or scattering objects, we should separate the target sound from the disturbed field. However, pressure is a scalar quantity, and it is difficult to separate the two components directly. The method of the present paper is to measure the pressures on two closely spaced parallel measurement surfaces and then realize the separation by ESM.

According to Sec. I, the field on or near a measurement surface can be approximated by the superposition of fields produced by a number of equivalent sources placed on a surface near the measurement surface. If there exist sources on the two sides of the measurement surface, the pressure on the measurement surface is produced by the sources on the two sides. Figure 2 depicts the geometry of measurement surfaces S_1 and S_2 and fictitious surfaces S_1^* and S_2^* on which the equivalent sources are distributed. Here, source 1 is the target source, and source 2 represents the disturbing source. The distances σ_1 and σ_2 are equal to 1/3 to 1/10 wavelength, the distance δh is less than a half wavelength, the equavalent sources are located on the fictitious surfaces and on the normal lines of measurement points on the corresponding measurement surfaces, and the numbers of equavalent sources on the fictitious surfaces S_1^* and S_2^* are equal to the numbers of measurement points on the measurement surfaces S_1 and S_2, respectively. The pressure on the surface S_1 is

$$\mathbf{p}_{S_1} = \mathbf{p}_{S_1}^1 + \mathbf{p}_{S_1}^2, \tag{8}$$

where $\mathbf{p}_{S_1}^1$ is the pressure produced by source 1, and $\mathbf{p}_{S_1}^2$ is the pressure produced by source 2. Similarly, the pressure on the surface S_2 is

$$\mathbf{p}_{S_2} = \mathbf{p}_{S_2}^1 + \mathbf{p}_{S_2}^2, \tag{9}$$

where $\mathbf{p}_{S_2}^1$ is the pressure produced by source 1, and $\mathbf{p}_{S_2}^2$ is the pressure produced by source 2.

According to Sec. I, the pressure $\mathbf{p}_{S_1}^1$ produced by source 1 on the surface S_1 can be approximated by a number of equivalent sources placed on the fictitious surface S_1^* which is near the surface S_1 and on the side of source 1. By Eq. (7), the source strength vector $\mathbf{W}_1$ of the equivalent sources on S_1^* can be determined from $\mathbf{p}_{S_1}^1$ as

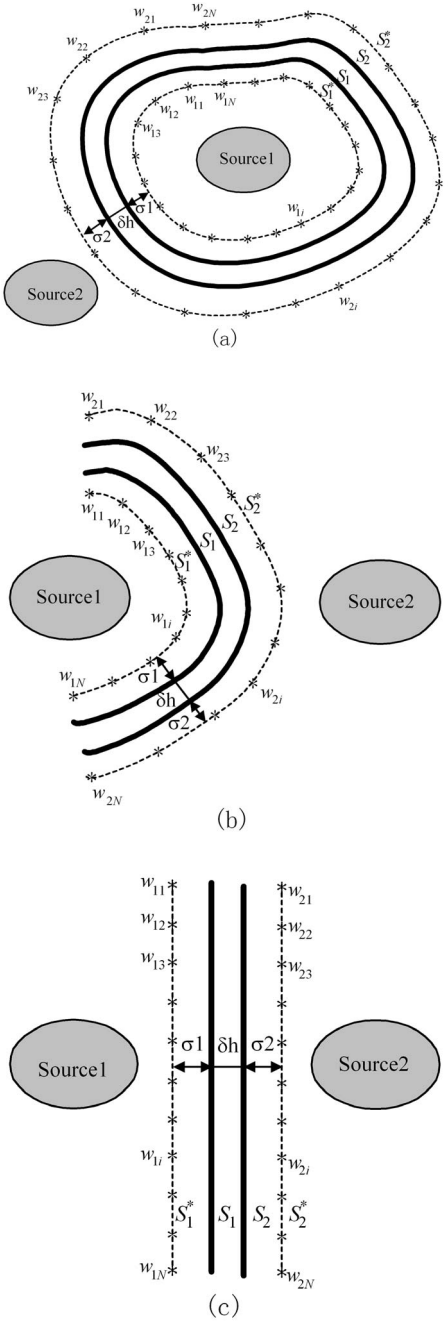


FIG. 2. Geometry of measurement surfaces S_1 and S_2 and fictitious surfaces S_1^* and S_2^* on which the equivalent sources are distributed. (a) Closed surfaces; (b) curved surfaces; (c) flat surfaces.

$$\mathbf{W}_1 = [(\mathbf{p}_{S_1}^1)^*]^+ \mathbf{p}_{S_1}^1, \tag{10}$$

where $(\mathbf{p}_{S_1}^1)^*$ is the transfer matrix from the equivalent sources on S_1^* to the surface S_1.

After obtaining the source strength vector $\mathbf{W}_1$, the pressure $\mathbf{p}_{S_2}^1$ produced by source 1 on the surface S_2 can be expressed as

$$\mathbf{p}_{S_2}^1 = (\mathbf{p}_{S_2}^1)^* \mathbf{W}_1 = (\mathbf{p}_{S_2}^1)^* [(\mathbf{p}_{S_1}^1)^*]^+ \mathbf{p}_{S_1}^1, \tag{11}$$

where $(\mathbf{p}_{S_2}^1)^*$ is the transfer matrix from the equivalent sources on S_1^* to the surface S_2.

In the same way, the pressures $\mathbf{p}_{S_2}^2$ produced by source 2 on the surface S_2 can be approximated by a number of equivalent sources placed on the fictitious surface S_2^* which is near the surface S_2 and on the side of source 2. The source strength vector $\mathbf{W}_2$ of the equivalent sources on S_2^* can be determined from $\mathbf{p}_{S_2}^2$ as

$$\mathbf{W}_2 = [(\mathbf{p}_{S_2}^2)^*]^+ \mathbf{p}_{S_2}^2, \tag{12}$$

where $(\mathbf{p}_{S_2}^2)^*$ is the transfer matrix from the equivalent sources on S_2^* to the surface S_2. Then, the pressure $\mathbf{p}_{S_1}^2$ produced by source 2 on the surface S_1 can be expressed as

$$\mathbf{p}_{S_1}^2 = (\mathbf{p}_{S_1}^2)^* \mathbf{W}_2 = (\mathbf{p}_{S_1}^2)^* [(\mathbf{p}_{S_1}^2)^*]^+ \mathbf{p}_{S_2}^2, \tag{13}$$

where $(\mathbf{p}_{S_1}^2)^*$ is the transfer matrix from the equivalent sources on S_2^* to the surface S_1. Substituting Eqs. (11) and (13) into Eqs. (9) and (8), respectively, yields

$$\mathbf{p}_{S_1} = \mathbf{p}_{S_1}^1 + (\mathbf{p}_{S_1}^2)^* [(\mathbf{p}_{S_2}^2)^*]^+ \mathbf{p}_{S_2}^2, \tag{14}$$

$$\mathbf{p}_{S_2} = (\mathbf{p}_{S_2}^1)^* [(\mathbf{p}_{S_1}^1)^*]^+ \mathbf{p}_{S_1}^1 + \mathbf{p}_{S_2}^2. \tag{15}$$

Combining Eqs. (14) and (15), we can obtain

$$\mathbf{p}_{S_1}^1 = (\mathbf{I} - \mathbf{G}_1 \mathbf{G}_2)^+ (\mathbf{p}_{S_1} - \mathbf{G}_1 \mathbf{p}_{S_2}), \tag{16}$$

$$\mathbf{p}_{S_2}^2 = (\mathbf{I} - \mathbf{G}_2 \mathbf{G}_1)^+ (\mathbf{p}_{S_2} - \mathbf{G}_2 \mathbf{p}_{S_1}), \tag{17}$$

where

$$\mathbf{G}_1 = (\mathbf{p}_{S_1}^2)^* [(\mathbf{p}_{S_2}^2)^*]^+, \tag{18}$$

$$\mathbf{G}_2 = (\mathbf{p}_{S_2}^1)^* [(\mathbf{p}_{S_1}^1)^*]^+. \tag{19}$$

Then substituting Eqs. (16) and (17) into Eqs. (8) and (9), respectively, we can obtain $\mathbf{p}_{S_2}^1$ and $\mathbf{p}_{S_1}^2$.

It is obvious that, by the above-mentioned method, the pressures produced by the sources on the two sides of the measurement surfaces can be separated. Then, the pressures produced by the target source on the measurement surfaces can be used to reconstruct the target source through NAH. The method to realize the NAH can be determined by the shape of the measurement surface. For a regular measurement surface, the NAH based on the SFT method can be used. For an arbitrarily shaped measurement surface, the NAH based on the BEM, etc., can be used.

IV. EXPERIMENT

To examine the validity of the proposed field separation technique and its application in NAH, an experiment has been carried out in a semianechoic chamber, where the background noise is lower than 16 dB(A). The experimental photograph is shown in Fig. 3.

In the experiment, two speakers with the same size of 0.14 m × 0.15 m × 0.32 m were chosen as target and disturbing sources, respectively. The two measurement surfaces were planar surfaces with the same size of 0.7 m(x) × 1.0 m(z), the grid space was 0.05 m along the horizontal (x) and vertical (z) directions. The centers of the speaker cones of the target and disturbing sources were located at

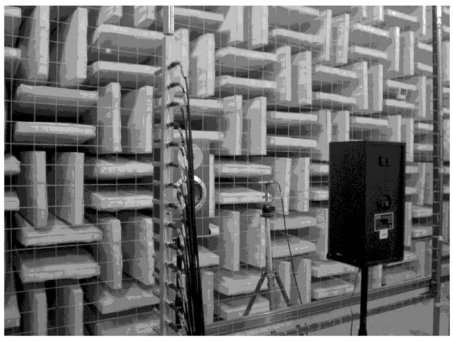


FIG. 3. Experimental photograph.

(0.0 m, −0.125 m, and −0.15 m) and (0.0 m, 0.225 m, and 0.15 m), respectively. Because the cleaned pressure on the measurement surface S_1 would be used to reconstruct the acoustic quantities of the target source, the measurement surface S_1 should be set to be close to the target source to include enough evanescent waves as needed in NAH. Here, the distance between the target source and the measurement surface S_1 was 0.125 m, and the distance between the disturbing source and the measurement surface S_2 was 0.15 m. The gap between the two measurement surfaces was 0.075 m. The distances σ_1 and σ_2 were all $1/8\lambda$ (λ is the wavelength). The geometric description of experimental sources and measurement surfaces is depicted in Fig. 4.

The pressures on the two measurement surfaces were measured by one fixed reference microphone and a scanning microphone array composed of 15 microphones. The amplitudes of pressures were obtained by the auto spectra of the sampled signals on the measurement surfaces. The phases of pressures were obtained by the cross spectra between the sampled signals on the measurement surfaces and the synchronous reference signal. The signal was generated by a signal generator and amplified by a power amplifier, and the amplified signal was connected to the two speakers to generate sounds, respectively. Here, the in-phase excitation was used. The sampling frequency was 5 kHz, the sampling block size was 2048, and ten blocks of data were measured at each measurement point.

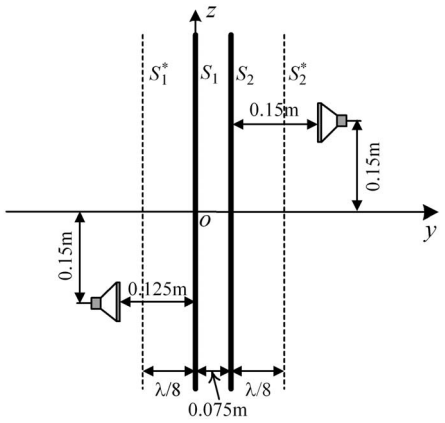


FIG. 4. Geometric description of experimental sources and measurement surfaces.

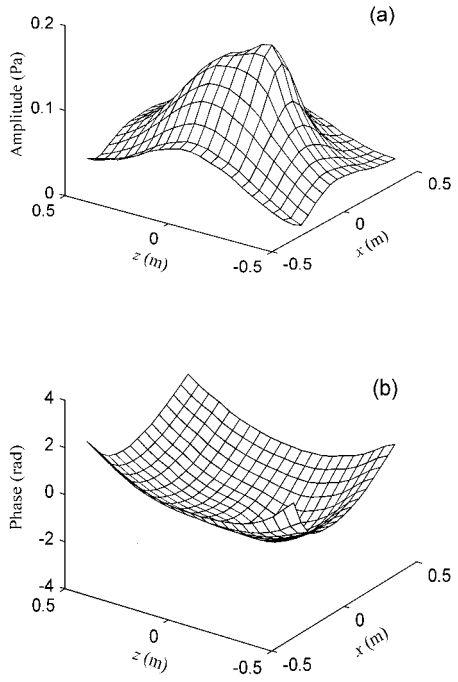


FIG. 5. Disturbed pressure measured on the measurement surface S_1 at 400 Hz. (a) The amplitude; (b) the phase.

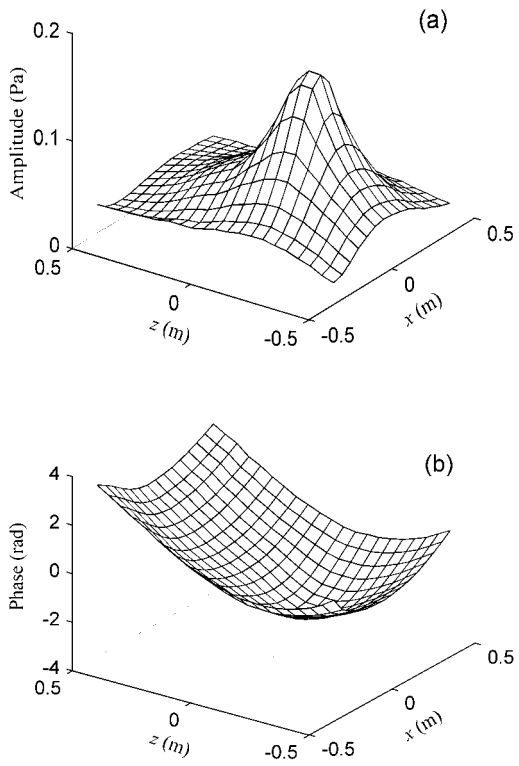


FIG. 6. Pressure measured on the measurement surface S_1 at 400 Hz after physically removing the disturbing source. (a) The amplitude; (b) the phase.

A. Effect of separation

In order to quantify the accuracy of the field separation technique, the separation error in percentage is defined by

$$\eta = \sqrt{\sum_{i=1}^{M} (|p_i| - |\bar{p}_i|)^2} \Bigg/ \sqrt{\sum_{i=1}^{M} |\bar{p}_i|^2} \times 100(\%), \quad (20)$$

where M is the number of measurement points, and p_i, $\bar{p}_i$ are the cleaned pressure and its measured value at the ith measurement point, respectively.

At first, an experiment at a single frequency of 400 Hz was investigated. Figures 5–7 depict the distrurbed pressure measured on the measurement surface S_1, the pressure measured on the measurement surface S_1 after physically removing the disturbing source, and the cleaned pressure on the measurement surface S_1 in the proposed field separation technique.

Figures 5(a) and 5(b) show that the amplitude and phase of the disturbed pressure measured on the measurement surface S_1 are influenced seriously by the disturbing source, and it cannot provide the correct information of the target source on the measurement surface as depicted in Figs. 6(a) and 6(b). There exists large difference between the disturbed pressure and the pressure produced by the target source.

Figures 7(a) and 7(b) show the amplitude and phase of the cleaned pressure on the measurement surface S_1in the proposed field separation technique, respectively, which show good agreement with those of the pressure measured on the measurement surface S_1 after physically removing the disturbing source as depicted in Figs. 6(a) and 6(b). Here, the separation error is only 6.81%. It is obvious that, by using the proposed field separation technique, the pressure produced by the target source on the measurement surface S_1 can be separated out correctly.

In order to further testify the performance of the proposed field separation technique, the separation errors at different frequencies (from 100 to 1600 Hz) were studied by experiments. Figure 8 shows the relationship between the separation error and the frequency. Here, the separation error is all below 15%. It is obvious that the error is acceptable. It further proved the validity of the proposed separation method.

B. Effect of reconstruction

When the pressure produced by the target source was separated out, it could be used to reconstruct the source through NAH. In the following, the disturbed and cleaned pressures on the measurement surface S_1 at 400 Hz, respectively, were used for reconstructing the active sound intensity on the source surface, and the validity of the application of the field separation technique in NAH was demonstrated by comparisons.

Here, the source surface was the same as the measurement surfaces with a 0.125 m distance away from the measurement surface S_1 and near the target source. Because the measurement surfaces were planar, the NAH based on the SFT method was used to realize the reconstruction on the source surface. In the reconstruction process, the measure-

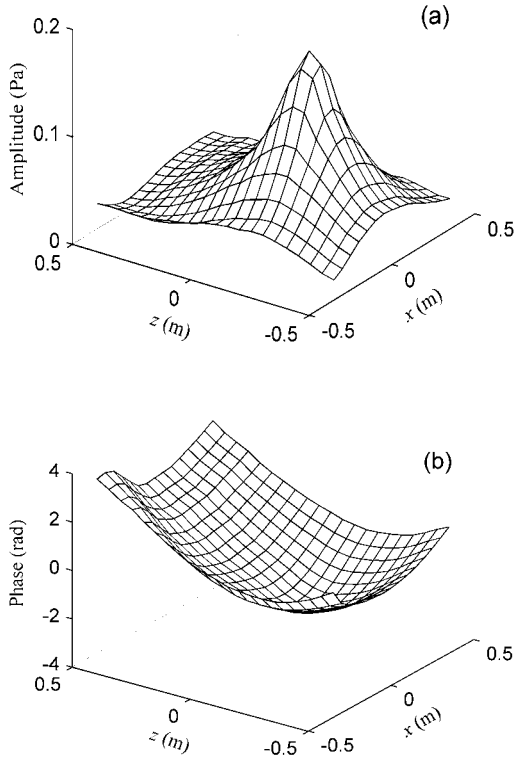


FIG. 7. Cleaned pressure on the measurement surface S_1 at 400 Hz in the proposed field separation technique. (a) The amplitude; (b) the phase.

ment array was padded with zeros to 41×41, and the filter function was used for reducing the error. The filter function is[22]

$$W(k_x,k_y)=\begin{cases}1-1/2\exp[(k_r/k_c-1)/\alpha], & k_r\leqslant k_c,\\ 1/2\exp[(1-k_r/k_c)/\alpha], & k_r>k_c,\end{cases} \tag{21}$$

where $k_r^2=k_x^2+k_y^2$, and k_c and α are adjustable parameters of the filter. In the experiment, α was equal to 0.2 and k_c was 37.6991, which were determined from Ref. 22.

Figures 9(a)–9(c) show the active sound intensities on the source surface reconstructed using the disturbed pressure measured on S_1, the cleaned pressure on S_1, and the pressure

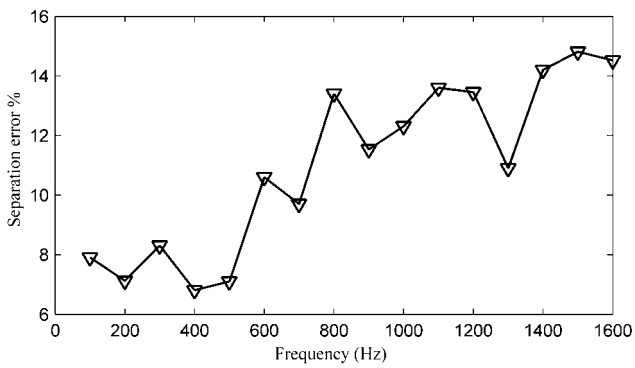


FIG. 8. Relationship between the separation error and the frequency.

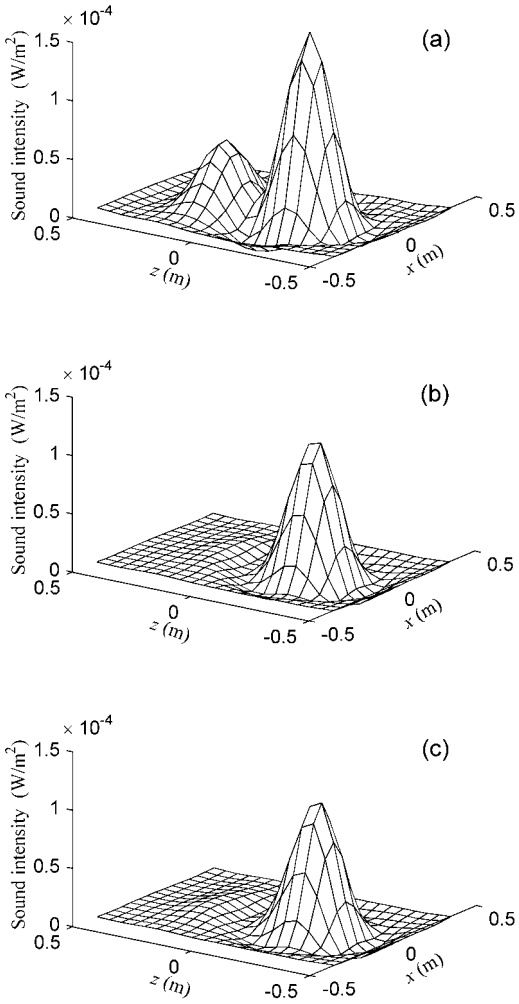


FIG. 9. Comparison of the active sound intensities on the source surface reconstructed using different pressures on the measurement surface S_1 at 400 Hz. (a) The result reconstructed using the disturbed pressure; (b) the result reconstructed using the cleaned pressure; (c) the result reconstructed using the pressure measured after physically removing the disturbing source.

measured on S_1 after physically removing the disturbing source, respectively. It can be easily seen that the reconstructed result in Fig. 9(b) removes the influence of the disturbing source and provides the source information correctly. However, Fig. 9(a) shows two peaks in the reconstructed result. One is generated by the target source, and the other is just generated by the disturbing source. It is obvious that the wrong reconstructed results are obtained, which could mislead noise control engineers in practical applications. By comparing Figs. 9(b) and 9(c), it can be found that the result reconstructed using the cleaned pressure is almost the same as that after physically removing the disturbing source. Here, the reconstruction error is only 7.21%. Therefore, the influence of the disturbing source on the reconstructed results has been removed by using the proposed field separation technique. The experimental results prove the validity of the application of the proposed method in NAH for reconstructing the target source in a field where there exist sources on the two sides of the hologram surface, which makes NAH applications in practical industrial environments possible.

V. CONCLUSIONS

In the conventional NAH, all sources are assumed on one side of the hologram surface. When there exist sources on the two sides of the hologram surface, the reconstructed results may be influenced by the noise coming from the "wrong" side. In order to remove the influence of the disturbing source, a technique for separating sound fields using two closely spaced parallel measurement surfaces and based on the ESM is proposed in the paper. In the method, measurement is performed on two closely spaced parallel surfaces, and the sound fields generated by the sources from the two sides of the measurement surfaces are separated by establishing the relationships between the two measurement surfaces and the equivalent sources placed on the two fictitious surfaces. The cleaned pressures can then be used to reconstruct the field produced by the target source. By using the proposed field separation technique, NAH can be applied in practical industrial environments.

The validity of the proposed field separation technique and its application in NAH are demonstrated by experiments with two speakers located on the two sides of the measurement surfaces. The results show that the pressure of the target source on the measurement surface S_1 can be separated out from the disturbed pressure correctly by the proposed field separation technique, and the cleaned pressure can be used to reconstruct the target source through NAH effectively.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China Grant Nos. 10504006 and 50675056.

[1]J. Pachner, "Investigation of scalar wave fields by means of instantaneous directivity patterns," J. Acoust. Soc. Am. **28**(1), 90–92 (1956).
[2]G. Weinreich and E. B. Arnold, "Method for measuring acoustic radiation fields," J. Acoust. Soc. Am. **68**(2), 404–411 (1980).
[3]E. G. Williams and J. D. Maynard, "Holographic imaging without the wavelength resolution limit," Phys. Rev. Lett. **45**, 554–557 (1980).
[4]G. V. Frisk, A. V. Oppenheim, and D. R. Martinez, "A technique for measuring the plane-wave reflection coefficient of the ocean bottom," J. Acoust. Soc. Am. **68**(2), 602–612 (1980).
[5]M. Tamura, "Spatial Fourier-transform method for measuring reflection coefficients at oblique incidence. I. Theory and numerical examples," J. Acoust. Soc. Am. **88**(5), 2259–2264 (1990).
[6]Z. Hu and J. S. Bolton, "The measurement of plane-wave reflection coefficients by using two-dimensional spatial transforms," J. Acoust. Soc. Am. **88**(S1), S173 (1990).
[7]M. Tamura, J. F. Allard, and D. Lafarge, "Spatial Fourier-transform method for measuring reflection coefficients at oblique incidence. II. Experimental results," J. Acoust. Soc. Am. **97**(4), 2255–2262 (1995).
[8]M. T. Cheng, J. A. Mann III, and A. Pate, "Wave-number domain separation of the incident and scattered sound field in Cartesian and cylindrical coordinates," J. Acoust. Soc. Am. **97**(4), 2293–2303 (1995).
[9]M. T. Cheng, J. A. Mann III, and A. Pate, "Sensitivity of the wave-number domain field separation methods for scattering," J. Acoust. Soc. Am. **99**(6), 3550–3557 (1996).
[10]F. Yu, J. Chen, W. B. Li, and X. Z. Chen, "Sound field separation technique and its applications in near-field acoustic holography," Acta Phys. Sin. **54**(2), 789–797 (2005).
[11]F. Jacobsen and V. Jaud, "Statistically optimized near field acoustic holography using an array of pressure-velocity probes," J. Acoust. Soc. Am. **121**(3), 1550–1558 (2007).
[12]C. Langrenne, M. Melon, and A. Garcia, "Boundary element method for the acoustic characterization of a machine in bounded noisy environment," J. Acoust. Soc. Am. **121**(5), 2750–2757 (2007).
[13]M. E. Johnson, S. J. Elliott, K. H. Baek, and J. Garcia-Bonito, "An equivalent source technique for calculating the sound field inside an enclosure containing scattering objects," J. Acoust. Soc. Am. **104**(3), 1221–1231 (1998).
[14]Y. I. Bobrovnitskii and T. M. Tomilina, "General properties and fundamental errors of the method of equivalent sources," Acoust. Phys. **41**, 649–660 (1995).
[15]G. H. Koopmann, L. Song, and J. Fahnline, "A method for computing acoustic fields based on the principle of wave superposition," J. Acoust. Soc. Am. **86**(5), 2433–2438 (1989).
[16]L. Song, G. H. Koopmann, and J. Fahnline, "Numerical errors associated with the method of superposition for computing acoustic fields," J. Acoust. Soc. Am. **89**(6), 2625–2633 (1991).
[17]M. Ochmann, "The source simulation technique for acoustic radiation problems," Acustica **81**(6), 512–527 (1995).
[18]A. T. Abawi and M. B. Porter, "Calculation of scattering from underwater targets using the equivalent source technique," J. Acoust. Soc. Am. **117**(4), 2612 (2005) (Abstract).
[19]C. X. Bi, X. Z. Chen, J. Chen, and R. Zhou, "Nearfield acoustic holography based on the equivalent source method," Sci. China, Ser. E: Technol. Sci. **48**(3), 338–353 (2005).
[20]A. Sarkissian, "Extension of measurement surface in near-field acoustic holography," J. Acoust. Soc. Am. **115**(4), 1593–1596 (2004).
[21]A. Sarkissian, "Method of superposition applied to patch near-field acoustic holography," J. Acoust. Soc. Am. **118**(2), 671–678 (2005).
[22]W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH) II: Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**(5), 1307–1322 (1987).

Two-tone suppression of stimulus frequency otoacoustic emissions[a)]

Douglas H. Keefe,[b)] John C. Ellison, Denis F. Fitzpatrick, and Michael P. Gorga
Boys Town National Research Hospital, 555 North 30th Street, Omaha Nebraska 68131

(Received 16 July 2007; revised 4 December 2007; accepted 4 December 2007)

Stimulus frequency otoacoustic emissions (SFOAEs) measured using a suppressor tone in human ears are analogous to two-tone suppression responses measured mechanically and neurally in mammalian cochleae. SFOAE suppression was measured in 24 normal-hearing adults at octave frequencies (f_p=0.5–8.0 kHz) over a 40 dB range of probe levels (L_p). Suppressor frequencies (f_s) ranged from −2.0 to 0.7 octaves *re*: f_p, and suppressor levels ranged from just detectable suppression to full suppression. The lowest suppression thresholds occurred for "best" f_s slightly higher than f_p. SFOAE growth of suppression (GOS) had slopes close to one at frequencies much lower than best f_s, and shallow slopes near best f_s, which indicated compressive growth close to 0.3 dB/dB. Suppression tuning curves constructed from GOS functions were well defined at 1, 2, and 4 kHz, but less so at 0.5 and 8.0 kHz. Tuning was sharper at lower L_p with an equivalent rectangular bandwidth similar to that reported behaviorally for simultaneous masking. The tip-to-tail difference assessed cochlear gain, increasing with decreasing L_p and increasing f_p at the lowest L_p from 32 to 45 dB for f_p from 1 to 4 kHz. SFOAE suppression provides a noninvasive measure of the saturating nonlinearities associated with cochlear amplification on the basilar membrane.
 [DOI: 10.1121/1.2828209]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Bt [BLM] Pages: 1479–1494

I. INTRODUCTION

Evoked otoacoustic emissions (OAEs) provide a noninvasive tool to test theories of auditory processing at the mechanical level within the cochlea. The present study was designed to measure two-tone suppression of stimulus frequency (SF) OAE responses in normal-functioning human ears over broad ranges of probe frequency and level, and suppressor frequency and level. Motivated by the many similarities in auditory processing at the cochlear, neural and behavioral levels, the approach was to describe suppression of SFOAEs and to compare these data to previous reports of effects of suppression and simultaneous masking at other levels of auditory processing. This introduction summarizes concepts from cochlear mechanics and, to a lesser extent, psychoacoustics that are directly relevant to our analyses of SFOAEs. Such concepts include cochlear nonlinearity, two-tone suppression, rate of growth of suppression, tuning curves and their parameterization using quality factor (or Q) and tip-to-tail difference.

Many psychoacoustic models are based on functional models of basilar-membrane (BM) mechanics, in which a frequency-selective nonlinear filter is an essential component (Moore and Oxenham, 1998). As reviewed in Robles and Ruggero (2001), BM mechanics provide sharp tuning at low stimulus levels similar to that of neural tuning, and less sharp tuning at higher levels. Sharp tuning was explained in terms of a cochlear-amplifier mechanism (Gold, 1948; Davis, 1983), and active cochlear models having a source of internal energy show sharper tuning than passive models (Neely and Kim, 1983). The endocochlear potential serves as a source of internal energy used by outer hair cells (OHCs) to amplify low-level vibrations on the BM. OHC electromotility (Brownell *et al.*, 1985) contributes to sharp tuning, providing a cycle-by-cycle cochlear feedback amplifier that increases the motion of the BM (Mountain *et al.*, 1983).

A neural tuning curve shows the stimulus threshold sound pressure level (SPL) at the tympanic membrane producing a criterion increment above the spontaneous firing rate of an auditory-nerve fiber as a function of frequency (Liberman, 1982). Its tip region is the narrow frequency range where the threshold is lowest, and its tail region at lower frequencies is where the thresholds are elevated, but relatively constant. Its tip-to-tail difference is the threshold SPL difference between the tip and tail regions, and its quality factor, or Q, assesses frequency selectivity.

Tuning curves measured from auditory fibers, depolarization of inner hair cell (IHC) potentials, and BM motion are similar, which suggests that the IHC and auditory-nerve fibers each function as a detector of the mechanical motion of the BM. OHCs function *in vivo* in a feedback loop as bi-directional transducers from mechanical to electrical energy, and from electrical to mechanical energy to amplify the low-level motions of the BM (Patuzzi, 1996). Cochlear dysfunction that results in a reduction in endocochlear potential would reduce OHC motion, reduce cochlear amplification of low-level BM motion, and thereby reduce the depolarization of IHCs, which would result in reduced auditory-nerve firing rates. Mild and moderate sensorineural hearing losses are

[a)] A preliminary report of this research was presented at the 26th Midwinter Research Meeting of the Association for Research in Otolaryngology (February 2003), page 681(A).
[b)] Author to whom correspondence should be addressed. Electronic mail: keefe@boystown.org.

often due to OHC dysfunction, although more severe sensorineural hearing losses are expected to involve IHC or neural dysfunction. Animal experiments based on administration of furosemide confirm this interpretation. Furosemide reduces the endocochlear potential and elevates the thresholds of auditory-nerve fibers by large amounts in the tip region, where cochlear amplification is predominant, and by smaller amounts in the tail region, where the BM motion is more nearly linear (Sewell, 1984). Thus, the tip-to-tail difference of auditory-nerve fibers and of BM motion is reduced by furosemide (Ruggero and Rich, 1991).

Evoked OAE responses can be measured using stimulus conditions that are correlates of those used in behavioral, neural and BM studies of spectral and temporal processing. Evoked OAE types have been classified in several ways, one of which is based on generation mechanism (Kemp, 1986; Shera and Guinan, 1999). A SFOAE is a quasi-sinusoidal evoked cochlear response at a probe frequency f_p, and is thought to be generated mainly within the tonotopic region of f_p on the BM (Kemp, 1979; Wilson, 1980; Kemp and Chum, 1980; Zwicker and Schloth, 1984; Kemp and Brown, 1983a). SFOAE and click-evoked (CE)OAE are thought to be generated in human ears via a linear coherent reflection emission mechanism (Zweig and Shera, 1995; Kalluri and Shera, 2007). Distortion-product (DP) OAE (Kemp and Brown, 1983a; Harris *et al.*, 1989) are generated via a nonlinear distortion mechanism (Kemp and Brown, 1983b; Kemp, 1986). Nonlinear distortion observed in the ear canal as a DPOAE originates in the mechano-electrical transducer of the OHC, which relates the input hair-bundle displacement to the output receptor potential. This receptor-potential nonlinearity is also responsible for two-tone suppression (Weiss and Leong, 1985; Geisler *et al.*, 1990; Lukashkin and Russell, 1998). While place-fixed emissions are linear at low stimulus levels, most observations of SFOAEs occur at moderate levels at which the same cochlear nonlinearity attributed to the OHC is present.

SFOAEs are of particular interest because many auditory experiments measure: (1) the response to excitation by a tone, or (2) the detection of a tone in the presence of a "suppressor" or "masker." While BM and other physiologically invasive measurements in animals have proven useful in developing psychoacoustical models, these measurements are impossible to perform in humans for ethical reasons. Conversely, most psychoacoustics experiments have been performed using human listeners. SFOAEs may serve as an OAE correlate to nonlinear sinusoidal response growth at the BM, neural and behavioral levels, and thus serve as a way to link data from humans to observations made through invasive measurements in nonhuman mammals.

While cochlear nonlinearity can be assessed using single tones in BM input/output (I/O) responses, it can also be assessed using pairs of tones in suppression and intermodulation distortion responses (Ruggero and Rich, 1991; Ruggero *et al.*, 1992). Two-tone suppression has been observed in neural responses (Sachs and Kiang, 1968; Arthur *et al.*, 1971) and mechanical responses, as reviewed in Robles and Ruggero (2001). Mechanical suppression results from the saturation of the active-feedback mechanism, which leads to a more nearly linear BM I/O function (Zwicker, 1986).

Two-tone suppression of SFOAEs has been measured in cat (Guinan, 1990), chinchilla (Siegel *et al.*, 2003; Rhode, 2007), and human ears (Kemp and Chum, 1980; Brass and Kemp, 1993), in which human data were reported for three ears over a one-octave range of f_p. Two-tone suppression of SFOAEs in chinchillas was similar to responses in humans for low-frequency suppressors, but more complex for high-frequency suppressors (Siegel *et al.*, 2003). Unlike neural, mechanical, or behavioral measurements of two-tone suppression, in which the initial unsuppressed signal is directly measured, two-tone suppression of SFOAEs uses a two-step procedure. First, the total SFOAE in human ears is indirectly inferred by fully suppressing the SFOAE (Guinan *et al.*, 2003; Schairer *et al.*, 2003; Schairer and Keefe, 2005). The *total SFOAE* is the complex pressure phasor between the fully suppressed and unsuppressed conditions. Second, the tone eliciting the SFOAE is combined with a suppressor tone at a level less than that needed for full suppression. The resulting partially suppressed SFOAE is interpreted with respect to a decrement of the total SFOAE. The bandwidth of two-tone suppression of SFOAEs in human ears extends approximately 2.1 octaves about a f_p of 1 kHz at L_p=40 dB SPL (Backus and Guinan, Jr., 2006).

Associated with the concepts of frequency selectivity and the cochlear amplifier is the concept of cochlear gain. Aside from issues related to indirect measurements of gain (Allen and Fahey, 1992; de Boer *et al.*, 2005; Ren and Nuttall, 2006; Shera and Guinan, Jr., 2007; Shera *et al.*, 2007), cochlear gain is higher when the OHC feedback loop is functioning normally and lower when OHC dysfunction is present. Several definitions of cochlear gain are in current use (Robles and Ruggero, 2001). Assuming that BM responses to characteristic-frequency (CF) tones grow linearly at very low and very high stimulus levels, cochlear gain can be defined as the sensitivity difference at the lowest and highest levels, where sensitivity is the ratio of BM displacement to sound pressure at the tympanic membrane. A second definition is that cochlear gain is the difference between the maximum sensitivities *in vivo* and postmortem. Unfortunately, neither definition is applicable to defining cochlear gain in human ears, so that other approaches are required.

Cochlear gain has been indirectly assayed in human ears using suppression of DPOAE responses (with stimulus frequencies f_2 and $f_1<f_2$) by a third sinusoid. Using a simplified nonlinear model of cochlear mechanics, Mills (1998) concluded that the DPOAE tip-to-tail difference, based on the level difference at the stimulus f_2 frequency and one octave lower, would be approximately proportional to the cochlear gain at f_2. As long as even a weak DPOAE response is present, this model predicts that the suppression of such a DPOAE response may be used to infer characteristics of the underlying cochlear gain. Using human subjects with normal hearing and mild levels of hearing loss, a negative correlation of DPOAE tip-to-tail difference with hearing loss was obtained, which explained 29–45% of the variance in thresholds at f_2 frequencies of 2, 3 and 4 kHz (Pienkowski and Kunov, 2001). This finding supports the Mills model based

on the expected high prevalence of mild sensorineural hearing loss due to impaired OHC function. These results, including the tendency to explain more of the variance in thresholds at a f_2 of 4 kHz, were replicated and extended across a broader range of DPOAE stimulus and suppression parameters (Gorga *et al.*, 2002). Focusing on a f_2 of 4 kHz, the correlation of tip-to-tail difference with sensation level explained more variance than did its correlation with threshold SPL (Gorga *et al.*, 2003). Gorga *et al.* concluded that subjects with mild-to-moderate hearing loss have reduced cochlear gain compared to normal-hearing subjects, even though they have similar tuning around the tip of the DPOAE suppression tuning curve (STC). The properties of cochlear gain assessed indirectly using SFOAE STCs have not been studied.

One goal of this study was to determine the extent to which noninvasive measurements of SFOAE suppression reveal properties of cochlear processing in humans similar to knowledge gained from invasive experiments in other mammals. Another goal was to compare such measurements of SFOAE suppression to measurements of psychoacoustic masking and suppression, which have been explained using theories incorporating a compressive nonlinearity in the auditory periphery (Oxenham and Bacon, 2004). SFOAE suppression was measured based on SFOAE residual responses (SPL and phase) obtained over wide ranges of frequencies and levels of both probe and suppressor signals. From these measurements, SFOAE STCs were constructed as a function of suppressor frequency and level. Parameters of these STCs relevant to understanding cochlear nonlinearity and selectivity were analyzed, including a quality factor Q_{ERB}, which is based on the equivalent rectangular bandwidth (ERB) of the STC, the rate of the growth of suppression, cochlear compression, and the tip-to-tail difference. These measurements in individual ears were analyzed into group results and compared to published results of corresponding mechanical and neural measurements.

II. METHODS

A. Subjects

Data from subjects participating in this study were included only if their air-conduction thresholds were ≤15 dB at each half octave from 0.25 to 8 kHz and their air-bone gaps were ≤10 dB. SFOAE data were acquired in one ear of each subject on one day at a particular f_p, and data were collected at five f_p frequencies. For each f_p, the air- and bone-conduction data were collected on the same day as were the SFOAE data. A group of 24 normal-hearing subjects participated in the present study (21 females and 3 males), and data were collected in 12 right ears and 12 left ears. The age range was 20–39 years with a mean age of 28 years. The original goal was to test each of 20 subjects at all frequencies, but only 15 subjects attended all five test sessions. Additional subject recruitment was performed until 20 subjects had been tested at each of the five f_p frequencies.

B. Data collection and analysis

Data were collected using a probe assembly with two receivers and one microphone (Eytmotic ER10C, modified at the factory to allow a 20 dB increase in output from each receiver). A sound card (CardDeluxe) under computer control outputted an electrical stimulus through each of a pair of digital-to-analog converters (DACs), with the output of each DAC driving a receiver. The output response signal from the microphone was recorded using an analog-to-digital converter (ADC). These converters were part of the sound card installed in a Windows computer operating with custom software. Data were converted synchronously at a sample rate of 22.050 kHz per channel. Each sinusoidal tone was presented through a separate receiver to minimize intermodulation distortion within the probe system. The probe was inserted into the ear canal using a soft foam tip to ensure the subject's comfort during the experiment. Acoustic pressure signals were measured at the probe tip.

The stimulus set included a presentation of a sinusoidal probe tone of frequency f_p, which was varied over a set of five octave frequencies from 0.5 to 8 kHz, a presentation of a sinusoidal suppressor tone of frequency f_s and level L_s, and the joint presentation of probe and suppressor tones. Data were acquired at each f_p for probe levels L_p ranging from 20 to 70 dB SPL in 10 dB steps. The f_s stimuli ranged from two octaves below f_p to 0.7 octaves above f_p. For example, the complete range at f_p=1 kHz extended from approximately 0.25 up to 1.6 kHz. The particular values of f_s/f_p for each f_p were 0.260, 0.348, 0.522, 0.609, 0.695, 0.825, 0.956, 1.043, 1.130, 1.174, 1.260, 1.347, 1.434, 1.477, and 1.61. The range was slightly reduced at the highest f_p=8 kHz: the maximum f_s/f_p was limited to 1.33 so that f_s would not exceed the Nyquist criterion.

The microphone response to the probe stimulus was p_p, to the suppressor stimulus was p_s, and to the joint presentation of the probe and suppressor stimuli was p_{ps}. The *SFOAE residual* was defined in terms of the nonlinear residual p_{oae} to these three stimuli by (Keefe, 1998):

$$p_{\mathrm{oae}} = p_p + p_s - p_{\mathrm{ps}}. \tag{1}$$

If this SFOAE residual corresponded to full suppression of the SFOAE, then this SFOAE residual would equal the total SFOAE as defined in the Introduction. The data collection and averaging procedures were similar to those described in Schairer *et al.* (2003). The low-level data acquisition procedures used to measure SFOAE residuals are further described in the Appendix.

Responses were always measured over the full range of L_p, but data were acquired over subranges of f_s and L_s that varied according to f_p and L_p. Preliminary test data were used to identify these subranges, based on the premise that data need not be acquired if no SFOAE residual would be detected. The choices of subranges were checked after all data in the present study were acquired. The range of L_s did not exceed 20–92 dB SPL at f_p=0.5 kHz or 14–80 dB SPL for f_p>0.5 kHz. The L_s was varied in 6 dB steps.

Data were collected while the subject was comfortably seated in a sound-attenuated room. The number and range of

parameter values was constrained by the practical limitation of a nominal 2 h duration for each f_p, which included a break in the middle. Data were collected for each subject over five separate sessions, with each session on a different day. It was never the case that the subrange was too restricted to measure a low-level SFOAE residual. The only restriction that was encountered occurred at f_p=4 kHz, for which the maximum L_s of 82 dB SPL was too low. This is further described in the Discussion.

For each L_p and f_p, a SFOAE residual was judged valid at each f_s and L_s only if each of the following was satisfied: (1) its signal to noise ratio (SNR) was at least 6 dB, (2) its noise SPL at each f_s and L_s did not exceed the median noise SPL (at that f_s and across its L_s range) by more than 6 dB for any L_s, and (3) its noise SPL at each f_s and L_s remained within 10 dB of the grand median noise SPL calculated across both L_s and f_s. Condition 2 was intended to exclude intermittent noise. Condition 3 ensured that intermittent noise spikes for individual f_s suppressors were excluded, so that the construction of the STC curve was not contaminated by such artifact. It was not surprising that some subject-generated artifacts occurred in an experimental session lasting 2 h, and the above objective criteria were helpful in excluding such artifacts.

The biological origin of the SFOAE residual was confirmed by comparisons with recordings in an artificial ear (IEC 711 reference coupler, Bruel & Kjaer Type 4157). SFOAE residuals were not measured at f_p=0.5 kHz for L_p <40 dB SPL because preliminary studies showed that the SNR was too small at lower levels. Coupler distortion was present at 8 kHz in the range of some of the biological responses, and especially at the highest L_p of 70 dB SPL. Thus, the range of probe levels in the analyzed results was restricted to 20–60 dB SPL.

III. RESULTS

The individual-ear responses plotted below were selected from recordings in several subjects so that the variety and overall consistency of responses can be appreciated. An example of SFOAE residual SPLs for an individual ear (subject 1) is shown in Fig. 1 for a SFOAE at f_p=1 kHz as a function of f_s and L_s. Each row shows data at increasing L_p levels from 20 to 60 dB SPL in 10 dB steps. Each SFOAE residual SPL curve is plotted versus L_s (solid lines), and the corresponding noise SPL is shown as a dashed line of the same thickness. The left column of panels show data for low-side suppressors $f_s<f_p$ and the right column for high-side suppressors $f_s>f_p$. Only those SFOAE residuals with a 6 dB SNR are plotted, except that the SPL is also shown for the next lower value of L_s that did not satisfy this SNR criterion so that the shape of the curve emerging from the noise floor is evident. Only a few of the suppressor frequencies produced sufficiently high SPLs at low L_p, e.g., see data for L_p=20 dB SPL. This occurred because the SFOAE level had insufficient dynamic range relative to the noise level.

At L_p=40 and 50 dB SPL, the SFOAE residual curves were well defined for many of the suppressors. Low-side SFOAE suppression was characterized by a relatively steep

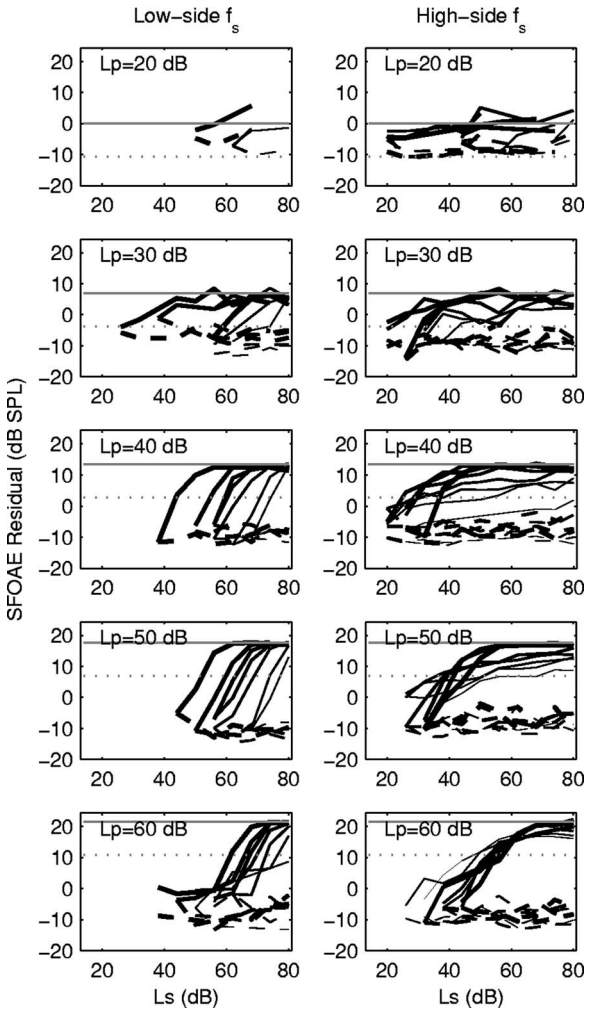


FIG. 1. For subject 1 at f_p=1 kHz, SFOAE residual SPL (solid lines) and noise SPL (dashed lines) are plotted vs L_s at suppressor frequencies $f_s<f_p$ in the left column, and $f_s>f_p$ in the right column. Lines are thicker for f_s closer to f_p. L_p increases from top to bottom row. The set of possible values of $f_s<f_p$ are specified in Sec. II, but data are plotted only when the SNR criterion was satisfied. A saturation region (horizontal solid line) shows when SFOAE is fully suppressed. Note that this value is independent of f_s and provides the SPL of the total SFOAE. STCs are constructed at a criterion SPL that is 10.7 dB below the saturation level (shown by horizontal dotted line).

slope (>1 dB/dB) that was nearly constant across f_s. High-side SFOAE suppression was characterized by a similarly steep slope for f_s just above f_p, but smaller slopes (<1 dB/dB) for higher f_s. These slopes are further described below (see Fig. 8). The saturation region (identified by a horizontal line in each panel) was well defined for L_p >20 dB SPL. The maximum residual SPL in the saturation region was the total SPL of the SFOAE (corresponding to full suppression of the SFOAE), and in this example was 13 dB SPL at L_p=40 dB SPL and 18 dB SPL at L_p=50 dB SPL. These total SFOAE SPLs are not plotted inasmuch as they were consistent with previous reports based on I/O functions (Schairer *et al.*, 2003; Schairer and Keefe, 2005). The saturation region was poorly defined at L_p=20 dB SPL because the SNR was small, as noted above. In such cases, the saturation region was defined as the maximum SFOAE residual SPL with a slope not exceeding 0.12 dB/dB, which

ensured that the region so defined had a shallow slope consistent with a saturation effect. A typical high-side suppression pattern showed incomplete suppression of the SFOAE at the highest f_s (i.e., the thin lines for L_p of 40 and 50 dB in Fig. 1), inasmuch as their residual SPLs were smaller than those measured at lower f_s. This is because the L_s levels used for these high-side suppressors were insufficient to completely suppress the SFOAEs.

For a specific probe condition parametrized by f_p and L_p, the SFOAE STC was the set of L_s values that produced iso suppression, i.e., the same SFOAE residual SPL, as a function of f_s. The SFOAE STC at each f_s was constructed using a cubic-spline interpolation of the data in Fig. 1 to calculate the L_s for which the SFOAE residual SPL was 10.7 dB below the total SPL. This criterion is shown by the horizontal dotted line that is 10.7 dB below the total SPL (horizontal solid line) in each panel of Fig. 1. This criterion SPL is defined as the *SFOAE decrement* and is measured relative to the SPL of the total SFOAE. SFOAE suppression of the probe tone produced a decrement at f_p. This decrement was observed as an increment in the SFOAE residual; i.e., no SFOAE residual meant no suppression of the SFOAE.

A 3 dB increment in the SFOAE residual corresponded to an amplitude increase by a factor of $1/\sqrt{2}$. This 3 dB increment was equivalent to a 10.7 dB decrement with respect to the SPL of the total SFOAE, because $20\log_{10}(1-1/\sqrt{2})=-10.7$ dB. STCs were also constructed using an alternative criterion decrement of 6 dB to study the effect of varying the criterion. A 6 dB decrement in the SFOAE residual was equivalent to a 6 dB decrement with respect to the SPL of the total SFOAE, because −6 dB corresponds to an amplitude change of $1/2$ and $20\log_{10}(1-1/2)=-6$ dB. The criterion of 10.7 dB was preferred for constructing a SFOAE STC, because it was closer to the onset-of-suppression threshold. STCs constructed using the alternative criterion of 6 dB were sufficiently similar to those constructed using the 10.7 dB criterion that results are reported only using the 10.7 dB criterion.

The SFOAE residual data had to be valid (as described in Sec. II) for at least 70% of the f_s frequencies in order to calculate a STC. The data in Fig. 1 illustrate that interpretable SFOAE STCs would be constructed for L_p at and above 30 dB SPL, but not at 20 dB SPL because of an insufficient number of valid data (mainly due to the low SNR at L_p =20 dB SPL).

The SFOAE residual plots showed evidence at 0.5 kHz in more than half of the ears of a high-level region above the saturation region, in which the SFOAE SPL residual increased rapidly with L_s. Although the example in Fig. 2 (subject 15) showed a well-defined saturation region for low-side suppression (left column), an additional high-level region for L_s above 80–86 dB SPL was present for high-side suppression when L_p=50 dB SPL. Such a rapid increase in the SFOAE SPL residual is not predicted by the two-tone suppression mechanism. This was probably not a middle-ear muscle reflex effect because it was present only for high-side suppression. Any reflex effect elicited by a suppressor f_s slightly larger than f_p would also have been elicited by a suppressor slightly smaller than f_p. It was likely a cochlear-

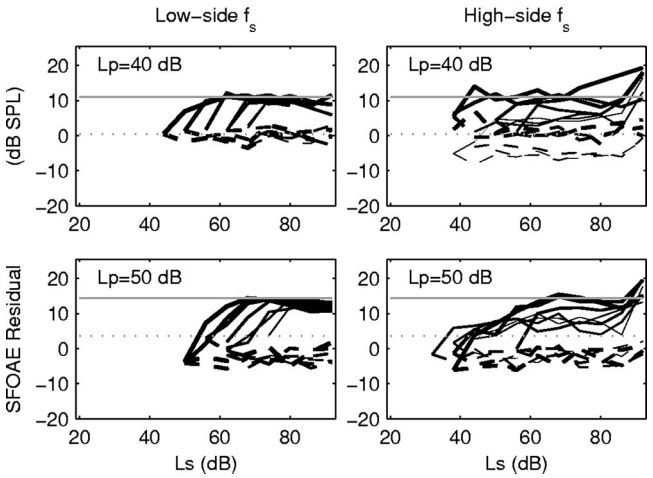


FIG. 2. SFOAE residual plots for subject 15 at f_p=0.5 kHz. Plotting conventions are the same as those in Fig. 1.

distortion effect acting near the tonotopic location of f_s and slightly basal to the f_p place. A possible mechanism for such distortion has been described by Talmadge *et al.* (2000). This increase was also observed on low-side suppression in a few ears (not shown), for which an activitation of the middle-ear muscle reflex by a high-level suppressor was a possibility.

The SFOAE residual at f_p=4 kHz for subject 5 in the left column of Fig. 3 shows the SFOAE SPL at fixed L_p as a function of f_s/f_p, with each curve plotted at fixed L_s. High-side suppression dominated at low L_s, with a most sensitive region approximately one-quarter octave above f_p ($f_s/f_p \approx 1.19$) for L_p in the range from 20 to 50 dB SPL. The most sensitive region increased to f_s/f_p=1.6 at L_p=60 dB SPL, a trend also evident in the STC at L_p=50 dB SPL. Saturation of the SFOAE residual occurred at high L_s, and the STC broadened towards lower f_s/f_p at higher L_s. As expected, the maximum suppression, when saturated, was independent of L_s and f_s, but varied with L_p, because it depended on the probe response alone.

An alternative representation of the SFOAE residual SPL is in terms of a SFOAE suppression sensitivity function. This is defined as the amplitude ratio of the SFOAE pressure residual (produced by the suppressor) to the suppressor pressure. When expressed as a level, the suppression sensitivity level is the difference of the SFOAE residual SPL relative to L_s. This is plotted as a function of f_s/f_p in the right column of Fig. 3, such that each curve is an iso-suppressor response at a particular L_s. Either representation shows that SFOAE suppression is most sensitive at low levels for $f_s/f_p>1$. The suppression sensitivity level shows the increasing breadth of the suppression peak with increasing L_s as a function of f_s/f_p. The panels with L_p=50 dB show one case of artifact when f_s was close to 0.125 kHz, which lies below the minimum frequency plotted. The maximum suppression sensitivity levels in Fig. 3 ranged from −25 dB at low L_p to −33 dB at high L_p.

The altered shape of the suppression sensitivity at a L_p of 50–60 dB SPL is an indicator that cochlear nonlinearity is qualitatively changing in this moderate range. The peak sensitivity shifted to slightly higher f_s/f_p at high L_p. The shift

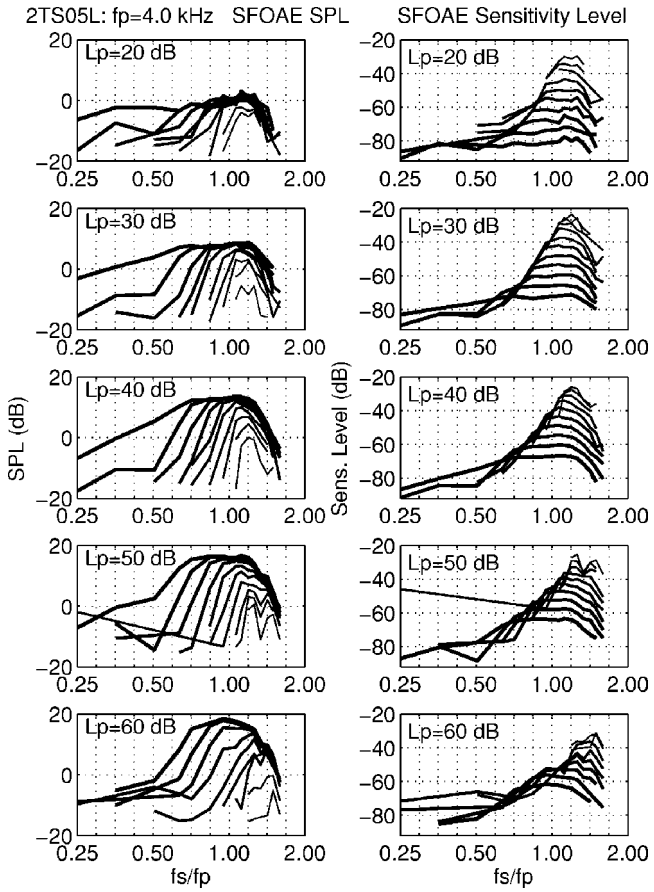


FIG. 3. SFOAE suppression results for subject 5 at f_p=4 kHz, shown as SFOAE SPL vs. f_s/f_p (left column) and the same data shown as SFOAE Suppression Sensitivity Level vs. f_s/f_p (right column). L_p increases from 20 to 60 dB SPL in 10 dB steps from top to bottom row. Thicker contour lines show higher L_s decreasing in 6 dB steps from 80 down to 14 dB SPL.

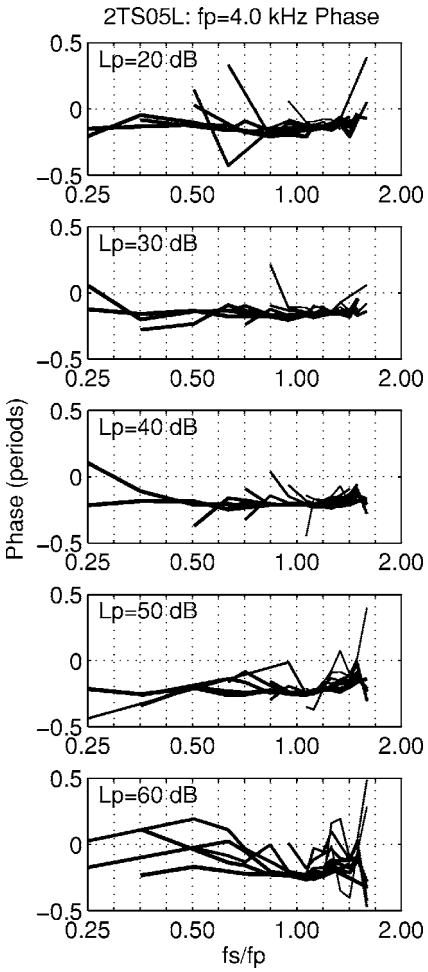


FIG. 4. SFOAE suppression results for subject 5 at f_p=4 kHz. Phase change vs. f_s/f_p. L_p increases from 20 to 60 dB SPL in 10 dB steps from top to bottom row. Thicker contour lines show higher L_s decreasing in 6 dB steps from 80 to 14 dB SPL.

appeared discontinuous inasmuch as the sensitivity at L_p =50 dB had two peaks, one similar to the single-peak frequency at L_p=40 dB and the other at higher f_s similar to the single-peak frequency at L_p=60 dB. One possibility is that this change is related to an onset of distortion at higher L_s that occurs at f_s/f_p close to 1.5, or just outside the region of cochlear-amplification around the tonotopic place. More study is needed to understand the underlying reasons for this effect.

The shape of the SFOAE suppression sensitivity somewhat resembles the vibration sensitivity (velocity per unit pressure) reported for mechanical measurements on the BM of a mammalian cochlea (Ruggero *et al.*, 1992), inasmuch as both transfer functions reveal the frequency selectivity associated with the cochlear-amplifier function of a healthy cochlea. Nevertheless, the pair of sensitivity functions are defined in terms of different variables: the input variable to the SFOAE sensitivity transfer function is the suppressor SPL at each suppressor frequency, whereas that for the vibration sensitivity at a particular location on the BM is the stimulus SPL.

The corresponding SFOAE phase suppression plot (SFOAE phase measured relative to the probe phase) is shown in Fig. 4. Each panel shows results for a single L_p, which increases from top to bottom in the figure. The individual curves show the response for a particular L_s, with increasing line thickness representing increasing L_s in 6 dB steps. SFOAE phase suppression was usually small, i.e., close to 0 periods. More phase variation was observed at L_p=60 dB SPL than at lower L_p. Averaged across all low-side suppressors, the average SFOAE phase lagged by approximately 0.15 periods (5°).

A SFOAE STC constructed from SFOAE residual data (such as those shown in Figs. 1 and 2) is shown for subject 23 in Fig. 5. The SFOAE STC was first measured (A, top five panels) and re-measured 30 days later (B, bottom five panels). Each panel shows the STC at an octave f_p from 0.5 to 8 kHz, with increasing line thickness representing increasing L_p. The absence of a curve at a particular L_p means that a valid STC was not obtained. Tuning in individual ears was better represented at 1, 2 and 4 kHz than at 0.5 or 8 kHz. Nevertheless, a tip region was present at most f_p. The mean and standard deviation (SD) of the difference of the first STC relative to the second STC was calculated at each f_p, by averaging over all f_s/f_p and all L_s common to both STCs. The mean difference in suppression thresholds varied from −1.7 to 5.6 dB with f_p, with a grand average of 0.6 dB. The SD of the difference varied from 6.0 to 8.7 dB with f_p, and was characterized by similar suppression thresholds for most

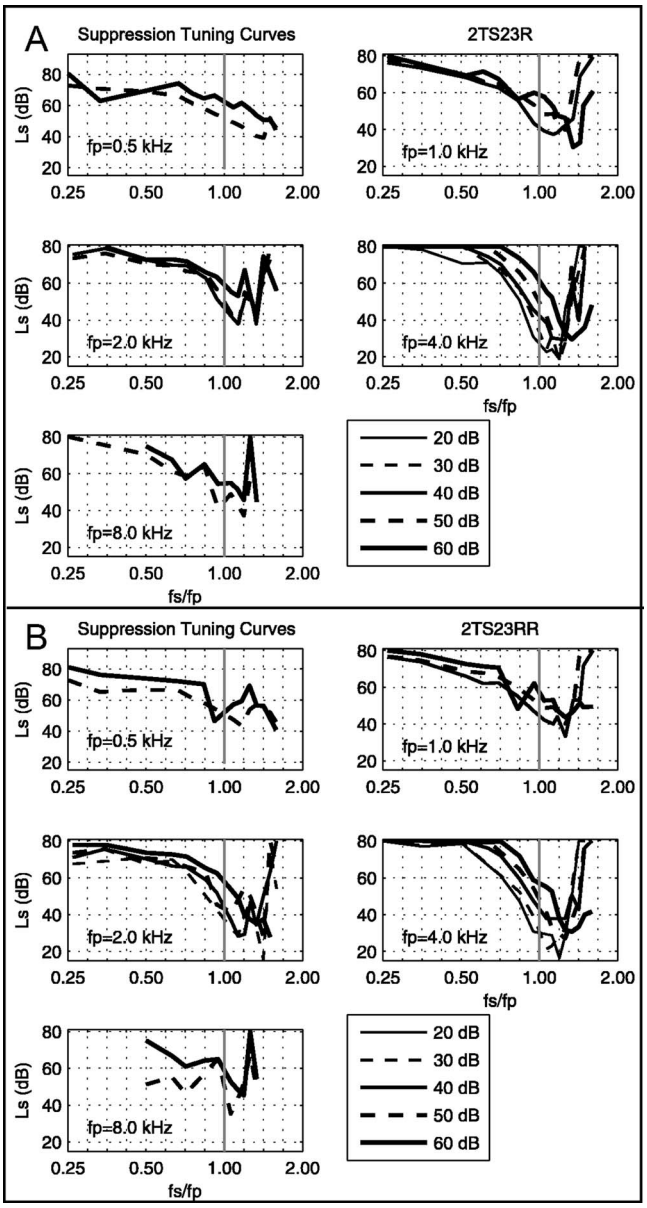


FIG. 5. Original and repeated tests (panels A and B, respectively) of STCs in subject 23 measured 30 days apart. SFOAE STCs are plotted at each octave of f_p from 0.5 to 8.0 kHz. Line thickness increases with increasing L_p, with adjacent L_p's coded with alternating solid and dashed lines (thickest solid line is L_p=60 dB SPL, thinnest is 20 dB SPL).

data with a few outliers increasing the SD. Thus, the SFOAE STC of an individual ear was repeatable to within 0.6 dB overall.

Group STC responses were analyzed by averaging the SFOAE SPL and phase STCs across subjects. The SFOAE STC mean and standard error of the mean (SE) are plotted in Fig. 6. The mean STC had a tip region of maximum sensitivity at a frequency slightly higher than f_p ($f_s/f_p \approx 1.19$), which agrees with the observations of Brass and Kemp (1993). For $L_p \geq 30$ dB SPL, the tip L_s typically increased with increasing L_p. This effect is combined in the figure with the effect of shifts with increasing L_p in the most sensitive f_s. The mean and SE of the SFOAE phase STC are plotted in Fig. 7. The phase varied within ± 0.3 cycles with variations in f_s/f_p and L_p.

IV. DISCUSSION

Properties of SFOAE suppression responses are described and compared to measured mechanical and neural responses, beginning with the mean rate of growth of suppression.

A. Rate of growth of suppression

The mean rate of growth of SFOAE suppression is the increase in the SPL of the SFOAE residual for an increase in L_s. It was measured at each f_s on the STC using a SFOAE residual suppression function such as that shown in Fig. 1. For each curve parametrized by f_s, the rate of growth of suppression was calculated as the slope of this function at the intersection of the SPL that was ΔL=10 dB down from the SPL of the total SFOAE. This is illustrated by the slopes of the SFOAE residual responses plotted in Figs. 1 and 2 at a level 10 dB below the dotted line indicating full suppression (i.e., the SPL of the total SFOAE). The slope was calculated from the derivative of the spline fit. The resulting group mean and SE of the rate of growth of suppression are shown in Fig. 8.

The SFOAE rate of growth of suppression was approximately 1.5 dB/dB for low-side suppression. This compares well with measurements reported in mammalian ears at the mechanical and neural level. For example, the rate of growth of two-tone suppression in BM mechanical responses in chinchilla ranged from 0.65 to 1.42 dB/dB, with a mean of 0.97 dB/dB (Ruggero *et al.*, 1992) for suppressor frequencies below the probe frequency, when measured at a characteristic frequency (CF) place close to the probe frequency. Another study on two-tone suppression in chinchilla in the 6 kHz region showed a rate of growth of suppression on the BM larger than 2.5 dB/dB for low-frequency suppressors, which decreased with increasing suppressor frequency to less than 1 dB/dB above the characteristic frequency (Rhode, 2007). Based on measurements of two-tone suppression in auditory-nerve fibers in cat, Delgutte (1990) found that the rate of growth of suppression varied from 1 to 3 dB/dB for low-side suppression. Rhode (2007) concluded that rate of growth of suppression produced by low-frequency suppressors was similar for BM and auditory-nerve measurements in mammalian ears.

For high-side suppression at the lowest L_p levels and independent of f_p, the SFOAE rate of growth of suppression decreased with increasing f_s/f_p to approximately 1 dB/dB at f_s/f_p=1, and 0.3 dB/dB at the largest f_s/f_p=1.7. In chinchilla BM measurements, the rate of growth of suppression in two individual ears ranged from 0.30 to 0.35 dB/dB near the CF place for f_s/f_p of 1.23 and 1.25, based on f_p of 8.6 and 8 kHz, respectively (Ruggero *et al.*, 1992). In the auditory-nerve fiber measurements in cat, the rate of growth of suppression ranged between 0.15 and 0.7 dB/dB for high-side suppressors (Delgutte, 1990). Both the SFOAE and neural growth rates decreased with increasing f_s/f_p for high-side suppression.

The fact that low-side suppression growth rates were occasionally higher in neural than BM mechanical measurements, and the fact that the growth rates were slightly higher

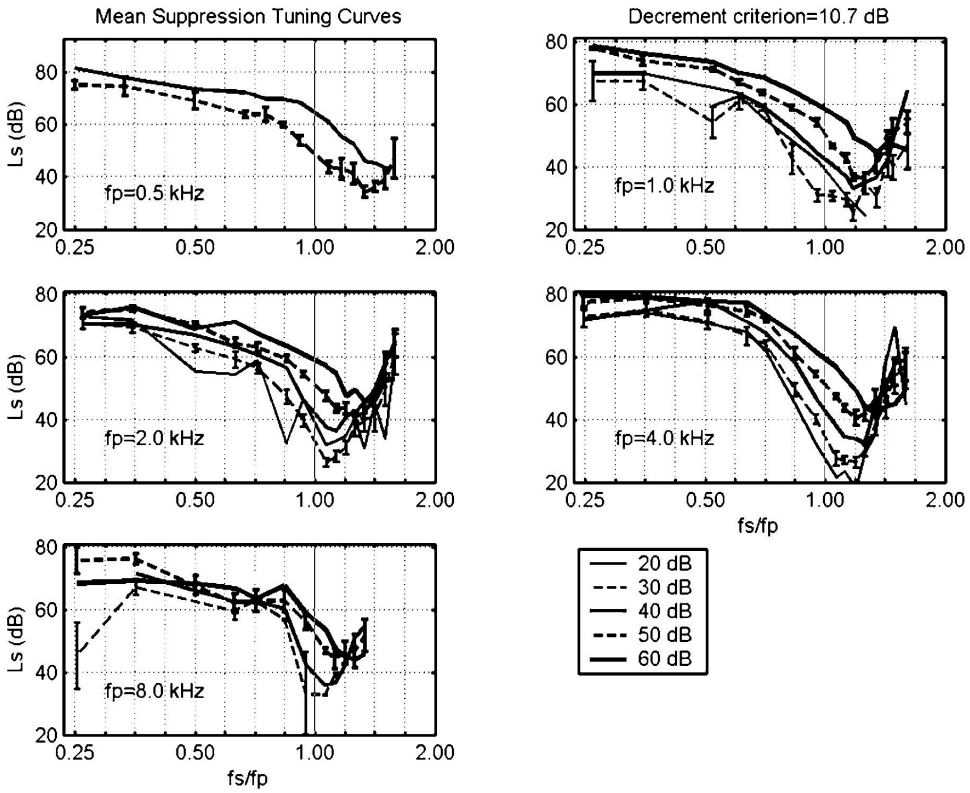


FIG. 6. Mean group SFOAE STC at each f_p from 0.5 to 8.0 kHz averaged across subjects based on the decrement criterion of 10.7 dB. Line thickness increases with increasing L_p, with adjacent L_p's coded with alternating solid and dashed lines. Every other STC shows ± 1 SE as error bars.

in SFOAE than BM mechanical measurements may reflect species differences, a reduction in the BM growth rate due to effects of surgery, or intrinsic differences in measurement techniques. For the latter, the mechanical BM and neural measurements are related to the motion over a particular place in the cochlea whereas the SFOAE is thought to in-

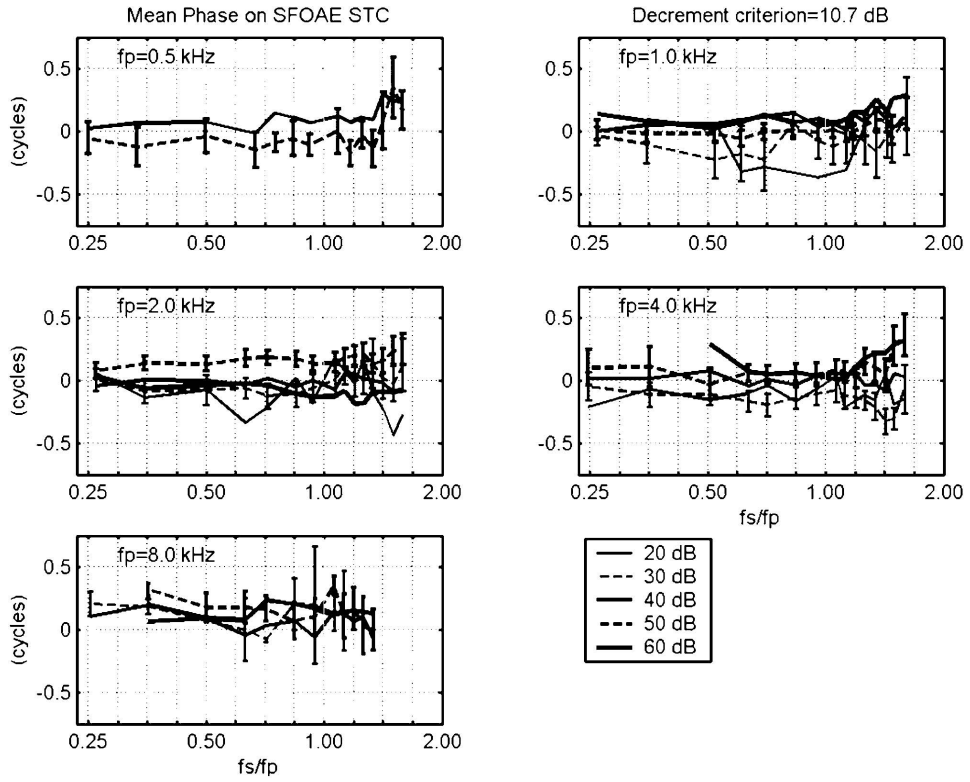


FIG. 7. Mean group SFOAE STC phase based on the decrement criterion of 10.7 dB. Similar plot conventions as those in Fig. 6.

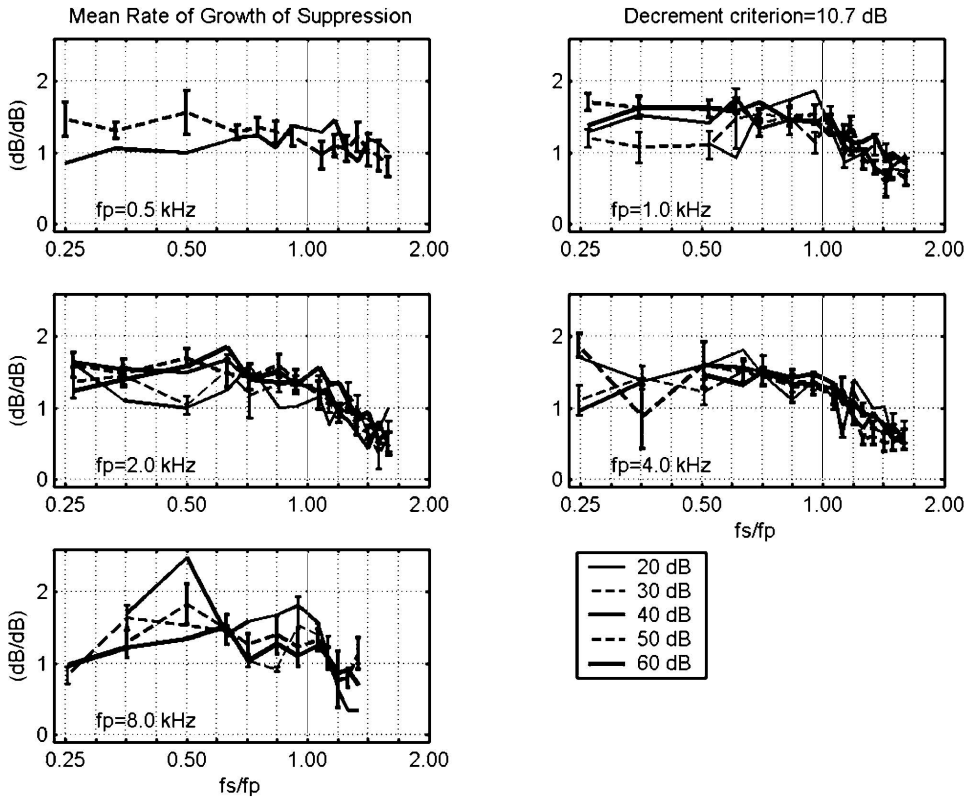


FIG. 8. Mean of the rate of growth of SFOAE suppression vs. f_s/f_p, at f_p from 0.5 to 8.0 kHz in different panels, and at L_p from 20 to 60 dB SPL (thicker lines for higher levels) within each panel. Every other slope curve shows ±1 SE as error bars.

clude contributions over the breadth of the tuning around the tonotopic place of f_p. The low-side suppression growth rate based on SFOAE measurements may underestimate the nonlinear compression of the cochlear amplifier because SFOAE results from a superposition of excitations over locations along the BM for which the cochlear amplification is less than near the CF place. Aside from slight differences in magnitude, however, the qualitative pattern of two-tone suppression is similar in mechanical, neural and SFOAE measurements.

B. Suppression tuning via Q_{ERB}

The frequency selectivity of suppression tuning was assessed in terms of a quality factor Q_{ERB} of each STC. The STC was inverted and its level converted to squared amplitude. At each f_p, the area under the inverted squared-amplitude function was calculated over the range of f_s, and the ERB was calculated as the ratio of this area to the peak squared amplitude at the tip of the original STC. The Q_{ERB} was calculated as $Q_{ERB}=f_p/\text{ERB}$. Each group mean ±1 SE was calculated based on the set of individual Q_{ERB} at each f_p and L_p, and plotted in Fig. 9. However, an individual Q_{ERB} response was excluded if the tip value occurred at the highest f_s for which data were otherwise valid. This excluded STCs lacking a well-defined tip region. An example of such an excluded STC is that at $L_p=60$ dB SPL and $f_p=0.5$ kHz in Fig. 5, for which the minimum L_s occurred at the highest f_s.

The Q_{ERB} varied little with L_p, and was approximately equal to the behavioral Q_{ERB} derived from psychophysical measurements of critical bandwidth based on simultaneous masking (Glasberg and Moore, 1990). These psychophysical measurements are also shown in Fig. 9 and include both excitation and suppression effects. The approximate equivalence of Q_{ERB} in SFOAE and psychophysical measurements suggests that excitation effects in simultaneous masking may not add significant differences to the critical bandwidths accounted for by suppression effects measured acoustically.

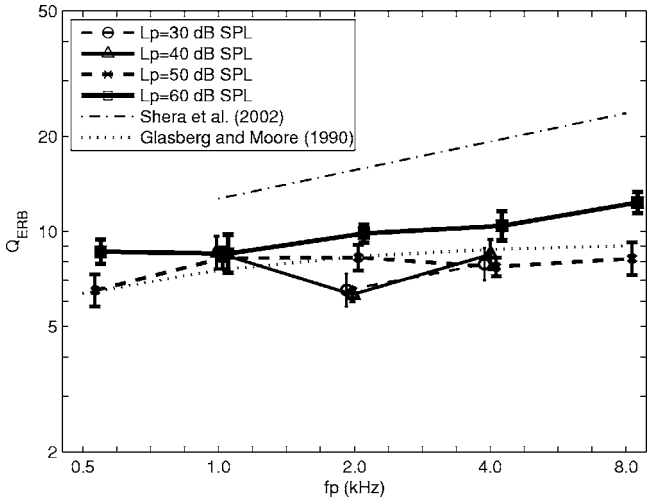


FIG. 9. The mean and ±1 SE of Q_{ERB} of the STC is plotted as a function of f_p for L_p's ranging from 20 to 60 dB SPL (thicker lines for larger L_p). The short dotted black line is the behavioral Q_{ERB} based on simultaneous masking using notched noise (Glasberg and Moore, 1990). The dashed-dotted black line is a measure of the behavioral Q_{ERB} based on nonsimultaneous masking and is a curve derived from SFOAE latency data assuming that SFOAEs are generated by a minimum-phase system (Shera *et al.*, 2002).

This is consistent with Oxenham and Plack (1998), who concluded that suppression plays a major role in the nonlinear upward spread of simultaneous masking for low-side maskers.

Figure 9 also shows Q_{ERB} estimated from SFOAE latencies for L_p=40 dB SPL, which is based on a proposed equivalence of BM latency and tuning via an assumption of cochlear scaling (Shera *et al.*, 2002). It has been suggested that this derived estimate of Q_{ERB} is similar to Q_{ERB} measured psychoacoustically using a forward-masking procedure with a low-level probe (Shera *et al.*, 2002; Oxenham and Shera, 2003). Schairer *et al.* (2006) reported SFOAE latencies at L_p=40 dB SPL that were similar at 1 kHz but shorter at 2 and 4 kHz compared to results in Shera *et al.*, which would suggest a smaller derived Q_{ERB} at 2 and 4 kHz. The discrepancy is currently unresolved. Schairer *et al.* also reported that SFOAE latency decreased with increasing L_p in the range from 40 to 70 dB SPL, which suggests a decrease in the derived Q_{ERB} with increasing L_p. Thus, the SFOAE-derived Q_{ERB} would approach the simultaneous-masking measurement of Q_{ERB} from above, and would also approach the present measurements of Q_{ERB} based on SFOAE suppression. It is interesting that the present Q_{ERB} results have no systematic variation with changes in L_p, whereas the SFOAE-latency derived Q_{ERB} decrease with increasing L_p. This is evidence that the SFOAE suppression and latency measures assess different attributes of the underlying BM mechanics. The present estimates of Q_{ERB} based on SFOAE suppression were similar to those associated with behavioral simultaneous masking, but lower than those associated with behavioral forward masking.

C. On-frequency and off-frequency SFOAE suppression

The SFOAE suppression thresholds L_s are plotted in Fig. 10 as a function of L_p for on-frequency (dashed lines) and off-frequency (solid lines) suppressors. This is termed a growth of suppression (GOS) response. The on-frequency GOS response represents the threshold L_s required when f_s was slightly higher (by 2–3%) than f_p, depending on the specific spectral bin frequencies available. With f_s close to f_p, the BM response is compressed by approximately the same amount at probe and suppressor frequencies, which means that the slope of the on-frequency GOS responses should be approximately 1 dB/dB. The off-frequency GOS response is the L_s required when f_s was one octave below f_p. This suppressor frequency was chosen because the entire excitation pattern at f_p is basal to the active region of the excitation pattern at f_s an octave below f_p. The BM response at the tonotopic place of f_p responds linearly to a unit growth in the off-frequency suppressor an octave below f_p, and this is taken as a criterion amount of suppression in the probe tone at f_p. Thus, a larger than unit growth in the on-frequency suppressor is required to produce the same criterion amount of suppression in the probe tone at f_p. For example, if the compressive slope c of the BM I/O function at f_p were c =0.25 dB/dB, then the on-frequency suppressor would need

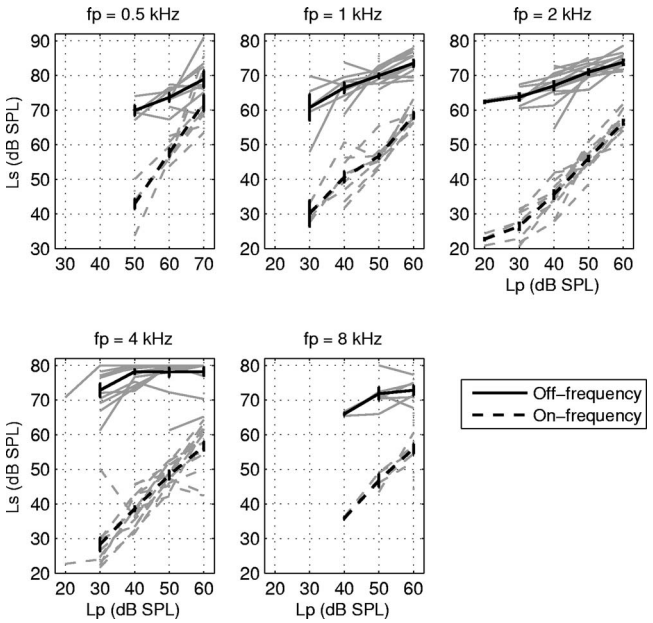


FIG. 10. Plots of threshold L_s (dB SPL) vs. L_p, with each panel showing results for f_p at octaves from 0.5 to 8 kHz. Dashed gray lines show individual-subject data for on-frequency suppressors, i.e., suppressor frequencies just above the probe frequencies ($f_s/f_p \approx 1.03$). Solid gray lines show individual-subject data for off-frequency suppressors, i.e., for f_s an octave below f_p. The group mean ±1 SE are plotted using black lines for all stimulus conditions with data from at least two subjects.

to be increased by $1/c$=4 dB to produce the same amount of suppression as the off-frequency suppressor would produce with an increase of 1 dB.

In order to apply this BM mechanical theory to SFOAE suppression, one needs the additional assumption, which is consistent with interpretations of SFOAE latency measurements and confirmed by SFOAE residual phase data described below, that SFOAEs are predominantly generated in the tonotopic region of the probe at f_p for the range of f_s and L_s used in the present experiments. To the extent that the BM compression is uniform over the tonotopic region, then SFOAE measurements can reveal the underlying saturating compressive nonlinearity of BM mechanics. If BM compression is nonuniform, then it is likely that the SFOAE GOS technique will underestimate the amount of compression.

The GOS results in Fig. 10 show individual-ear results with valid data as well as the group mean ±1 SE of the threshold L_s. The on-frequency GOS response extends to low threshold values of L_s and has a slope close to one. The off-frequency GOS response is limited to relatively high threshold values of L_s and has a slope much less than one, consistent with the above model of compressive nonlinearity. The off-frequency GOS response at f_p=4 kHz has a ceiling effect inasmuch as most ears did not have a threshold L_s within the measurement range (≤80 dB SPL). This ceiling effect had only slight impact on the results at L_p of 30–40 dB SPL. The GOS responses at f_p=4 kHz also included a few unexpected individual-ear responses with negative slopes, which may be a result of artifact, but the overall mean appears reasonable.

The slope of each individual-ear GOS response was calculated over its range of valid data. Because these data based

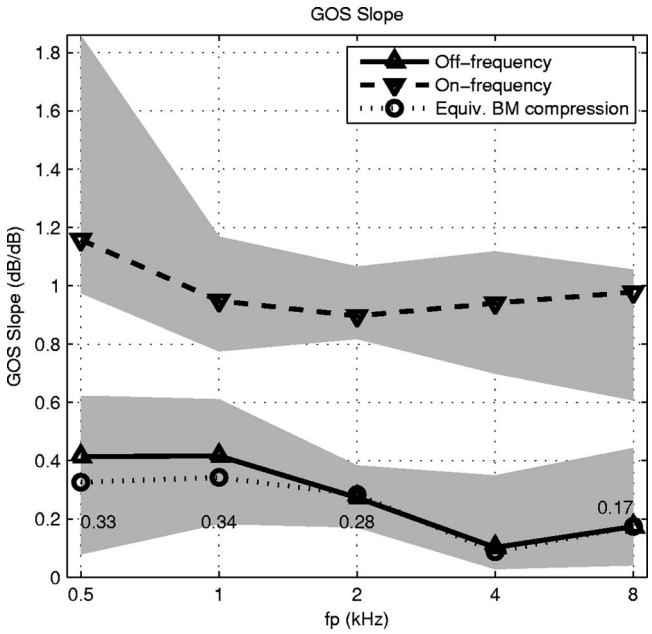


FIG. 11. Slopes for GOS on-frequency and off-frequency responses and equivalent BM compression.

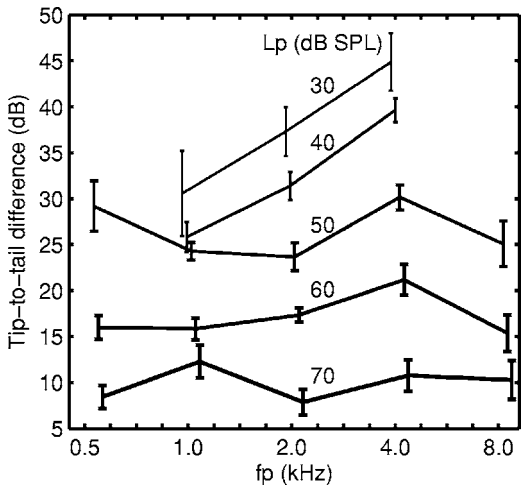


FIG. 12. Mean ±1 SE of the tip-to-tail difference of the SFOAE STC vs. probe frequency. Each curve represents different L_p. Tip-to-tail difference functions are slightly displaced along the horizontal axis to improve clarity. The variation of the tip-to-tail difference with L_p is implicitly represented at each f_p by the tip-to-tail differences along each vertical slice through the individual curves.

on responses in Fig. 10 were somewhat noisy, the median and inter-quartile range of the slopes of the individual GOS responses were calculated (as an alternative to the mean and SE), and are plotted in Fig. 11. The on-frequency median slope was close to 1 dB/dB for f_p=1, 2, 4 and 8 kHz. The on-frequency median slope was close to 1.2 dB/dB and more variable at f_p=0.5 kHz, most likely due to fewer valid data at 0.5 kHz and perhaps the higher noise levels accompanying measurements at this frequency. The off-frequency median slope ranged from 0.2 to 0.4 dB/dB, except that it was 0.1 dB/dB at f_p=4 kHz. This latter result does not indicate more compression at 4 kHz, but, rather, is a result of the ceiling effect apparent in Fig. 10 at 4 kHz. The equivalent BM compression is defined as the ratio of the off-frequency to the on-frequency GOS slopes. Because the on-frequency GOS slope was usually close to 1 dB/dB (i.e., for $f_p \geq 1$ kHz), the equivalent BM compression was approximately equal to the off-frequency GOS slope. This compression is plotted in Fig. 11 along with its numerical values at f_p other than 4 kHz. The equivalent BM compression was approximately 0.3 dB/dB at 1–2 kHz, but large variability in individual responses precluded a detailed comparison across f_p. This slope is typical of compression measured psychophysically, as summarized in Yasin and Plack (2007), but slightly larger (i.e., less compressive) than the range of 0.13–0.16 dB/dB in BM mechanical measurements (Yates *et al.*, 1990; Ruggero *et al.*, 1992).

Regarding slopes estimated in other OAE studies, the compression inferred from suppression of DPOAEs at f_2 =4 kHz was 0.26 dB/dB (Gorga *et al.*, 2002), as compared to 0.3 dB/dB in the present study at 1 and 2 kHz. The slope of the SFOAE I/O function, which is the slope of the total SFOAE SPL as a function of L_p, is compressive with values in the range of 0.44–0.61 dB/dB between f_p of 1–4 kHz (Schairer *et al.*, 2003). These slopes are higher than the present slopes obtained from GOS data. Along with the presence of notches in SFOAE I/O functions (Schairer *et al.*, 2003; Schairer and Keefe, 2005), these data suggest that the SFOAE I/O function is not simply related to the BM I/O function.

In contrast, using a method originally described in Epstein *et al.* (2004), Epstein and Florentine (2005) reported a mean slope of click-evoked (CE) OAE and tone-burst evoked (TB) OAE I/O functions of 0.19 dB/dB in human ears near 1 kHz, which they concluded were similar to slopes of the underlying BM I/O function. CEOAEs and SFOAEs are thought to be generated by the same underlying mechanism in the human cochlea (Konrad-Martin and Keefe, 2003, 2005; Kalluri and Shera, 2007), so that the difference between the OAE I/O function slopes (0.19 and 0.44–0.61, respectively) is problematical. Epstein *et al.* (2004) stated that the TBOAE was "measured within a 21.48 ms window beginning 13–20 ms after the onset of the tone burst." However, the SFOAE (and thus CEOAE or TBOAE) delay at 1 kHz is approximately 11 ms at low levels, and decreases with increasing stimulus level (Schairer *et al.*, 2006). Thus, it would appear that the onsets of windows used by Epstein *et al.* were later than the latency of the CEOAE energy, so that most of the OAE energy occurred prior to the window onset. Since latency decreases with increasing stimulus level, proportionally less of the TBOAE energy would have been contained within their analysis window at higher stimulus levels. That might explain why their reported I/O function slope was so small. The present compressive slope derived from on-frequency and off-frequency GOS appears more closely related to the underlying BM mechanics.

D. Tip-to-tail difference of the STC and cochlear gain

The tip-to-tail difference of each SFOAE STC was calculated as the level difference between the on-frequency GOS threshold relative to the off-frequency GOS threshold at each f_p. The resulting group mean ±1 SE of the tip-to-tail difference is plotted in Fig. 12. Tip-to-tail differences de-

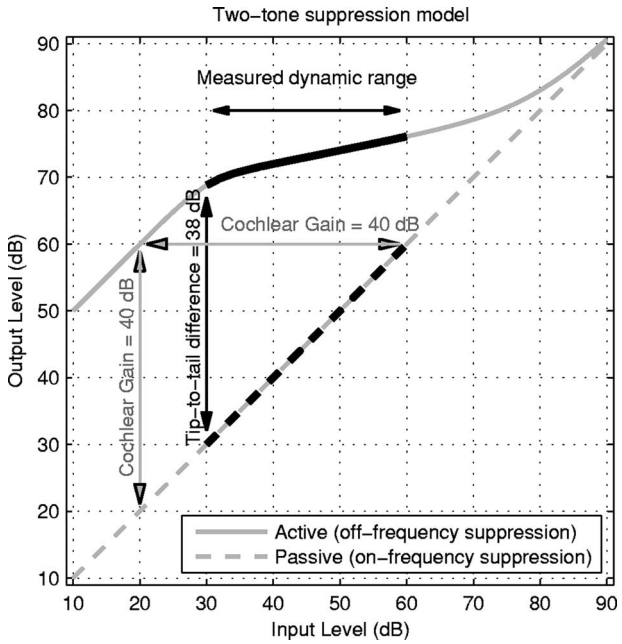


FIG. 13. Model of a compressive and saturating nonlinear growth response in the cochlea.

creased with increasing L_p at each f_p, as expected from cochlear-amplifier theory. At the lower L_p's (30–40 dB SPL) for which the active mechanism is strongest, the tip-to-tail differences increased with increasing f_p from 1 to 4 kHz. At the lowest L_p=30 dB SPL, the mean tip-to-tail difference increased from 32 dB at 1 kHz to 45 dB at 4 kHz. The tip-to-tail differences at f_p=4 kHz at $L_p \geq 40$ dB SPL were underestimated due to the ceiling effect illustrated in Fig. 10, so that the conclusion derived from Fig. 12 that cochlear gain is higher at 4 kHz than other frequencies is further strengthened. The tip-to-tail differences at f_p=0.5 kHz were difficult to interpret because of higher variability. The tip-to-tail differences at 8 kHz appeared reduced relative to results at 4 kHz at L_p of 50 and 60 dB SPL. These 8 kHz data are further discussed below.

Two-tone suppression processes at the BM and neural levels have been explained by a positive feedback process that is linear at low stimulus levels and saturates at high stimulus levels (Zwicker, 1979). In particular, the saturating nonlinearity of OHC receptor currents as a stage in mechano-electrical transduction is mainly responsible for two-tone suppression in the mammmalian cochlea (Geisler *et al.*, 1990). Full saturation of this nonlinearity at high suppressor levels results in a passive, or unity-gain, cochlea. A simple model of a saturating nonlinear growth of cochlear response is shown in Fig. 13 by the solid gray curve, such that the input and output levels (in dB) are scaled in arbitrary units. The plot represents the BM input/output (I/O) function recorded at the best-frequency place of the input probe tone. The I/O function has linear growth at low input levels, compressive growth at moderate levels, and linear growth at input levels sufficiently high that the active gain mechanism has saturated. The onset of compressive growth in this example occurs at an input level of 30 dB. The corresponding BM I/O function for a passive cochlea, i.e., one without the active mechanism providing gain at low levels, is shown by the dashed curve in Fig. 13. The horizontal level difference between the low-level linear-growth region of the active model and passive model is one definition of cochlear gain described earlier. The cochlear gain in this example is 40 dB.

In the two-tone suppression model of Geisler *et al.*, the active cochlear response in the absence of suppression is represented by the solid gray curve, and the fully suppressed cochlear response is represented by the dashed gray curve. Because the passive I/O function has a slope of one, the cochlear gain may be measured equally well by the vertical level difference between the active and passive I/O functions at an input level for which the active cochlear model is in the low-level linear-growth regime. This is shown in Fig. 13 by the output level difference at an input level of 20 dB, for which the cochlear gain is again 40 dB. When these I/O functions are compared at the input level corresponding to the onset of compression (30 dB input level in this example), the vertical level difference is 38 dB. Thus, the level difference between the unsuppressed and fully suppressed cochlear responses at the onset of compression is a slight underestimate of cochlear gain.

Suppose that the unsuppressed and fully suppressed responses are measured over a limited range of input levels from 30 to 60 dB, as shown by the black lines in Fig. 13. The unsuppressed and fully suppressed model responses are similar to the off-frequency and on-frequency GOS responses, respectively, in Fig. 10. The link between this model of BM nonlinearity and the present SFOAE measurements is the hypothesis that the SFOAE at f_p is generated as a result of two-tone suppression near the tonotopic place of f_p. The tip-to-tail difference in the SFOAE suppression response thus serves as a conservative measure of cochlear gain because the tip-to-tail difference is analogous to the vertical difference in I/O functions at the onset of compression.

The ceiling effect at f_p=4 kHz (see Fig. 10) means that tip-to-tail differences at $L_p \geq 40$ dB SPL were likely even larger than those shown in Fig. 12 (at f_p=4 kHz). This suggests that the cochlear gains were also larger in these stimulus conditions. While data are limited at f_p=8 kHz to higher L_p levels, it appears that the tip-to-tail differences and, thus, the cochlear gain is reduced at 8 kHz compared to 4 kHz.

Our results suggest that cochlear-amplifier gain in the human cochlea increases with increasing f_p in the range from 1 to 4 kHz. This is qualitatively similar to single-unit neural data described in the Introduction. This frequency trend was present for higher L_p of 50 and 60 dB SPL, but possibly not at L_p=70 dB SPL. The tip-to-tail difference was reduced at 8 kHz for L_p of 50 and 60 dB SPL, but there is lower confidence in these data because of increased measurement-system distortion at 8 kHz. For example, in Fig. 10, the slope of the off-frequency GOS responses was negative, indicative of an artifact such as distortion or intermittent noise, in a larger proportion of individual ears at f_p=8 kHz than at lower f_p. Of course, the absolute number of individual ears with valid data at f_p=8 kHz was also small.

Relative to the tail region where the BM mechanics are linear, the fact that the greatest amount of suppression near the tip occurred at the lowest L_p levels is consistent with BM

mechanical measurements in chinchilla (Ruggero *et al.*, 1992) and in neural measurements in cat (Delgutte, 1990). The maximum tip-to-tail ratios in the neural measurements were approximately 40–50 dB as compared to a range of 30–45 dB in the SFOAE measurements at the lowest L_p =30 dB SPL. This relation between two-tone suppression and cochlear gain has also been used to account for growth-of-masking responses in auditory-nerve fibers in cat (Pang and Guinan, 1997). The mean tip-to-tail difference in studies of DPOAE suppression in human ears lies in the range of 42–45 dB at a stimulus frequency f_2 of 4 kHz and stimulus level of L_2=20 dB SPL (Gorga *et al.*, 2002, 2003), which is similar to the present mean measurement of 45 dB at f_p =4 kHz.

As stated earlier, results were reported using a SFOAE decrement criterion of 10.7 dB, but were also analyzed using a decrement criterion of 6 dB, which corresponded to a greater amount of suppression. The resulting STCs generated using the two criteria were qualitatively similar, and the quantitative differences were in the expected directions. Using the decrement criterion of 6 dB, the STC tip was slightly elevated (i.e., larger L_s) and the tip-to-tail difference was slightly smaller. For example, the tip-to-tail difference was, on average, 2.3 dB higher across frequency and level using the decrement criterion of 10.7 dB than for the criterion of 6 dB. At the lowest probe level (30 dB SPL) for which there were data using both criteria, the tip-to-tail difference was 3.4 dB higher with the 10.7 dB criterion. It was possible to measure the tip-to-tail difference at L_p of 20 dB SPL for f_p of 2 and 4 kHz, but only using the 6 dB decrement criterion. These tip-to-tail differences were 41 dB at 2 kHz, and 47 dB at 4 kHz. This overall pattern of results was consistent with suppression acting in the tonotopic region of the probe tone at f_p.

E. Low-frequency slope of SFOAE STCs

Brass and Kemp (1993) predicted that the inverse of the SFOAE STC at low-side suppression frequencies was approximately equal to the traveling-wave envelope shape on the BM. This was based on the assumption that most of the low-side suppression occurred near the peak of the traveling-wave envelope and that the traveling wave had an approximately constant shape over the range of f_s. The long-wavelength theory of cochlear mechanics predicts that the traveling-wave envelope increases at 6 dB/octave at low frequencies (Lighthill, 1991). A previous analysis of SFOAE STCs showed a steeper traveling-wave envelope than 6 dB/octave, but the lowest f_s was only 0.6 octaves below f_p (Brass and Kemp, 1993). Brass and Kemp commented that f_s would need to be lower, probably so that the f_s peak would lie outside the region of cochlear amplification near the f_p place.

The results of the present study included SFOAE STC data at lower f_s/f_p than in the study of Brass and Kemp, which enabled an examination of this method to estimate the traveling-wave envelope. The low-frequency tail of the SFOAE STC was selected for each subject for f_s/f_p in the range from 0.25 up to 0.61. The upper cutoff at 0.61 was to exclude f_s values close to the tip region of the STC. The STC was averaged over the low-side suppressor range, with the restriction that individual STCs asymptoting at the highest L_s were excluded. The need for such exclusion is evident at 4 kHz in individual STCs (e.g., the STC in Fig. 5) and at f_p of 0.5 kHz in the mean STC (e.g., Fig. 6), because the resulting slope would otherwise be zero. Restricting attention to $f_p<4$ kHz, the slopes at f_p=0.5, 1 and 2 kHz were all in the range of 5–7 dB/dB for moderate L_p values of 40–50 dB SPL. These results support the hypothesis that the inverse of the STC at low-side suppression frequencies is approximately equal to the traveling-wave envelope shape (Brass and Kemp, 1993). A confounding factor is the frequency dependence of middle-ear transmission, which would be expected to produce deviations from this prediction.

F. SFOAE STC phase effects

In mechanical BM measurements in chinchilla, Ruggero *et al.* (1992) reported "small-to-moderate" phase lags of above-CF suppressors relative to the probe near CF as a function of increasing suppressor level. A constant SFOAE phase in Fig. 7 means that the latency of the suppressed SFOAE was appproximately independent of f_s. The variability in the SFOAE phase was larger for high-side than low-side suppressors, and no particular level effect was observed. This is evidence that the place of generation of the SFOAE residual did not vary as f_s varied, and supports the hypothesis that the SFOAE residual was generated on the BM in the region of the tonotopic place of f_p. For the purposes of the present study, this is an important finding because this lack of variation was assumed to establish a link beween SFOAE tip-to-tail difference and the gain of the cochlear amplifier.

This constant SFOAE phase is similar to low-side SFOAE suppression measured in a cat (Guinan, 1990). The high-side SFOAE suppression in cat showed that SFOAE phase varied rapidly, e.g., by 1.5 periods as f_s varied over one octave (for f_p=1.5 kHz, L_p=41 dB SPL, L_s=60 dB SPL). Guinan interpreted this rapid phase variation as associated with SFOAE components at f_p originating from the cochlear region tuned to f_s. These responses in cat were measured up to f_s/f_p=6.7, whereas the present human measurements were limited to $f_s/f_p \leq 1.6$.

Our mean SFOAE STC phase responses were approximately constant to within the SEs shown in Fig. 7, but the trends showed more variation for high-side suppressors. For example, at f_p=4 kHz, the SFOAE STC phase was slightly negative at lower L_p and slightly positive at higher L_p. The individual-ear SFOAE residual phase responses plotted in each panel of Fig. 4 for particular values of L_s had increased variability at the largest f_s/f_p values at the highest L_p of 50–60 dB SPL.

The discrete change in phase slope in the Guinan cat data near f_s/f_p of 1.3–1.4 suggests a transition between a two-tone suppression mechanism acting near the f_p place on the BM at lower f_s and an alternative distortion mechanism acting near the f_s place, e.g., the mechanism described by Talmadge *et al.* (2000). When such a distortion mechanism is present, the labeling of responses as *probe* and *suppressor* is

incomplete and possibly misleading. It is possible that such an alternative mechanism may have influenced our human data at the highest f_s/f_p values, but the interpretation of the SFOAE phase data was limited by a relatively small SNR, which had the effect of increasing the SEs. The subject of SFOAEs measured in human ears at f_p in the presence of a second tone of much higher frequency has been a subject of recent interest (Shera *et al.*, 2004; Siegel *et al.*, 2004). The present study limited the maximum f_s/f_p, as well as the maximum stimulus levels, to ranges in which the two-tone suppression mechanism was expected, and subsequently confirmed, to be the dominant source mechanism for SFOAEs.

G. Implications

OAEs are useful clinically for detecting ears at risk for sensorineural hearing loss. The present results in normal-hearing adults suggest the potential use of SFOAE suppression to evaluate cochlear compression and gain associated with normal OHC function. While the complete research protocol took 10 h per subject, it is possible that cochlear compression and gain might be assessed using a significantly shorter protocol with only two suppressor frequencies, one slightly larger than f_p and another an octave below, combined with a restricted range of suppressor levels. A relatively short test to measure cochlear compression and cochlear gain associated with normal OHC functioning might have clinical benefits. An analogous test based on DPOAE suppression already exists, but interpretable results in normal-hearing and hearing-impaired subjects were obtained only for an f_2 of 4 kHz (Gorga *et al.*, 2003). The current study obtained results on SFOAE suppression at probe frequencies of 1, 2 and 4 kHz, with results in some ears obtained at 0.5 and 8 kHz. More research is needed to assess SFOAE suppression in subjects with OHC dysfunction.

V. CONCLUSIONS

SFOAEs provide a noninvasive measure of cochlear mechanics in adult human ears. Two-tone suppression of SFOAEs is similar to suppression effects reported in basilar-membrane measurements, neural recordings, and simultaneous behavioral masking in terms of the level and frequency dependence of STCs. The SFOAE STC was parametrized in terms of rate of growth of suppression, Q_{ERB}, the compressive slope of off-frequency growth-of-suppression responses, and tip-to-tail difference, which serves as a conservative measure of cochlear gain. The SFOAE parameter values were generally similar to the corresponding parameters in physiological studies of nonhuman mammalian cochlea and in psychoacoustical studies of normal-hearing humans.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. DC003784 with core support from NIH Grant No. DC004662.

APPENDIX: SFOAE MEASUREMENT PROCEDURES

This Appendix describes details of the SFOAE measurements. Each stimulus frequency was constrained to have an integral number of periods within the elementary buffer length of $N=512$ samples, so that the nominal frequencies differed from actual test frequencies. For example, the nominal f_p of 0.5 kHz was a tone generated with a frequency of 0.517 kHz.

DAC1 and DAC2 denote channels 1 and 2, respectively, of the DACs on the sound card, with each channel driving one probe receiver. Different data acquisition procedures were used for f_p at 0.5 and 1 kHz than for f_p at 2, 4 and 8 kHz because of the longer SFOAE latencies at lower f_p. A stimulus block was constructed of N_s elementary buffers with the f_p stimulus presented through DAC1 (with zeros presented through DAC2), N_s buffers with f_s presented through DAC2 (with zeros presented through DAC1), and N_s buffers with f_p presented through DAC1 and f_s presented through DAC2. The stimulus block thus contained $3N_s$ elementary buffers, meaning that the signal recorded through the ADC comprised three N_s buffers.

An acoustic discontinuity was created each time the stimulus changed (from f_p alone, to f_s alone, to their joint presentation, and so forth). In addition, a system delay of the ADC relative to the DACs, which included effects of digital filters in the converters, meant that the initial response buffer for f_s contained some of the response from f_p, and, similarly, for the other discontinuities.

These artifacts were eliminated by discarding the first M_s elementary buffers of each block of N_s elementary buffers ($M_s<N_s$). The number of saved responses associated with each stimulus block was N_s-M_s for each response type (f_p alone, f_s alone, f_p and f_s jointly presented). For low probe frequencies, the sizes $N_s=7$ and $M_s=2$ were chosen, so that five buffers were saved for each response type. This resulted in an effective waiting time of $M_s=2$ elementary buffers, or 46 ms, between saved buffers. The transient OAE response associated with any acoustic discontinuity would have sufficient time to decay with respect to the continuous, synchronous-evoked response that was the signal of interest. For high probe frequencies, the sizes $N_s=5$ and $M_s=1$ were chosen, so that the number of saved buffers for each response type was $N_s-M_s=4$ buffers. The 23 ms waiting time associated with discarding one elementary buffer was sufficiently long with respect to system delays and OAE latencies at these higher frequencies to ensure that the response in the last four buffers was stationary (system delays were shorter than OAE latencies). Pilot data analyzed in Schairer *et al.* (2003) and additional new pilot data confirmed that these choices of numbers of discarded buffers gave substantively identical results when compared to tests with larger numbers of discarded buffers.

One advantage of interpolating between the different stimulus types during data collection is that the effects of slowly varying noise signals in the environment tended to cancel out, e.g., any effect of slowly varying static pressure within the middle-ear cavities.

Data collection continued until R sets of responses were acquired, with $R=15$ for low probe frequencies and $R=16$ for high probe frequencies. The time to collect data at particular stimulus condition (i.e., for fixed f_p, L_p, f_s, L_s) was $R*N_s*3*N/22050$, or 7.3 s at low probe frequencies and 5.6 s at high probe frequencies. The number of independent buffers saved for each stimulus type was $R*(N_s-M_s)$, or 75 buffers for low probe frequencies and 60 buffers for high probe frequencies.

Allen, J. B., and Fahey, P. (**1992**). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," J. Acoust. Soc. Am. **92**, 178–188.

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). "Properties of two-tone inhibition in primary auditory neurones," J. Physiol. (London) **212**, 593–609.

Backus, B. C., and Guinan, Jr., J. J. (**2006**). "Time-course of the human medial olivocochlear reflex," J. Acoust. Soc. Am. **119**, 2889–2904.

Brass, D., and Kemp, D. T. (**1993**). "Suppression of stimulus frequency otoacoustic emissions, J. Acoust. Soc. Am. **93**, 920–939.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). "Evoked mechanical responses of isolated cochlear outer hair cells," Science **227**, 194–196.

Davis, H. (**1983**). "An active process in cochlear mechanics," Hear. Res. **9**, 79–90.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (**2005**). "The Allen–Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1220–1226.

Delgutte, B. (**1990**). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Epstein, M., Buus, S., and Florentine, M. (**2004**). "The effects of window delay, delinearization, and frequency on tone-burst otoacoustic emission input/output measurements," J. Acoust. Soc. Am. **116**, 1160–1167.

Epstein, M., and Florentine, M. (**2005**). "Inferring basilar-membrane motion from tone-burst otoacoustic emissions and psychoacoustic measurements," J. Acoust. Soc. Am. **117**, 263–274.

Geisler, C. D., Yates, G. K., Patuzzi, R. B., and Johnstone, B. M. (**1990**). "Saturation of outer hair cell receptor currents causes two-tone suppression," Hear. Res. **44**, 241–256.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gold, T. (**1948**). "Hearing. II. The physical basis of the action of the cochlea," Proc. R. Soc. London, Ser. B **135**, 492–498.

Gorga, M. P., Neely, S. T., Dierking, D., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. (**2003**). "Distortion product otoacoustic emission tuning curves in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **114**, 263–278.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Konrad-Martin, D. (**2002**). "The use of distortion product otoacoustic emission suppression as an estimate of response growth," J. Acoust. Soc. Am. **111**, 271–284.

Guinan, Jr., J. J. (**1990**). "Changes in stimulus frequency otoacoustic emissions produced by two-tone suppression and efferent stimulation in cat," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. Geisler, J. Matthews, M. Ruggero, and C. Steele (Springer-Verlag, Madison, WI), pp. 170–177.

Guinan, Jr., J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). "Medial olivocochear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," J. Assoc. Res. Otolaryngol. **4**, 521–540.

Harris, F.P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., and Martin, G. K. (**1989**). "Acoustic distortion products in humans: Systematic changes in amplitude as a function of f_2/f_1 ratio," J. Acoust. Soc. Am. **85**, 220–229.

Kalluri, R., and Shera, C. (**2007**). "Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions," J. Acoust. Soc. Am. **121**, 2097–2110.

Keefe, D. H. (**1998**). "Double-evoked otoacoustic emissions: I, Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.

Kemp, D. T. (**1979**). "The evoked cochlear mechanical response and the auditory microstructure—Evidence for a new element in cochlear mechanics," in *Models of the Auditory System and Related Signal Processing Techniques*, edited by M. Hoke and E. de Boer (Scand. Audiol. Suppl., Stockholm), Vol. **9**, pp. 35–47.

Kemp, D. T. (**1986**). "Otoacoustic emissions, traveling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Kemp, D. T., and Brown, A. (**1983a**). "A comparison of mechanical nonlinearities in the cochleae of man and gerbil from ear canal measurements," in *Hearing—Physiological Basis and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin), pp. 82–88.

Kemp, D. T., and Brown, A. (**1983b**). "An integrated view of cochlear mechanical nonlinearities observable from the ear canal," in *Cochlear Mechanics*, edited by E. deBoer and M. A. Viergever (Delft University Press, Delft).

Kemp, D. T., and Chum, R. A. (**1980**). "Observations on the generator mechanism of stimulus frequency acoustic emissions–two tone suppression," in *Psychophysical, Physiological and Behavioral Studies in Hearing*, edited by E. deBoer and M. A. Viergever (Delft University Press, Delft), pp. 34–41.

Konrad-Martin, D., and Keefe, D. H. (**2003**). "Time-frequency analyses of transient-evoked stimulus-frequency and distortion-product otoacoustic emissions: Testing cochlear model predictions," J. Acoust. Soc. Am. **114**, 2021–2043.

Konrad-Martin, D., and Keefe, D. H. (**2005**). "Transient-evoked stimulus-frequency and distortion-product otoacoustic emissions in normal and impaired ears," J. Acoust. Soc. Am. **117**, 3799–3815.

Liberman, M. C. (**1982**). "Single-neuron labeling in the cat auditory nerve," Science **216**, 1239–1241.

Lighthill, J. (**1991**). "Biomechanics of hearing sensitivity," ASME J. Vib., Acoust., Stress, Reliab. Des. **113**, 1–13.

Lukashkin, A. N., and Russell, I. J. (**1998**). "A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer," J. Acoust. Soc. Am. **103**, 973–980.

Mills, D. (**1998**). "Interpretation of distortion product otoacoustic emission measurements. II. Estimating tuning characteristics using three stimulus tones," J. Acoust. Soc. Am. **103**, 507–523.

Moore, B. C. J., and Oxenham, A. J. (**1998**). "Psychoacoustic consequences of compression in the peripheral auditory system," Psychol. Rev. **105**, 108–124.

Mountain, D. C., Hubbard, A. E., and McMillen, T. A. (**1983**). "Electromechanical processes in the cochlea," in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever (Delft University Press, Delft), pp. 119–126.

Neely, S. T., and Kim, D. O. (**1983**). "An active cochlear model showing sharp tuning and high sensitivity," Hear. Res. **9**, 123–130.

Oxenham, A. J., and Bacon, S. P. (**2004**). "Psychophysical manifestations of compression: Normal-hearing listeners," in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 62–106.

Oxenham, A. J., and Plack, C. J. (**1998**). "Suppression and the upward spread of masking," J. Acoust. Soc. Am. **104**, 3500–3510.

Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554, erratum in: "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **5**, 459 (**2004**).

Pang, X. D., and Guinan, Jr., J. J. (**1997**). "Growth rate of simultaneous masking in cat auditory-nerve fibers: Relationship to the growth of basilar-membrane motion and the origin of two-tone suppression," J. Acoust. Soc. Am. **102**, 3564–3575.

Patuzzi, R. (**1996**). "Cochlear micromechanics and macromechanics," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 186–257.

Pienkowski, M., and Kunov, H. (**2001**). "Suppression of distortion product otoacoustic emissions and hearing thresholds," J. Acoust. Soc. Am. **109**, 1496–1502.

Ren, T., and Nuttall, A. L. (**2006**). "Cochlear compression wave: An implication of the Allen–Fahey experiment," J. Acoust. Soc. Am. **119**, 1940–1942.

Rhode, W. S. (**2007**). "Mutual suppression in the 6 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **121**, 2805–2818.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A., and Rich, N. C. (**1991**). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," J. Neurosci. **11**, 1057–1067.

Ruggero, M. A., Robles, L., and Rich, N. C. (**1992**). "Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-

nerve rate suppression," J. Neurophysiol. **68**, 1087–1099.
Sachs, M. B., and Kiang, N. Y. S. (**1968**). "Two-tone inhibition in auditory nerve fibers," J. Acoust. Soc. Am. **43**, 1120–1128.
Schairer, K. S., Ellison, J., Fitzpatrick, D. F., and Keefe, D. H. (**2006**). "Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear and middle-ear mechanics in human ears," J. Acoust. Soc. Am. **120**, 901–914.
Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**). "Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears," J. Acoust. Soc. Am. **114**, 944–966.
Schairer, K. S., and Keefe, D. H. (**2005**). "Simultaneous recording of stimulus-frequency and distortion-product otoacoustic emission input-output functions in human ears," J. Acoust. Soc. Am. **117**, 818–832.
Sewell, W. F. (**1984**). "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats," Hear. Res. **14**, 305–314.
Shera, C. A., and Guinan, Jr., J. J. (**1999**). "Evoked otoacoustic emissions arise by fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., Guinan, Jr., J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shera, C. A., and Guinan, Jr., J. J. (**2007**). "Cochlear traveling-wave amplification, suppression, and beamforming probed using noninvasive calibration of intracochlear distortion sources," J. Acoust. Soc. Am. **121**, 1003–1016.
Shera, C. A., Tubis, A., Talmadge, C.L., de Boer, E., Fahey, P. F., and Guinan, Jr., J. J. (**2007**). "Allen–Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.
Shera, C. A., Tubis, A., Talmadge, C. L., and Guinan, Jr., J. J. (**2004**). "The dual effect of 'suppressor' tones on stimulus-frequency otoacoustic emissions," in *Abstracts of the 27th Midwinter Research Meeting of the ARO* (Association for Research in Otolaryngology, Des Moines), p. 181.
Siegel, J. H., Cerka, A. J., Temchin, A. N., and Ruggero, M. R. (**2004**). "Similar two-tone suppression patterns in SFOAEs and the cochlear microphonics indicate comparable spatial summation of underlying generators," in *Abstracts of the 27th Midwinter Research Meeting of the ARO* (Association for Research in Otolaryngology, Des Moines), p. 181.
Siegel, J. H., Temchin, A. N., and Ruggero, M. A. (**2003**). "Empirical estimates of the spatial origin of stimulus-frequency otoacoustic emissions," in *Abstracts of the 26th Midwinter Research Meeting of the ARO* (Association for Research in Otolaryngology, Des Moines), p. 679(A).
Talmadge, C. L., Long, G. R., Tubis, A., and Tong, C. (**2000**). "Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure," J. Acoust. Soc. Am. **108**, 2911–2932.
Weiss, T. F., and Leong, R. (**1985**). "A model for signal transmission in an ear having hair cells with free-standing stereocilia. IV. Mechanoelectric transduction stage," Hear. Res. **20**, 175–195.
Wilson, J. P. (**1980**). "Evidence for cochlear origin for acoustic re-emissions, threshold fine structure, and tonal tinnitus," Hear. Res. **2**, 233–252.
Yasin, I., and Plack, C. J. (**2007**). "The effects of low- and high-frequency suppressors on psychophysical estimates of basilar-membrane compression and gain," J. Acoust. Soc. Am. **121**, 2832–2841.
Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hear. Res. **45**, 203–220.
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. (**1979**). "A model describing nonlinearities in hearing by active processes with saturation at 40 dB," Biol. Cybern. **35**, 243–250.
Zwicker, E. (**1986**). "Suppression and $(2f_1-f_2)$-difference tones in a nonlinear cochlear preprocessing model with active feedback," J. Acoust. Soc. Am. **80**, 163–176.
Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different oto-acoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

Comparison between otoacoustic and auditory brainstem response latencies supports slow backward propagation of otoacoustic emissions

Arturo Moleti[a)]
Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

Renata Sisto[b)]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

(Received 15 November 2007; revised 23 December 2007; accepted 28 December 2007)

Experimental measurements of the latency of transient evoked otoacoustic emission and auditory brainstem responses are compared, to discriminate between different cochlear models for the backward acoustic propagation of otoacoustic emissions. In most transmission-line cochlear models otoacoustic emissions propagate towards the base as a slow transverse traveling wave, whereas other models assume fast backward propagation via longitudinal compression waves in the fluid. Recently, sensitive measurements of the basilar membrane motion have cast serious doubts on the existence of slow backward traveling waves associated with distortion product otoacoustic emissions [He *et al.*, Hear. Res. **228**, 112–122 (2007)]. On the other hand, recent analyses of "Allen–Fahey" experiments suggest instead that the slow mechanism transports most of the otoacoustic energy [Shera *et al.*, J. Acoust. Soc. Am. **122**, 1564–1575 (2007)]. The two models can also be discriminated by comparing accurate estimates of the otoacoustic emission latency and of the auditory brainstem response latency. In this study, this comparison is done using human data, partly original, and partly from the literature. The results are inconsistent with fast otoacoustic propagation, and suggest that slow traveling waves on the basilar membrane are indeed the main mechanism for the backward propagation of the otoacoustic energy.
 [DOI: 10.1121/1.2836781]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Ri [BLM] Pages: 1495–1503

I. INTRODUCTION

Otoacoustic emissions (OAEs) are the by-product of the active filter mechanism located in the inner ear, which provides both low auditory threshold and good frequency discrimination. OAEs can be recorded in the ear canal, either spontaneously or as a response to an acoustic stimulus (Probst *et al.*, 1991). Spontaneous OAEs (SOAEs) are recorded without any stimulus, TEOAEs (transient evoked OAEs) are measured using a broadband, short duration, acoustic transient stimulus, stimulus frequency OAEs (SFOAEs) are evoked by a pure tone, and distortion product OAEs (DPOAEs) are the response to two primary tones, recorded at a third frequency, which is a linear combination of the primary frequencies. From a theoretical point of view, the OAE energy can be transmitted back to the cochlear base by different mechanisms, and in this study we will try to use available experimental data to discriminate between them. Most theoretical cochlear models and strong experimental evidence support the idea that the acoustic stimulus propagates forward along the basilar membrane (BM) as a slow transverse traveling wave (TW). A family of transmission-line cochlear models (e.g., Furst and Lapid, 1988; Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera *et al.*, 2005) also predicts that, for each frequency component, a slow backward TW packet is generated near the resonant place. The sum of all these backward transverse waves is transmitted back to the ear canal and eventually recorded as an OAE. Other cochlear models assume that most of the otoacoustic energy propagates back via fast longitudinal compression waves in the fluid (Wilson, 1980; Vetesnik *et al.*, 2006). Compression-wave models are supported by recent measurements of the phase of the BM motion at the distortion product frequency (He *et al.*, 2007), which, repeating with greater accuracy similar previous experiments (Ren, 2004), seem to rule out the existence of a significant backward TW on the BM. The key result of He *et al.* (2007) is that the phase of the BM motion at the distortion product (DP) frequency $f_{DP}=2f_1-f_2$ increases (as expected for a forward wave) with the distance from the base x in the region basal to the cochlear place resonant at frequency f_2, where a backward TW at the DP frequency would be expected instead. The authors claim that this result implies that a TW at the DP frequency moves forward along the BM, whereas no backward TW exists, within the experimental sensitivity. The interpretation of these results may be not that straightforward, as suggested by Shera *et al.* (2007), because the experiment does not identify the cochlear position of the source of the measured DP BM motion. Moreover, one should account for the complex

[a)]Electronic mail: arturo.moleti@roma2.infn.it
[b)]Electronic mail: r.sisto@dil.ispesl.it

mechanism of DPOAE generation, which involves at least two different sources at different cochlear places, with different phase/frequency relations (e.g., Shera and Guinan, 1999). This complexity is generally assumed to be responsible for the observed DPOAE fine structure (Kalluri and Shera, 2001), which is associated with the interference between the two backward waves, generated at the f_2 and at the f_{DP} place. Recently, Shera *et al.* (2007) provided instead indirect evidence against the predominance of compression-wave OAEs in mammalians, by analyzing data of "Allen–Fahey" experiments available in the literature (Allen and Fahey, 1992; de Boer *et al.*, 2005; Shera and Guinan, 2007). Allen–Fahey experiments measure the DP otoacoustic response at fixed f_{DP} as a function of the ratio $r=f_2/f_1$, keeping constant the excitation level of the BM at $x(f_{DP})$. The experimentally observed sharp decrease of the "Allen–Fahey function" at close primary-frequency ratios ($r\rightarrow 1$) is qualitatively incompatible with the general predictions of compression-wave models (Shera *et al.*, 2007).

Experimental evidence supporting fast backward propagation of OAEs had also been shown by Siegel *et al.* (2005). Comparing the BM group delay to the SFOAE phase-gradient delay in chinchillas, they concluded that the SFOAE phase-gradient delay is of the same order, or even shorter, than the BM group delay. Assuming that the SFOAE delay is a measure of the physical round trip cochlear delay, it should have been instead equal to twice the BM group delay. This experimental evidence was therefore interpreted as a demonstration that the OAE backward propagation occurs through some fast mechanism, and that the coherent reflection filtering theory (Shera *et al.*, 2005) was flawed. This interpretation has been questioned by Shera *et al.* (2006), who showed that the expected SFOAE phase-gradient delay is not simply twice the BM delay, but significantly less, and that the presence of a second, place-fixed SFOAE source must be taken into account.

In this study, we will not enter in the discussion about the interpretation of the above-mentioned experiments. Rather, we will investigate another possible way to discriminate between fast and slow propagation mechanism for the backward transmission of OAEs, based on the available TEOAE and auditory brainstem response (ABR) estimates of cochlear transmission delays. This method is applied here to TEOAE latency, but could also be applied to the estimates of the cochlear transmission delay based on measurements of the DPOAE or SFOAE phase-gradient delay. This study focuses on TEOAEs because the interpretation of the SFOAE and DPOAE phase-gradient delays in terms of cochlear transmission delays is neither straightforward nor model independent (Sisto *et al.*, 2007).

The time delay between the transient stimulus and the TEOAE response is defined as the TEOAE latency (τ_{OAE}), and it is a function of frequency and stimulus level. Time–frequency techniques based either on the wavelet transform (WT) (e.g., Tognola *et al.*, 1997; Sisto and Moleti, 2002, 2007), or on matching pursuit (MP) algorithms (Jedrzejczak *et al.*, 2004, 2005; Notaro *et al.*, 2007) permit to get a measure of the latency of each frequency component of the response. Time–frequency analysis of the TEOAE wave forms provides a direct estimate of the physical delay between the transient stimulus and the OAE response, because the square of the wavelet coefficient amplitude at a given time and frequency measures the energy density within the corresponding "tile" in the time–frequency plane. Therefore, the time of the maximum of the wavelet coefficient of a given frequency is directly associated with the physical delay of the OAE response at that frequency, provided that the contributions from noise and linear artifact are sufficiently small.

The phase-gradient delay is a conceptually different OAE characteristic time, defined as the derivative of the phase with respect to the angular frequency, and can be measured for TEOAEs, SFOAEs, and DPOAEs. The characteristic times associated with the phase gradient depend not only on the transmission properties of the BM, but also on the nature of the OAE generation mechanisms. Place-fixed and wave-fixed mechanisms indeed produce different predictions about the frequency dependence of the phase of the OAE response. For TEOAEs, a close correspondence between time–frequency and phase-gradient latency estimates has been recently demonstrated by Sisto *et al.* (2007), suggesting that most of the TEOAE signal is generated by some place-fixed mechanism, with noteworthy exceptions, which will not be discussed here. For DPOAEs, the situation is further complicated because the stimulus is composed by two tones at frequencies f_1 and f_2, whereas the response is recorded at a third frequency, $f_{DP}=2f_1-f_2$. Moreover, the response is usually assumed to be the vector superposition of two waves coming from two different BM locations, generated by two different mechanisms, nonlinear distortion near $x(f_2)$ and linear reflection at $x(f_{DP})$. Even assuming that the first mechanism is dominant, which is probably true at very high stimulus levels, one should consider that the backward path of this DP component does not start from its tonotopic place, but from a more basal cochlear location. As a consequence, the delay associated with its backward path is neither equal to that of an SFOAE (or TEOAE) of frequency f_2, nor to that of one of frequency f_{DP}.

ABR are evoked electrical signals, consisting of a series of characteristic peaks, named waves I, II, III, etc., corresponding to activation of different sections of the auditory neural pathway. The latency of wave V, $\tau_{ABR\text{-}V}$, is that measured with the best accuracy, and it is made up of a cochlear contribution, which is dependent on frequency and stimulus level, and of a noncochlear contribution, which is approximately constant, $\tau_n \approx 5$ ms. Most of this delay corresponds to that between waves I and V, which has been proved to be independent on frequency and stimulus levels, $\tau_{ABR\text{-}I\text{-}V} \approx 4-4.5$ ms (Eggermont ant Don, 1980; Don *et al.*, 1993, Abdala and Folsom, 1995), whereas an additional synaptic delay of ≈ 0.8 ms is generally assumed.

The frequency dependence of the cochlear latency is obviously expected from the tonotopic structure of the BM (Greenwood, 1990), which implies that the lower frequency components of the stimulus must travel a longer path to reach their resonant place, so their latency is longer. The dependence on the stimulus level is associated with two well-known facts: cochlear tuning is lower at higher stimulus levels and the slowing down of each wave packet approach-

ing its resonant place is more relevant for high-quality-factor (high-Q) resonances. This fact introduces an additional dependence on frequency, as the cochlear tuning Q is also dependent on frequency, at a given stimulus level. A further complication arises from the interplay between the cochlear amplifier nonlinearity and the frequency dependence of the middle ear input/output transmission matrix, which implies that the stimulus level measured in the ear canal is converted into an effective stimulus level at the stapes through a frequency dependent function. As a consequence, experimental differences between cochlear tuning across subjects or frequency, at a given stimulus level (in the ear canal), may also be due to different middle ear forward transmission properties, which cause differences in the effective stimulus level at the stapes (Abdala and Keefe, 2006).

The comparison between the OAE and ABR latencies can provide information on the cochlear mechanisms that propagate the OAE energy backwards to the stapes. Regardless of their generation mechanism, if OAEs propagate backwards via slow transverse waves, their latency must be approximately twice the frequency dependent part of the acoustic brainstem response latency, which is an estimate of the cochlear forward latency. Instead, if most of the OAE energy were transmitted by fast waves in the fluid, the OAE latency should be approximately equal to the ABR estimate of the cochlear forward latency, because the delay associated with a backward propagation via compression waves in the fluid would be negligibly short. The two predictions are very different, and the accuracy of OAE and ABR delay estimates is amply sufficient to discriminate between them. Due to the dependence of the TW velocity on the level of BM excitation, the comparison must be made as a function of the stimulus level. In this study, we compare time–frequency estimates of the OAE latency in humans at different stimulus levels with ABR latency measurements from the literature. Previous TEOAE latency estimates from the literature (Tognola *et al.*, 1997; Jedrzejczak *et al.*, 2004, 2005) have been added to the comparison, also to show that the uncertainties in the TEOAE latency estimates associated with different data analysis procedures and arbitrary choices of the data analysts do not affect significantly the results.

Such a comparison had already been attempted by Neely *et al.* (1988) who showed, using tone bursts between 0.5 and 2 kHz, that the cochlear component of the ABR latency (identified with that dependent on frequency and stimulus level) was approximately equal to half the TEOAE latency, and by Tognola *et al.* (1997), who also obtained very similar results comparing their estimates of click-evoked OAE latency with previous ABR latency measurements. A similar comparison has also been done by Schoonoven *et al.* (2001), who compared DPOAE phase-gradient delays to compound action potential (CAP) and ABR estimates of the cochlear forward latency, τ_{CAP} and τ_{ABR}. DPOAE phase-gradient delays were measured with two different paradigms: τ_{DP1}, sweeping f_1 at fixed f_2 and τ_{DP2}, sweeping f_2 at fixed f_1. The experimental relation between τ_{DP1} and τ_{DP2} agrees with general predictions of transmission line models (Prijs *et al.*, 2000) for wave-fixed generations, whereas the ratio of 2 between τ_{DP1} and both τ_{CAP} and τ_{ABR} seems in agreement with slow propagation of backward TW along the BM. Much care is necessary in the analysis of these results because the DPOAEs stimulus level (65–55 dB SPL) was lower than that used for ABR and CAP measurements (98 dB pSPL), and because, as already pointed out, the DP phase-gradient delay is not straightforwardly associated with the cochlear transmission delays.

II. METHODS

TEOAE signals were measured in both ears of a sample of 14 young subjects (22–30 years, 8 males and 6 females) at seven different click-stimulus levels, from 60 to 90 dB pSPL. The ILO Echoport 292 (Otodynamics, Ltd.) system was used in the nonlinear acquisition mode. Part of these subjects had also been analyzed in a recent study on TEOAE latency and cochlear tuning (Sisto and Moleti, 2007).

The latency of each OAE component was evaluated by a time–frequency technique based on the wavelet transform, as described, e.g., in Sisto *et al.* (2007). The wavelet transform (Mallat, 1989) decomposes the wave form $s(t)$ is into a linear superposition of wavelets:

$$WT(\alpha,\tau) = \int s(t) h_{\alpha,\tau}(t) dt. \tag{1}$$

All wavelets are scaled and time-shifted copies of a "mother" wavelet $h_0(t)$, an oscillating function with zero mean:

$$h_{\alpha,\tau}(t) = \frac{1}{\sqrt{|\alpha|}} h_0\left(\frac{t-\tau}{\alpha}\right). \tag{2}$$

In this study, we have used the mother wavelet (Tognola *et al.*, 1997):

$$h_0(t) = \frac{\cos(20t)}{(1+t^4)}, \tag{3}$$

with t in milliseconds.

Each wavelet is localized both in time, by the value of the parameter τ, and in frequency, by the value of the scale parameter α, which is directly proportional to the time duration of the wavelet, and inversely proportional to its central frequency. Therefore the central frequency and the bandwidth of each wavelet are proportional to each other. At fixed α, the wavelet coefficient $WT(\alpha,\tau)$ represents the time evolution of the signal amplitude within the corresponding frequency band.

The continuous wavelet transform was integrated over 500 Hz bandwidths, centered at frequencies f_i from 1 to 4 kHz. The wavelet latency estimate for the jth ear in the ith frequency band, $\tau_j(f_i)$, is defined as the time when the band-averaged wavelet coefficient reaches its maximum absolute value. This time is that associated with the return time of the most intense energy pulse within the ith frequency band. The average of the latencies over all the ears was defined as the ear-averaged band latency, $\tau^*(f_i)$, including only the individual latencies from frequency bands where the TEOAE signal-to-noise ratio (SNR) was larger than unity. The data selection was necessary to avoid including in the estimate of $\tau^*(f_i)$ those ears whose latency estimates could be probably determined by noise. In this case the estimated

latency is a random number between 2.5 and 20 ms (the acquisition window width), which affects the estimate of $\tau^*(f_i)$, particularly at the higher frequencies, whose latency is lower. This selection rule does not protect against another systematic error coming from the presence of intense SOAEs, which also leads to individual latency estimates close to the center of the acquisition window, independently of frequency. Again, this systematic error is most important at the highest frequencies, and, therefore, much care must be used when studying neonate ears, which frequently show a significant number of intense high-frequency OAEs. Although the TEOAE response is recorded from 0 to 6 kHz, latency estimates below 1 kHz and above 4 kHz are not considered in this study, because they may be not fully reliable, mainly due to the finite acquisition window length. In fact, the typical latency of frequency components above 4 kHz is close to the onset time of the acquisition window (between 2.5 and 5 ms) and the latency of components around or below 500 Hz may be close to 20 ms. For this reason, when averaging over a sample of ears, low-frequency latency may be systematically underestimated and high-frequency latency overestimated.

Another possible definition of average latency, $\tau(f_i)$, is the time of the maximum of the average of the individual normalized wavelet coefficients (Sisto and Moleti, 2007). In this case no selection based on SNR, which could always bias the results, is applied. If both the noise level and the number of SOAEs are both low, as for the young adult subjects of this study, the two procedures yield similar results, as we will show in Sec. IV.

III. THEORETICAL MODELS

Transmission-line models of the cochlea predict forward propagation of the acoustic stimulus as a TW on the BM. The linearized equations in the frequency domain for a one-dimensional transmission line with constant series impedance (box model) can be written as

$$\frac{\partial^2 P_d(x,\omega)}{\partial x^2} + k^2 P_d(x,\omega) = 0, \tag{4}$$

where ω is the angular OAE frequency, k is the wave vector, P_d is the differential pressure applied to the BM, and x is the longitudinal coordinate measured on the BM from the base. The cochlear resonance frequency is related to the position x along the BM by the Greenwood map (Greenwood, 1990):

$$\omega(x) = \omega_1 + \omega_{\max} \exp(-k_\omega x), \tag{5}$$

where $\omega_{\max}=2\pi\cdot 20\,655$ rad/s, $\omega_1=2\pi\cdot 145$ rad/s, and $k_\omega=1.382$ cm^{-1}.

Following the acoustic stimulation used to evoke OAEs or ABRs, each frequency component ω of the stimulus propagates as a forward traveling wave packet, whose wave number k is a function of the angular frequency ω and of the longitudinal position x along the BM:

$$k^2 = \frac{k_0^2 \omega^2}{\Delta(x,\omega)}, \tag{6}$$

where $k_0=31$ cm^{-1} is a constant, dependent on geometrical and mechanical properties of the BM and of the cochlear duct. The analytical form of the denominator $\Delta(x,\omega)$ depends on the particular model. It may correspond to a very simple resonance with linear damping (Sisto and Moleti, 2002, 2007):

$$\Delta(x,\omega) = \omega_0^2(x) - \omega^2 + j\omega\Gamma(x), \tag{7}$$

or, as in more complex models, include delayed-stiffness (Zweig, 1991) terms (for the numerical values of these parameters see Talmadge *et al.*, 1998):

$$\Delta(x,\omega) = \omega_0^2(x) - \omega^2 + j\omega\Gamma(x) - \rho_f \omega_0^2(x) e^{j\psi_f \omega/\omega_0(x)} - \rho_s \omega_0^2(x) e^{j\psi_s \omega/\omega_0(x)}. \tag{8}$$

Each wave packet propagates at its group velocity, defined by

$$v_g^{-1} = \frac{\partial\, Re(k(x,\omega))}{\partial\omega}, \tag{9}$$

which, for a given frequency, is a function of the position x along the BM.

The analytical form of the function $\Delta(x,\omega)$ may include even more elaborate and nonlinear terms related to stiffness and/or damping, yet, for any reasonable model, it is always characterized by a tonotopically resonant form. The modulus of the wave vector squared of a wave packet of frequency ω increases approaching its resonant place $x(\omega)$, whereas its phase changes from zero to π. For high-Q resonances, both these changes are sharper. Two main physical effects correspond to these changes: The group velocity decreases approaching the resonant place because the relation between k and ω becomes steeper, and resonant absorption takes place, as the wave vector k gets a non-null imaginary part. The cochlear quality factor, or tuning, defined here as $Q=\omega/\Gamma$, is a function of frequency, evaluated near the corresponding tonotopic place.

The forward delay of a Fourier component of frequency ω of a transient stimulus is given by (Sisto and Moleti, 2002):

$$\tau_{fw}(\omega) = \int_0^{\hat{x}(\omega)} v_g^{-1}(x,\omega)dx = \int_0^{\hat{x}(\omega)} \frac{\partial\, Re(k(x,\omega))}{\partial\omega} dx. \tag{10}$$

This is the physical delay associated with the path of a traveling wave packet centered at frequency $f=\omega/2\pi$ propagating along the BM at its local group velocity, from the base to the cochlear place $\hat{x}(\omega)$ where OAE generation occurs, which is assumed to be close to the resonant place $x(\omega)$.

This forward delay contributes to both the ABR response delays and to the TEOAE delays that are experimentally measured. In the ABR case, the measured delay also includes a neural contribution, which for the wave V, amounts to $\tau_n \approx 5$ ms, quite independent of stimulus level and frequency (Eggermont and Don, 1980; Neely *et al.*, 1988; Don *et al.*, 1993; Donaldson and Ruth, 1993; Abdala and Folsom, 1995). In the TEOAE case, the measured delay

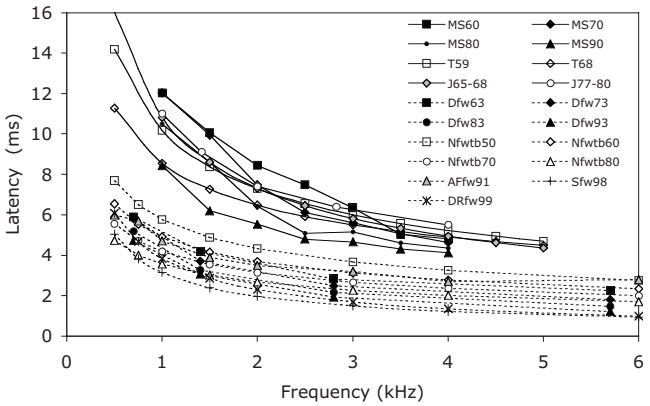


FIG. 1. Global comparison between the TEOAE latency measured in this study, by Tognola *et al.* (1997) and by Jedrzejczak *et al.* (2004, 2005), and the estimates of the cochlear forward latency derived from ABR wave V latency measurements found in the literature. Only the fit functions provided by the authors are displayed, except for the data of this study and for those of Don *et al.* (1993) and Abdala and Folsom (1995). The large vertical spread of the two data sets is due to the different stimulus level, varying between 60 and 90 dB pSPL for OAEs and between 63 and 99 dB pSPL for ABR). Nevertheless, it is quite evident that the OAE latency is systematically longer than the ABR forward latency.

includes instead the delay associated with the backward propagation of the acoustic energy from the OAE generation place to the base, along with a very small noncochlear contribution τ_{nc}. In previous studies (e.g., Sisto and Moleti, 2002) this contribution had been erroneously estimated of order 0.5 ms, whereas a more accurate analysis has shown that it is instead of order 0.1 ms, and so quite negligible at the level of sensitivity of our measurements.

TEOAE generation models consider several mechanisms, both linear and nonlinear, but a common feature is that TEOAEs are generated, for each frequency, near the place of maximum response, with the only exception of intermodulation distortion components. In fact, a necessary condition to get a significant backward acoustic emission is the strong activation of the BM, which is possible only close to the resonance. If this prerequisite is achieved, different models postulate different reflection mechanisms. Some mechanisms are associated with cochlear nonlinearity and/or roughness (Talmadge *et al.*, 1998; Shera *et al.*, 2005), others with the sharp change of transverse impedance near the resonance (Sisto and Moleti, 2005). Transmission line models generally assume that this energy is transmitted back to the stapes via a TW on the BM, implying therefore a strong symmetry between the forward and backward transmission delays. Other models suggest instead that the forward stimulus travels mainly along the BM as a transverse wave, but that the OAE energy is transmitted back through fast longitudinal compression waves in the fluid (Wilson 1980; Vetesnik *et al.*, 2006). Fast longitudinal waves consist of symmetric (between the two sides of the BM) pressure perturbations. They do not involve motion of the BM, and are therefore insensitive to its local resonant properties, whereas the transmission of the transverse slow wave is driven by differential pressure and, as noted earlier, its dynamics is strongly dependent on the local mechanical properties of the BM.

If OAEs were transmitted backwards by fast compression waves, the backward transmission would give no significant contribution to the TEOAE delay, and therefore:

$$\tau_{\mathrm{OAE}} = \begin{cases} 2\tau_{\mathrm{fw}} \approx 2(\tau_{\mathrm{ABR}-V} - \tau_n) & \text{for slow backward transverse waves} \\ \tau_{\mathrm{fw}} \approx \tau_{\mathrm{ABR}-V} - \tau_n & \text{for fast backward compression waves.} \end{cases} \tag{11}$$

Another important issue is the following: if the TEOAE latency were associated with the forward transmission delay only, the estimates of cochlear tuning based on the TEOAE latency should be corrected by a rather large factor. In fact, according, e.g., to Sisto and Moleti (2007), cochlear tuning would be given, in the two cases, by

$$Q(\omega) = \begin{cases} \dfrac{1}{\beta^2}\left(\dfrac{\omega k_\omega}{k_0}(\tau_{\mathrm{OAE}} - \tau_{\mathrm{nc}}) + \dfrac{2\omega}{\omega_{\max}}\right)^2 \approx \left(\dfrac{\omega k_\omega}{\beta k_0}\right)^2 \tau_{\mathrm{OAE}}^2 & \text{for slow waves} \\ \dfrac{1}{\beta^2}\left(\dfrac{2\omega k_\omega}{k_0}(\tau_{\mathrm{OAE}} - \tau_{\mathrm{nc}}) + \dfrac{2\omega}{\omega_{\max}}\right)^2 \approx 4\left(\dfrac{\omega k_\omega}{\beta k_0}\right)^2 \tau_{\mathrm{OAE}}^2 & \text{for fast waves,} \end{cases} \tag{12}$$

where β is a model-dependent dimensionless parameter of order unity (1.2). Although these tuning estimates are indirect and model dependent, this issue deserves to be added to the discussion, because an increase by a factor of 4 would lead to exceedingly high tuning estimates. This may be seen as a sign of a possible flaw in the simple model proposed by Sisto and Moleti (2007), or as a further suggestion that the fast-wave scenario does not agree with the cochlear phenomenology.

IV. RESULTS AND DISCUSSION

Figure 1 shows the global comparison between time–frequency estimates of the TEOAE latency (solid lines, with different symbols for different stimulus level intervals and authors: τ^* from this study; Tognola *et al.*, 1997; Jedrzejczak *et al.*, 2004, 2005) and the available ABR estimates of the cochlear forward latency (dotted lines, with different symbols for different stimulus level intervals and authors: Neely *et al.*, 1988; Don *et al.*, 1993; Donaldson and Ruth, 1993; Abdala and Folsom, 1995; Schoonhoven *et al.*, 2001), as a

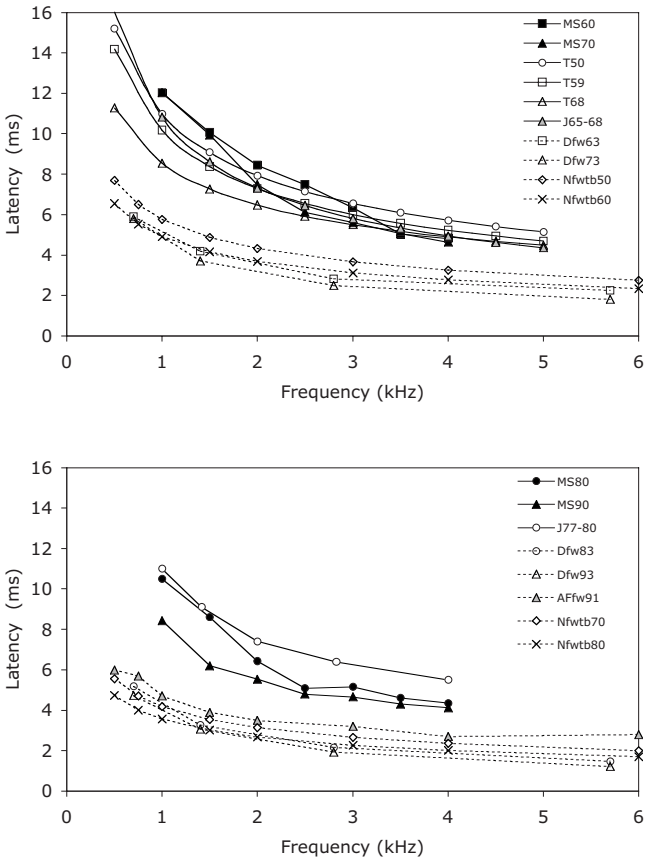


FIG. 2. Comparison between time–frequency TEOAE latencies and ABR cochlear forward latency estimates in different stimulus level ranges. The comparison is shown in the top panel for moderate stimulus levels (<75 dB pSPL), and in the bottom panel for higher stimulus levels (>80 dB pSPL). This figure shows better than Fig. 1 that the data are not compatible with the notion of fast OAE backward propagation, which would predict that the two latencies were approximately equal [see Eq. (1)].

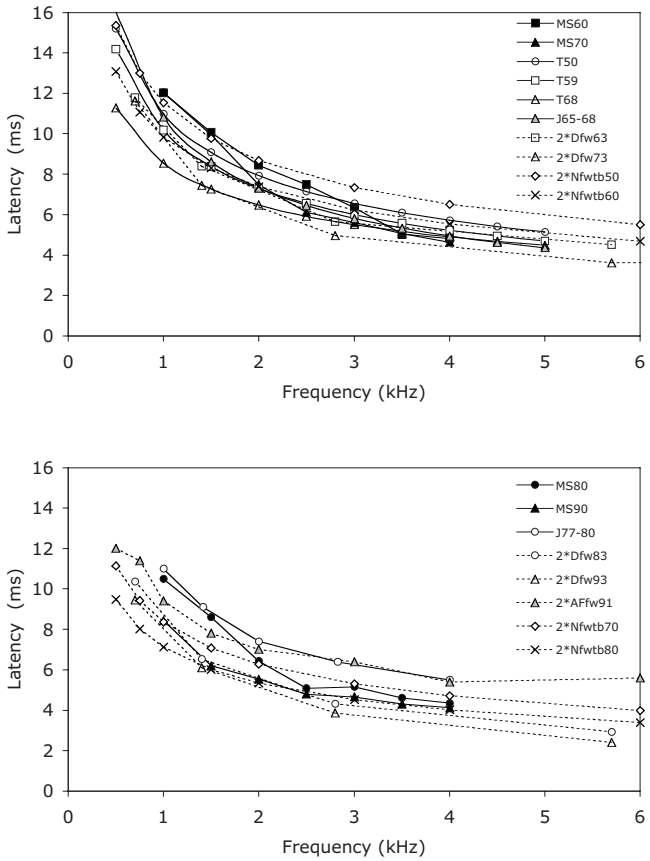


FIG. 3. Comparison between time–frequency TEOAE latencies and twice the ABR estimates of the cochlear forward latency. As in Fig. 2, this comparison is shown in the top panel for moderate stimulus levels (<75 dB pSPL), and in the bottom panel for higher stimulus levels (>80 dB pSPL). Although the vertical spread of the data is rather large, there is good agreement with the prediction of transmission-line models that assume slow OAE backward propagation as a TW along the BM. The fact that the TEOAE latency is approximately twice the ABR estimate of the forward latency, at any stimulus level, suggests indeed that the forward transmission of the stimulus and the backward transmission of the OAE rely on the same physical mechanism.

function of frequency and for different stimulus levels. Only the fit functions provided by the authors are displayed, except for the data of this study and for those of Don *et al.* (1993) and Abdala and Folsom (1995), who reported data tables with the results. Neely *et al.* (1988) used tone bursts of frequency-dependent duration, whereas all other experiments used clicks of duration $\Delta t \approx 100$ μs. The energy density in each frequency band is comparable if the click stimulus, expressed in dB pSPL is 10–20 dB higher (depending on the burst duration, which varied with frequency) than the tone burst level of the Neely *et al.* (1988) experiment, expressed in dB SPL. Due to the dependence of latency on the stimulus level, there is a large vertical spread in both sets of measurements. Recently, Ruggero and Temchin (2007) suggested that the estimates by Neely *et al.* (1988) should also be corrected for not having kept constant the burst onset-ramp duration across frequencies. Figure 1 is only meant to show that, with no exception, the time–frequency latency estimates are much longer than the estimates of the cochlear forward delay reported by ABR experiments. A more meaningful comparison, keeping into account the variation associated with the stimulus level, is shown in Fig. 2. In the top panel we show lower-stimulus level data and in the bottom panel higher-stimulus level data. It is quite evident that, independently of the stimulus level, the OAE latency data are not compatible with the ABR estimates of the cochlear forward latency, implying that a substantial contribution to the OAE delay comes from their backward propagation. In particular, as shown in Fig. 3, the OAE latency is in very good agreement with twice the ABR estimate of the forward cochlear latency. To get a rough quantitative estimate of the significance of this result, we interpolated all the available experimental data to get best estimates of the OAE and ABR latencies on a two-dimensional frequency-stimulus level 7×7 grid (between 1 and 4 kHz with 500 Hz steps and between 60 and 90 dB with 5 dB steps). We computed the ratio between the OAE latency and the ABR estimate of the cochlear forward latency for each of these 49 points, obtaining the distribution shown in Fig. 4, with an average ratio of 2.08 and a standard deviation of 0.19. This result is clearly compatible with a ratio of 2 between the two latencies and seems not compatible with any ratio close to unity, although we cannot give an estimate of the probability associated with the two hypotheses, because the experimental uncertainties associated with random and systematic errors (partly unknown to us, for the

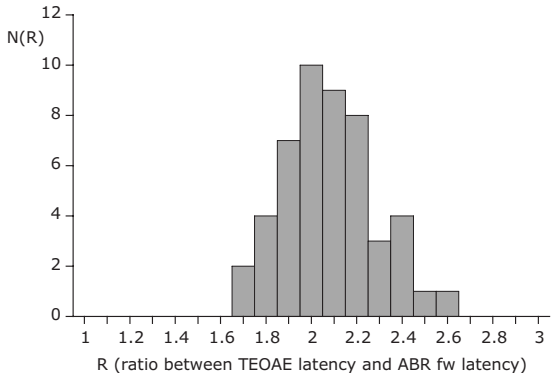


FIG. 4. Distribution of the ratios between the best estimates of the TEOAE latency and of the ABR estimate of the cochlear forward latency obtained by interpolating all the data of Fig. 1 on a 7 by 7 grid for stimulus levels between 60 and 90 dB pSPL (with steps of 5 dB) and for frequencies between 1 and 4 kHz (with steps of 500 Hz). The average of the distribution is 2.08, with a standard deviation of 0.19. This result is clearly not compatible with the prediction of equality between the two latencies implied by the hypothesis of fast OAE backward transmission, whereas it is fully compatible with the hypothesis of slow OAE backward transmission, with exact symmetry between the forward and backward delays.

ABR data) were not considered at all in the analysis shown in Fig. 4. This evidence suggests that a strong symmetry holds indeed between the forward and backward OAE propagation delays. The simplest interpretation is that the two mechanisms are the same, i.e., both the forward and the backward propagation of acoustic energy along the cochlea occur via slow traveling waves on the BM.

Both ABR and OAE latency estimates suffer from uncertainties associated with noise and with systematic bias dependent on the data analyst's choices. To understand to which extent some analyst's choices may affect the previous results, the TEOAE latency estimates obtained using two different definitions are compared in Fig. 5, for three different stimulus levels. Black symbols refer to the definition of average latency $\tau^*(f_i)$ as the average of the individual ear latencies, defined as the time of the maximum of the corresponding band wavelet coefficient (having selected only the bands with SNR$>$1). White symbols refer to the alternative definition of average latency $\tau(f_i)$ as the time of the maximum of the ear-averaged band wavelet coefficient. The difference is not negligible, also due to the small number of subjects, but it is typically of the same order of the standard error, and does not change at all the physical meaning of the results.

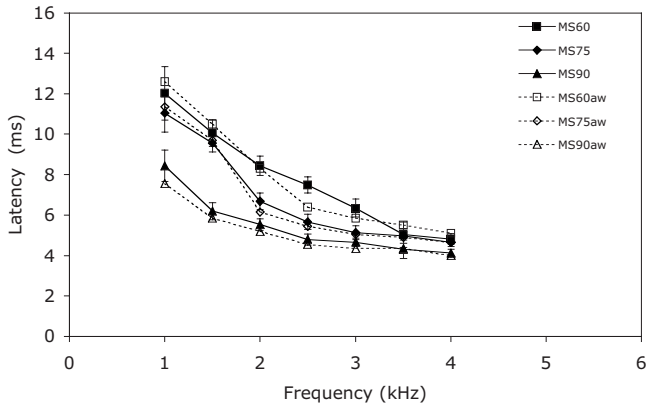


FIG. 5. Average TEOAE latency estimates obtained using two different definitions, for three different stimulus levels (60, 75, and 90 dB pSPL). Black symbols refer to $\tau^*(f_i)$, the average of the individual ear latencies, which are defined as the time of the maximum of the corresponding band wavelet coefficient (having selected only the bands with SNR$>$1). White symbols refer $\tau(f_i)$, the time of the maximum of the ear-averaged band wavelet coefficient. The difference is not negligible, but it is typically of the same order of the statistical standard error, and does not change the physical meaning of the results shown in Figs. 1–4.

As mentioned in Sec. II, and more amply discussed in Sisto and Moleti (2007) and in Notaro *et al.* (2007), both the wavelet and the MP estimates of the individual TEOAE latency are affected by noise, by the presence of SOAEs, of intermodulation distortion components, of multiple intracochlear reflections. The need for windowing the data to cancel the residual linear ringing artifact implies that the high-frequency latency estimates may begin to be systematically overestimated above 3–4 kHz, particularly at high stimulus levels.

ABR estimates of the cochlear forward latency obtained with the derived-band technique are generally referred to a frequency that is the arithmetic or the geometric mean frequency of the band. This choice may cause some bias, particularly for wide frequency bands (1/2 octave or octave), because it is not obvious that the observed delay of the neural response is associated with the center of the frequency band.

For the TEOAE band latency estimates, this kind of bias could be significantly different, because the OAE response is very frequency sensitive, and may drop rather sharply below 1 kHz and above 2–3 kHz. As a consequence, it is possible that, in the bands where the TEOAE response is a steep function of frequency, the measured latency is more representative of the higher-frequency edge of the band at low frequency and of its lower-frequency edge at high frequencies. This problem may be solved using narrow band analysis (as narrow as permitted by the intrinsic bandwidth of the wavelet analysis). It could be responsible for slightly underestimating the slope of the latency-frequency relation, but it should not affect significantly the average value of the TEOAE latency.

An uncertainty in the interpretation of the comparison is relative to the actual length of the ABR neural delay, which we have assumed to be constant and equal to 5 ms. The actual neural delay could be shorter, but never shorter than 4–4.5 ms, which is the experimental value of the difference between the wave I and wave V latency. Therefore the evaluation of the forward cochlear delay that we have made cannot be systematically underestimated by more than 0.5–1 ms. Such an underestimate of the forward delay would slightly weaken the conclusions against fast OAE backward transmission mechanisms.

All the above-mentioned experimental uncertainties suggest using some caution in the interpretation of the previous comparison between TEOAE and ABR latency. To reduce the uncertainties, it would be useful to design an experiment to measure both latencies on the same subjects with the same stimuli in the same laboratory. Nevertheless, all the data

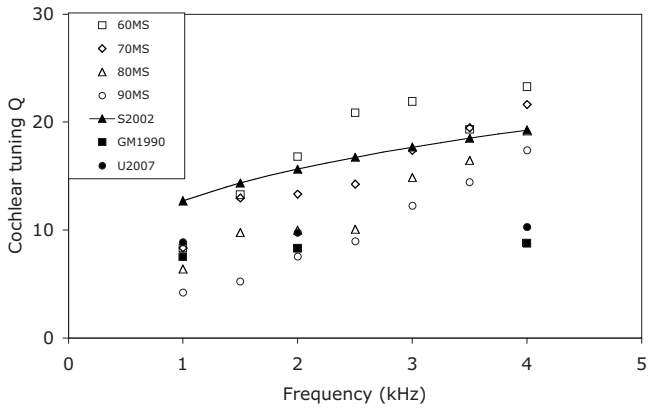


FIG. 6. Cochlear tuning estimates based on the experimental TEOAE latency data of this study, between 1 and 4 kHz, for different stimulus levels, obtained assuming slow backward propagation of OAEs (white symbols and solid lines). The tuning estimates obtained assuming fast backward propagation would be approximately four times higher. Other cochlear tuning estimates are shown for comparison, from Glasberg and Moore (1990), from Unoki *et al.* (2007) (behavioral, average among three different masking conditions) and from Shera *et al.* (2002) (average between behavioral estimate and estimates based on the SFOAE phase-gradient delay). The reasonably good agreement shown here implies that an increase by a factor of 4 of the TEOAE-based tuning estimates would make them fully incompatible with the other behavioral and SFOAE-based tuning estimates.

seem to be coherently (across frequency and stimulus level) supporting the idea that a strong symmetry holds between the forward and backward propagation of, respectively, the stimulus and the OAE response along the cochlea.

An additional issue may be added at this point of the discussion: In case of fast backward propagation of OAEs, the TEOAE-latency-based estimates of tuning reported by Moleti and Sisto (2003), Moleti *et al.* (2005), and by Sisto and Moleti (2007) should be increased by approximately a factor of 4, as shown in Eq. (12). Using the latency data of this study and assuming slow backward propagation, quality factors between 4 and 24 are found, between 1 and 4 kHz, for stimulus levels between 60 and 90 dB pSPL, increasing with frequency and decreasing with stimulus level, as shown in Fig. 6 (white symbols and solid lines). These tuning estimates are rather indirect and model dependent, yet they are not very different from other OAE-based (Shera *et al.*, 2002; Shera and Guinan, 2003) and psychoacoustic (Glasberg and Moore, 1990; Shera *et al.* 2002; Unoki *et al.*, 2007) estimates. The comparison suggests that the TEOAE latency-based tuning estimates shown in Fig. 6 (white symbols) could even be overestimated at high frequency, which is what would happen if the high-frequency latency were overestimated due to the systematic effect of the acquisition window. In case of fast backward propagation of OAEs, an increase of the cochlear tuning estimates by a factor of 4 would lead to unreasonably high values of the quality factor, of order 50–100. Of course, the simple cochlear model used here is not accurate enough to draw conclusions based on this observation alone, yet this consideration fits with the previously discussed, and much stronger, evidence suggesting that the backward OAE propagation is probably associated with a slow mechanism.

V. CONCLUSIONS

This study confirms that, at moderate to high stimulus levels, the OAE latency is approximately twice the ABR estimate of the forward cochlear latency, and therefore supports slow propagation of the otoacoustic energy as backward TWs on the BM. These results contradict the hypothesis that the OAE acoustic energy is transmitted back through some fast mechanism in the fluid, and support a view in which a symmetrical forward and backward propagation along the BM accounts for the traveling time of both acoustic stimulus and OAE response. Further experimental and theoretical efforts are necessary to reduce the experimental uncertainties, and to provide a coherent framework for the explanation of the results coming from BM motion and OAE studies.

Abdala, C., and Folsom, R. C. (**1995**) "Frequency contribution to the click-evoked auditory brain-stem response in human adults and infants," J. Acoust. Soc. Am. **97**, 2394–2404.

Abdala, C., and Keefe, D. (**2006**) "Effects of middle-ear immaturity on distortion product otoacoustic emission suppression tuning in infant ears," J. Acoust. Soc. Am. **120**, 3832–3842.

Allen, J. B., and Fahey, P. F. (**1992**). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," J. Acoust. Soc. Am. **92**, 178–188.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (**2005**). "The Allen-Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1260–1266.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1993**). "Gender differences in cochlear response time: an explanation for gender amplitude differences in the unmasked auditory brainstem response," J. Acoust. Soc. Am. **94**, 2135–2148.

Donaldson, G. S., and Ruth, R. A. (**1993**). "Derived band auditory brainstem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects," J. Acoust. Soc. Am. **93**, 940–951.

Eggermont, J. J., and Don, M. (**1980**). "Analysis of the click-evoked brainstem potentials in humans using high-pass noise masking. II. Effect of click intensity," J. Acoust. Soc. Am. **68**, 1671–1675.

Furst, M., and Lapid, M. (**1988**). "A cochlear model for acoustic emissions," J. Acoust. Soc. Am. **84**, 222–229.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Greenwood, D. D. (**1990**). "A cochlear frequency position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

He, W., Nuttall, A. L., and Ren, T. (**2007**) "Two-tone distortion at different longitudinal locations on the basilar membrane," Hear. Res. **228**, 112–122.

Jedrzejczak, W. W., Blinowska, K. J., and Konopka, W. (**2005**). "Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise," Hear. Res. **205**, 249–255.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emission components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.

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.

Mallat, S. G. (**1989**). "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693.

Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.

Moleti, A., Sisto, R., Tognola, G., Parazzini, M., Ravazzani, P., and Grandori, F. (**2005**) "Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates," J. Acoust. Soc. Am. **118**, 1576–1584.

Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.

Notaro, G., Al-Maamury, A. M., Moleti, A., and Sisto, R. (**2007**). "Wavelet and Matching Pursuit estimates of the transient-evoked otoacoustic emission latency," J. Acoust. Soc. Am. **122**, 3576–3585.

Prijs, V. F., Schneider, S., and Schoonhoven, R. (**2000**). "Group delays of distortion product otoacoustic emissions: relating delays measured with

f1- and f2-sweep paradigms," J. Acoust. Soc. Am. **107**, 3298–3307.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.
Ruggero, M. A., and Temchin, A. N. (**2007**). "Similarity of traveling-wave delays in the hearing organs of humans and other tetrapods," J. Assoc. Res. Otolaryngol. **8**, 153–166.
Schoonhoven, R., Prijs, V. F., and Schneider, S. (**2001**). "DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears," J. Acoust. Soc. Am. **109**, 1503–1512.
Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise from two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Guinan, J. J., Jr. (**2003**). "Stimulus-frequency emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A., and Guinan, J. J., Jr. (**2007**). "Cochlear traveling-wave amplification, suppression, and beamforming probed using noninvasive calibration of intracochlear distortion sources," J. Acoust. Soc. Am. **121**, 1003–1016.
Shera, C. A., Guinan, J. J., Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**(5), 3318–3323.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2006**) "Delays of SFOAEs and cochlear vibrations support the theory of coherent reflection filtering," Assoc. Res. Otolaryngol. 2006 Meeting Poster, http://web.mit.edu/apg/posters/shera-tubis-talmadge-ARO2006.pdf.gz.
Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahey, P. F., and Guinan, J. J., Jr. (**2007**). "Allen-Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.
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.
Sisto, R., and Moleti, A. (**2002**). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297–308.
Sisto, R., and Moleti, A. (**2005**). "On the large scale spectral structure of otoacoustic emissions," J. Acoust. Soc. Am. **117**, 1234–1240.
Sisto, R., and Moleti, A. (**2007**). "Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels," J. Acoust. Soc. Am. **122**, 2183–2190.
Sisto, R., Moleti, A., and Shera, C. A. (**2007**). "Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays," J. Acoust. Soc. Am. **122**, 3554–3561.
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.
Tognola, G., Ravazzani, P., and Grandori, F. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
Unoki, M., Miyauchi, R., and Tan, C.-T. (**2007**). "Estimates of tuning of auditory filter using simultaneous and forward notched-noise masking," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Berlin), pp. 19–26.
Vetesnik, A., Nobili, R., and Gummer, A. (**2006**). "How does the inner ear generate distortion product otoacoustic emissions?. Results from a realistic model of the human cochlea," ORL **68**, 347–352.
Wilson, J. P. (**1980**). "Model for cochlear echoes and tinnitus based on an observed electrical correlate," Hear. Res. **2**, 527–532.
Zweig, G. (**1991**). "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am. **89**, 1229–1254.
Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Ear asymmetries in middle-ear, cochlear, and brainstem responses in human infants

Douglas H. Keefe,[b] Michael P. Gorga, and Walt Jesteadt
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

Lynette M. Smith
College of Public Health, 984375 Nebraska Medical Center, Omaha, Nebraska 68198-4375

(Received 22 October 2007; revised 11 December 2007; accepted 15 December 2007)

In 2004, Sininger and Cone-Wesson examined asymmetries in the signal-to-noise ratio (SNR) of otoacoustic emissions (OAE) in infants, reporting that distortion-product (DP)OAE SNR was larger in the left ear, whereas transient-evoked (TE)OAE SNR was larger in the right. They proposed that cochlear and brainstem asymmetries facilitate development of brain-hemispheric specialization for sound processing. Similarly, in 2006 Sininger and Cone-Wesson described ear asymmetries mainly favoring the right ear in infant auditory brainstem responses (ABRs). The present study analyzed 2640 infant responses to further explore these effects. Ear differences in OAE SNR, signal, and noise were evaluated separately and across frequencies (1.5, 2, 3, and 4 kHz), and ABR asymmetries were compared with cochlear asymmetries. Analyses of ear-canal reflectance and admittance showed that asymmetries in middle-ear functioning did not explain cochlear and brainstem asymmetries. Current results are consistent with earlier studies showing right-ear dominance for TEOAE and ABR. Noise levels were higher in the right ear for OAEs and ABRs, causing ear asymmetries in SNR to differ from those in signal level. No left-ear dominance for DPOAE signal was observed. These results do not support a theory that ear asymmetries in cochlear processing mimic hemispheric brain specialization for auditory processing.
 [DOI: 10.1121/1.2832615]

PACS number(s): 43.64.Jb, 43.64.Ri, 43.64.Ha, 43.71.Rt [BLM] Pages: 1504–1512

I. INTRODUCTION

Hemispheric specialization is evident in a variety of tasks, including those that involve auditory function, but its relation to reports of ear asymmetry in auditory measurements is unclear. Transient-evoked (TE) otoacoustic emissions (OAEs) have higher energy in the right than the left ear in twins (McFadden *et al.*, 1996) and in infants (Aidan *et al.*, 1997). There is a higher prevalence of spontaneous (S) OAEs in the right ears of adults, 5–12 year old children (Bilger *et al.*, 1990), and infants (Burns *et al.*, 1992). The signal-to-noise ratio (SNR) of distortion-product (DP) OAEs is larger in the right ear of children (age range 5.2–7.9 years) at DPOAE f_2 test frequencies of 1.9, 3, 3.8, and 6 kHz (Keogh *et al.*, 2001). Each of these reports is consistent with right-ear dominance at the level of cochlear mechanics. Larger auditory brainstem response (ABR) amplitudes and lower ABR thresholds have been reported for right-ear stimulation (Levine and McGaffigan, 1983; Levine *et al.*, 1988; Sininger *et al.*, 1998), although the neural asymmetry may not always favor the right ear (Ballachanda *et al.*, 1994). Sininger and Cone-Wesson (2006) found larger wave V amplitudes and mostly shorter latencies in the right ear of infants. Thus, the prevailing trends in both cochlear and lower brainstem neural response data support the view that auditory responses, in general, favor the right ear.

In contrast, Sininger and Cone-Wesson (2004) reported evidence of stimulus-specific specialization at the cochlear level in neonates, in which the right ear produced larger TEOAE SNRs in response to a click stimulus, and the left ear produced larger DPOAE SNRs in response to two tonal stimuli with slightly different frequencies. This left-ear DPOAE dominance contrasts with the above-mentioned studies, in which TEOAEs, SOAEs, and DPOAEs all appeared to favor the right ear. They considered a possible brainstem role via feedback to the outer hair cells in the cochlea by the medial olivocochlear efferent system. This role was not studied for OAE, inasmuch as the OAE measurements did not include conditions to elicit a change in the functional feedback provided by the medial olivocochlear efferent system.

Sininger and Cone-Wesson proposed that hemispheric specialization is facilitated at the cochlea by providing "greater amplification to stimuli that will also be preferentially processed in the auditory areas of the contralateral hemisphere."

One ambiguity in their study is that their conclusions regarding stimulus-specific, ear dominance was based only on measurements of OAE SNR; they did not describe the extent to which ear asymmetries were present in either the OAE signal or the noise. One might hypothesize that SNR is

[a] Portions of this work were presented in "Ear asymmetries in middle ear, cochlear and brainstem responses in infants," 2005 Midwinter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, 2005.
[b] Electronic mail: keefe@boystown.org

the relevant "signal" variable in single-fiber neural responses to the central nervous system. However, OAE noise is dominated by electrical noise in the measurement system and physiologic noise due to respiration, subject movement, and perhaps blood circulation that manifests itself in the measured acoustic ear-canal response. Because the OAE noise is unlikely to reflect the noise in neural responses, OAE SNR is unlikely to reflect neural SNR. Parsing of the OAE SNR into OAE signal and OAE noise components may be important as the conclusion that asymmetric cochlear processing mimics hemispheric specialization would appear to be based on the expectation that the DPOAE signal is stronger in the left ear.

Additionally, it may be important to consider ear asymmetries in the middle-ear functioning of neonates in order to understand ear asymmetries in cochlear or neural responses. Ear effects of middle-ear functioning were examined in infants based on acoustic reflectance and admittance measurements in the ear canal (Keefe *et al.*, 2000). For example, the equivalent volume, which is proportional to the imaginary part of the admittance, and the energy reflectance were larger in left ears below 1.4 kHz, whereas energy reflectance was larger in right ears between 2 and 5.7 kHz. Any ear asymmetries in middle-ear function might act as a confounding factor on measurements of ear asymmetries at all levels of the auditory pathway, including the cochlea.

Sininger and Cone-Wesson (2006) measured ABRs with and without contralateral stimulation and observed that the response was reduced in the left ear during contralateral noise presentation to a greater extent when compared to similar conditions for the right ear. They interpreted this asymmetric activation following contralateral stimulation to mean that the medial olivocochlear system may be the mechanism for the asymmetry in brainstem responses.

The purpose of the present study was to further explore the factors underlying these neonatal ear asymmetries by (1) replicating the Sininger and Cone-Wesson (2004, 2006) studies, (2) extending the study by separately evaluating OAE signal and noise levels averaged across frequency but also for individual $\frac{1}{2}$ octave frequencies from 1.5 to 4 kHz, (3) evaluating ear asymmetries in ABR from the same subjects, and (4) comparing measures of acoustic reflectance and admittance in efforts to describe the influence of middle-ear function on asymmetries in sensory and neural responses.

II. METHODS

A. Measured neonatal responses

The data described in this paper were part of a large multicenter project that focused on identification of hearing loss in neonates (Norton *et al.*, 2000a). This is the same project from which the data were drawn for the Sininger and Cone-Wesson (2004) study, although the data reported here were from a different sample of subjects. The protocol for data collection was identical to that used in the Sininger and Cone-Wesson studies, with the exception that the protocol for the present study included measures of middle-ear function on a subset of infants. Briefly, using a computer-controlled paradigm in which test and ear order were randomized, DPOAE, TEOAE, and ABR data were collected from well babies and graduates from intensive-care nurseries. The data reported here were originally collected at three different test sites (Boys Town National Research Hospital and the University of Nebraska Medical Center in Omaha, NE and Women's and Infant's Hospital in Providence, RI).

OAE data were evaluated for $\frac{1}{2}$ octave frequencies from 1.5 to 4 kHz. DPOAE data were collected with primary levels (L_1, L_2) of 65 and 50 dB SPL, whereas TEOAE data were collected in response to a click at 80 dB pSPL, using the nonlinear mode implemented on a commercial OAE measurement system. OAE data were not analyzed below 1.5 kHz because of high noise levels (and, therefore, poor SNR) at lower frequencies. Despite the fact that response measurements were restricted to the same range of stimulus frequencies for both DPOAEs and TEOAEs, it may be important to keep in mind that DPOAE measurements were made at a frequency ($2f_1-f_2$) that was about $\frac{1}{2}$ octave lower than the stimulus frequency (f_2). In contrast, TEOAE responses were measured at the same frequencies as the stimulus. Although this would be expected to affect estimates of signal and noise across OAE measurements, it should have no influence on comparisons across ears. DPOAEs and TEOAEs were also tested at higher stimulus levels, but the data for those stimulus conditions were not analyzed in the present study. Complete methodological details for TEOAE and DPOAE data collection can be found in Norton *et al.* (2000c) and Gorga *et al.* (2000), respectively. Signal SPL, noise SPL, and SNR were evaluated separately, and were either averaged across frequency (as in Sininger and Cone-Wesson, 2004), or analyzed separately at each frequency.

OAE responses were acquired in the original study using synchronous, time-domain averaging procedures. During acquisition of DPOAE responses, the minimum averaging included at least four 1.6-s averages at each frequency. Testing terminated at each frequency when the signal level was at least 3 dB above the mean noise level plus two standard deviations above the mean noise level. Overall, the test stopped when this definition of SNR was met at four of five frequencies (1, 1.5, 2, 3, and 4 kHz). During acquisition of TEOAE responses, at least 60 averages were obtained. Data collection continued until the SNR (defined in the more traditional way relative to mean noise level) was 3 dB or greater in the $\frac{1}{2}$ octave bands centered at 1 and 1.5 kHz, and 6 dB in the $\frac{1}{2}$ octave bands centered at 2, 3, and 4 kHz. For a test to stop in a given ear, the above-mentioned criteria had to be met in four of five $\frac{1}{2}$ octave bands. If these stopping rules were not met, then additional rules were evaluated, but such additional rules were not relevant for data included in the present study.

ABR data were collected in response to clicks presented at both 30 and 69 dB nHL [see Sininger *et al.* (2000) for details on ABR data collection]. Absolute ABR latencies (waves I and V), interpeak latency (IPL) differences defined as (wave V latency—wave I latency), wave V amplitude (peak to following trough at 30 and 69 dB), noise amplitude (at 30 and 69 dB), and F_{sp} (at 30 and 69 dB) were compared between ears. F_{sp} is closely related to SNR, and was originally described in the context of ABR detection by Don *et al.* (1984). It was used as the stopping criterion during ABR data

collection as part of the multicenter clinical trial, from which the present data and those described by Sininger and Cone-Wesson (2004, 2006) were extracted.

We also collected reflectance and admittance data (Keefe *et al.*, 2000) in a subset of the present sample of subjects (subjects at the Providence site and one of the two Omaha sites). The measurements describing middle-ear (and ear-canal) status always followed the TEOAE measurements. Middle-ear variables were selected based on previous research in terms of factors calculated using a principal components analysis (Keefe *et al.*, 2003a). A factor accounting for the most variance in that study was highly correlated with the equivalent volume averaged from 0.25 to 1 kHz (vLo); a second factor accounting for the next most variance was highly correlated with energy reflectance averaged from 2 to 8 kHz (rHi), and was an important factor in predicting transient middle-ear dysfunction in neonates (Keefe *et al.*, 2003b). For these reasons, these two middle-ear variables were used in the statistical analyses described in the following.

Two inclusion criteria were used for the analyses on the full data set. Criterion 1 in OAE analyses required a SNR of at least 3 dB at each of the four $\frac{1}{2}$ octave frequencies in both ears for both OAE measures, and in ABR analyses required an ABR $F_{sp}30$ of at least 3.0 in both ears. Sininger and Cone-Wesson (2004) used a similar 3-dB SNR criterion for their analysis of OAE responses. Sininger and Cone-Wesson (2006) subsequently considered ear asymmetries in ABRs using a slightly more stringent inclusion criterion of $F_{sp}=3.1$, based on data collected under the same conditions as in the present study. It is not expected that the slight differences in ABR inclusion criteria would impact comparisons between the two data sets.

Our study also used an additional inclusion criterion, which was designed to be similar across OAE and ABR test types. Inclusion criterion 2 in OAE analyses equated each of the DPOAE and TEOAE tests on test specificity (90%), which was similar to a criterion on test specificity used by Norton *et al.* (2000b). Inclusion criterion 2 in ABR analyses used a $F_{sp}30$ that also resulted in a test specificity of 90%. This approach facilitated comparisons across measurements because it was based on the same point on the response distributions of ears with hearing within normal limits for each test type. Inclusion criterion 2 resulted in the following critical test scores: SNR $\leqslant$ 1.82 dB for DPOAE, SNR $\leqslant$ 5.81 dB for TEOAE, and $F_{sp}30 \leqslant 2.44$. The apparent difference in SNR inclusion criteria for DPOAEs and TEOAEs is a consequence of the fact that, during data collection, noise levels for DPOAEs were defined as the mean noise level plus two standard deviations. The SNR criteria for the two OAE measurements are much more similar when mean noise levels are used for DPOAE measurements.

For analyses on the smaller data set that included reflectance and admittance responses along with the OAE and ABR responses, our study used an additional criterion 3, which was based on the following critical test scores: SNR $\leqslant$ 3.2 dB for DPOAE, SNR $\leqslant$ 6.4 dB for TEOAE, and $F_{sp}30 \leqslant 3.27$ dB. Imposing criterion 3 resulted in a database of 693 subjects.

B. Statistical design

The first set of OAE analyses was based on the frequency-averaged SNR, OAE level (signal), and OAE noise level measured in 2,640 subjects. In this test group, 2,093 subjects were included after selection based on criterion 1 and 2,150 subjects were included based on the slightly less stringent criterion 2. Separate repeated-measures analyses of variance (ANOVA) examined associations between SNR, signal and noise levels, and a pair of independent predictors, Ear (right or left) and OAE Type (DPOAE or TEOAE). A random subject effect was included in this and subsequent models to take into account the correlation of measurements within a subject. A random site effect was considered but was not a significant source of variability; thus, it was excluded from this and subsequent models. The summaries to follow were based on the combined data from the three sites. Fixed effects included Ear, OAE Type, and their interaction. In this and other ANOVAs, effects with $p < 0.05$ were considered statistically significant, and Tukey's method was used to adjust p-values for multiple comparisons. The corresponding 95% confidence intervals (CI) were calculated for each mean.

The second set of OAE analyses was based on the SNR, signal and noise at each of four frequencies. Separate repeated-measures ANOVAs examined associations at each frequency. Otherwise, these analyses were the same as those used for frequency-averaged data.

Separate repeated-measures ANOVAs examined associations between ABR variables listed above and the independent predictor of ear (except that wave I latency and the IPL at 30 dB were not analyzed because wave I could not be measured reliably in many subjects at this level). Analyses of ABR data were performed using only inclusion criterion 2.

The statistical analyses on OAE responses using criteria 1 and 2 resulted in nearly identical outcomes. For this reason, the first set of analyses that compared the present results to those reported by Sininger and Cone-Wesson (2004) was based on criterion 1. The remaining results described below used criterion 2 in order to equate OAE and ABR measures on test specificity.

The analyses on OAE and ABR responses that included the admittance and reflectance test results used the middle-ear variables vLo and rHi defined earlier. These variables were included as covariates in a set of repeated-measures ANOVAs that were otherwise similar to those described previously. Criterion 3 was used to select data for inclusion in these analyses.

All statistical analyses were performed using SAS software (SAS Institute Inc., Cary, NC), Version 9.1.

III. RESULTS

A. Ear effects in frequency-averaged OAEs

The SNR data from Sininger and Cone-Wesson (2004) showed a (statistically) significant ear $\times$ type interaction and significant ear differences for TEOAE SNR (right ear larger) and DPOAE SNR (left ear larger). Figure 1 shows the mean and CI of the frequency-averaged DPOAE and TEOAE SNR reported by Sininger and Cone-Wesson (left panels), and the

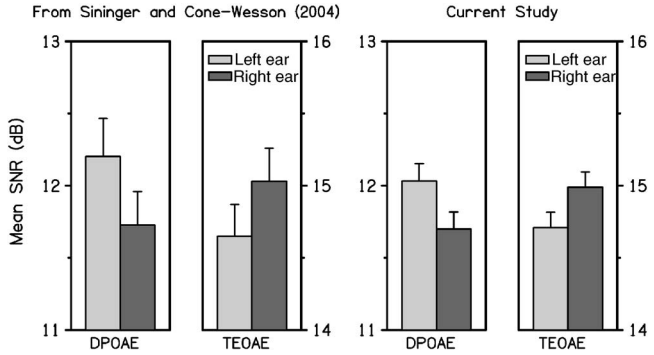


FIG. 1. The mean and upper 95% confidence interval (CI) of the frequency-averaged DPOAE SNR and TEOAE SNR reported by Sininger and Cone-Wesson (2004) are plotted in the left-hand panels, and those obtained in the present study are plotted in the right-hand panels using similar inclusion criteria. Each error bar denotes the upper half of the CI. In each pair of bar plots, the left bar (lightly shaded) represents the left-ear response and the right bar (darkly shaded) represents data from the right ear. In the current study, there was no significant difference between ears, but there was a significant ear × type interaction, with a significant difference between ears for DPOAE SNR.

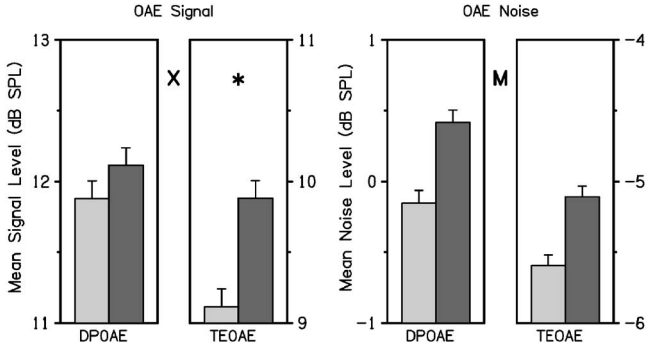


FIG. 2. The frequency-averaged OAE signal and noise SPL are plotted in the left- and right-hand halves, respectively, following the same conventions as those used in Fig. 1. The DPOAE and TEOAE SPL are separately plotted for signal and noise. Each error bar denotes the upper half of the 95% CI. Between each pair of bar plots in each panel, the letter **X** denotes that the ear × type interaction was significant for OAE signal, and the letter **M** represents that the main effect of ear was significant for OAE noise. When the interaction was significant, an asterisk denotes that ear was significant for a particular ear type; in this case, the OAE signal for TEOAE was higher in the right ear.

present study (right panels) using similar inclusion criteria. The interaction in our analyses between OAE type and ear was significant ($p=0.0004$), just as it was in the Sininger and Cone-Wesson study. The DPOAE SNR was significantly larger in the left than the right ear. The mean TEOAE SNR was larger in the right than the left ear, but this difference was not significant in the present analyses. Consistent with results in Sininger and Cone-Wesson, we observed that TEOAE SNR was significantly larger than DPOAE SNR in both ears (note the difference in scales on the vertical axis for TEOAE and DPOAE SNRs). As stated earlier, however, this observation relates mainly to differences in the way noise level was defined for the two OAE measures. In one case (TEOAE), it was defined as the mean noise level, whereas in the other (DPOAE), it was defined as the mean noise level plus two standard deviations.

To better understand the source of the ear asymmetry in SNR, signal and noise levels were evaluated separately. Figure 2 summarizes the mean signal and noise levels in left and right ears of the subjects included in the present study, following the conventions used in Fig. 1. As in Fig. 1, these results represent averages across the four frequencies included in this analysis, only here signal and noise are presented separately. The significant effects of primary interest are shown in Fig. 2 and subsequent figures using symbols described in the figure caption. The ear × type interaction was significant for OAE signal. The OAE signal for TEOAEs was higher in the right ear, but the OAE signal for DPOAEs did not vary with ear. The interaction between OAE ear and type was not significant for OAE noise. There was a main effect for OAE noise, which was significantly higher in the right ear (by approximately 0.5 dB) regardless of OAE type. There was no significant interaction between ear and type in the OAE noise results. Thus, the higher noise level reduced the SNR in the right ear for both OAE types. The fact that DPOAE noise was significantly higher (by approximately 5.5 dB) than TEOAE noise is of lesser interest, because it relates to differences in how noise was defined for the OAE types. There was also a main effect for OAE signal of Type, with DPOAE levels larger than TEOAE levels, and this asymmetry was significant in each ear. This effect was not of interest in the present study because differing source mechanisms may result in different DPOAE and TEOAE signal levels.

B. Ear effects in frequency-specific OAEs

Ear asymmetries in OAEs were next investigated on a frequency-specific basis (DPOAEs at $f_2=1.5$, 2, 3, and 4 kHz, and TEOAEs in $\frac{1}{4}$ octave bands centered at the same frequencies) by studying the significance of the ear × type interaction, and the significance of main effects of OAE SNR, signal, and noise levels. We report main effects for ear only when the ear × type interaction lacks significance, but do not report results indicating main effects for OAE type, which would simply identify frequencies at which DPOAE and TEOAE variables were unequal. Figure 3 summarizes the mean and the upper half of the 95% confidence interval data for SNR, signal, and noise in left, middle, and right columns, respectively, with data for a different frequency shown in each row.

In OAE SNR analyses (left column), the ear × type interaction was significant at 1.5, 2, and 3 kHz. The ear asymmetries at these frequencies that were significant for either DPOAE SNR or TEOAE SNR were as follows: the DPOAE SNR was larger in the left ear at 1.5 kHz, and the TEOAE SNR was larger in the right ear at 2 kHz. The ear × type interaction was not significant at 4 kHz. The corresponding main effect at 4 kHz was significant, with the SNR larger in the left ear. The averaging across frequency (as shown in Fig. 1), obscured some of these differences, e.g., the results for SNR at 4 kHz. However, the general trends in the frequency-specific analyses were similar to those observed when data were collapsed across frequency.

In OAE signal analyses (middle column), the ear × type interaction was significant at 1.5, 2, and 3 kHz. The significant ear asymmetries in signal at these frequencies

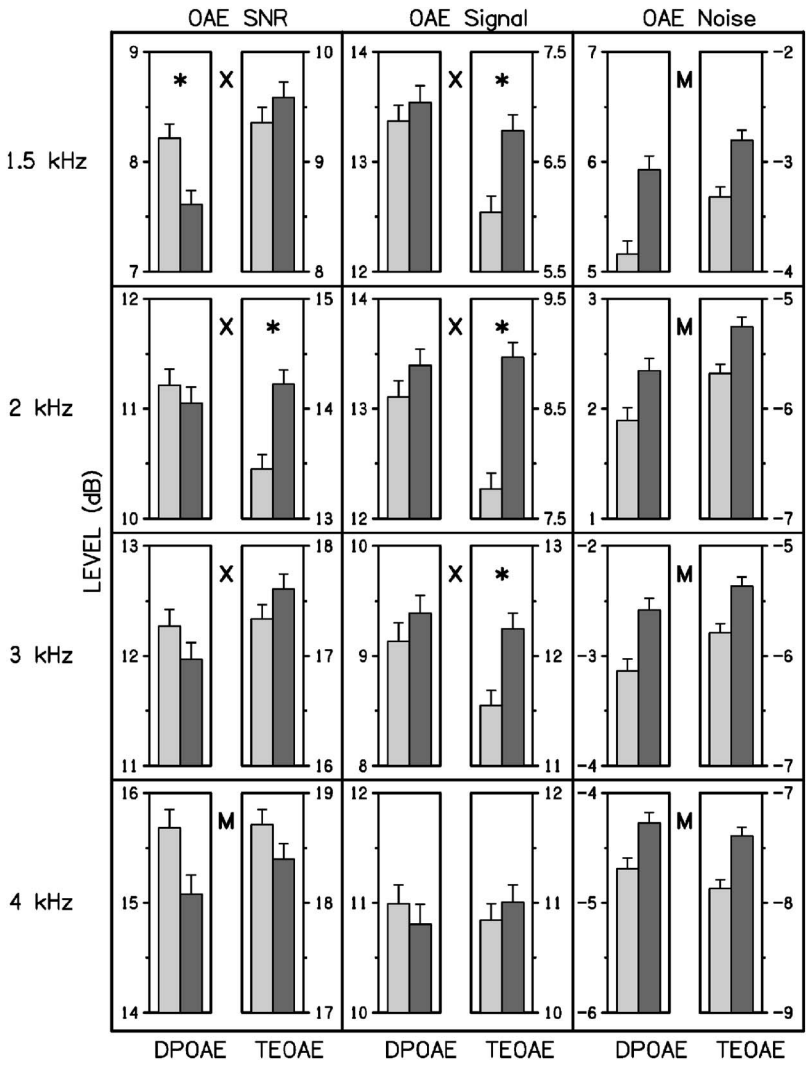


FIG. 3. The mean and 95% CI of the OAE SNR, and the OAE signal and noise SPL are plotted in the left, middle, and right columns of plots, respectively. The DPOAE and TEOAE SPL are separately plotted for signal and noise for each ear, with a plotting convention similar to that in Fig. 2. Between each pair of bar plots in each panel, the letter **X** denotes that the ear × type interaction was significant, and the letter **M** represents that the main effect of ear was significant. When the interaction was significant, an asterisk denotes that ear was significant for a particular ear type (DPOAE or TEOAE).

were as follows: the TEOAE signal was larger in the right ear at 1.5, 2, and 3 kHz. At 4 kHz, there was no ear × type interaction and no main effect for ear asymmetry. Despite the trends favoring DPOAE level in the right ear as well at 1.5, 2, and 3 kHz, no statistically significant ear asymmetries were observed.

In OAE noise analyses (right column), the ear × type interaction was not significant at any frequency. Consistent with frequency-averaged noise results, OAE noise was significantly larger in the right ear at every frequency, regardless of OAE type. The possible reasons for the observation of higher noise levels for stimulus conditions involving the right ear are further discussed below.

Like the previous findings in which data were averaged across frequency, these results bring into question the view that stimulus-driven ear asymmetries facilitate hemispheric specialization. Rather, the TEOAE signal at 1.5, 2, and 3 kHz in the present results is consistent with previous TEOAE (and SOAE) results showing a right-ear advantage.

C. Effect of middle-ear covariates on OAEs

With only two exceptions, the OAE ANOVA results including vLo and rHi as middle-ear covariates had the same pattern of significance as the ANOVAs described previously without middle-ear covariates. One exception was that TEOAE SNR was higher in the right ear at 2 kHz ($p = 0.021$) when the middle-ear covariates were included but was not significant otherwise. Second, a marginal main effect for ear in the SNR data at 4 kHz for the basic model (in this reduced data set of 693 subjects) was not significant when middle-ear covariates were included. The middle-ear covariates did not change the pattern of significance in any of the OAE signal and noise level analyses. The fact that the OAE models including middle-ear covariates performed nearly identically to OAE models without covariates means that middle-ear functioning, as encoded in vLo and rHi, did not explain ear differences in OAEs, even though ear differences existed in vLo and rHi. Such inter-ear differences in admittance and reflectance responses were described in Keefe *et al.* (2000). Further, rHi was important for identifying middle-ear dysfunction in infants who did not pass a newborn hearing screening test based on OAE/ABR testing, yet who were subsequently found to have hearing within normal limits (Keefe *et al.*, 2003a, b).

D. Ear effects in ABR results

Imposing the ABR criterion 2 resulted in an inclusion of 2221 subjects. The ABR results are organized into two

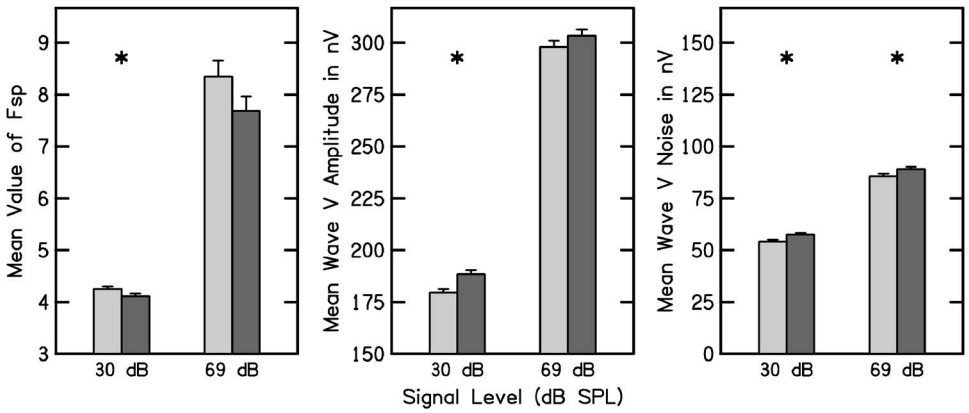


FIG. 4. The mean and 95% CI of the ABR F_{sp}, wave V signal amplitude, and wave V noise amplitude are plotted in the left, middle, and right panels, respectively. The left pair of bar plots in each panel shows data obtained at 30 dB nHL click stimulus level, and the right pair are data obtained at 69 dB nHL. In each pair of bar plots, ear is designated following the conventions used in previous figures. An asterisk denotes that ear was significant for a particular ABR response.

groups: (1) F_{sp}, and response and noise amplitudes; and (2) absolute response latencies and interpeak latency differences. Figure 4 plots F_{sp}, wave V amplitude and noise amplitude measured using click stimulus levels of 30 and 69 dB nHL. Figure 4 follows the convention used to designate ears in other figures. Consistent with previously reported data (Sininger and Cone-Wesson, 2006), the F_{sp} at 30 dB nHL was significantly smaller in the right ear, but F_{sp} at 69 dB did not differ across ears. Note that the error bars were smaller for the F_{sp} at 30 dB, compared to the 69 dB condition. This occurred because the stopping rule during data collection at 30 dB nHL included both the F_{sp} and observer-based criteria, whereas at 69 dB, only the observer-based criterion was used, resulting in greater F_{sp} variability at the end of testing for the 69 dB condition. Although the absolute difference in F_{sp} was smaller at 30 dB, it is likely that it was more than counterbalanced by the reduced variability, which resulted in a significant outcome at this level.

The wave V amplitude at 30 dB was larger for the right ear, but wave V amplitude at 69 was not statistically significantly different between ears. The ABR noise levels at both 30 and 69 dB were significantly higher in the right ear. These results indicate that the larger wave V amplitude in the right ear at 30 dB was associated with a larger noise amplitude in this ear, so that the Fsp, which functions as a SNR measure in an ABR test, was smaller in the right ear. The wave V amplitudes at 30 dB are consistent with some of the OAE findings, namely that DPOAE and TEOAE signals at 1.5 kHz and the TEOAE signal at 2 kHz were greater in the right ear. The noise sources for OAE and ABR measurements are different and perhaps statistically independent. Our observations are similar to those of Sininger and Cone-Wesson (2006), although they found statistically significant differences at both stimulus levels, with larger responses following right-ear stimulation. Sininger and Cone-Wesson (2006) did not report ABR noise levels.

Figure 5 plots absolute latencies and IPL differences (i.e., wave V latency minus wave I latency). Wave I latency, which was analyzed only at a 69 dB click level, was longer in the right ear. This result is further discussed in the next section, which considers the effects of middle-ear covariates on ABR responses. Wave V latency was significantly shorter on the right side at both stimulation levels, which is an effect in the opposite direction, compared to Wave I latency at 69 dB nHL. As a consequence, the IPL (which is reported only at the 69-dB click level for reasons described above) was significantly shorter on the right side.

E. Effect of middle-ear covariates on ABRs

When middle-ear covariates were included in the model for the subsample of responses in the 693 subjects on whom middle-ear testing was performed and inclusion criterion 3 was satisfied, there was no ear asymmetry in wave I latency. Removing the middle-ear covariates from the ANOVA in this subsample resulted in a longer wave I latency for the right ear, which is the same result obtained for the full sample (as described above). As wave I reflects the first neural response along the auditory pathway, these results are consistent with an absence of inter-ear differences at the lowest neural level of the auditory system if the influence of the middle ear is taken into account. In this same subsample, the wave V latencies at both stimulus levels were shorter in the right ear whether the middle-ear covariates were present or absent from the ANOVA. Further, the IPL was shorter in the right ear for both cases. These wave V latency and IPL results in the subsample replicated the results in the full sample, and were insensitive to presence or absence of the middle-ear covariates.

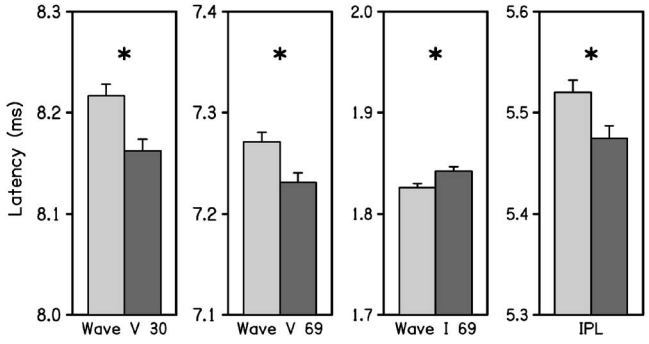


FIG. 5. The mean and 95% CI of the ABR wave latencies are plotted in each panel. The latency responses from left to right across these panels are for wave V at 30 dB nHL, wave V at 69 dB nHL, wave I at 69 dB nHL, and the IPL (i.e., wave V latency minus wave I latency) at 69 dB nHL. In each pair of bar plots, ear is designated following the conventions used in previous figures. An asterisk denotes that ear was significant for a particular ABR latency.

The absence of ear asymmetry for wave I latencies when middle-ear covariates were included suggests that the ear-asymmetry effect, which was represented by the shorter IPL for the right ear, occurred above the level of the cochlea. It is likely that the effect of forward transmission through the middle ear were not fully described by the middle-ear covariates. We might expect an ear asymmetry in middle-ear transmission to result in an asymmetry in wave I latency with no asymmetry in IPL. Middle-ear transmission would influence wave I and wave V latencies in a similar manner, and, because the IPL is the difference in these latencies, the IPL would not be affected.

IV. DISCUSSION

These results show that the interaction of OAE type and ear was significant for both criteria 1 and 2. There was no ear asymmetry in the TEOAE SNR for either criterion, but the DPOAE SNR for each criterion was larger in the left ear. This pattern in OAE SNR asymmetry was at least qualitatively similar to that reported by Sininger and Cone-Wesson (2004). Based on ear asymmetries in OAE SNR, they stated that the cochlea provides greater amplification to stimuli that would also be preferentially processed in the contralateral hemisphere of the brain. However, when SNR was separated into signal and noise level components, our frequency-averaged results showed no ear asymmetry in DPOAE signal, but a larger TEOAE signal level in the right ear. This suggests that, regardless of stimulus, OAE signal levels were either equal between ears or greater in the right ear. In the frequency-averaged results, there was no stimulus condition for which the left ear produced the larger OAE signal level.

Frequency-averaged results revealed an ear asymmetry in OAE noise of ~0.5 dB. It is difficult to understand how an asymmetry in peripheral acoustic noise relates to differential processing in the auditory areas of the cortex (i.e., favoring tones eliciting DPOAEs or clicks eliciting TEOAEs) or how this might relate to stimulus-specific ear asymmetries in cochlear amplification. Because OAE signal is related to the strength of cochlea excitation that also drives neural encoding, OAE signal level has a more direct representation of stimulus encoding within the cochlea than does OAE SNR. OAE noise levels were higher in the right ear at each frequency and in the average across frequency. TEOAE signal was higher at 1.5, 2, and 3 kHz in the right ear, with no ear asymmetry at any frequency in DPOAE signal. OAE signal levels following right-ear stimulation were either equal to or greater than those observed following left-ear stimulation regardless of OAE type and, therefore, the stimulus eliciting the response. No left-ear advantage for OAE signal was observed for any combination of OAE type and stimulus frequency.

The absence of a left-ear advantage for OAE signal across OAE type and stimulus frequency contrasts with the larger left-ear DPOAE SNR reported by Sininger and Cone-Wesson (2004) and replicated in our results. Because Sininger and Cone-Wesson reported results only for SNR, the extent to which this SNR effect was influenced by ear asymmetries in DPOAE signal, noise, or some combination thereof cannot be known from their report. In our analyses, the dominant effect producing the SNR result was revealed by the presence of a higher noise level in the right ear (of approximately 0.5 dB) that was independent of OAE type. This is because TEOAE signal was larger in the right ear for several conditions, whereas DPOAE signal was similar across ears. The effect of higher noise in the right ear produced an ear asymmetry in OAE SNR, controlled by the reduction of the right-ear dominance of TEOAE signal compared to TEOAE SNR, and by the presence of a left-ear dominance for DPOAE SNR due only to the larger DPOAE noise in the right ear. One possible explanation for the higher noise level in the right ear is that a potential source of the noise, related to blood flow, is closer in the right than the left ear. This is consistent with Rubens *et al.* (2007), who described effects on TEOAE measurements of an anatomical asymmetry in newborns for which the right innominate vein is closer to the right inner ear than is the left innominate vein to the left inner ear (Rohen *et al.*, 2002).

One potential confounding issue is the extent to which the measurement-based stopping rules in the DPOAE and TEOAE tests, which were fully described in the original studies that were the source of the data (Gorga *et al.*, 2000; Norton *et al.*, 2000a, b, c; Sininger *et al.*, 2000) and briefly summarized in Sec. II, might affect the interpretation of OAE signal, noise, and SNR. Specifically, given a stopping rule based on SNR, is it valid to evaluate levels of signal and noise as well as SNR? If it were not possible to interpret the measured OAE signal SPL as a measure of the magnitude of the cochlear source strength due to the presence of measurement noise, it would also not be possible to interpret the measured SNR as a measure of the magnitude of the cochlear source strength. In that case, it would not be possible to utilize OAE SNR as a measure of asymmetric cochlear processing. In fact, our use of SNR and Signal and Noise separately is identical to how these quantities were described in the aforementioned original studies.

Further, the OAE measurements were not unduly contaminated by the SNR-based stopping rules used to acquire the data. Suppose the noise level was stationary. A longer averaging time would increase the SNR and decrease the noise level, but the time-synchronous measurement of signal level would only be slightly affected if the SNR were sufficiently large. The stopping rules were relevant to responses with SNRs in the range of 3–6 dB or less, but would have less of an effect for responses with larger SNRs. The inclusion criteria excluded ears with small SNR, i.e., criterion 1 required a SNR of at least 3 dB; criterion 2 required a SNR of at least 1.82 dB for DPOAEs, and 5.81 dB for TEOAEs. As stated earlier, the statistical results for ear asymmetries in OAEs were nearly identical using either criterion 1 or 2. The insensitivity of the analyses to choice of criterion suggests that any confounding factor of ears with low SNR was minimal. This is evident in the measured SNRs across frequency, which were on average 12 dB for DPOAEs and 15 dB for TEOAEs (see Fig. 1). Most ears included in the analysis had SNRs higher than the limits required by the stopping rules.

Because of these factors, it was possible to interpret OAE signal levels as providing information on the cochlear source strengths.

The ABR results showed right-ear dominance for some of the response and noise amplitudes, and never showed left-ear dominance, findings that, on balance, are consistent with other reports (e.g., Sininger and Cone-Wesson, 2006). A left-ear dominance for F_{sp} at 30 dB nHL was observed, but it is likely, although not certain, that this effect resulted from larger ear differences in noise amplitude than in wave V amplitude. In addition, a right-ear dominance in the full sample was observed for the IPL and for the wave I and wave V latencies. The observation of right-ear dominance for IPL, but not for wave I latency, when the middle-ear covariates were taken into account suggests that the source(s) of asymmetry are at least partly neural in origin, rather than residing solely in the peripheral mechanics. These results also show that the middle-ear covariates encode some information relevant to middle-ear transmission. The fact that energy reflectance at high frequencies is larger in the right ear of neonates (Keefe *et al.*, 2000) is consistent with a reduced forward transmission through the right middle ear, and a reduced middle-ear transmission would be expected to increase the wave I latency in the right ear. This prediction is consistent with our data.

Ear asymmetries in OAEs, which are pre-neural responses, were also present in the latency of wave I, which is the first neural response. Because the underlying ear-canal measurements did not directly measure forward transmission through the middle ear, a detailed explanation of this result would be confounded by the possibility of an ear asymmetry in forward transmission through the middle ear that was not represented by the middle-ear covariates. However, the right-ear dominance of IPL would not be expected to be confounded by middle-ear transmission, and is consistent with the right-ear dominance of TEOAE signal for some conditions. Our results support the findings referenced in the Introduction of a right-ear dominant effect in cochlear mechanics and neural responses, but do not support the finding of a left-ear dominance in DPOAE signal. Thus, they do not support the view that there is a stimulus-driven, peripherally based factor in these interaural effects of auditory processing that might influence the functional maturation of hemispheric specialization in the brain.

Our data set of OAE and ABR responses did not allow an evaluation of the potential contribution to ear asymmetries due to feedback by the medial olivocochlear efferent system, although this was explored by Sininger and Cone-Wesson (2006) in their evaluation of ABR asymmetries. Further study with more refined stimulus conditions is needed to assess the extent to which these efferent effects contribute to ear asymmetries in OAE responses.

The significant ear asymmetries in OAE and ABR responses reported herein have too small an effect size to carry clinical importance in newborn hearing screening programs, even though they do reveal subtle aspects of cochlear and neural processing. Based on the present results, the OAE and ABR test criteria, which are used in such hearing screening programs to identify infants at risk for sensorineural hearing loss, should not incorporate information on response asymmetry between right and left ears.

V. CONCLUSIONS

Our observations are consistent with the preponderance of earlier studies showing a right-ear dominance for TEOAE signal and ABR measures. Both OAE and ABR analyses revealed that noise was larger in the right than the left ear. The middle-ear covariates had little effect on the OAE analyses but were useful in interpreting the ear asymmetries in ABR wave I latencies. Our observations are inconsistent with the conclusions of Sininger and Cone-Wesson (2004), inasmuch as we found no evidence of left-ear dominance in DPOAE measures, and, in particular, no left-ear dominance in DPOAE signal. We did not observe a stimulus-guided asymmetry in OAE responses (i.e., different ear dominance for click-elicited TEOAE and tone-elicited DPOAE signals). Such an asymmetry would appear to be essential to a theory that asymmetric cochlear processing mimics hemispheric specialization.

ACKNOWLEDGMENTS

We thank Betty Vohr, M.D., for allowing us to include data from Women and Infants Hospital, Providence, RI, which was one of the sites for the multicenter clinical trial. We appreciate the assistance of James C. Lynch, Ph.D. of the University of Nebraska Medical Center, who consulted on the statistical analyses. The present research was supported by the NIH grant nos. DC003784 and DC002251 using a database of responses collected with support by the NIH grant no. DC01958. We thank Associate Editor Brenda Lonsbury-Martin and two anonymous reviewers for comments on an earlier version of this report.

Aidan, D., Lestang, P., Avan, P., and Bonfils, P. (**1997**). "Characteristics of transient-evoked otoacoustic emissions (TEOEs) in neonates," Acta Oto-Laryngol. **117**, 25–30.

Ballachanda, B. B., Rupert, A., and Moushegian, G. (**1994**). "Asymmetric frequency-following responses," J. Am. Acad. Audiol **5**, 133–137.

Bilger, R. C., Matthies, M. L., Hammel, D. R., and Demorest, M. E. (**1990**). "Genetic implications of gender differences in the prevalence of spontaneous otoacoustic emissions," J. Speech Hear. Res. **33**, 418–432.

Burns, E. M., Arehart, K. H., and Campbell, K. (**1992**). "Prevalence of spontaneous otoacoustic emissions in neonates," J. Acoust. Soc. Am. **91**, 1571–1575.

Don, M., Elberling, C., and Waring, M. (**1984**). "Objective detection of averaged auditory brainstem responses," Scand. Audiol. **13**, 219–228.

Gorga, M. P., Norton, S. J., Sininger, Y. S., Cone-Wesson, B., Folsom, R. C., Vohr, B. R., and Widen, J. E. (**2000**). "Identification of neonatal hearing impairment: Distortion product otoacoustic emissions during the perinatal period," Ear Hear. **21**, 400–424.

Keefe, D. H., Folsom, R. C., Gorga, M. P., Vohr, B. R., Bulen, J. C., and Norton, S. J. (**2000**). "Identification of neonatal hearing impairment: Ear-canal measurements of acoustic admittance and reflectance in neonates," Ear Hear. **21**, 443–461.

Keefe, D. H., Gorga, M. P., Neely, S. T., Zhao, F., and Vohr, B. (**2003b**). "Ear-canal acoustic admittance and reflectance effects in human neonates. II. Predictions of middle-ear dysfunction and sensorineural hearing loss," J. Acoust. Soc. Am. **113**, 407–422.

Keefe, D. H., Zhao, F., Neely, S. T., Gorga, M. P., and Vohr, B. (**2003a**). "Ear-canal acoustic admittance and reflectance effects in human neonates. I. Predictions of otoacoustic emission and auditory brainstem responses," J. Acoust. Soc. Am. **113**, 389–406.

Keogh, T., Kei, Rl., Criscoll, C., and Smyth, V. (**2001**). "Distortion-product otoacoustic emissions in schoolchildren: Effects of ear asymmetry, handedness, and gender," J. Am. Acad. Audiol **12**, 506–513.

Levine, R. A., Liederman, J., and Riley, P. (**1988**). "The brainstem auditory evoked potential asymmetry is replicable and reliable," Neuropsychologia **26**, 603–614.

Levine, R. A., and McGaffigan, P. M. (**1983**). "Right-left asymmetries in the human brainstem: Auditory evoked potentials," Electroencephalogr. Clin. Neurophysiol. **55**, 532–537.

McFadden, D., Loehlin, J. C., and Pasanen, E. G. (**1996**). "Additional findings on heritability and prenatal masculinization of cochlear mechanisms: Click-evoked otoacoustic emissions," Hear. Res. **97**, 102–119.

Norton, S. J., Gorga, M. P., Widen, J. E., Folsom, R. C., Sininger, Y. S., Cone-Wesson, B., Vohr, B. R., and Fletcher, K. (**2000a**). "Identification of neonatal hearing impairment: A multicenter investigation," Ear Hear. **21**, 348–356.

Norton, S. J., Gorga, M. P., Widen, J. E., Folsom, R. C., Sininger, Y. S., Cone-Wesson, B., Vohr, B. R., Mascher, K., and Fletcher, K. (**2000b**). "Identification of Neonatal Hearing Impairment: Evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem evoked response test performance," Ear Hear. **21**, 508–528.

Norton, S. J., Gorga, M. P., Widen, J. E., Vohr, B. R., Folsom, R. C., Sininger, Y. S., Cone-Wesson, B., and Fletcher, K. (**2000c**). "Identification of neonatal hearing impairment: Transient evoked otoacoustic emissions during the perinatal period," Ear Hear. **21**, 425–442.

Rohen, J. W., Yokochi, C., and Lutjen-Drecoll, E. (**2002**). *Color Atlas of Anatomy*, 5th ed. (Lippincott, Williams & Wilkins, Philadelphia), p. 257.

Rubens, D. D., Vohr, B. R., Tucker, R., O'Neil, C. A., and Chung, W. (**2007**). "Newborn oto-acoustic emission hearing screening tests," Early Hum. Dev..

Sininger, Y. S. and Cone-Wesson, B. (**2004**). "Asymmetric cochlear processing mimics hemispheric specialization," Science **305**, 1581.

Sininger, Y. S., and Cone-Wesson, B. (**2006**). "Lateral asymmetry in the ABR of neonates: Evidence and mechanisms," Hear. Res. **212**, 203–211.

Sininger, Y. S., Cone-Wesson, B., and Abdala, C. (**1998**). "Gender distinctions and lateral asymmetry in the low-level auditory brainstem response of the human neonate," Hear. Res. **126**, 58–66.

Sininger, Y. S., Cone-Wesson, B., Folsom, R. C., Gorga, M. P., Vohr, B. R., Widen, J. E., Ekelid, M., and Norton, S. J. (**2000**). "Identification of neonatal hearing impairment: Auditory brain stem responses in the perinatal period," Ear Hear. **21**, 400–424.

Inverted direction of wave propagation (IDWP) in the cochlea

Egbert de Boer
Academic Medical Center, University of Amsterdam, Room D2-225/226, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands and Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098

Jiefu Zheng and Edward Porsov
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098

Alfred L. Nuttall[a)]
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098; Kresge Hearing Research Institute, University of Michigan, 1301 E. Ann Street, Ann Arbor, Michigan 48109-0506, and Department of Otolaryngology, Shanghai Jiao Tong University, 1954 Shanghai Huansan Road, Shanghai, China

(Received 17 July 2007; revised 27 November 2007; accepted 3 December 2007)

The "classical" view on wave propagation is that propagating waves are possible in both directions along the length of the basilar membrane and that they have identical properties. Results of several recently executed experiments [T. Ren, Nat. Neurosci. **2**, 333–334 (2004) and W. X. He, A. L. Nuttall, and T. Ren, Hear. Res., **228**, 112–122 (2007)] appear to contradict this view. In the current work measurements were made of the velocity of the guinea-pig basilar membrane (BM). Distortion products (DPs) were produced by presenting two primary tones, with frequencies below the characteristic frequency f_0 of the BM location at which the BM measurements were made, with a constant frequency ratio. In each experiment the phase of the principal DP, with frequency $2f_1 - f_2$, was recorded as a function of the DP frequency. The results indicate that the DP wave going from the two-tone interaction region toward the stapes is not everywhere traveling in the reverse direction, but also in the forward direction. The extent of the region in which the forward wave occurs appears larger than is accounted for by classical theory. This property has been termed "inverted direction of wave propagation." The results of this study confirm the wave propagation findings of other authors. The experimental data are compared to theoretical predictions for a classical three-dimensional model of the cochlea that is based on noise-response data of the same animal. Possible physical mechanisms underlying the findings are discussed.
 [DOI: 10.1121/1.2828064]

PACS number(s): 43.64.Kc, 43.64.Bt [BLM] Pages: 1513–1521

I. INTRODUCTION

Ever since otoacoustic emissions were detected (Kemp, 1978), particular attention has been given to reverse traveling waves in the cochlea. This became more relevant when distortion product otoacoustic emissions—where more parameters are involved—were studied in detail. Many authors have held the view that the round-trip delay of such distortion products (DPs) is composed of a forward and a backward delay, and can approximately be decomposed into two almost equal parts (among others, Goodman *et al.*, 2004; Shera and Guinan, 1999; Kalluri and Shera, 2001). Recently, findings by Ren (2004a, b) and Ren *et al.* (2006) have cast doubt on this opinion. For these studies measurements of DPs were made in the gerbil at several locations along the basal part of the basilar membrane. Unexpectedly, reverse traveling DP waves could not be detected over the entire basilar-membrane (BM) region between the presumed source of the DP and the stapes that these authors were able to inspect. In contrast, in part of this region a forward instead of a reverse traveling wave was found. We propose to call this phenomenon "inverted direction of wave propagation" (IDWP), and will consider it against the background of "classical" cochlear mechanics. Moreover, others have provided a new interpretation of the existing round-trip data (Ruggero, 2004; Siegel *et al.*, 2005). A possible explanation of the findings, in terms of compression waves, was advanced (Ren, 2004a, b, 2006) but this idea has been criticized (Shera, 2006; de Boer and Nuttall, 2006a; Shera *et al.* 2007).

In this paper we describe a set of experiments directed at clarifying the question of wave travel of DPs in the cochlea. In particular, we want to answer the question whether or not there exists an inverted direction of wave propagation as it has been reported by Ren and associates when an experiment is performed with a different experimental technique and in a different animal model. Our basic experiment consisted of presenting two pure tones, with frequencies f_1 and f_2, to the cochlea of a guinea pig and measuring the DP at a *fixed* location L_0 of the BM, whereby the frequencies f_1 and f_2

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

were varied with *a constant frequency ratio*. In contrast to Ren (2004a, b, 2006), we did not keep the frequencies constant while varying the position of observation, but we used a fixed observation point and varied the tone frequencies. Viewed in terms of location along the length of the basilar membrane (BM), the tone pair is moving along the BM with a constant separation—this corresponds to the constant frequency ratio because of the tacitly assumed logarithmic scaling of frequency to location. Our measurement technique is similar to the technique that we have employed to study the Allen–Fahey puzzle (Allen and Fahey, 1992, de Boer *et al.*, 2005). The main difference is that all frequencies involved are *lower* than the best frequency (BF) of the location L_0 at which we measure BM velocity. A further difference is that we concentrate on the phase of the DP component in the BM response instead of the amplitude.

Results of the experiments unequivocally demonstrate that between the f_2 site and the stapes there exists a region where the direction of wave travel is inverted, i.e., where there is a forward and not a reverse traveling wave. As stated above, we have termed this property inverted direction of wave propagation, abbreviated: IDWP. We also show theoretical predictions on DP wave travel for a linear cochlear model based on noise-response data. Computation is performed in analogy to earlier work (de Boer *et al.*, 2007), except that we now compare predictions and data from the *same* animal, and we selected tone frequencies below the BF of the measurement location L_0. Prediction and data clearly diverge and IDWP, in the form of an extended range of a forward wave, is evident. This more theoretical topic is briefly expanded upon.

II. EXPERIMENTAL TECHNIQUES

We have performed our experiments in guinea pigs. The protocols of the experiments described in this paper were approved by the Committee on the Use and Care of Animals, Oregon Health & Science University. Details about experimental animals and surgical techniques have been described in several previous publications (e.g., Nuttall *et al.*, 1991, 2004; de Boer and Nuttall, 2000). We give a shortened description here, but focus on the new aspects of our present series of experiments.

The experimental animal was deeply anesthetized, the bulla was opened and a small opening was made in the first-turn scala tympani wall of the cochlea. Glass beads ($\sim$20 μm diameter and gold coated) were dropped through this opening upon the basilar membrane (BM). The laser beam of a Doppler velocimeter (Polytec Corp. OFV 1102) was focused on one of the beads, and produced a signal representing the bead's motion. Two miniature condenser microphones were used as loudspeakers for the two primary tones; they were embedded in a plastic coupler placed in the external ear canal of the experimental animal. Stimulus generation and data acquisition were performed by a custom-made computer program based on LABVIEW© software and TDT© hardware (this is the main difference between the instrumentation of these and earlier experiments). Each of the two tones was sent to one of the loudspeakers. We ensured that the system's intermodulation distortion was well below the cochlear DP level.

In order to obtain a database on frequency selectivity and nonlinearity of the animal under study, responses were obtained for stimuli consisting of noise signals with various bandwidths and various levels. Composite-spectrum files were constructed and served as bases for computations to be described further on. At the spectral level of 20 dB per octave the best frequencies (BFs) of the animals for this study ranged between 17 and 21 kHz. During experiments the sensitivity of the experimental animal was checked from time to time. Hearing losses at the BF due to surgery varied from 1 to 15 dB (8.4 ± 5.5 dB, mean $\pm$ SD, $n=8$) among animals. During data acquisition, significant deterioration of sensitivity (i.e., Compound Action Potential (CAP) threshold elevation greater than 10 dB near the BF) was not observed in this study.

Tone pairs were chosen so that both frequencies, f_1 and f_2, were *below* the frequency f_0, which is the estimated BF of the measurement location L_0. In that way we can consider L_0 as a "station on the way" of the DP wave from its generation site to the stapes. By varying the pair of frequencies we can observe different parts of that DP wave. We limit our discussion here to the DP with frequency $2f_1-f_2$. For the interpretation it is important that for every frequency the phase of each of the primary tones at the peak of its response is assumed to be (nearly) constant with respect to the phase at the stapes. Locally, the DP is thus also generated with (nearly) constant phase. Note that this property is another expression of "scaling symmetry" of which the projection of log frequency to location is the basic property.

In closing this section, we describe the physical signals used in this work and the process of data analysis. Stimulus signals were generated with a sample frequency of 200 kHz, the stationary parts had a duration of 100 ms and were preceded and followed by the same signals with gradual onset and offset envelopes lasting 5 ms. The full 100 ms of stimulus and response signals, signal onset and offset excluded, in total comprising 20 000 data samples, was used for data analysis. By making all frequencies integral multiples of 10 Hz we ensured that the 100 ms interval always contained an integral number of periods. The stimuli were repeated 400 times and the waveform of the response was averaged over these repetitions. Fourier transformation of the recorded waveform yielded the corresponding spectrum.

In the processing of the data we have to correct for the delay inherent in the onset period (1000 time samples, equivalent to 5.0 ms) and for the delay from transducer to stapes (in time approximately 180 μs). When we use the symbol τ for the onset delay of 5.0 ms, we have to compensate for phase delays of $\Theta_1=2\pi f_1\tau$, and $\Theta_2=2\pi f_2\tau$ radians, respectively, for the two frequencies f_1 and f_2. This constitutes a straightforward and invariant correction. Compensation for the travel time from transducer to stapes is based on the cross-correlation spectrum from transducer to stapes as measured with a broad-band noise stimulus signal, for every animal individually, and sampled at the frequencies f_1 and f_2. These phase shifts were *included* in the phase values Θ_1 and Θ_2. From considering third-order nonlinear distortion as giv-

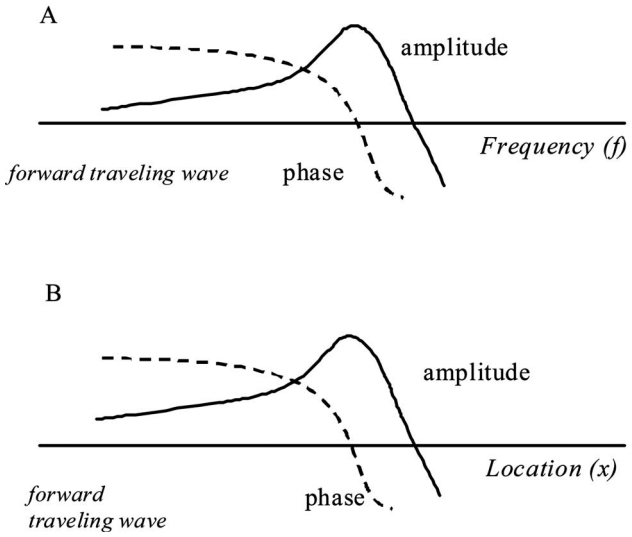


FIG. 1. Panel A: Frequency response, the response of the cochlea measured at one location (L_0) as a function of (log) frequency. Panel B: Cochlear pattern, the response for one frequency (f_0) as a function of location x (distance from the stapes). Solid curves: amplitude plotted logarithmically, dashed curves: phase plotted linearly.

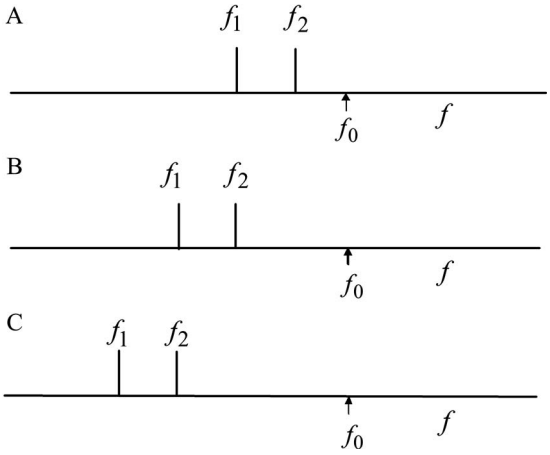


FIG. 2. Experimental protocol, frequency scale is logarithmic. Frequencies f_1 and f_2. Frequency ratio R_1 equals f_2/f_1. The frequency f_0 is equal to 17 kHz, close to the BF of the observation location. Panel A, istage=1: The higher frequency f_2 is equal to f_0/R_1, the lower frequency f_1 is equal to f_0/R_1^2. Panel B, istage=2: Both frequencies are reduced by the factor R_1. Panel C, istage=3: Both frequencies are again reduced by the factor R_1. For each value of R_1 nine of such conditions are used in the experiments.

ing rise to the DP with frequency $2f_1-f_2$ it follows that we have to compensate the observed DP response for a phase shift of $2\Theta_1-\Theta_2$.

III. A ROOM WITH A VIEW

Before turning to the actual results obtained, let us briefly discuss the ways we can look at data. Figure 1(A) shows the frequency response of the cochlea in stylized form. Amplitude and phase of the velocity of the BM are imagined to be measured at a *fixed location* (L_0) for a sinusoidal stimulus, and are plotted as a function of frequency f. The frequency scale is chosen to be logarithmic. The amplitude (assumed to be plotted logarithmically) grows, first slowly and then more rapidly, toward a peak and rapidly decreases thereafter. The frequency corresponding to the peak will be referred to as the BF (best frequency). The stimulus is assumed to be presented at the stapes with zero phase. The response phase (plotted linearly) has a monotonic course, starting at a value close to $+\pi/2$, the lag increasing with frequency (the phase becoming more negative). Consider next the response for a *fixed frequency* plotted as a function of cochlear location x, i.e., the distance to the stapes measured along the length of the BM. The "cochlear pattern" as we call it, is shown in Fig. 1(B) in the location domain. The abscissa scale is linear rather than logarithmic and in this form the two graphics resemble one another very much. This is due to the property of scaling symmetry referred to in the previous section.[1] In what follows we will have occasion to refer to these two representations.

How do these figures vary when we consider variations of parameters? Suppose we consider the frequency response at a location closer to the stapes. The response as shown by Fig. 1(A) will move to the right, the BF is higher. In contrast, when we look at the cochlear pattern and increase the frequency, the pattern of Fig. 1(B) will shift to the left. Such considerations are useful for the interpretation of various conditions.

Actual data are collected in the view of Fig. 1(A), with frequency as abscissa. It is useful to realize that a forward traveling wave has the same phase course in the two views: a *phase lag* that is *increasing (phase becoming more negative)* when we move to the right in the figure. Conversely, in a reverse traveling wave the *phase lag decreases (phase is becoming less negative)* when going to the right. This, again, is true in both representations. Because of the correspondence between Figs. 1(A) and 1(B) it will be easy to convert phase data as a function of frequency to data as a function of place and to draw conclusions about their nature—forward *versus* reverse traveling waves. This will be done later in this paper.

IV. RESULTS

The experimental protocol is illustrated by Fig. 2. We have designated the various tone pairs by a variable that we have called *istage*. Panel A shows the stimulus conditions in the first presentation of the set (*istage* being equal to 1). The frequency f_0 is the BF of the observation location L_0. Actually, because the BF is not known until after (a part of) the experiment has been performed, we artificially set f_0 equal to 17 kHz. The frequency ratio is R_1, and is constant during one run of the experiment. The higher frequency f_2 is, more or less arbitrarily, chosen to be equal to f_0/R_1, the lower frequency f_1 is then equal to f_0/R_1^2. Panel B shows the condition for *istage* equal to 2. Both frequencies are reduced by the factor R_1. In panel C *istage* is equal to 3 and both frequencies are again reduced by the factor R_1. In one run we have used *nine* of such presentations, *istage* goes from 1 to 9. As argued in the previous section responses will always be plotted as functions of frequency. In every one of our experiments we have used maximally *twelve* runs with different

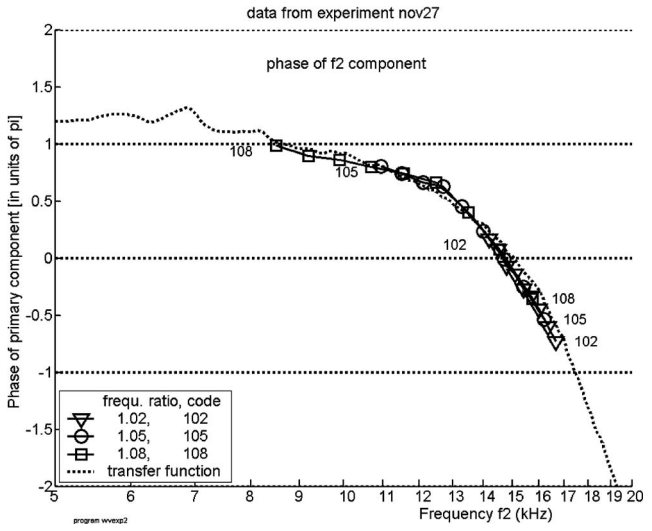


FIG. 3. The phase of the primary component with frequency f_2 in the recorded spectrum, plotted against its own frequency, for animal "nov27." Each of the curves encompasses the set of values of istage from 1 to 9, whereby begin and end are marked with a code signifying the frequency ratio f_2/f_1 (see legend inside figure). Only three ratios were used in this experiment. The dotted curve, marked "transfer function," shows the phase as a function of frequency derived from the composite-spectrum file measured at the level of 80 dB per octave, for the same animal.

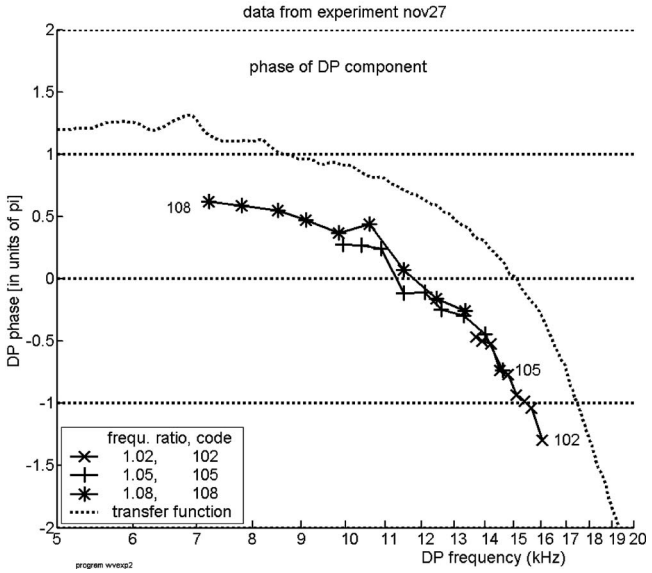


FIG. 4. Experiment "nov27." The phase of the DP with frequency $2f_1-f_2$ produced by presentation of two tones with frequencies f_1 and f_2, plotted against the DP frequency f_{DP}. Tone levels: 70 dB SPL. The DP tone's phase shows an increasing lag with increasing frequency (f_2), denoting a forward traveling wave. The dotted curve, marked "transfer function," shows the phase as a function of frequency, for the same animal, in the same way as in Fig. 3. This curve is used as a guide; it does not run parallel to the DP-data curves.

values for the ratio R_1. For clarity we will plot results for only a few values of R_1.

In the experiments described in this paper the primary tones were presented at a level of 70 dB sound pressure level (SPL), a relatively low level for which we could expect measurable and reliable DPs. Figure 3 shows the phase response of one of the two *primary tones*, the one with frequency f_2, corrected for its individual phase shift Θ_2 as described in Sec. II, for one experiment. The abscissa is the tone's frequency f_2. Each curve of Fig. 3 shows the phase for one value of the frequency ratio R_1 and connects symbols indicating the nine "stages," i.e., positions of the tone pair (where *istage*=1 corresponds to the rightmost point in each curve). For the ratio R_1 we have selected three values from the set of frequency ratios employed, see the legend inside the figure. As expected, the phase curve of the primary tone demonstrates a forward traveling wave. The dotted curve shows the single-tone phase as function of frequency as it has been derived from a composite-spectrum file (see Sec. II) at the level of 80 dB per octave. The phase of the single primary tone deviates slightly from the tone phase derived from the composite-spectrum file, the difference is not consistent enough to be attributed to two-tone interaction.

Figure 4 displays the course of the *phase of the DP with frequency* $2f_1-f_2$ as a function of the DP frequency f_{DP}, for the same experiment. The DP phase is corrected with $2\Theta_1-\Theta_2$ as described above. Each curve shows the phase for nine positions of the tone pairs; the different curves relate to different values of the frequency ratio R_1, see the inside legend. This figure answers the principal question of the present paper. If the DP wave were a pure reverse traveling wave, its phase lag at location L_0 would be smaller for higher than for lower frequencies and its phase function would thus have a positive slope, it would rise upwards. Obviously, this is not the case: the curves bend *downwards*. In terms of the x-domain representation of Fig. 1(B) (the cochlear pattern) this indicates a forward traveling wave. This result demonstrates that the DP wave—which theoretically would be starting approximately from the x_2-region—is not a pure reverse traveling wave, but it appears here in the form of a *forward traveling* wave. We have termed this effect "inverted direction of wave propagation," abbreviated: IDWP. We note that IDWP is similar to the effect found by Ren *et al.* (2006) and thus the data serve as an independent confirmation of the earlier findings. To the figure has been added as a dotted curve the nearly continuous phase response for a pure tone, as in Fig. 3. The general course is the same as that of the data but note: the curves are not parallel.

In this figure—as well as in later figures—we have plotted only data in which the peak in the spectrum at the DP frequency is more than 8 dB above the level of the surrounding components. The DP peaks generally become smaller for lower f_2 and larger R_1 values, and often cease to meet this criterion. When this is the case we have not included these points, and that is why in later figures we see isolated points or groups of points.

Figure 4 is for the experiment labeled "nov27." Figures 5 and 6 show the phase of this type of DP for two other experiments, "nov24" and "nov30," respectively. All three figures demonstrate the same basic feature, IDWP. The latter two figures also indicate the transition from a reverse traveling wave—a small positive slope on the left side of the curves—to the more pronounced negative slope of the forward traveling wave. Figure 7, finally, includes a few results obtained with stronger primary tones, where we may expect clearer evidence of that transition. However, we are unable to show exactly where the forward wave changes into a reverse

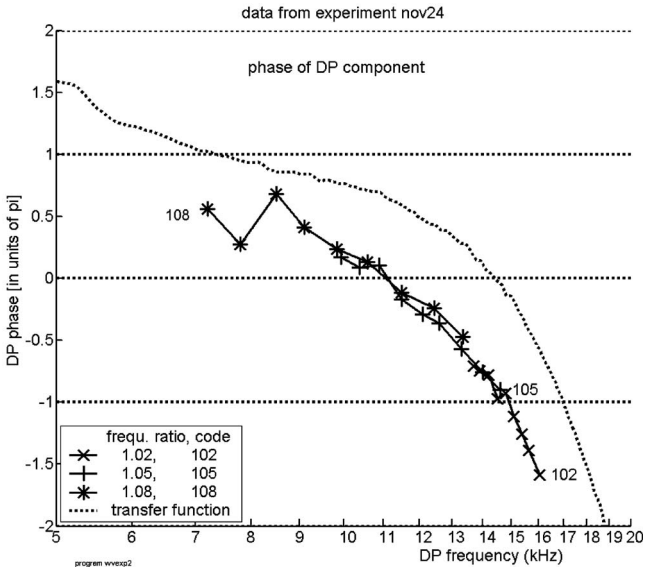


FIG. 5. The same as Fig. 4, but from a different animal, experiment "nov24."

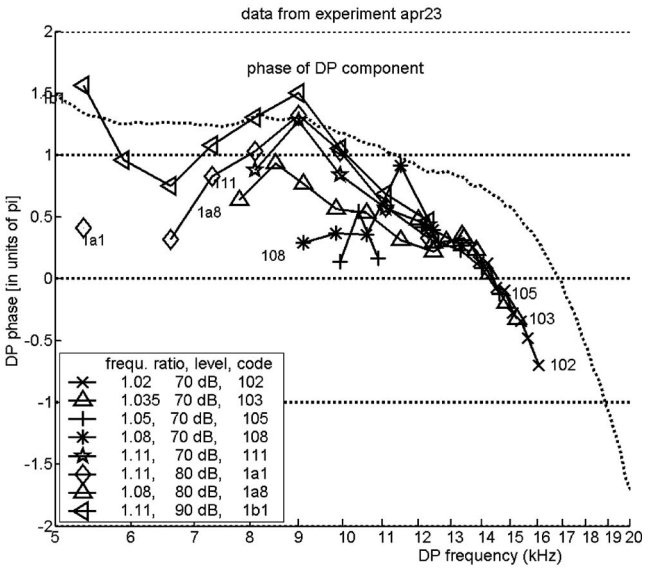


FIG. 7. The same as Fig. 4, but from a different animal, "apr23." In an attempt to display the transition from reverse to forward traveling wave better, a few records with stronger primary tones are included. This attempt was not very successful.

traveling wave because even with the stronger stimuli the signal-to-noise ratios are too small in and below the transition region. Concurrently, deviations and errors in this region are quite large. In every one of these four figures a dotted curve derived from the same experiment has been added. These curves are slightly different but can serve as guidelines.

V. DISCUSSION OF EXPERIMENTAL RESULTS

In our experiments the laser beam was reflected from a gold-coated bead placed on the BM. This allowed us to use fairly weak primary tones (note that the two tones were given the same level). In contrast, in some of his early experiments (done without beads) Ren was forced to use higher levels, and doubts have risen that his results were perhaps

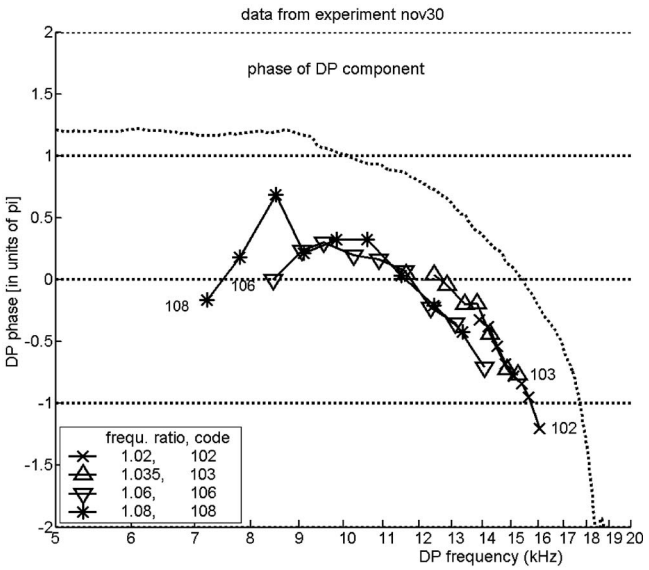


FIG. 6. The same as Fig. 4, but from a different animal, "nov30," with different ratio values. In this case a hint of the transition between a reverse and a forward traveling wave is visible.

due to those high levels. That this is not so has been proven in the most recent series of experiments in his laboratory (He *et al.*, 2007) in which very low levels—down to 40 dB—were used. We have chosen our stimulus level in the intermediate range (basically, 70 dB) in the hope to be able to see both the reverse and forward segments of the DP wave. This hope was not fulfilled; even the use of higher levels did not help us.

In the interpretation of our results we need the assumption that at their peak points the two primary tones have the same phase if they are started at the stapes with the same phase. This condition is equivalent to the condition that the cochlea scales log frequency to location, a property that is generally accepted to be true in the basal turn. In the data of Figs. 4 and 5 the ratio of the highest to the lowest frequency is 2.8, or not more than 1.5 octaves, sufficiently small to justify this assumption.

Another subject involved in interpretation of our results is the influence of reflection at the stapes. Suppose the original reverse traveling wave starting from the x_2-region undergoes substantial reflection at the stapes. Can the resulting forward traveling wave ever be larger in amplitude than the original reverse traveling wave? The answer is yes but this can only occur within (or apical to) a region where appreciable amplification *for the DP frequency* exists. That is, we could only expect a dominant forward traveling DP wave in the "active" region associated with the DP frequency. In practice this is the region apical to that occupied by the primary tones. Clearly, this is not what our data tell us; hence the observed forward traveling wave is not a reflected wave.

In connection with the transition from reverse to forward traveling wave one of the conclusions of Dong and Olson (2007) should be recalled. They found that, in the path from the interaction region to the stapes, there is a net reverse traveling wave. Furthermore, they detected evidence of place-fixed DPs. In their work they report a considerable

degree of variability in the data (possibly due to multiple causes). We also found much variability in our data and we believe this is due to random errors. Nevertheless, the trend of the forward traveling wave is a robust feature of the DP waves in our findings. We will come back to the subject of data errors, place-fixed DPs and variability later on.

A few more general observations can be made. Consider the figures showing the DP phase for different frequency ratios. In almost all cases the curves for different frequency ratios align neatly. Why would that be so? Consider one point on the abscissa, one frequency. When the frequency ratio is increased, keeping the DP frequency the same, the effective location of the DP source moves to the left in the x domain. This shift entails only a small shift of the overall phase pattern, and thus the phase recorded at location L_0 remains nearly the same. In addition, the dotted curves are not exactly parallel to the data curves, which indicates that the various forward DP waves cannot be regarded as if they originate from an invariant location.

VI. GREAT EXPECTATIONS

In a previous paper (de Boer *et al.*, 2007) theoretical predictions for DP waves were presented. The computations were based on a "classical" model of the cochlea, a model in which the BM impedance expresses the precise relation between the BM velocity and the across-membrane pressure difference, at the *same* location. In that case the BM impedance is a *local* parameter, a driving-point impedance which has all the system-theoretical properties of driving-point impedances (Guillemin, 1957; Diependaal and Viergever, 1983). By way of the "inverse solution" the BM impedance can be derived from a given BM response function, see de Boer *et al.* (2007) and its references for details. For the computations in that paper one experiment, coded 19922, was selected in view of the very smooth form of the BM impedance functions, both for low and for high levels. The computations produced the theoretical course of the DP wave with frequency $2f_1-f_2$ in the model, as a function of the location variable x. For the purpose of that paper the frequencies f_1 and f_2 of the two primary tones were chosen to be above the reference frequency f_0 (which simulates the BF of the measurement location L_0). In that earlier work we found that 9 out of 55 experiments produced a BM impedance function that was smooth enough for the purpose of that study.

We now turn to the series of eight experiments performed for the present study. Fortunately, we found two animals of which the data can be used for a similar purpose. Accordingly, the computer program for Figs. 5–7 of that earlier paper (de Boer *et al.*, 2007) was inspected whether it would work with primary-tone frequencies well below f_0. The stimulus-response data on which the figures were based were obtained from stimulation with broadband noise, measured in the same animal. In the earlier program only the frequency range from 6.42 kHz upward was included in the projection to the x axis. To accommodate frequencies below 17 kHz better the model was modified to work upward from lower frequencies. It later turned out that this expedient was not strictly necessary; hence the earlier program was retained with only minor modifications. The abscissa in the figures to follow goes from 0 to 7 mm; the model had a length of 12 mm. The result, the theoretical DP wave derived on the basis of data from experiment 'apr23,' is shown by Fig. 8. Response and underlying impedance functions for this animal are not as smooth as in the earlier paper (data not shown). The thick gray dashed curves show the shapes of the cochlear amplitude patterns produced by two relatively strong tones, presented at approximately 70 dB SPL[2] and having a frequency ratio equal to 1.08. This value is in the middle range of the values used in the preceding figures. The two frequencies f_1 and f_2 are then 14.57 and 15.74 kHz, respectively, and the DP frequency $2f_1-f_2$ is 13.4 kHz. The locations corresponding to these three frequencies are indicated by triangles on the bottom of the figure. The thin dotted curve in the figure, labeled "amplitude of hypothetical $2f_1$-f_2 response," can be seen as a *reference response*. It is the response that a pure tone of 13.4 kHz would produce if it were generated at the stapes. Note that its peak is somewhat sharper than those of the primary tones because the hypothetical response corresponds to a very low level (20 dB), comparable to that of DPs.

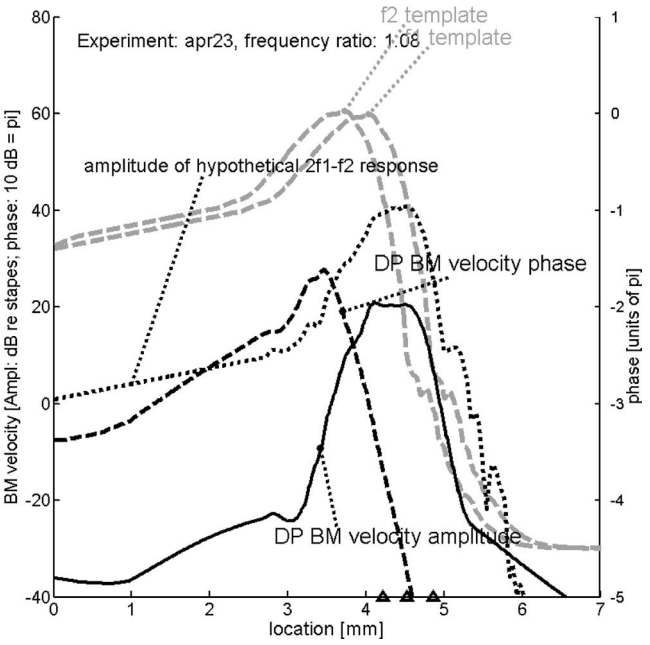


FIG. 8. Solid and coarsely dashed curves: Predicted cochlear pattern of the distortion product (DP) with frequency $2f_1-f_2$ where f_1 and f_2 are the frequencies of the primary tones. Computed with a "classical" model of cochlear mechanics (de Boer *et al.*, 2007) based on responses to noise signals in experiment "apr23"—with low-level BF equal to 19.23 kHz. Reflection of the reverse traveling wave at the stapes has been reduced. Actual length of the model: 12 mm. The triangles on the abscissa denote the locations corresponding to the frequencies f_2, f_1 and $2f_1-f_2$ (from left to right). The location corresponding to 17 kHz is at 3.9 mm. Gray dashed curves: cochlear response patterns produced by primary tones with frequencies of 14.57 and 15.74 kHz, respectively. The DP wave has a frequency of 13.4 kHz. Finely dotted curve: relative amplitude of a hypothetical tone with the frequency $2f_1-f_2$ as if it were generated at the stapes.

The DP wave (with frequency $2f_1-f_2$) is generated by nonlinear distortion in the region where the two patterns generated by the primary tones overlap. The manner in which

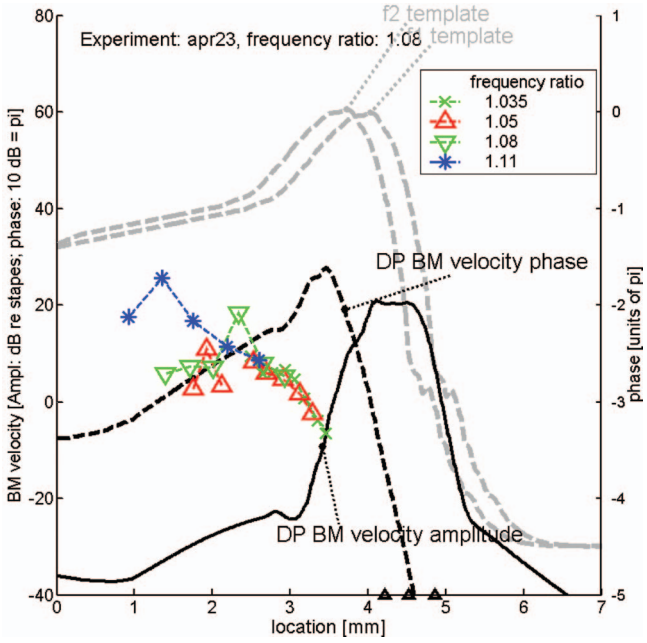


FIG. 9. Partial copy of Fig. 8, with actual data (from the same animal) added in color. Dashed lines with crosses and symbols: selected DP phase responses (Fig. 7) transformed from the (DP) frequency to the location domain. This figure shows the true nature and extent of inverted direction of wave propagation (IDWP) in the cochlea.

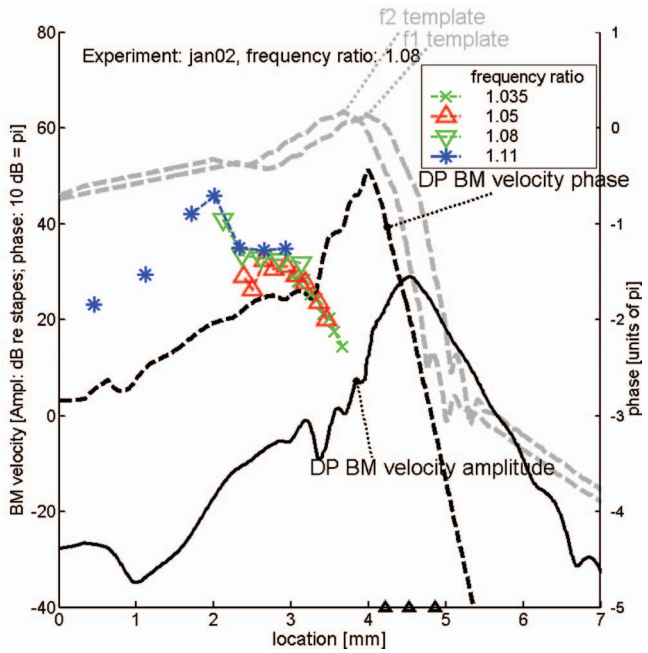


FIG. 10. Predicted amplitude and phase of DP wave, based on responses to noise signals in experiment "jan02" (low-level BF equal to 19.53 kHz). Measured DP phase values for the same experiment are added in color. Layout as in Fig. 9.

the resulting DP wave is computed, from primary-tone excitation patterns via nonlinear distortion in OHCs to the creation of a cochlear DP wave, has been described in detail in de Boer *et al.* (2007). The amplitude of the DP wave, labeled "DP BM velocity amplitude," is shown by the thick solid curve; it shows a peak at the location corresponding to its "own" frequency (13.4 kHz). The thick black dashed curve in the figure, labeled "DP BM velocity phase," shows the computed phase pattern of the DP wave. Note that in the present paper the phase scale has been *expanded* with respect to the figures in our earlier paper. This DP phase curve shows what we intuitively expect the DP wave's phase to be. On the left side of the overlap region the DP wave travels to the left—toward the stapes—and to the right of this region the DP wave travels to the right—in the direction of the helicotrema. In the center of the overlap region the computed phase is more or less constant over a small stretch. In summary, according to the theory of the classical model, a reverse traveling wave travels from the overlap region to the left (towards the stapes) and a forward traveling wave to the right (towards the helicotrema).

Figure 9 is a copy of Fig. 8; it shows the same predicted DP wave (the 13.4 kHz reference response curve has been omitted) for the frequency ratio of 1.08. Added to this figure are selected phase data (for three frequency ratios 1.05, 1.08 and 1.11) *from the same animal* (see Fig. 7), converted from the (DP) frequency to the place domain in the manner indicated by Figs. 1(A) and 1(B), and plotted in color. It is obvious that the measured phase of the DP wave goes in the opposite direction of the predicted response, demonstrating IDWP in the experimental data. Note that the average negative slope in the experimental data (the measured forward traveling wave) is of the same order of magnitude as the positive slope (the reverse traveling wave) in the prediction. From the comparison it becomes clear what is the size and extent of the IDWP effect. At first sight it may appear as a minor deviation, but it cannot be dismissed as a result of errors in the data or a shortcoming of the theory. This means that we should try to discover the origin of this effect. Figure 10 has the same layout as Fig. 9 but shows prediction and data from another one of the animals used in the present study. This figure fully confirms our conclusions from the preceding figure.

VII. GLOBAL THEORETICAL CONSIDERATIONS

The findings of Ren *et al.* (2006) and our present confirmation of them constitute an obstacle to the classical theory of cochlear mechanics. A classical model of the cochlea has a BM of which the mechanics is a *local* property. It has been shown, by de Boer *et al.* (2007) and Shera *et al.* (2007), that in a classical model with smoothly varying parameters *only* a reverse traveling wave arises between the response peak of the f_2 tone and the stapes. Our Fig. 8 confirms this prediction for the present set of conditions. Therefore, IDWP in the sense as we mean it here is not compatible with current theory.

Classical and nonclassical models and their interrelation have been studied by the first-named author (de Boer, 1997). In a classical model forward and reverse traveling waves undergo the same amount of amplification or attenuation and have the same velocities of propagation. In a nonclassical model this premise is not true. Indeed it has been found that a feed-forward model does not have the same properties for forward and reverse traveling waves (de Boer, 2007). If the feed-forward model is designed in such a way as to show wave *amplification* for forward traveling waves, it will show

wave *attenuation* for reverse traveling waves. A similar property is valid for a feed-backward model: if the model is designed to produce amplification for forward traveling waves, it has attenuation for reverse traveling waves. On the other hand, the *propagation velocity* is the same for forward and reverse traveling waves.

Examples of nonclassical models are feed-forward models (Steele *et al.*, 1993; Geisler and Sang, 1995; Fukazawa, 1997; Lim and Steele, 2003) and feed-backward models. A combination of these two has been analyzed, too (Steele *et al.*, 1993; de Boer and Nuttall, 2003; Wen and Boahen, 2003), because neither form is realistic by itself. Other examples of nonclassical models are models with more than two compartments (e.g., Hubbard, 1993). Because a general unifying theory of these models does not exist we are forced to leave these models out of consideration.

It will not be surprising that a feed-forward model designed to amplify forward traveling waves can show IDWP for the DP wave. This is so because the reverse traveling DP wave starting in the f_2-region will be rapidly attenuated so that the forward traveling wave which also arises in that region will dominate. However, a pure feed-forward model is not satisfactory because such a model would show exceedingly weak otoacoustic emissions. Furthermore, the process of coherent reflection (Zweig and Shera, 1995; Shera *et al.*, 2005; de Boer and Nuttall, 2006b), a prerequisite for a valid cochlear model, would be severely hampered.

As an alternative, we can consider corrugations in the mechanics of the cochlear partition. In Fig. 8 of the present paper we have used data from a particular animal to derive predictions on wave travel. We have selected that animal on the basis of smoothness of response and BM impedance functions, in other words, on the basis of minimal corrugations. The theory that we applied to derive the predicted response includes the possibility of internal reflections from BM corrugations. In the case of Fig. 8 the computation does not predict IDWP. In other animals, as in some of those of the present series of experiments, response and BM impedance functions are not as smooth, and the corrugations could have given rise to local reflection. However, it has not been found possible to force IDWP to occur by imposing *random* cochlear irregularity on a model with a smooth BM impedance function (data not shown, for the technique see de Boer and Nuttall, 2006b). As a last possibility, we should consider inadequacy of our method of data processing, i.e., the "inverse" solution. It may be that the actual "active" process in the cochlea extends over a larger distance towards the stapes than is revealed by our BM impedance results. We have little or no evidence on such a shortcoming. In summary, a satisfactory explanation of the phenomenon of IDWP has not yet been found.

VIII. CONCLUSIONS

We have executed a series of experiments on the phase of distortion products (DPs), directed at finding the nature of the DP-associated wave inside the cochlea. We focused on the DP with frequency $2f_1-f_2$ where f_1 and f_2 are the frequencies of the two stimulus tones. The movements of the BM were measured with a laser interferometer, at a location L_0 with a best frequency around 17 kHz. The frequencies f_1 and f_2 always were below 17 kHz and were varied over more than one octave, keeping their frequency ratio constant. In this way the cochlear DP wave was monitored between the generation location (assumed to be near the f_2-associated location) and the (fixed) observation location L_0. It was found that in this region a segment of the DP wave runs in a counter-intuitive direction: it is a forward and not a reverse traveling wave. Our data confirm the findings of Ren *et al.* (2006) and He *et al.* (2007) in this respect. We conclude that for the DP with frequency $2f_1-f_2$ a substantial forward traveling wave exists in the region around the location associated with the frequency f_2, and that a reverse traveling wave develops more towards the stapes. These findings are presenting an obstacle to present-day cochlear theory. Further study is necessary to find out under which conditions this anomaly of wave propagation (which we call IDWP) exists before it can be attempted to explain this phenomenon.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the continual feedback provided by Christopher A. Shera. This study received support from Grant Nos. NIH NIDCD, R01 DC00141, and 1-P30-DC005983.

[1]There is a small difference in shape resulting from the frequency dependence of the boundary condition at the stapes but in the figure this has been neglected.

[2]As in the cited paper, amplitudes of primary tones are not drawn to scale.

Allen, J. B., and Fahey, P. F. (**1992**). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," J. Acoust. Soc. Am. **92**, 178–188.

de Boer, E. (**1997**). "Classical and nonclassical models of the cochlea," J. Acoust. Soc. Am. **101**, 2148–2150.

de Boer, E. (**2007**). "Forward and reverse waves in nonclassical models of the cochlea," J. Acoust. Soc. Am. **121**, 2819–2821.

de Boer, E., and Nuttall, A. L. (**2000**). "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507.

de Boer, E., and Nuttall, A. L. (**2003**). "Properties of amplifying elements in the cochlea," in *Biophysics of the Cochlea: From Molecules to Model*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 331–342.

de Boer, E., and Nuttall, A. L. (**2006a**). "Amplification via "compression waves" in the cochlea—a parable," *Association for Research in Otolaryngology, Mid-Winter-Meeting*, Baltimore, MD, Abstracts **29**, 349.

de Boer, E., and Nuttall, A. L. (**2006b**). "Spontaneous basilar-membrane oscillation (SBMO) and coherent reflection," J. Assoc. Res. Otolaryngol. **7**, 26–37.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (**2005**). "The Allen–Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1260–1267.

de Boer, E., Nuttall, A. L., and Shera, C. A. (**2007**). "Wave propagation patterns in a "classical" three-dimensional model of the cochlea," J. Acoust. Soc. Am. **121**, 352–362.

Diependaal, R. J., and Viergever, M. A. (**1983**). "Point-impedance characterization of the basilar membrane in a three-dimensional cochlea model," Hear. Res. **11**, 33–40.

Dong, W., and Olson, E. S. (**2008**). "Evidence for reverse cochlear traveling waves," J. Acoust. Soc. Am. **123**, 222

Fukazawa, T. (**1997**). "A model of cochlear micromechanics," Hear. Res. **113**, 182–190.

Geisler, C. D., and Sang, C. (**1995**). "A cochlear model using feed-forward outer-hair-cell forces," Hear. Res. **86**, 132–146.

Guillemin, E. A. (**1957**). *Synthesis of Passive Networks* (Wiley, New York).

Goodman, S. S., Withnell, R. H., de Boer, E., Lilly, D. J., and Nuttall, A. L. (**2004**). "Cochlear delays measured with amplitude-modulated tone-burst evoked OAEs," Hear. Res. **188**, 57–69.

He, W. X., Nuttall, A. L., and Ren, T. (**2007**). "Two-tone distortion at different longitudinal locations on the basilar membrane," Hear. Res. **228**, 112–122.

Hubbard, A. E. (**1993**). "A traveling wave-amplifier model of the cochlea," Science **259**, 68–71.

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 emission from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Lim, K.-M., and Steele, C. R. (**2003**). "Response suppression and transient behavior in a nonlinear active cochlear model with feed-forward," Int. J. Solids Struct. **40**, 5097–5107.

Nuttall, A. L., Dolan, D. F., and Avinash, G. (**1991**). "Laser Doppler velocimetry of basilar membrane vibration," Hear. Res. **51**, 203–214.

Nuttall, A. L., Grosh, K., Zheng, J., de Boer, E., Zou, Y., and Ren, T. (**2004**). "Spontaneous basilar membrane oscillation and otoacoustic emission at 15 kHz in a guinea pig," J. Assoc. Res. Otolaryngol. **5**, 337–349.

Ren, T. (**2004a**). "Propagation direction of the otoacoustic emission along the basilar membrane," *Association for Research in Otolaryngology, Mid-Winter-Meeting*, Daytona Beach, FL, Abstracts, **27**, 343.

Ren, T. (**2004b**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Ren, T., He, W. X., and Nuttall, A. L. (**2006**). "Backward propagation of otoacoustic emissions in the cochlea," in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 79–85.

Ruggero, M. A. (**2004**). "Comparison of group delay of $2f_1-f_2$ distortion product otoacoustic emissions and cochlear travel times," ARLO **5**, 143–147.

Shera, C. A. (**2006**). "Four counter-arguments for slow-wave OAEs," in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall, T. Ren, P. Gillespie, K. Grosh, and E. de Boer (World Scientific, Singapore), pp. 449–455.

Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.

Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahey, P. A., and Guinan, J. J. (**2007**). "Allen-Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.

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.

Steele, C. R., Baker, G., Tolomeo, J., and Zetes, D. (**1993**). "Electromechanical models of the outer hair cell," in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 207–214.

Wen, B., and Boahen, K. (**2003**). "A linear cochlear model with active bi-directional coupling," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Cancun, Mexico), Vol. **3**, pp. 2013–2016.

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.

Forward-masked spatial tuning curves in cochlear implant users

David A. Nelson[a)]
Clinical Psychoacoustics Laboratory, Department of Otolaryngology, University of Minnesota, MMC396, 420 Delaware Street S.E., Minneapolis, Minnesota 55455

Gail S. Donaldson
Department of Communication Sciences and Disorders, University of South Florida, PCD 1017, 4202 E. Fowler Avenue, Tampa, Florida 33620

Heather Kreft
Clinical Psychoacoustics Laboratory, Department of Otolaryngology, University of Minnesota, MMC396, 420 Delaware Street S.E., Minneapolis, Minnesota 55455

(Received 19 March 2007; revised 1 January 2008; accepted 2 January 2008)

Forward-masked psychophysical spatial tuning curves (fmSTCs) were measured in twelve cochlear-implant subjects, six using bipolar stimulation (Nucleus devices) and six using monopolar stimulation (Clarion devices). fmSTCs were measured at several probe levels on a middle electrode using a fixed-level probe stimulus and variable-level maskers. The average fmSTC slopes obtained in subjects using bipolar stimulation (3.7 dB/mm) were approximately three times steeper than average slopes obtained in subjects using monopolar stimulation (1.2 dB/mm). Average spatial bandwidths were about half as wide for subjects with bipolar stimulation (2.6 mm) than for subjects with monopolar stimulation (4.6 mm). None of the tuning curve characteristics changed significantly with probe level. fmSTCs replotted in terms of acoustic frequency, using Greenwood's [J. Acoust. Soc. Am. **33**, 1344–1356 (1961)] frequency-to-place equation, were compared with forward-masked psychophysical tuning curves obtained previously from normal-hearing and hearing-impaired acoustic listeners. The average tuning characteristics of fmSTCs in electric hearing were similar to the broad tuning observed in normal-hearing and hearing-impaired acoustic listeners at high stimulus levels. This suggests that spatial tuning is not the primary factor limiting speech perception in many cochlear implant users.
 [DOI: 10.1121/1.2836786]

PACS number(s): 43.64.Me, 43.66.Fe, 43.66.Dc, 43.66.Mk [AJO] Pages: 1522–1543

I. INTRODUCTION

A multichannel cochlear implant is designed to take advantage of the tonotopic arrangement of auditory nerve fibers within the cochlea. Under ideal circumstances, the current delivered through a given electrode stimulates a discrete group of auditory nerve fibers that reside close to that electrode. This allows information in adjacent frequency bands of the acoustic stimulus to be transmitted to adjacent groups of auditory nerve fibers, roughly mimicking tonotopic stimulation patterns in acoustic hearing. In practice, however, at least two factors can alter the desired "electrotopic" (electrode-to-place) mapping. First, cochlear implant users may have irregular patterns of auditory nerve survival (Hinojosa and Lindsay, 1980; Spoendlin and Schrott, 1988; Zappia *et al.*, 1991; Nadol *et al.*, 2001). If no nerve fibers exist near a particular electrode, then current delivered through that electrode will necessarily stimulate auditory fibers that are apical and/or basal to the intended location. Second, structural changes to the cochlea following deafness and cochlear implantation may result in irregularities in the impedance pathways that govern current flow (Spelman *et al.*, 1982; Black *et al.*, 1983; Shepherd *et al.*, 1994; Jolly, 1998; Hughes *et al.*, 2001; Saunders *et al.*, 2002). This may result in stimulation of auditory nerve fibers remote from the electrode contact in either the apical or basal direction. In either case, neural excitation may be displaced from the intended neural targets.

Sound perception through a cochlear implant may also be affected by the spatial extent of current fields delivered through individual electrode contacts. The "spatial selectivity" of these current fields determines the degree of overlap between adjacent frequency channels, thereby influencing the maximum spectral resolution that can be supported by the device. The spatial selectivity of current fields is primarily determined by the mode of electrical coupling, with bipolar coupling producing narrower current fields than monopolar coupling. However, the corresponding neural activation patterns are influenced by additional factors, such as the density of neurons in particular regions of the cochlea, and the radial distance between the electrode array and neural targets in the modiolus (Kral *et al.*, 1998; Liang *et al.*, 1999; Cohen *et al.*, 2001; Saunders *et al.*, 2002; Skinner *et al.*, 2002; Cohen *et al.*, 2003). Since these parameters vary across individuals, they can be expected to result in varying degrees of neural spatial selectivity across individual cochlear implant users, even when the mode of electrode coupling is held constant.

[a)]Electronic mail: dan@umn.edu

Together the two factors just described—aberrations in the electrotopic map and limited spatial selectivity—reduce spectral resolution by restricting the amount and accuracy of spectral information that can be transmitted to cochlear implant listeners. Given the critical role that spectral cues play in sound perception, these factors also represent key limitations to overall device benefit. It may be possible to partly compensate for reduced spectral resolution by using alternative modes of stimulation, such as tripolar stimulation (Jolly *et al.*, 1996; Kral *et al.*, 1998; Morris and Pfingst, 2000; Bierer, 2007) and dual-electrode stimulation (McDermott and McKay, 1994; Wilson *et al.*, 1994; Busby and Plant, 2005; Donaldson *et al.*, 2005; Dingemanse *et al.*, 2006), or by devising custom frequency-to-electrode mappings for individual users. However, robust measures of spatial selectivity are needed to evaluate the effectiveness of such approaches.

In acoustic hearing, the forward-masked psychophysical tuning curve (fmPTC) has been used to evaluate the frequency selectivity of the auditory system and to identify irregularities in the tonotopic map. Frequency selectivity has been quantified by measuring the slopes of the apical and basal sides of the fmPTC and by computing the relative bandwidth (Q value) of the fmPTC at some criterion level above its tip (e.g., Nelson and Freyman, 1984; Nelson, 1991). Irregularities in the tonotopic map have been documented by evaluating fmPTC tip frequencies. Tips that are displaced in frequency signify "dead regions" in the cochlea where hair cells are no longer functional and amplification may not be effective (Moore *et al.*, 2000; Moore, 2001; Moore and Alcantara, 2001).

In electric hearing, forward masking has been used to evaluate spatial selectivity in the form of spatial masking patterns. However, these masking patterns involve relatively high masker amplitudes (Lim *et al.*, 1989; Fu *et al.*, 1997; Chatterjee and Shannon, 1998; Collins and Throckmorton, 1998; Cohen *et al.*, 2001; Boex *et al.*, 2003; Dingemanse *et al.*, 2006; Kwon and van den Honert, 2006), which preclude the assessment of tuning in localized regions of the cochlea. Therefore, in this paper, we introduce a new measure that is a variation of the acoustic fmPTC. This measure, which we call the *forward-masked spatial tuning curve* (fmSTC), utilizes low-level probes to quantify spatial selectivity and assess electrotopic aberrations in localized regions of the surviving neural array within the cochlea. The measurement paradigm is similar to that used to measure fmPTCs, allowing us to directly compare spatial tuning in electric hearing with frequency tuning in acoustic hearing. In acoustic hearing, tuning curves are extremely narrow at low stimulus levels, owing to nonlinear processes associated with cochlear outer hair cells (e.g., Nelson *et al.*, 2001; Nelson and Schroder, 2004). At higher stimulus levels, the cochlea is dominated by passive mechanical processes, and tuning curves are substantially broader (Nelson, 1991). Since electrical stimulation bypasses the cochlea, such nonlinear behaviors should not be observed in electrically stimulated fmSTCs. Instead, electrical stimulation should produce tuning curves that are essentially constant with level. The slopes and bandwidths of these tuning curves should reflect the attenuation characteristics of the current stimulus and patterns of surviving neurons, rather than mechanical properties of the cochlea.

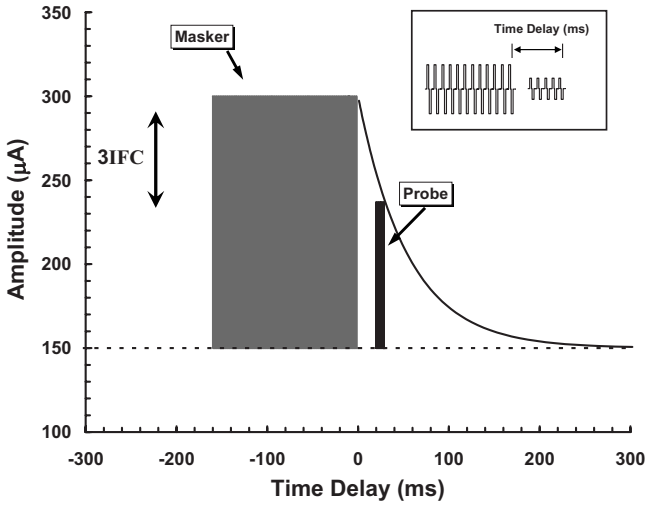


FIG. 1. Schematic diagram of the forward-masking paradigm used to obtain spatial tuning curves. Masker/probe amplitude (μA) is shown as a function of time (ms) relative to the offset of the masker pulse train. Masker and probe biphasic pulse trains are depicted by the inset. Probe threshold is shown by the horizontal dashed line. The masker pulse train is shown by the shaded rectangle, and the probe pulse train is shown by the black narrow rectangle. Expected recovery from forward masking, as a function of time delay, is depicted by the exponential curve following the masker. The probe pulse train is fixed in level at a predetermined time delay and the amplitude of the masker pulse train is adjusted (with a 3IFC adaptive procedure) until the masker just masks the probe.

The goals of the present study are: (1) to characterize fmSTCs for probe electrodes near the middle of the electrode array in cochlear implant users stimulated with either a bipolar electrode configuration (Nucleus N22 subjects) or a monopolar electrode configuration (Clarion C-I or C-II subjects); (2) to confirm that probe level has no systematic effect on fmSTC slopes; and (3) to compare fmSTC-based measures of spatial selectivity in cochlear implant listeners to previously reported fmPTC-based measures of frequency selectivity in normal-hearing and hearing-impaired acoustic listeners.

II. METHODS

A. Stimuli and procedures

1. Forward-masking paradigm

Our procedure for obtaining fmSTCs uses a forward-masking paradigm (Shannon, 1983a, 1986, 1990a; Chatterjee and Shannon, 1998; Chatterjee, 1999; Nelson and Donaldson, 2002) to measure a constant amount of forward masking on a specific *probe* electrode, for multiple *masker* electrodes surrounding that probe. This paradigm is illustrated in Fig. 1. A short train of probe pulses is presented at a fixed time delay following a longer train of masker pulses. The masker is presented in three sequential listening intervals, and the probe follows the masker in one randomly chosen interval. The subject's task is to choose the interval that contains the probe stimulus. The amplitude of the probe stimulus is fixed in level, and the amplitude of the masker stimulus is adjusted (vertical double arrow) using an adaptive procedure (de-

scribed in the following), until the amount of forward masking produced is just sufficient to mask the probe. In Fig. 1, the time course of recovery from forward masking is represented by the black curve. In this study, the masker duration was 160 ms, the time delay was 20 ms, and the probe duration was 10 ms. As indicated in the inset to Fig. 1, time delay was specified as the time between masker offset and probe offset, because the final portion of the probe primarily determines the probe detection threshold in forward masking (Nelson and Donaldson, 2002).

For Nucleus subjects, experiments were carried out using a computer linked to a specialized cochlear implant interface (Shannon *et al.*, 1990). For Clarion subjects, experiments were carried out using a computer with a special-purpose microprocessor that controlled a dedicated speech processor (Clarion Research Interface). The stimuli used for Nucleus subjects were trains of 500-Hz biphasic current pulses. The biphasic pulses were cathodic first (relative to the active electrode), with a per-phase pulse duration of 200 μs and an interphase delay of 44 μs. Stimulus amplitudes were specified in clinical amplitude steps, called *current step units* or CSUs, which are uneven amplitude steps that vary between 0.07 and 0.30 dB for the range of current amplitudes used here. CSUs were converted to calibrated current amplitudes (μA) using subject-specific tables provided by Cochlear Corporation. One Nucleus subject (N09) was tested as part of an earlier study that used different stimulation and forward-masking parameters. For this subject, stimuli were trains of 125-Hz biphasic pulses, masker duration was 256 ms, probe duration was 32 ms, and the time delay between masker and probe was 42 ms.

For the Clarion subjects, stimuli were trains of 500-Hz biphasic current pulses, with a per-phase duration of 77 μs and no delay between phases. Stimulus amplitudes were specified in clinical amplitude steps, called *stimulus units* or SUs, which are logarithmic amplitude steps of 0.15–0.30 dB. SUs were translated to calibrated amplitudes (μA) using a set of look-up tables developed in our laboratory. These tables compensate for nonlinearities in the current source that depend upon electrode impedance and pulse rate. Electrode impedances for Clarion subjects were measured at the beginning and end of each data collection session. The impedances measured for individual subjects typically varied less than 10% across sessions. Variance associated with impedance measurements can produce up to 10% error in calibrated amplitudes for Clarion C-I subjects, but has essentially no effect on calibrated amplitudes for Clarion C-II subjects, due to the improved current source incorporated in the C-II device. The variance associated with the impedance measurements will, of course, be reflected in the variability of the forward-masked thresholds, as represented by error bars shown for individual fmSTCs in later figures.

2. Absolute thresholds and maximum acceptable loudness levels

Prior to obtaining fmSTCs for a particular probe electrode, absolute detection threshold (THSp) and maximum acceptable loudness level (MALp) were determined for the 10-ms probe pulse train. THSp was measured with a three-interval forced choice (3IFC) adaptive procedure similar to that used for measuring masked thresholds (described in the following), but only one interval contained the probe stimulus and the other two contained silence. MALp was measured with an ascending method of limits procedure in which pulse trains, presented at a rate of 2/s, were slowly increased in amplitude until the subject indicated that loudness had reached a "maximum acceptable" level. Estimates for two consecutive ascending runs were averaged to obtain a single measure of MALp. THSp and MALp were measured at the start of each test session for the probe electrode to be evaluated in that session, and THSp was measured again at the end of the test session. Corresponding measures of threshold (THSm) and maximum acceptable loudness level (MALm) were obtained for all masker electrodes, using the 160-ms masker pulse train. These measures were obtained both before and after the measurement of a series of fmSTCs at different probe levels. The values of THSm and MALm reported here represent the average of measures obtained across all data collection sessions for a given subject.

3. Adaptive masked-threshold procedure

As mentioned earlier, forward-masked thresholds were obtained using a 3IFC adaptive procedure. The masker pulse train was presented in each of three listening intervals. The probe pulse train was presented in one of the three intervals, chosen randomly from trial to trial, at a fixed time delay following the masker. The subject's task was to choose the "different" interval by pressing the appropriate button on a three-button computer mouse or by selecting one of three colored squares on a video screen using a standard mouse. Stimulus intervals were cued by the appearance of the three squares on the video screen, and correct-answer feedback was provided after each trial. The amplitude of the masker pulse train was initially set to a level 2–4 dB below the anticipated masked threshold. Masker amplitude was adjusted on subsequent trials using a two-up, one-down stepping rule that estimated the stimulus level corresponding to 70.7% correct discrimination (Levitt, 1971). For the first four reversals in the direction of amplitude changes, masker level was varied in l-dB steps (0.5-dB steps in a few subjects with very small dynamic ranges). These initial reversals quickly moved the adaptive track into the target region for masked threshold. After the fourth reversal, step size was reduced, typically to one-fourth of the initial step size, and remained constant for all remaining trials. Trials continued until a total of 12 reversals occurred. The mean of the final 8 reversals was taken as the masked threshold.

Masked thresholds were determined in this manner for maskers on electrodes surrounding and including the probe electrode. Testing began with the masker on the probe electrode and proceeded to sequentially more apical or more basal maskers. The order of testing electrodes (apical-direction-first versus basal-direction-first) was alternated on consecutive retests. Retests continued in the initial and subsequent test sessions until three or more masked thresholds

TABLE I. Subject information. Gender, age when tested, etiology of deafness, duration of bilateral severe-to-profound hearing loss prior to implantation, duration of implant use prior to the study, implanted device with electrode type (SPRL=spiral; HF=HiFocus; HFP=HiFocus with positioner) for Clarion users only, and insertion depth.

Subject	M/F	Age	Etiology	Duration (years)	CI Use (years)	Device (electrode type)	Depth (mm)
C03	F	53.4	Progressive, familial	27	4.2	C-I (SPRL)	25
C05	M	47.8	Unknown, sudden	<1	5.6	C-I (SPRL)	25
C16	F	50.1	Progressive	13	2.6	C-I (HF)	25
C18	M	71.4	Otosclerosis	33	4.6	C-I (HFP)	25
C23	F	42.7	Progressive	27	1.1	C-I (HFP)	25
D08	F	54.7	Otosclerosis	13	3.8	C-II (HF)	25
N09	M	67.4	Meniere's disease	1	10.6	N22	22
N13	M	65.2	Progressive, familial	4	12.8	N22	24
N14	M	58.0	Progressive	1	8.4	N22	25
N28	M	65.6	Meningitis	<1	8.6	N22	25
N32	M	37.2	Maternal rubella	<1	7.4	N22	23
N34	F	57.9	Mumps, progressive	9	4.4	N22	22

were obtained for each masker electrode. Each point on the fmSTC was based on the average of three to five forward-masked thresholds.

B. Subjects and electrodes

Subjects were 12 postlingually deafened adults, 6 implanted with a Nucleus 22 device (Patrick and Clark, 1991), 5 implanted with a Clarion C-I device (Schindler and Kessler, 1993; Kessler, 1999), and 1 implanted with a Clarion C-II device (Frijns *et al.*, 2002). The Nucleus users were implanted with a 22-electrode straight array. The Clarion users were implanted with a 16-electrode Spiral array (SPRL), a 16-electrode HiFocus array (HF), or a 16-electrode HiFocus array with an electrode positioning system (HFP). Table I displays relevant information for each subject including gender, age, etiology of deafness, duration of hearing loss prior to implantation, duration of implant use prior to participation in the study, electrode type (Clarion users), and insertion depth.

For each of the 12 subjects, a fmSTC was obtained for a probe electrode near the middle of the array. Two to four probe levels were assessed at current amplitudes between 8% and 33% of the probe dynamic range in microamperes (DRμA).

Nucleus subjects were stimulated in bipolar mode, using an electrode separation of 0.75 mm (BP), or the narrowest separation greater than 0.75 mm [either 1.5 mm (BP+1) or 2.25 mm (BP+2)] that allowed MALp and MALm to be achieved at current amplitudes that could be produced by the device. Nucleus electrodes were numbered sequentially from 1 to 22, beginning with the most apical electrode. This numbering scheme is in reverse order from the one used clinically.

Clarion subjects were stimulated in monopolar mode. The Clarion Spiral electrode array (SPRL) incorporates 8 pairs of lateral and medial electrodes. For this array, electrodes were numbered sequentially from 1 to 16, beginning with the most apical electrode; thus all odd-numbered electrodes were lateral electrodes and all even-numbered electrodes were medial electrodes. For subjects with the Clarion C-I receiver and SPRL array (C03 and C05), only medial electrodes were stimulated, which were separated by 2.0 mm. The Clarion HiFocus electrode array (HF) incorporates 16 electrodes in a linear arrangement; these electrodes were numbered 1–16, beginning with the most apical electrode. Subjects with the Clarion C-I receiver and HF array (C16, C18, and C23) were stimulated on the same electrodes used in their clinical speech-processor programs (either all even-numbered or all odd-numbered electrodes), which were separated by 2.2 mm. The remaining electrodes were not used because lack of regular stimulation resulted in high electrical impedances. One subject with the Clarion C-II receiver and HF array (D08) was tested on all 16 electrodes, which were separated by 1.1 mm.

Because the electrode-numbering schemes used here differ in some cases from clinical numbering schemes, electrode numbers are specified using the prefix "rEL", to indicate research electrode numbering.

C. Parameters of the fmSTC

Figure 2 shows an example fmSTC from a Nucleus subject. Figure 2 will be used to familiarize the reader with our standard format for presenting fmSTCs and with the quantitative measures used to describe them.

Probe electrode: The heading on the graph specifies the subject (N14), the probe electrode pair (rEL12:11), and the electrode configuration of the probe (BPp) and the masker (BPm) used to generate the fmSTC. The *x* axis shows electrode number, with more apical electrodes toward the left and more basal electrodes toward the right. The *y* axis shows stimulus amplitude in microamperes. The horizontal positions of the open circle and arrow identify the active and reference electrodes, respectively, of the bipolar probe; the vertical level of these symbols indicates the probe level (in microamperes) used to obtain the fmSTC. The level of the horizontal bar connecting the probe and reference electrodes indicates the absolute threshold of the probe (THSp), which was 488 μA in this example. If monopolar stimulation had

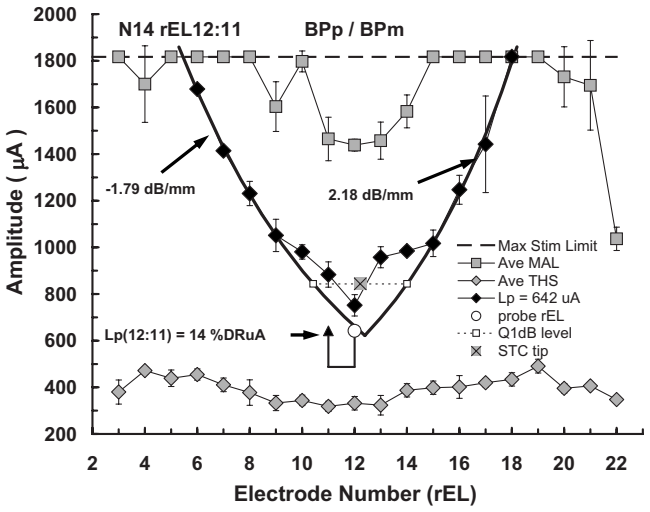


FIG. 2. Example forward-masked spatial tuning curve (fmSTC) from a Nucleus subject. Stimulus amplitude (μA) is shown on the ordinate; research electrode number (rEL) is shown on the abscissa. The rEL numbering system assigns 1 to the most apical electrode with consecutive numbering proceeding to the basal electrode. Average maximum acceptable loudness (Ave MAL) levels obtained with the masker stimulus are shown as shaded squares. Standard deviations are shown by the error bars. The dashed line across the top of the graph indicates the maximum current amplitude that can be generated by this subject's device. Average 3IFC thresholds (Ave THS) obtained with the masker stimulus are shown by the shaded diamonds, because this subject used a Nucleus device, stimuli were delivered in bipolar mode. The horizontal positions of the open circle and closed triangle identify the "active" and "reference" probe electrodes, respectively; probe level is indicated by the vertical position of these two symbols, with the height of the vertical lines indicating the sensation level of the probe; the horizontal line below the symbols indicates the subject's absolute threshold for the probe stimulus. For each fmSTC, masker levels required to forward-mask the probe stimulus, at each masker electrode, are shown by the closed black symbols. Masked thresholds were obtained for all testable electrodes, but those that reached the subject's MAL (or the maximum stimulation limit of the device) on any electrode are not plotted in the graphs. Heavy solid curves on each side of the fmSTC show the logarithmic fitted slopes. Slope values are often specified next to each curve. In this example fmSTC, the results of the bandwidth calculation are shown: The dashed line shows the level 1 dB above the tip, and the open squares show the apical and basal limits of the bandwidth measure. The fmSTC tip calculation is shown in this example by the shaded $\times$ symbol. Labels at the top of the graph indicate the subject code (N14), the probe electrode (rEL12:11) (active:reference research electrode numbers), and the mode of stimulation for the probe and masker (BPp/BPm).

been used, then the probe electrode would be indicated by a single open circle and a vertical bar extending to the level of the probe threshold. Because the probe electrode is fixed in level, and because the stimulus amplitude on the probe is relatively low, it is assumed that the probe is exciting a constant, relatively small population of neurons that may or may not be close to the probe electrode.

The absolute threshold of the probe (represented by the horizontal bar) was typically higher than the absolute threshold for the masker on the same electrode. This can be attributed to temporal integration (Donaldson *et al.*, 1997), since the masker duration was approximately ten times greater than the probe duration.

Masker electrodes: The plot of forward-masked thresholds across electrodes defines a fmSTC. In this example, forward-masked thresholds are represented by the solid diamonds. The corresponding legend entry indicates the level of the probe used to obtain these thresholds (Lp=642 μA). Forward-masked thresholds were always determined for all test electrodes in the array; however, for clarity, forward-masked thresholds that reached the MALm on a particular masker electrode are not plotted. The fmSTC defines a spatial region over which masker stimuli interfere with the perception of the probe stimulus. Masker stimuli within and above the fmSTC function mask the probe; those outside and below the function do not.

Average values of MALm and THSm are shown by the shaded squares across the top of the graph and the shaded diamonds across the bottom of the graph, respectively. Error bars indicate 1 s.d. above and below each value. The dashed line across the top of the graph indicates the maximum current amplitude that can be generated by this subject's device. Maximum stimulation limits are relevant only for subjects with the Nucleus implant, and vary across individuals according to the calibration tables provided by Cochlear Corp.

Slopes, bandwidths, and tip shifts: In order to quantify spatial tuning characteristics across subjects and probe levels, each fmSTC was fitted with two logarithmic functions, one on the apical side and one on the basal side. As indicated by the heavy solid curves in Fig. 2, only the steepest portion of each side was included in the least-squares fit. Typically, the function included three or more masked thresholds. However, in cases of extremely steep slopes, or when the tested electrodes had larger physical separations, only two masked thresholds were sometimes included. The fitted slopes were expressed in units of dB/mm. The spatial bandwidth of the fmSTC was calculated as the distance (in millimeters) between the apical and basal fitted slopes at an amplitude 1 dB above the lowest masker level. We refer to this measure as the fmSTC *bandwidth*. In Fig. 2, bandwidth is represented by the two open squares connected by a dashed line. The fmSTC *tip place*, illustrated by the shaded $\times$ symbol, was specified at the midpoint of the bandwidth, and the *tip shift* was specified as the difference between the tip place and the probe electrode place (in millimeters). These procedures for quantifying fmSTCs focus on the steepest portion of the function and ignore irregularities that sometimes occurred close to the probe electrode. Such irregularities were common among the Nucleus subjects tested in bipolar mode, and will be described later for individual subjects.

Statistical analysis. The effect of probe level on each fmSTC parameter (apical or basal slope, bandwidth, or tip shift) was evaluated by performing a linear regression on that parameter as a function of probe amplitude, with probe amplitude specified as a percentage of probe dynamic range in microampere (%DRμA). Separate regression analyses were performed for monopolar (Clarion) and bipolar (Nucleus) subjects. The effect of subject group (monopolar versus bipolar) was evaluated by first computing the average parameter value across probe levels for each subject, and then performing a two-tailed T-test to assess differences between the two groups. The use of average data (collapsed across probe level) for individual subjects was justified because probe level was found to have no systematic effect on fmSTC parameters for either group.[1]

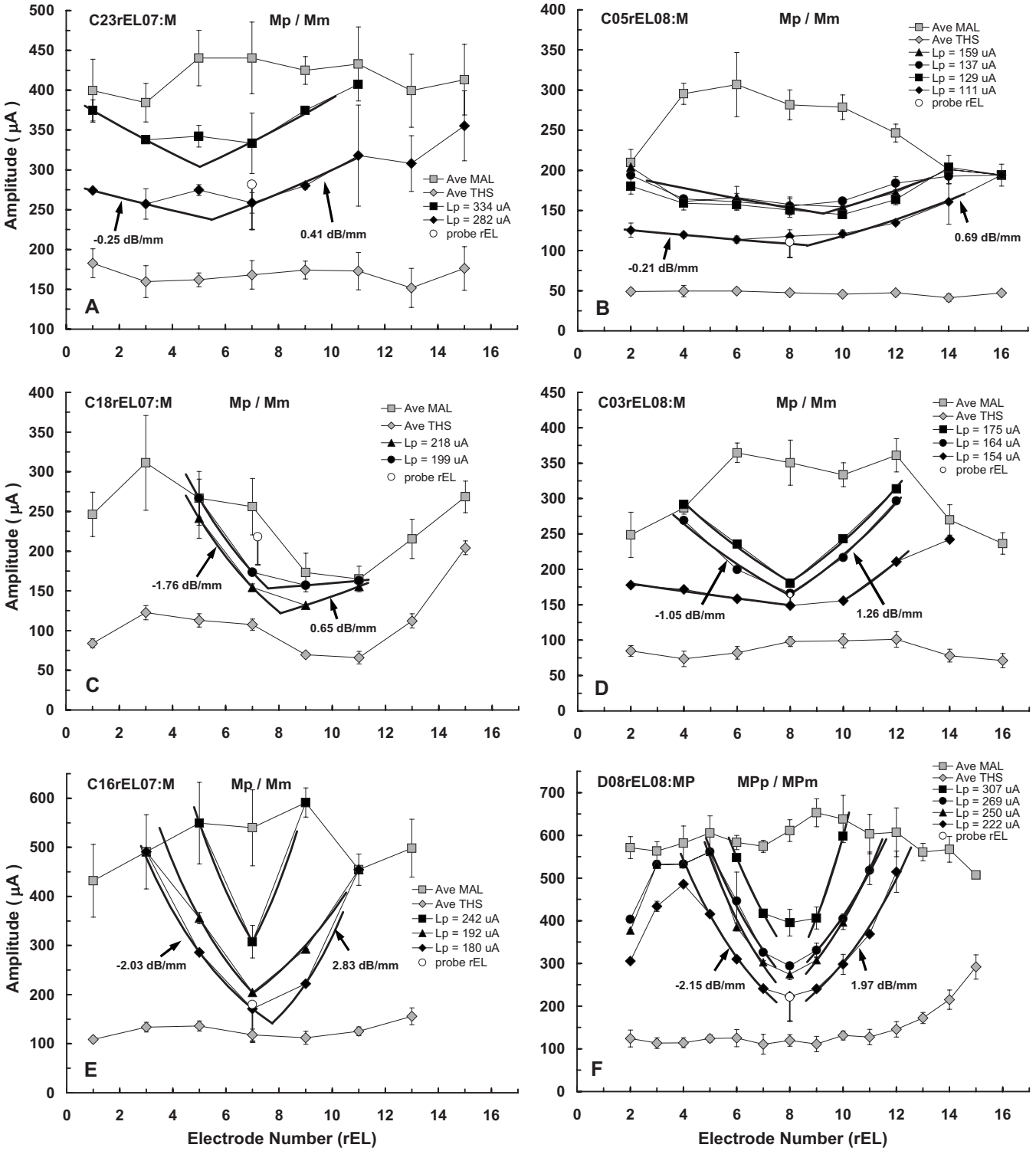


FIG. 3. fmSTCs for six Clarion cochlear-implant users who were stimulated using a monopolar electrode configuration. The horizontal position of the open circle in each graph identifies the probe electrode; probe level is indicated by the vertical position of this symbol, with the height of the vertical line indicating the sensation level of the probe; the horizontal line below the open circle indicates the subject's absolute threshold for the probe stimulus. For clarity, only the lowest probe level is plotted. Other features of the individual graphs are the same as those described for Fig. 2.

III. RESULTS AND DISCUSSION

A. Characteristics of fmSTCs in electrical hearing

Figure 3 shows fmSTCs obtained at two or more probe levels from each of the six Clarion subjects. These fmSTCs were obtained using a monopolar probe and a monopolar masker (MPp/MPm). Thresholds for the masker stimulus ranged between 50 and 200 μA, and corresponding MALs ranged between 200 and 650 μA. For purposes of clarity, only the lowest probe level is shown on each graph (open circle); however, all probe levels (Lp) are specified in the graph legends. The slopes, bandwidths, and tip-shift parameters for each of the fmSTCs are given in Table II.

TABLE II. Spatial tuning curve parameters in Clarion subjects.

Subject probe rEL	Ave Lp (%DRμA)	Slopes (dB/mm) Apical	Slopes (dB/mm) Basal	Bandwidth (mm)	Tip shift (mm)	Ap/Am (dB)
C23rEL07:M	34	−0.41	0.40	9.00	−2.11	0.0
	15	−0.25	0.41	9.09	−1.54	0.7
C05rEL08:M	31	−0.32	0.57	7.18	0.27	0.1
	21	−0.30	0.38	7.48	−0.34	−1.1
	17	−0.21	0.74	6.70	0.19	−1.4
	8	−0.21	0.69	9.64	−1.86	−0.5
C18rEL07:M	20	−1.76	0.65	3.51	1.97	3.0
	8	−1.70	0.14	5.03	2.44	1.2
C03rEL08:M	21	−1.04	1.20	1.67	−0.05	−0.3
	17	−1.05	1.26	1.96	−0.13	−0.1
	14	−0.27	1.31	6.05	−0.57	0.3
C16rEL07:M	24	−2.29	2.58	0.83	−0.02	−2.1
	13	−2.19	1.58	1.16	0.13	−0.6
	11	−2.03	2.83	2.25	0.63	0.4
D08rEL08:MP	29	−2.15	3.05	2.70	0.00	−2.2
	21	−2.15	1.77	2.22	0.02	−0.8
	17	−2.44	2.09	2.37	−0.06	−0.8
Ave (C)	18.9	−1.22	1.27	4.64	−0.06	−0.2
Sdev (C)	7.5	0.89	0.92	3.05	1.15	1.2
Min (C)	7.6	−0.21	0.14	0.83	−2.11	−2.2
Max (C)	33.5	−2.44	3.05	9.64	2.44	3.0

A range of fmSTC shapes was exhibited by the Clarion users. Subjects C23 and C05 [Figs. 3(A) and 3(B)] exhibited very shallow slopes and broad spatial tuning, with slopes ranging between 0.2 and 0.6 dB/mm, and bandwidths ranging between 7.0 and 9.6 mm (not shown in Fig. 3, but given in Table II). At the other extreme, subjects C16 and D08 [Figs. 3(E) and 3(F)] exhibited relatively steep slopes and narrow tuning, with slopes ranging between 1.6 and 3.0 dB/mm, and bandwidths ranging between 0.8 and 2.7 mm. Subjects C18 and C03 [Figs. 3(C) and 3(D)] demonstrated intermediate slopes and bandwidths.

For three of the Clarion subjects in Fig. 3 (C03, C16, D08), the tips of tuning curves (lowest masker levels) were located very close to the probe electrode, suggesting that neurons stimulated by the probe stimulus were proximal to the probe electrode. Note that subject D08 [Fig. 3(F)] demonstrated a downturn in masked thresholds for apical maskers rEL02 and rEL03. This may be attributable to a bending-over of the tip of her electrode array (visualized on x ray) and an associated pitch-reversal on rEL02 and rEL03 (Donaldson *et al.*, 2005). The other three subjects represented in Fig. 3 (C23, C05, and C18) exhibited varying degrees of mistuning, or displacement between the probe electrode and the tuning curve tip. Subjects C23 and C05 [Figs. 3(A) and 3(B)] exhibited minor mistuning with fmSTC tips that were shifted 1.5–2.1 mm in the apical direction and 0.6–1.9 mm in the basal direction, respectively. However, the extremely shallow slopes in these subjects resulted in some uncertainty as to "true" tip locations. Subject C18 [Fig. 3(C)] exhibited moderate mistuning, with the fmSTC tips shifted 2.4–3.6 mm in the basal direction. For this subject, fmSTC shape may have been influenced by irregular MALm values, which were considerably lower on electrodes rEL09 and rEL11 than on other electrodes.

For most subjects, the fmSTC tip level increased with increasing probe level, i.e., masker amplitude grew with probe amplitude. However, notice that fmSTC amplitudes actually decreased with increasing probe amplitude for subject C18 [Fig. 3(C)]. This is likely explained by fluctuations in probe thresholds over the long testing period required to collect the data for this subject. The probes used to generate these curves had amplitudes corresponding to 15% and 27% (triangles and circles, respectively) of the dynamic range measured on the days that testing was completed; thus, masker amplitudes grew with probe sensation level (expressed as %DRμA). Subject C03 [Fig. 3(D)] exhibited extremely broad spatial tuning at the lowest probe level tested, despite the fact that the probe level was well above threshold at 15 %DRμA. Such extremely broad tuning was also observed during pilot testing with several other subjects, but only when very low probe levels were used (e.g., see subject N32 in Fig. 4).

Figure 4 shows the fmSTCs obtained at two or more probe levels in each of the six Nucleus subjects. For three of these subjects (N14, N34, N32), fmSTCs were obtained using a bipolar probe and a bipolar masker (BPp and BPm; both with 0.75-mm spatial separation between active and reference electrodes). For two other subjects (N28 and N13), the BP maskers did not produce sufficient forward masking to mask the probe stimulus. Therefore, the electrode separation of the BP masker was increased to BP+1 or BP+2 (1.5 or 2.25 mm).[2] The remaining subject (N09) was tested with a

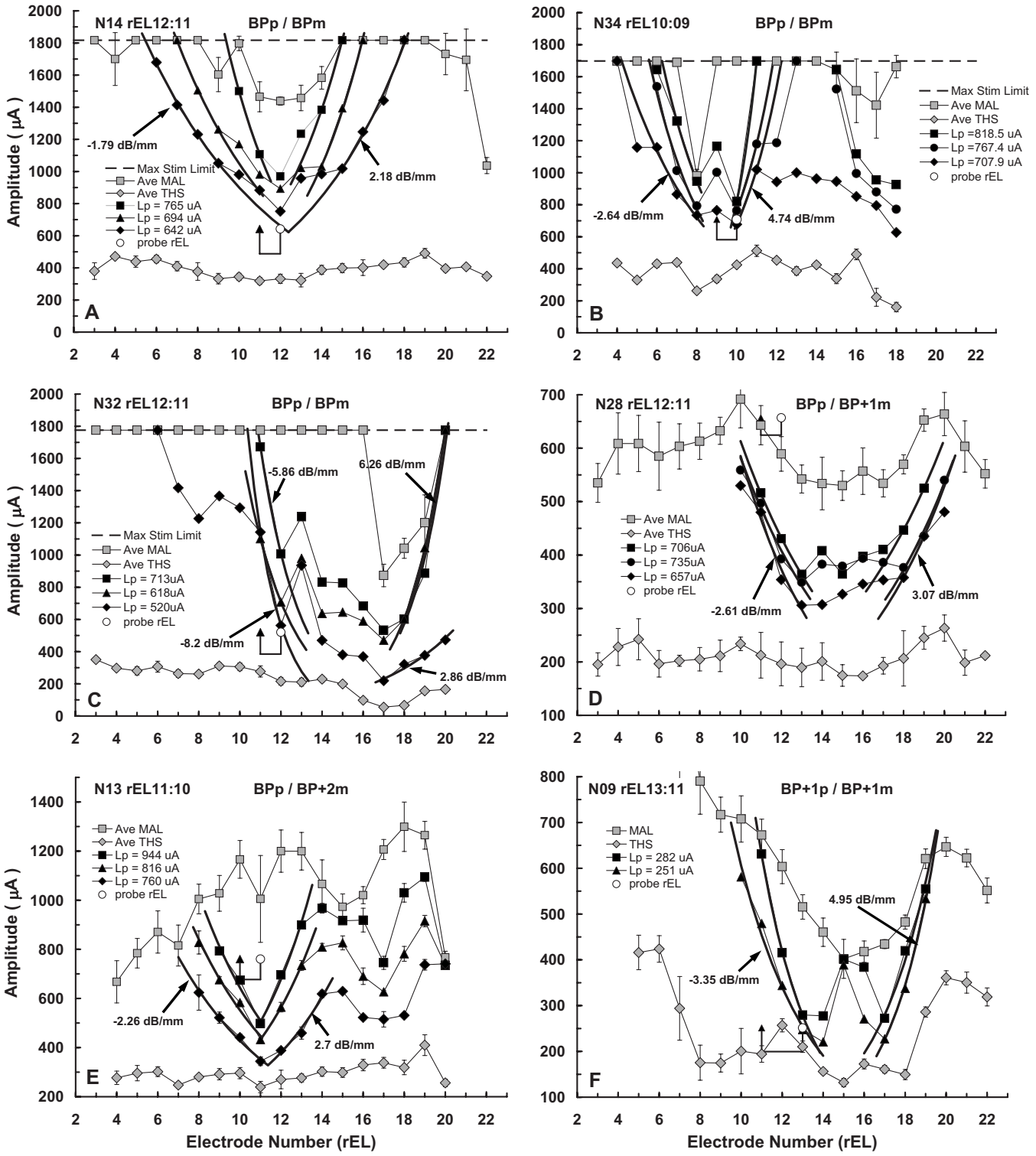


FIG. 4. fmSTCs from six Nucleus-22 cochlear-implant users who were stimulated using a bipolar electrode configuration (BP, BP+1 or BP+2, corresponding to 0.75, 1.5, or 2.25 mm separations between the active and reference electrodes, respectively). Other features of the individual graphs are the same as those described for Fig. 2, except that only the lowest probe level is plotted.

BP+1 probe and a BP+1 masker. Recall that this subject was also stimulated with 125-Hz stimuli, rather than the 500-Hz stimuli used for other subjects. Because bipolar (BP, BP+1, or BP+2) maskers were used for these Nucleus subjects, THSs and MALs for the maskers were considerably higher than those obtained from Clarion subjects using monopolar stimulation. In addition, there was greater variability across electrodes within an individual for the Nucleus subjects. As in Fig. 3, only the lowest probe level is indicated on each graph. Table III provides the slopes, bandwidths, and tip-shift parameters for the fmSTCs in Fig. 4 from Nucleus subjects.

A wide range of fmSTC characteristics was exhibited by the Nucleus users. Slopes ranged between 1.7 and 8.2

TABLE III. Spatial tuning curve parameters in Nucleus subjects.

Subject probe rEL	Ave Lp (%DRμA)	Slopes (dB/mm) Apical	Slopes (dB/mm) Basal	Bandwidth (mm)	Tip shift (mm)	Ap/Am (dB)
N14 rEL12:11	20	−3.52	3.15	1.54	0.82	−2.1
	14	−2.11	3.29	2.72	0.80	−2.2
	9	−1.79	2.18	2.69	0.91	−1.4
N34 rEL10:09	21	−3.87	8.44	1.55	0.09	0.0
	17	−3.84	5.04	1.94	−0.02	0.1
	11	−2.64	4.74	1.50	0.07	0.4
N32 rEL12:11	24	−5.86	6.26	3.79	3.42	−3.0
	17	−5.12	6.27	3.82	3.15	−1.2
	10	−8.20	2.86	3.06	3.16	−0.7
N28 rEL12:11	33	−2.05	2.09	4.53	3.19	5.4
	25	−2.02	1.86	3.83	2.91	4.3
	11	−2.61	2.26	4.06	3.06	5.4
N13 rEL11:10	29	−2.69	3.42	0.56	0.73	5.5
	17	−2.43	3.07	0.61	0.77	5.5
	12	−2.26	2.70	1.16	0.97	6.9
N09 rEL13:11	31	−4.71	4.13	3.36	3.01	0.1
	22	−3.35	4.95	3.10	3.12	0.1
Ave (N)	18.9	−3.48	3.92	2.58	1.77	1.4
Sdev (N)	7.6	1.70	1.81	1.26	1.35	3.3
Min (N)	8.6	−1.79	1.86	0.56	−0.02	−3.0
Max (N)	32.9	−8.20	8.44	4.53	3.42	6.9

dB/mm, and bandwidths ranged between 0.6 and 4.5 mm (not shown in Fig. 4, but given in Table III). In general, these fmSTCs obtained with bipolar stimulation were much sharper and included more irregularities than those obtained from the Clarion subjects with monopolar stimulation. Notice that MALm values varied across electrodes for all of the subjects, and that THSp values also varied across electrodes in a few subjects [especially N09, Fig. 4(F)]. In addition, there were irregularities in masked thresholds within the fmSTC tip regions (between the two slopes) for several subjects (N34, N32, N28, N09). These irregularities could reflect variations in neural survival (Kawano *et al.*, 1998; Saunders *et al.*, 2002; Pfingst and Xu, 2004; Pfingst *et al.*, 2004), nonuniform current pathways resulting from individual variations in cochlear impedances, or peaks and nulls in the current fields produced by the bipolar maskers. The inverted tips demonstrated by these subjects (especially N09) are reminiscent of "split-tip" neural tuning curves described by Kral *et al.* (1998). Kral *et al.* attributed the split tips to peaks and nulls that occur in bipolar current fields.[3]

Three of the Nucleus subjects (N14, N34, and N13), exhibited fmSTC tips close to the probe electrode in terms of longitudinal distance, suggesting that a functional population of neural elements exists in the region of the probe electrode. However, two of these subjects (N13 and N34), also demonstrated a secondary tip in the basal region of the electrode array. This implies that there is cross-talk between basal and middle regions of the cochlea. One possibility is that stimuli presented on basal electrodes, near the secondary tip, activate fibers of passage originating near the probe electrode as they pass through the modiolus to exit from the cochlea (Frijns *et al.*, 2001).

The remaining three Nucleus subjects (N32, N28, and N09) demonstrated displaced tuning curves, with the fmSTC tips located 2.9–3.4 mm basal to the probe electrode. These three subjects may have better nerve survival in the basal region of the cochlea than in the middle region. However, realistic conclusions about nerve survival (e.g., "dead regions") must await a more complete evaluation of the displaced spatial tuning in these subjects' cochleas. This will require the measurement of fmSTCs for additional probe electrodes, both in the basal and apical regions of the electrode array.

Figure 5 summarizes the fmSTC slope parameters from Tables II and III in graphical form. Figure 5(A) shows the slope data for Clarion subjects (monopolar stimulation), and Fig. 5(B) shows the corresponding data for Nucleus subjects (bipolar stimulation). Two key findings can be observed from these data. First, the fmSTC slopes of both Clarion and Nucleus subjects are essentially constant with probe level, for the range of probe levels sampled (approximately 8–33 %DRμA). Second, the slopes for Nucleus subjects are approximately three times as steep as those for Clarion subjects.

The latter finding is consistent with previous reports that current attenuates two to three times more rapidly for bipolar stimulation than for monopolar stimulation with increasing distance from the source (Black and Clark, 1980; Black *et al.*, 1981, 1983; O'Leary *et al.*, 1985; Hartmann and Klinke, 1990; Kral *et al.*, 1998). That finding is also consistent with most previous studies that employed psychophysical forward masking patterns to examine spread of excitation across the electrode array (Shannon, 1983b; Lim *et al.*, 1989; Boex *et al.*, 2003; Chatterjee *et al.*, 2006). In general, these studies

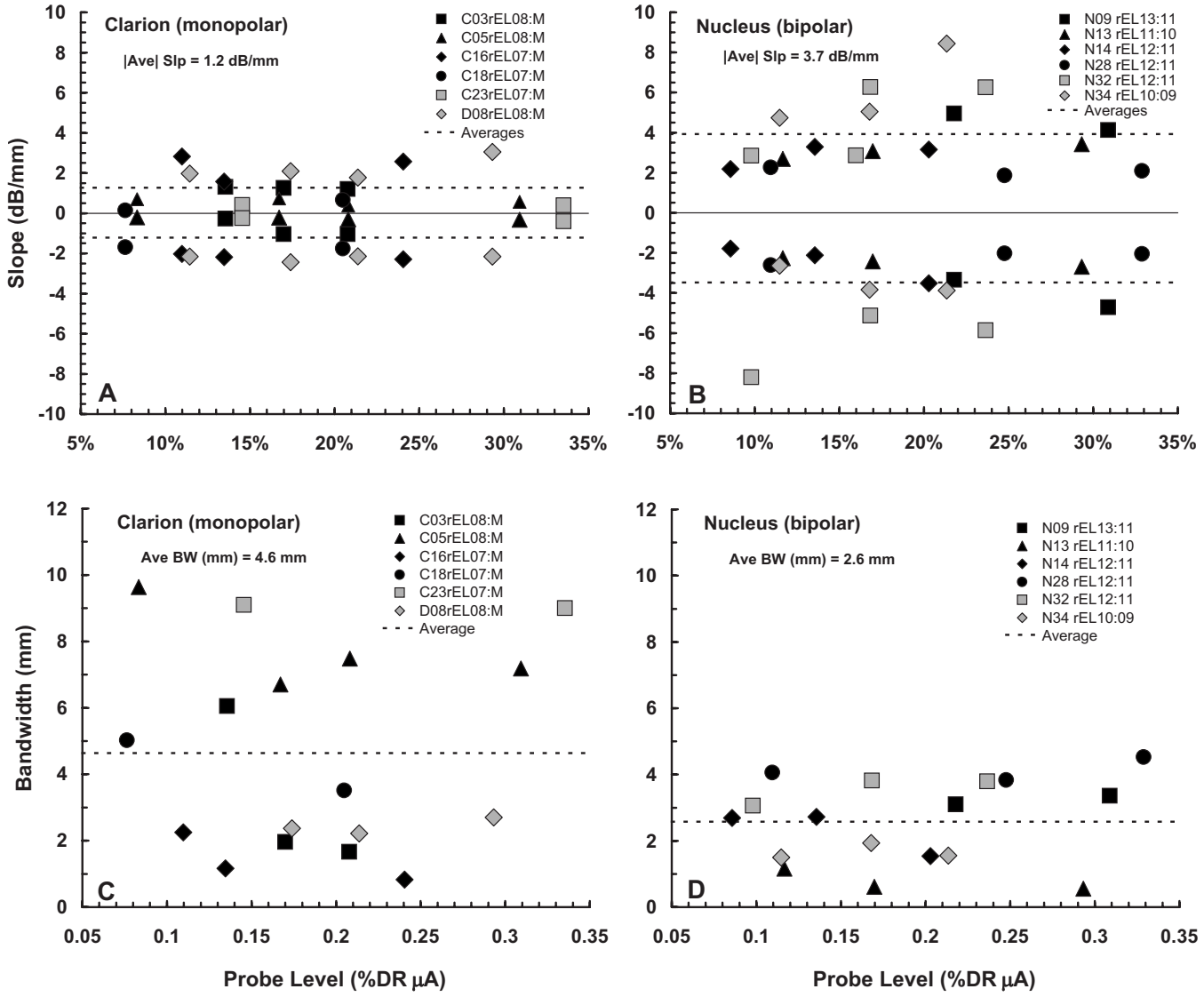


FIG. 5. fmSTC slopes and bandwidths. Slopes (dB/mm) of the fmSTCs obtained from Clarion subjects (A) and Nucleus subjects (B) are shown as a function of probe level, expressed as a percentage of the probe dynamic range in microamperes. In each panel, apical slopes are plotted as negative values and basal slopes are plotted as positive values. Different symbols represent different subjects, as indicated in the legends of each graph, and dashed lines indicate the average slopes for each group. The average of the absolute values (disregarding sign) of both apical and basal slopes is indicated under the heading of each graph. $Q_{1\ \mathrm{dB}}$ bandwidths (mm) of fmSTCs obtained from Clarion subjects (C) and Nucleus subjects (D) are shown as a function of probe level, which is expressed as a percentage of the probe dynamic range in μA. Different symbols represent different subjects, as indicated in the legends of each graph. The average $Q_{1\ \mathrm{dB}}$ bandwidth for each group is shown by the horizontal dashed line and is given under the heading of each graph.

have found that masking patterns are broader for monopolar stimulation than for bipolar stimulation, but that differences in slopes are not nearly so dramatic as those seen here for spatial tuning curves. An exception to this general finding was reported by Kwon and van den Honert (2006), who found little difference in the spread of forward masking across electrodes for monopolar and bipolar stimulation.

Note from Tables II and III that the absolute variability of slopes is approximately twice as large for the Nucleus (bipolar) subjects (1.75 dB/mm) than for the Clarion (monopolar) subjects (0.9 dB/mm). However, this difference is reduced substantially when the variability is normalized with respect to the mean slope (S.D./ mean), leading to values of 0.47 and 0.72, respectively.

Figure 5 also summarizes the bandwidth parameters from Tables II and III. Figure 5(C) shows the bandwidth data for Clarion subjects (monopolar stimulation), and Fig. 5(D) shows the corresponding data for Nucleus subjects (bipolar stimulation). Again, two key findings are apparent. First, although level effects occur for a few individual subjects (e.g., Clarion subject C03; Nucleus subject N14), there is no systematic effect of probe level on the bandwidths of fmSTCs for subjects in either group. Second, average bandwidths are about twice as large in Clarion (monopolar) subjects as they are in Nucleus (bipolar) subjects, consistent with the slope differences observed in Fig. 5. The absolute variability of fmSTC bandwidths is substantially larger for the Clarion subjects than for the Nucleus subjects. However, similar to the slope estimates, the normalized variability is fairly consistent across the two groups (0.49 and 0.66 for Nucleus and Clarion subjects, respectively).

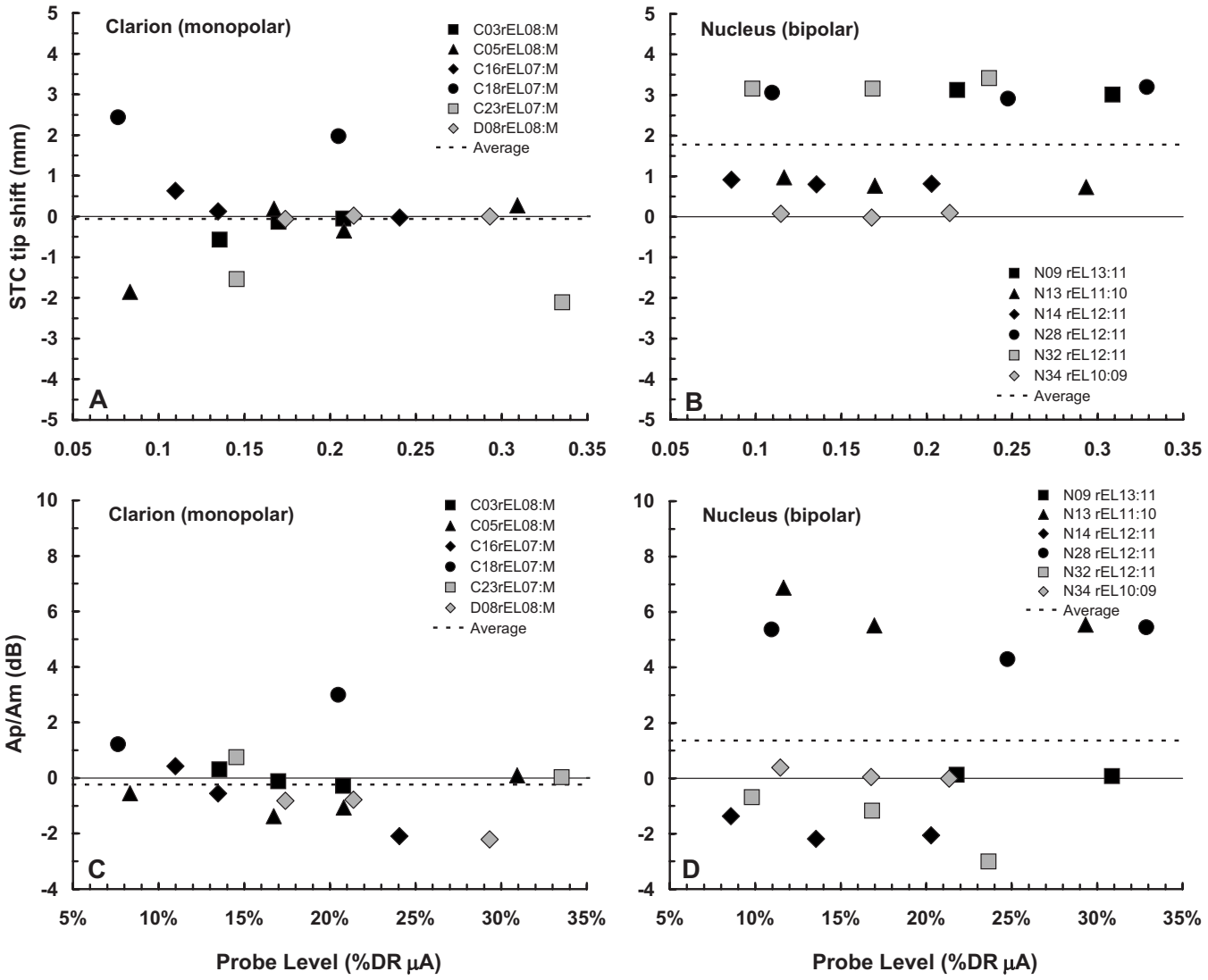


FIG. 6. fmSTC tip shifts and Ap/Am ratios. Mistuning of the fmSTCs obtained from Clarion subjects (A) and Nucleus subjects (B), expressed as the "tip shift" (mm) between the STC tip and the probe electrode, is shown as a function of probe level, which is expressed as a percentage of the probe dynamic range in microamperes. Positive values indicate a tip shift toward the base of the cochlea; negative values indicate a tip shift toward the apex. Different symbols represent different subjects, as indicated in the legends of each graph, and dashed lines indicate the average tip shift for each group. The Ap/Am ratio on the probe electrode from Clarion subjects (C) and Nucleus subjects (D) is shown as a function of probe level, expressed as %DRμA. The Ap/Am ratio is an indicator of the relative amount of forward masking produced in each subject. It shows the level of the probe relative to the level of the masker at masked threshold, in this case when both the masker and the probe are on the same electrode.

Figure 6 summarizes the tip-shift and Ap/Am ratio data from Tables II and III, using a similar format as Fig. 5. The tip-shift data are shown in Figs. 5(A) and 5(B). Evaluation of the individual subjects' data within each panel indicates that estimates of tip location are generally unaffected by probe level. Two Clarion subjects (C18, C23) and three Nucleus subjects (N09, N28, N32) showed consistent tip shifts of more/than 1 mm. While the two Clarion subjects had tip shifts in opposite directions (one apical and one basal), the three Nucleus subjects all demonstrated tip shifts in the basal direction. There is some evidence of asymmetrical flow of current toward the base of the cochlea in electrical stimulation (Lim *et al.*, 1989), which could be viewed as a possible explanation for the basalward shifts observed among Nucleus subjects. However, it seems unlikely that asymmetrical current flow could underlie the large tip shifts seen here for some subjects, especially given the low probe levels that were used.

Although it was not discussed earlier, absolute probe thresholds (THSp) varied considerably within the groups of Clarion and Nucleus subjects (Figs. 3 and 4). In particular, several subjects (C18, N13 and, especially, N28) demonstrated absolute probe thresholds that were substantially elevated in comparison to the masked probe threshold on the probe electrode. The last column of Table III summarizes the relation between probe amplitude (Ap) and masker amplitude (Am) at masked threshold, by computing the ratio Ap/Am in decibels when the masker and the probe are on the same electrode. Recall that the amplitude of the masker, at masked threshold represents the amplitude of a 160-ms masker pulse train required to forward mask the 10-ms probe

TABLE IV. Statistical results.

Linear regression analyses [$y=ax+b$, x=probe level (%DRμA)]						
y	a	b	R^2	F	df	p
Clarion (monopolar)						
Apical slope (dB/mm)	0.00	−1.28	0.00	0.01	15	0.92
Basal slope (dB/mm)	0.01	1.05	0.01	0.15	15	0.71
$Q_{1\ dB}$ bandwidth (mm)	0.00	4.55	0.00	0.00	15	0.96
Tip shift (mm)	−0.03	0.55	0.05	0.71	15	0.41
Ap/Am (tip)	−0.05	0.62	0.08	1.22	15	0.29
Nucleus (bipolar)						
Apical slope (dB/mm)	0.01	−3.65	0.00	0.02	15	0.88
Basal slope (dB/mm)	0.03	3.41	0.01	0.20	15	0.66
$Q_{1\ dB}$ bandwidth (mm)	0.03	1.98	0.04	0.56	15	0.47
Tip shift (mm)	0.05	0.76	0.09	1.49	15	0.24
Ap/Am (tip)	0.08	−0.10	0.03	0.49	15	0.50

T-tests [Clarion monopolar vs Nucleus bipolar]			
	t	df	p
Apical slope (dB/mm)	3.68	10	0.004
Basal slope (dB/mm)	−3.79	10	0.004
$Q_{1\ dB}$ bandwidth (mm)	1.623	10	0.14
Tip shift (mm)	−2.35	10	0.04
Ap/Am (tip)[a]	$T=30.0$[a]	$n=6,6$[a]	0.18

[a]Mann-Whitney rank sum test used because equal variance test failed.

pulse train. The Ap/Am ratio varies considerably across subjects (−3 to +7 dB), indicating that the masker stimulus produced different amounts of forward masking for different individuals. These differences in masker effectiveness likely reflect differences in temporal integration across subjects (Donaldson *et al.*, 1997), as well as differences in the rate of recovery from forward masking (Nelson and Donaldson, 2002; Dingemanse *et al.*, 2006). Temporal integration is a factor because the probe stimulus is considerably shorter than the masker stimulus and, therefore, has a higher absolute threshold (compare probe thresholds and masker thresholds for the fmSTCs in Figs. 3 and 4). Recovery from forward masking determines the extent to which the masker stimulus shifts the threshold for the probe stimulus at a fixed masker-probe time delay.

Table IV summarizes the statistical results that confirm the above-described findings. Linear regression analyses indicated that there were no significant effects of probe level on any of the fmSTC parameters examined (apical slope, basal slope, bandwidth, tip-shift or Ap/Am ratio). T-tests performed on the average parameter values (collapsed across probe levels) for each subject confirmed that both apical and basal fmSTC slopes were shallower in Nucleus (bipolar) subjects than in Clarion (monopolar) subjects. A t-test performed on the bandwidths did not show a significant difference between groups, even though mean bandwidths were about half as large for the Nucleus (bipolar) subjects as for the Clarion (monopolar) subjects. This can be attributed to the relatively large variability in bandwidths exhibited by the Clarion subjects (Fig. 5). Finally, t-tests showed no significant differences between groups for either the tip-shift or Ap/Am ratio parameters.

B. Normal-hearing and hearing-impaired fmPTCs in acoustic hearing

One goal of this investigation was to compare fmSTC-based measures of spatial selectivity in cochlear implant listeners with previously reported fmPTC-based measures of frequency selectivity in normal-hearing and hearing-impaired acoustic listeners. For this comparison, we used fmPTCs obtained by Nelson (1991) in normal-hearing listeners and listeners with sensorineural hearing loss, which were obtained with a forward-masking procedure similar to that used in the present study. The acoustic maskers were 200-ms (peak duration) tone bursts that forward-masked a 20-ms acoustic tone burst at 1000 Hz. The offset-to-offset delay was 42 ms. Twenty-six normal-hearing listeners and twenty-four hearing-impaired listeners with varying degrees of hearing loss were tested. Example fmPTCs for two probe levels in one normal-hearing listener [Fig. 7(A)] and one hearing-impaired listener [Fig. 7(B)] are shown. Acoustic levels on the ordinate are decibels relative to the threshold at 1000 Hz. Note that the ordinates span a different range of levels for the normal-hearing and hearing-impaired listeners, reflecting the different dynamic ranges available to the two listeners.

Figure 7(A) shows fmPTCs from a normal-hearing listener for two probe levels. Stimulus level is plotted on the ordinate in decibels relative to unmasked threshold at 1000 Hz. For this normal-hearing listener, threshold for the masker was 4.4 dB SPL and threshold for the probe was 12 dB SPL at 1000 Hz. At some of the remote masker frequencies, masker levels up to 105 dB SPL were required to mask the probe tone. The difference between absolute threshold for the masker and the maximum masker levels required to mask the probe reveals a dynamic range of about 100 dB.

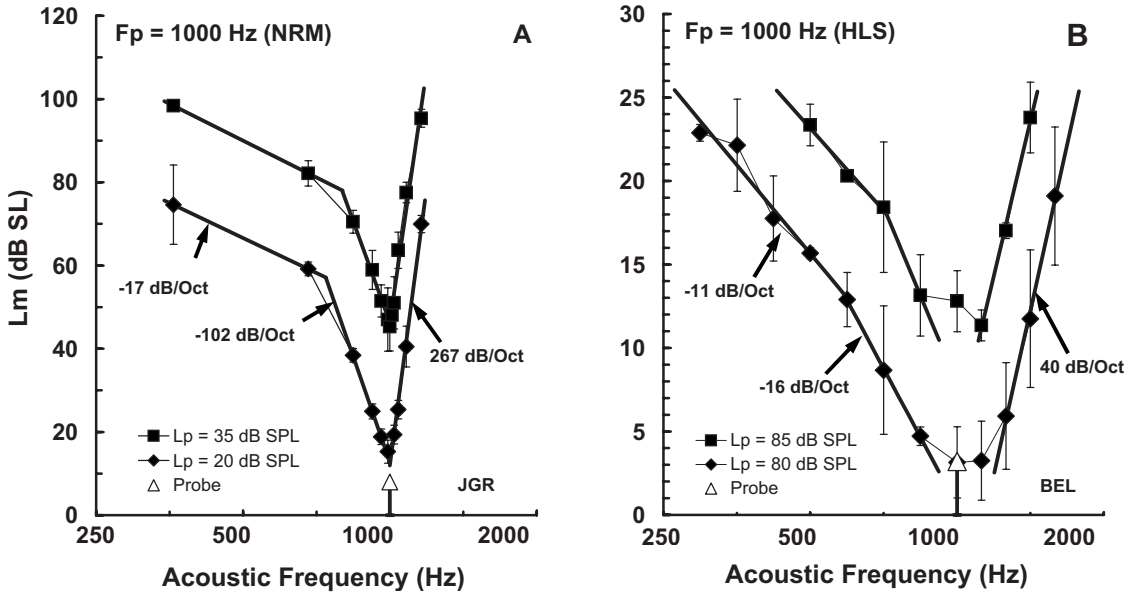


FIG. 7. Forward-masked psychophysical tuning curves (fmPTCs) from acoustic listeners obtained with a procedure comparable to that used to obtain the fmSTCs for the cochlear-implant subjects. (A) fmPTCs from a normal-hearing listener at two probe levels. (B) fmPTCs from a listener with sensorineural hearing loss at two probe levels. The ordinate shows decibels relative to thresholds at the probe frequency (1000 Hz).

The fmPTCs have three segments: a high-frequency (HF) segment with an extremely steep slope (e.g., 267 dB/octave), a low-frequency (LF) segment with a moderately steep slope (e.g., −102 dB/octave), and a "Tail" segment with a relatively flat slope (e.g., −17 dB/octave). Bandwidths measured 10 dB above the fmPTC tip corresponded to approximately 12% of the probe frequency. At the lowest probe level, the bandwidth was 116 Hz, yielding a Q value of 8.6. As probe levels increased, the HF and Tail slopes remained relatively constant, and the LF slope decreased. Consequently, Q values decreased at high probe levels.

Figure 7(B) shows fmPTCs from a listener with sensorineural hearing loss. For this hearing-impaired listener, threshold for the masker was 74 dB SPL and threshold for the probe was 77 dB SPL at 1000 Hz. Maximum masker levels were nearly 100 dB SPL at the lowest masker frequencies. The dynamic range between absolute threshold and the most intense maskers that could be tolerated was less than 25 dB, which is about one-fourth of the dynamic range available to normal-hearing listeners. At the lower probe level (80 dB SPL or 3 dB SL), the fmPTC from this hearing-impaired listener is still characterized by three segments, but the slopes are much more gradual and the bandwidths much wider than for a normal-hearing listener. Specifically, for the lower probe level, the HF slope is 40 dB/octave and the LF slope is −16 dB/octave, both much shallower than for the normal-hearing listener; the Tail slope is −11 dB/octave, which is not much different than for the normal-hearing listener. The bandwidth is 858 Hz (86% of the probe frequency), which corresponds to a Q value of only 1.16. As probe level is increased, the fmPTC parameters change very little.

C. Spatial tuning expressed as tonotopic frequency

In order to make meaningful comparisons between fmSTCs in electric hearing and fmPTCs in acoustic hearing, the fmSTCs must be displayed using the same frequency scale as that used in acoustic hearing. This requires a transformation between spatial distance along the electrode array and acoustic frequency. Previous research has provided a tonotopic map that relates distance along the basilar membrane to characteristic frequency in an adult human cochlea (Greenwood, 1961, 1990). Using Greenwood's equation, it is possible to approximate the nominal acoustic frequency region associated with each electrode in a given cochlear-implant listener's ear, and to replot electrically stimulated fmSTCs in terms of "spatial frequency" (Hertz) instead of spatial distance (millimeters).[4]

Example fmSTCs plotted in terms of spatial frequency are shown in Fig. 8 for data from Nucleus subject N14. Stimulus level is plotted on the ordinate, in decibels relative to the threshold for the probe electrode, as a function of spatial frequency on the abscissa, in Hertz using a logarithmic scale similar to that used in Fig. 7. Tuning curve parameters were calculated for the spatial-frequency fmSTCs using a procedure similar to that used earlier to calculate parameters for the spatial-distance fmSTCs. Specifically, curves were fitted to each fmSTC using a least-squares fitting procedure which related log (microamperes) on the ordinate to log (hertz) on the abscissa. The slopes were expressed in units of decibel/octave. Bandwidths were calculated (in hertz) at a masker level 1 dB above the minimum masker level, and were expressed as Q values (tip frequency divided by bandwidth). These are the same metrics used to describe acoustic tuning curves. The resultant fmSTC parameters, expressed in terms of spatial frequency, are given in Table V for Clarion subjects and Table VI for Nucleus subjects. The slopes in Tables V and VI are expressed in terms of decibels per octave, comparable to slopes of fmPTCs in acoustic hearing.

Absolute slopes for the fmSTCs from listener N14 (Fig. 8) ranged between 8 and 16 dB/octave. Bandwidths ranged

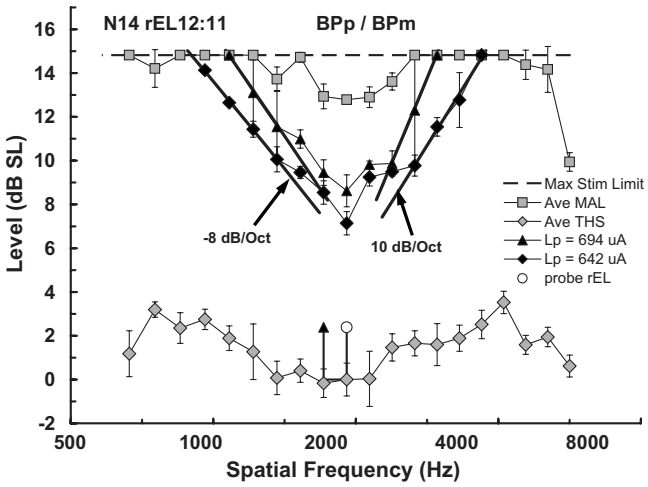


FIG. 8. Forward-masked spatial tuning curves (fmSTCs) from a cochlear implant listener, plotted in terms of spatial frequency calculated from Greenwood's (1961) place-frequency function. The fmSTCs are two of those shown in Fig. 4(A), but for comparison with acoustic fmPTCs, the abscissa has been converted to the spatial frequency corresponding to each electrode, and the ordinate is decibels relative to thresholds on the probe electrode.

between 23% and 40% of the spatial frequency corresponding to the probe electrode, yielding Q values between 4.4 and 2.5. For comparison with acoustic hearing, note that the dynamic range between absolute threshold for a masker stimulus on the probe electrode and the maximum acceptable stimulus level was only 14.8 dB. This dynamic range is less than one-sixth of that available to the normal-hearing listener shown previously. Also note that the low-frequency (apical) side of then fmSTC has only one distinct segment, as compared to the two segments that can be seen in acoustic fmPTCs.

The average tip frequency for fmSTCs was 1536 Hz (range 1105–2347 Hz) for the Clarion subjects and 2650 Hz (range 1838–4298 Hz) for the Nucleus subjects. The differences in average tip frequencies between groups was partially due to the selection of a range of middle electrodes for testing (rEL10 through rEL13 in Nucleus subjects, and rEL07 and rEL08 in Clarion subjects) and partially due to the basalward tip shifts seen in several Nucleus subjects, described earlier. Slopes of the fmSTCs averaged 6 dB/octave for the Clarion subjects and 17 dB/octave for the Nucleus subjects. Average Q values were 2.17 for the Clarion subjects and 3.86 for the Nucleus subjects.

D. Comparisons of tuning in acoustic and electric hearing

Figure 9 compares the tuning characteristics of normal-hearing and hearing-impaired acoustic listeners from Nelson (1991) with the tuning characteristics of the cochlear-implant subjects tested in the present study. Each panel shows a different tuning-curve parameter plotted against dB SPL for the acoustic listeners and percent dynamic range (calculated in microamperes) for the cochlear-implant listeners. Each data point represents a measurement from the tuning curve for an individual listener at a given probe level. Data for one or more probe levels are plotted for each subject.

Figure 9(A) shows that LF slopes decrease with increasing probe level in normal-hearing listeners (closed squares). LF slopes for hearing-impaired ears (closed diamonds and open circles) are similar to those for normal-hearing ears at high probe levels. This is true even for the hearing-impaired ears that exhibit abnormal tuning (open circles). The normal-hearing listeners have much steeper LF slopes than the

TABLE V. Spatial-frequency tuning curve parameters in Clarion subjects.

Subject probe rEL	Ave Lp (%DRμA)	Slopes (dB/octave) Apical	Basal	STCtip (Hz)	Bandwidth (Hz)	Bandwidth (% CF)	Q1 dB (CF/BW)
C23rEL07:M	34	−1.70	1.87	1184	1767	149.2%	0.67
	15	−1.04	1.91	1294	1798	139.0%	0.72
C05rEL08:M	31	−1.41	2.71	1533	1735	113.2%	0.88
	21	−0.81	1.80	1396	1665	119.3%	0.84
	17	−0.94	3.51	1515	1593	105.1%	0.95
	8	−0.89	3.32	1105	1792	162.1%	0.62
C18rEL07:M	20	−8.01	3.11	2193	1146	52.3%	1.91
	8	−7.69	0.68	2347	1767	75.3%	1.33
C03rEL08:M	21	−4.59	5.60	1459	370	25.4%	3.94
	17	−4.63	5.88	1442	431	29.9%	3.34
	14	−1.15	6.19	1348	1286	95.4%	1.05
C16rEL07:M	24	−10.37	12.02	1629	202	12.4%	8.05
	13	−9.94	7.43	1666	292	17.5%	5.71
	11	−9.31	13.46	1797	606	33.7%	2.96
D08rEL08:MP	29	−9.77	14.38	1375	979	71.2%	1.40
	21	−9.64	8.39	1351	1021	75.6%	1.32
	17	−10.98	9.90	1443	1212	84.0%	1.19
Ave (C)	18.9	−5.46	6.01	1534	1157	80.0%	2.17
Sdev (C)	7.5	4.11	4.31	325	587	47.1%	2.06
Min (C)	7.6	−10.98	0.68	1105	202	12.4%	0.62
Max (C)	33.5	−0.81	14.38	2347	1798	162.1%	8.05

TABLE VI. Spatial-frequency tuning curve parameters in Nucleus subjects.

Subject probe rEL	Ave Lp (%DRμA)	Slopes (dB/octave) Apical	Basal	STCtip (Hz)	Bandwidth (Hz)	Bandwidth (% CF)	Q1 dB (CF/BW)
N14 rEL12:11	20	−16.18	14.96	1932	443	22.9%	4.36
	14	−9.46	15.65	1927	784	40.7%	2.46
	9	−7.95	10.47	1959	786	40.1%	2.49
N34 rEL10:09	21	−17.76	39.69	1868	433	23.2%	4.31
	17	−17.52	23.71	1838	532	29.0%	3.45
	11	−12.06	22.29	1863	416	22.3%	4.48
N32 rEL12:11	24	−27.75	30.57	3777	2078	55.0%	1.82
	17	−24.25	30.62	3637	2020	55.5%	1.80
	10	−38.82	10.94	3639	1613	44.3%	2.26
N28 rEL12:11	33	−6.43	8.19	2740	1835	67.0%	1.49
	25	−9.42	9.01	2630	1487	56.5%	1.77
	11	−12.15	10.93	2686	1610	59.9%	1.67
N13 rEL11:10	29	−12.47	16.09	1980	164	8.3%	12.09
	17	−11.20	14.46	1990	181	9.1%	11.01
	12	−10.43	12.77	2052	353	17.2%	5.81
N09 rEL13:11	31	−22.57	20.19	4230	2049	48.4%	2.06
	22	−15.99	24.21	4298	1918	44.6%	2.24
Ave (N)	18.9	−16.02	18.51	2650	1100	37.9%	3.86
Sdev (N)	7.6	8.35	8.91	905	739	18.4%	3.15
Min (N)	8.6	−6.43	8.19	1838	164	8.3%	1.49
Max (N)	32.9	−38.82	39.69	4298	2078	67.0%	12.09

hearing-impaired listeners at low SPLs, where the hearing-impaired listeners cannot hear. The sharp tuning at low SPLs in normal-hearing listeners stems from cochlear processes related to the outer hair cells, which are dysfunctional in the hearing-impaired listeners. The LF (apical) slopes of fmSTCs in cochlear-implant listeners (Clarion=shaded triangles; Nucleus=shaded circles), are substantially shallower than the LF slopes measured in normal-hearing listeners at low probe levels; however, they are roughly equivalent to the LF slopes measured in normal-hearing and hearing-impaired listeners at high probe levels.

Figure 9(B) shows that the HF slope is extremely steep in normal-hearing listeners (closed squares), and in many hearing-impaired listeners (closed diamonds). Some of the hearing-impaired listeners (open circles) exhibited more gradual HF slopes, which led Nelson (1991) to classify them as having abnormal tuning. By comparison, the HF (basal) slopes of most of the cochlear-implant listeners are more gradual. Many of the Nucleus users (shaded circles), who were stimulated in a bipolar electrode configuration, exhibited HF slopes in the same range as the hearing-impaired acoustic listeners with abnormal tuning (open circles). All of the Clarion users (shaded triangles), who were stimulated in a monopolar mode, exhibited HF slopes that were shallower than those for the hearing-impaired listeners with abnormal tuning.

Figure 9(C) shows that the Tail slope is relatively level-independent in normal-hearing and hearing-impaired listeners. The Tail slope presumably reflects tuning in the auditory system without the sharply tuned amplification mechanisms of the normal cochlea. That is, it reflects passive mechanical tuning associated with the traveling wave. Comparisons with the LF and HF slopes from cochlear-implant listeners reveal that Nucleus listeners (shaded circles), who were stimulated in a bipolar mode, exhibited tuning-curve slopes (either LF or HF slopes) that were similar to Tail slopes in acoustic hearing. Clarion listeners (shaded triangles), who were stimulated in a monopolar mode, exhibited tuning-curve slopes that were shallower than Tail slopes in acoustic hearing.

Figure 9(D) compares bandwidth estimates obtained from acoustic listeners and cochlear-implant listeners expressed as Q values (CF/BW). Q values range from about 10 in the normal-hearing listeners at low probe levels, to about 1 in the normal-hearing and hearing-impaired listeners at very high probe levels. Q values are even smaller in hearing-impaired listeners with abnormal tuning.[5] The Q values for cochlear-implant listeners fall within the range of those obtained at high probe levels in the acoustic listeners.

The Nelson (1991) data for acoustic listeners were obtained using a probe frequency of 1000 Hz. In contrast, the present data for cochlear-implant listeners varied in terms of the spatial frequency corresponding to the tip of the fmSTC. Tip frequencies ranged between 1105 and 2347 Hz for the Clarion subjects and between 1838 and 4298 Hz for the Nucleus subjects. It is well known in acoustic hearing that tuning becomes sharper with increasing probe frequency (Glasberg and Moore, 1990; Oxenham and Shera, 2003), i.e., Q values increase with probe frequency. Therefore, a fair comparison of the acoustic and cochlear-implant data requires comparison at the same probe frequencies. Fortunately, PTCs have been investigated for a wide range of frequencies in normal-hearing subjects, and an equation has been published that defines changes in Q values with probe

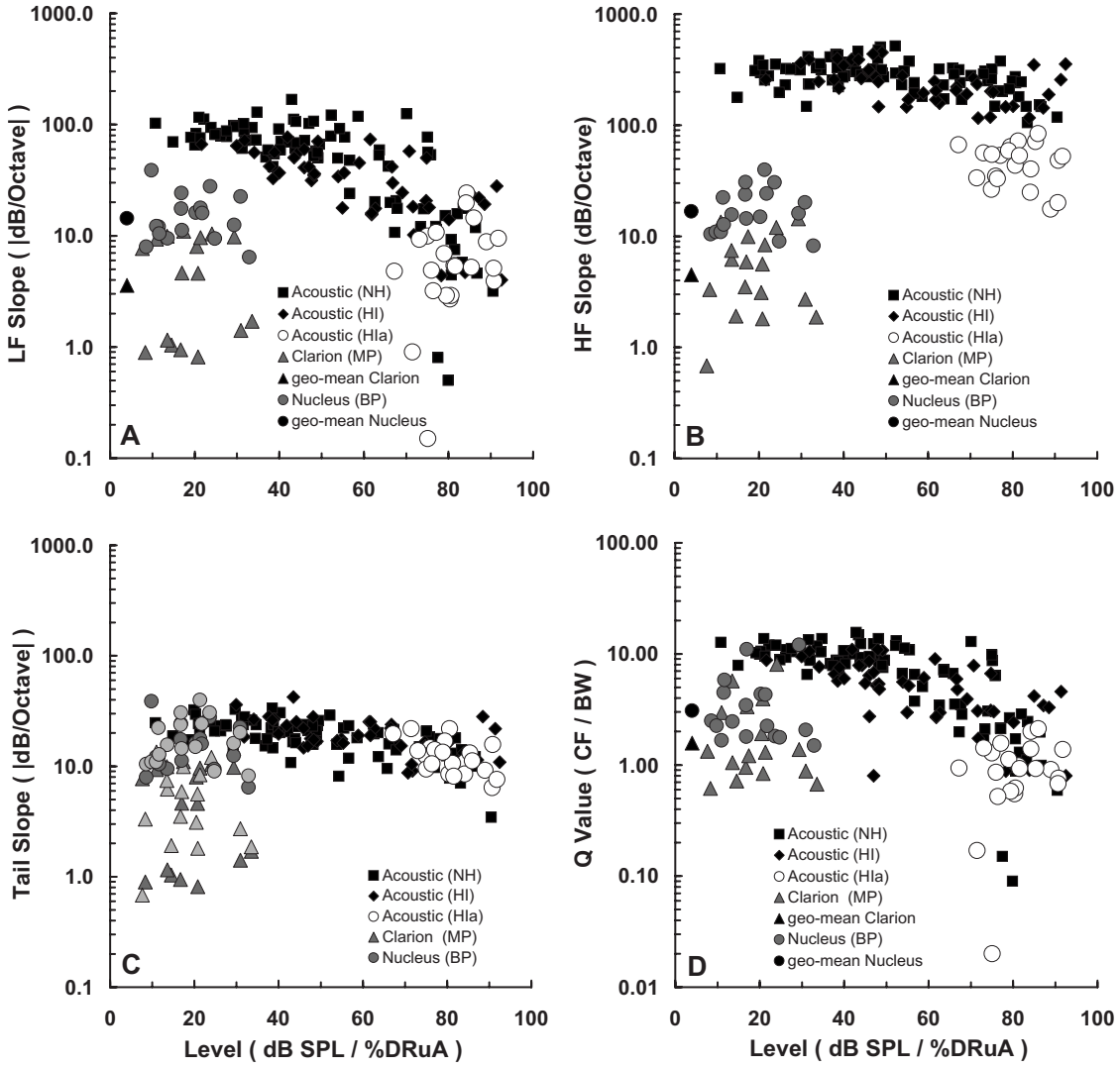


FIG. 9. Tuning characteristics of fmPTCs in normal-hearing and hearing-impaired acoustic listeners, taken from previous work by Nelson (1991), compared with tuning characteristics of fmSTCs obtained from Clarion (monopolar) and Nucleus (bipolar) cochlear implant listeners. Each panel plots a different tuning parameter for three groups of acoustic listeners along with two groups of cochlear-implant listeners. Normal-hearing (NH) acoustic listeners are represented by the black squares; hearing-impaired acoustic listeners with relatively normal tuning (HI) are represented by the black diamonds; hearing-impaired acoustic listeners with abnormal tuning (HIa) are represented by the open circles (abnormal tuning was judged by the HF slope). Cochlear-implant listeners are represented by the shaded circles (Clarion—monopolar) and the shaded triangles (Nucleus—bipolar). The abscissa differs for the acoustic listeners and cochlear-implant listeners. The tuning parameters for the acoustic listeners are plotted as a function of the level (dB SPL) of the acoustic probe used to collect the fmPTC. A wide range of probe levels were tested in both normal-hearing and hearing-impaired listeners but, of course, the range was smaller in hearing-impaired listeners. The tuning parameters for the cochlear-implant listeners are plotted as a function of percent dynamic range (%DRμA) for the probe pulse train. In addition, the geometric means for each group of cochlear-implant subjects are shown by the black circle (Clarion) and black triangle (Nucleus) to the left in (A), (B), and (D). In (C), both the LF slopes (dark-shaded triangles) and HF slopes (light-shaded triangles) of the cochlear implant tuning curves are compared with the Tail slopes (black and open symbols) of the acoustic tuning curves.

frequency (Oxenham and Shera, 2003). Using Eq. (5) from Oxenham and Shera (2003), the average Q values for normal-hearing acoustic listeners can be computed for the mean fmSTC tip frequencies measured here in the Clarion and Nucleus subjects (1536 and 2650 Hz). These values are 12.5 and 14.4, respectively. The average acoustic Q value at the highest fmSTC tip frequency evaluated in our cochlear-implant subjects (4298 Hz) is 16.5. If the acoustic hearing data in Fig. 9(D) were transposed so that the origin of the normal-hearing low-level data were positioned at 16.5 instead of 10, the Q values from the cochlear-implant listeners would still fall within the range of Q values obtained at high probe levels in acoustic listeners. Thus, differences in the tip frequencies tested in our acoustic (Nelson, 1991) and cochlear-implant subjects do not substantially alter the above-described comparisons.

It is also worth noting that the Q values for acoustic and cochlear-implant subjects were calculated at approximately equal levels within the dynamic range of hearing. Q values for the acoustic listeners were based on the bandwidths of the tuning curves 10 dB above the tip level. This corresponds to approximately 10% of the 100+-dB dynamic range available to acoustic listeners. Using the same 10-dB criterion to measure bandwidth in cochlear implant users was not realistic, since their dynamic ranges are roughly 1/10th as large as the normal-hearing acoustic dynamic range. In these cochlear-

implant users, the dynamic ranges for the masking stimuli ranged between 7 and 17 dB, averaging 11.3 dB for the Clarion (monopolar) users and 11.9 dB for the Nucleus (bipolar) users. Bandwidths in these subjects were measured 1 dB above the tip level, which also corresponds to approximately 10% of the average dynamic range.

If one were to calculate fmSTC slopes using the percentage of dynamic range in microamperes (%DRμA) as the dependent variable, instead of absolute microamperes, the values of the %DRμA/octave slopes would depend upon the size of the dynamic range and the constancy across electrodes of THS and MAL values. Slopes would become exceptionally steep or gradual when either the THS or the MAL varied dramatically across electrodes and the other variable (MAL or THS) did not (e.g., see N13 and N09 in Fig. 4). An analysis of fmSTC slopes in terms of %DRμA/octave may prove to be useful when predicting speech recognition through the implant, especially if the independent variable for the fmSTC (frequency in octaves) is recalculated according to the acoustic frequency mapping parameters for each individual implant user. However, at this point in time, there is little certainty about the physiological mechanisms underlying dramatic changes in MAL across electrodes. For this reason, we have chosen to limit our analyses of fmSTC slopes to units of absolute μA/mm and μA/octave.

We did, however, compare %DRdB/octave slopes from the normal-hearing and the hearing-impaired acoustic tuning curves in Fig. 7 with %DRμA/octave slopes from the cochlear-implant spatial tuning curves in Fig. 8 (N14). This comparison indicated that the hearing-impaired and cochlear implant slopes were less steep than the normal-hearing slopes, but the cochlear implant slopes were still within the same range as the hearing-impaired slopes. Thus, the general conclusion remained the same: The cochlear implant and the hearing-impaired tuning curve slopes were similar, and both were shallower than the normal-hearing slopes.

E. Factors influencing STC shapes

Cochlear-implant subjects in the present study demonstrated a wide range of fmSTC shapes. The factor with the strongest apparent influence on STC shapes was the mode of electrode coupling: Clarion subjects stimulated in monopolar mode demonstrated substantially shallower STC slopes (1.2 dB/mm, average for apical and basal slopes) than Nucleus subjects stimulated in bipolar mode (3.7 dB/mm). Average bandwidths were also larger for Clarion subjects (4.7 mm) than for Nucleus subjects (2.6 mm), although there was considerable variability in this parameter among Clarion subjects. Since fmSTC slopes are thought to reflect rates of current spread with spatial distance in the cochlea, the slope difference that we observed between Clarion (monopolar) and Nucleus (bipolar) subjects is most likely attributable to differences in rates of current attenuation associated with the two types of electrode coupling. However, some of the variability across subjects within groups may be related to the position of the electrode array within scala tympani. As described in the following, larger distances between the electrode contacts and the residual neurons would be expected to result in fmSTCs with shallower slopes and correspondingly larger bandwidths. Other factors that could potentially contribute to variability across individuals with similar electrode configurations include unique impedance pathways related to individual patterns of neural survival and bone growth in the cochlea (Liang *et al.*, 1999), and variations in the relative electrical resistivities for cochlear bone and fluid. It has been reported that individual differences in cochlear bone-to-fluid resistivity ratios have significant effects on rates of current attenuation in the human cochlea (Whiten and Eddington, 2007); however, to our knowledge, there are no data concerning the range of ratios that exist in human subjects.

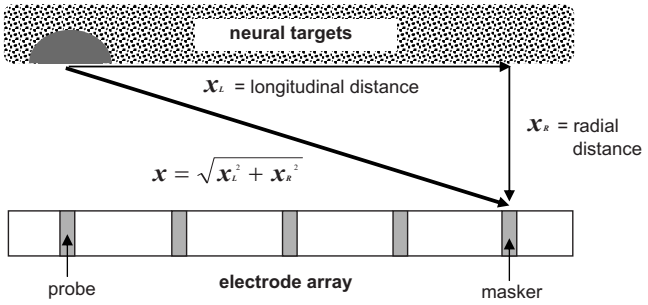


FIG. 10. Schematic demonstrating the spatial relationship between the masker electrode and the neural targets activated by the probe electrode. The horizontal dimension represents longitudinal distance along scala tympani; the vertical dimension represents radial distance between the electrode array and neural targets residing within the modiolus. The dark gray semicircle represents neural targets activated by the low-amplitude probe stimulus.

To demonstrate the potential influences of stimulation mode and radial distance on the shapes of fmSTCs, we used a simple model to generate hypothetical STCs. Scala tympani was approximated by a tube bordered by residual neurons along one side, as depicted in Fig. 10. The electrode array was displaced from the neural targets by a fixed distance, which corresponds to radial distance in a normal (spiral) cochlea. In Fig. 10, longitudinal distance along the cochlear duct ($\boldsymbol{x}_L$) is represented in the horizontal dimension, while radial distance ($\boldsymbol{x}_R$) is represented in the vertical dimension. The linear distance ($\boldsymbol{x}$) between the masker electrode contact and the neurons closest to the probe electrode was then computed as: $\boldsymbol{x}=\sqrt{\boldsymbol{x}_L^2+\boldsymbol{x}_R^2}$. $\boldsymbol{x}_R$ was allowed to range from 0 to 3 mm, reflecting the range of physical distances that is likely to exist between an electrode contact and residual nerve fibers or cell bodies within the cochlear modiolus (Cohen *et al.*, 2001; Saunders *et al.*, 2002; Cohen *et al.*, 2006).

The amplitude of current applied to the masker electrode was assumed to decay exponentially with linear distance, following

$$I_p = I_m e^{-\boldsymbol{x}/\lambda}, \tag{1}$$

where I_m is the current amplitude at the masker electrode, I_p is the current amplitude at the probe electrode, $\boldsymbol{x}$ is the distance between the masker and probe electrodes described earlier, and λ is the length constant of current attenuation. This exponential decay function provides a reasonable approximation of current flow, except at locations very close

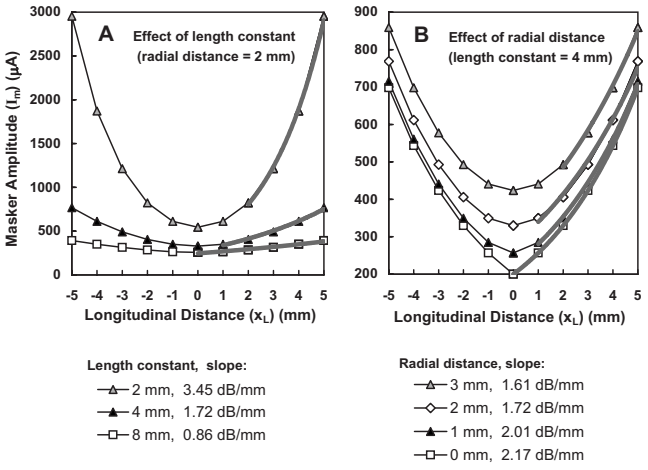


FIG. 11. Hypothetical STCs demonstrating the effect of length constant and radial distance on STC shapes. (A) As the length constant increases from 2 to 8 mm, the tuning curve slope decreases from 3.5 to 0.86 dB/mm. (B) As radial distance increases from 0 to 3 mm, the tip of the STC is elevated and rounded, and the slope is reduced by approximately 25%.

(e.g., <1 mm) to the stimulating electrode (Briaire and Frijns, 2000).

STCs were constructed by computing the level of I_m needed to produce a constant value of I_p, as a function of masker electrode. I_p was fixed arbitrarily at 200 μA, which approximates the average fmSTC tip levels observed in the present data for monopolar stimulation. Note that altering I_p causes the STC to shift vertically (in overall amplitude), but does not change STC slopes.

Figure 11 shows model STCs demonstrating the effects of stimulation mode and radial distance on STC shapes. In Fig. 11(A), STCs are shown for length constants of 2, 4, and 8 mm, for a fixed radial distance (x_R) of 2 mm. The STCs become progressively broader as the length constant increases from 2 to 8 mm, as expected. Slopes were quantified using the same procedure used for fmSTCs shown in earlier figures; that is, an exponential function was fit to the steepest segments, as indicated by the heavy gray lines. The resultant slopes decreased from 3.5 dB/mm for the 2 mm length constant to 0.86 dB/mm for the 8 mm length constant.

Figure 11(B) demonstrates the effect of radial distance on STC shapes. As radial distance increases, STC slopes become shallower and STCs become more rounded near the tip. In this example, STC slopes decrease approximately 25%, from 2.17 to 1.61 dB/mm, as radial distance increases from 0 to 3 mm. These model STCs were generated using a length constant of 4 mm; however, proportional effects would be observed for other length constants. Radial distance clearly has a much smaller effect on STC slopes than length constant; however, it may account for some portion of the within-group differences in STC slopes and shapes observed in our data.

In order to determine what length constants were most consistent with our average slope data for bipolar and monopolar stimulation (3.7 and 1.2 dB/mm, respectively), theoretical fmSTC slopes were calculated for combinations of length constants ranging from 2 to 10 mm and radial distances ranging from 0 to 3 mm. The resulting slopes are listed in Table VII. The average monopolar slope measured in the present study (1.2 dB/mm) falls within the range of model slopes (1.08–1.45 dB/mm) generated with a 6-mm length constant, and the average bipolar slope (3.7 dB/mm) falls within the range of model slopes (3.23–4.34) generated with a 2-mm length constant. Thus, our average fmSTC data are consistent with length constants of 2 mm for bipolar stimulation and 6 mm for monopolar stimulation.

TABLE VII. Slopes of theoretical fmSTCs (dB/mm), as a function of length constant (mm) and radial distance (mm).

	Radial distance			
Length constant (mm)	0 mm	1 mm	2 mm	3 mm
1	8.69	8.04	6.90	6.46
2	4.34	4.02	3.45	3.23
3	2.90	2.68	2.30	2.15
4	2.17	2.01	1.72	1.61
5	1.74	1.61	1.38	1.29
6	1.45	1.34	1.15	1.08
7	1.24	1.15	0.99	0.92
8	1.09	1.00	0.86	0.81
9	0.97	0.89	0.77	0.72
10	0.87	0.80	0.69	0.65

Previous studies have estimated current attenuation in the cochlea by measuring current levels directly or by measuring neural responses from the primary auditory nerve or IC as a function of the intracochlear position of the stimulating electrode (see Kral *et al.*, 1998, for a review). Early studies reported length constants of approximately 3 mm for bipolar stimulation and 10 mm for monopolar stimulation (Black and Clark, 1980; Black *et al.*, 1981, 1983; O'Leary *et al.*, 1985). More recent studies which measured primary auditory nerve responses in cat as a function of electrode position have reported attenuation slopes of about 8 dB/mm for bipolar stimulation and 3 dB/mm for monopolar stimulation (Hartmann and Klinke, 1990; Kral *et al.*, 1998). These slopes correspond to length constants of 1 and 3 mm, respectively. Thus, although there is a consistent ratio of monopolar-to-bipolar length constants (~3) across studies, the absolute length constants reported in previous studies vary threefold, presumably due to differences in the methodologies employed and species examined (Kral *et al.*, 1998). The average length constants derived from the present fmSTC data fall within the broad range of length constants reported previously. Specifically, our average length constant of 2 mm for bipolar stimulation falls within the range of length constants previously reported for bipolar stimulation (1–3 mm), and our average length constant for monopolar stimulation (6 mm) falls within the range of length constants previously reported for monopolar stimulation (3–10 mm). In addition, our data exhibit a similar ratio of monopolar-to-bipolar length constants (~3) as that described in earlier studies.

F. Implications for speech recognition

Given that a primary goal of cochlear implantation is to facilitate speech recognition, it is relevant to consider the

potential relationship between fmSTC parameters, such as those measured in the present study, and speech recognition in individual cochlear implant users. Previous studies have shown that implant users vary widely in their ability to perform various spectral-resolution tasks, and that performance on such tasks may predict speech recognition ability, in general, or the perception of spectral speech cues, in particular (Dorman *et al.*, 1990a, b; Nelson *et al.*, 1995; Dorman *et al.*, 1996; Donaldson and Nelson, 2000; Henry *et al.*, 2000, 2005; Won *et al.*, 2007).

Some of the strongest relations between psychophysical and speech tasks have been observed when both the psychophysical and speech stimuli are presented through the speech processor (Dorman *et al.*, 1990a, b; 1996; Henry *et al.*, 2005; Won *et al.*, 2007). Under those conditions, both tasks are limited by the sum total of constraints to spectral resolution existing for a given individual, which include constraints imposed by the speech processor (compression and basalward shifting of the frequency-place map; analysis bandwidths; filter slopes), constraints associated with the electrode array (spacing between electrode contacts; position of the electrode array within the cochlear duct), and constraints related to the cochlear status of the individual (degree and pattern of auditory nerve survival; bone growth; other factors that influence current pathways).

On the other hand, the fmSTC parameters reported here do not reflect constraints imposed by the speech processor or constraints associated with the electrode array. Instead, they reflect physiological limitations to spatial resolution at the level of the cochlea or above: rates of current attenuation (fmSTC slopes and bandwidths) and distortions in the tonotopic map due to "holes" in neural survival and aberrant current pathways (fmSTC tip locations). Thus, for a given individual and stimulation mode, fmSTC parameters can be considered to reflect the "hard" limits of spatial resolution—those that cannot be readily overcome by improving spatial resolution at the level of the speech processor and/or the electrode array.

Because fmSTC parameters do not reflect limitations to spatial resolution imposed by the speech processor or electrode array, we do not necessarily expect them to correlate strongly with measures of speech recognition obtained through the speech processor. Furthermore, any comparisons between STC parameters and speech recognition should evaluate fmSTCs from multiple probe electrodes per subject. The present study reports fmSTC parameters for only a single electrode near the center of each subject's electrode array. Thus, the current data are too limited to allow meaningful comparisons to speech recognition scores.

Some of the factors that limit spatial resolution by cochlear implant users, notably those associated with the speech processor and electrode array, can be mitigated by improvements in technology. For example, deeper insertion of electrode arrays allows for less compression of the basic frequency-to-place map (Baskent and Shannon, 2003, 2005; Gani *et al.*, 2007). Similarly, the implementation of current steering in speech processing strategies can potentially reduce limitations related to fixed electrode locations by allowing current peaks to be "steered" to varying locations between pairs of adjacent electrodes (McDermott and McKay, 1994; Busby and Plant, 2005; Donaldson *et al.*, 2005; Kwon and van den Honert, 2006; Firszt *et al.*, 2007; Koch *et al.*, 2007). However, little has been done thus far to address underlying irregularities in neural survival and aberrant current pathways, which may severely degrade the transmission of spectral information for some CI listeners (Shannon *et al.*, 2001). The development of strategies to compensate for such irregularities will require that "electrotopic" maps be constructed for individual ears. This can be accomplished by analyzing fmSTCs for probe electrodes spanning the entire implanted array, since regions of poor or absent neural survival should produce fmSTCs with shifted or broadened tips. In this way, fmSTCs may prove to be particularly useful for developing remapping strategies for enhancing spectral resolution in individual cochlear implant recipients.

IV. SUMMARY AND CONCLUSIONS

fmSTCs were obtained at several probe levels, for each of six Clarion cochlear implant users stimulated in a monopolar electrode configuration, and for each of six Nucleus cochlear implant users stimulated in a bipolar electrode configuration. Tuning-curve characteristics were calculated for each of the spatial tuning curves and were compared with characteristics of fmPTCs obtained previously in normal-hearing listeners and listeners with sensorineural hearing loss (Nelson, 1991). Analyses led to the following key findings and conclusions:

(1) fmSTCs measured in cochlear implant users can be characterized by apical and basal slopes (dB/mm) and by bandwidths (mm) measured 1 dB above the tuning-curve tip.
(2) Apical and basal slopes of fmSTCs are not significantly different from each other, i.e., spatial tuning curves are relatively symmetric for cochlear implant listeners.
(3) Slopes of fmSTCs are relatively independent of probe level for probe levels between 8% and 33% of the dynamic range of the probe stimulus. This finding is consistent with a linear growth of response in electric hearing. It contrasts with the nonlinear growth of response in normal acoustic hearing that is associated with a progressive decrease in tuning sharpness with stimulus level.
(4) Slopes of fmSTCs were steeper for subjects stimulated with a bipolar electrode configuration (avg=3.7 dB/mm) than for subjects stimulated with a monopolar electrode configuration (avg=1.2 dB/mm). This primarily reflects differences in the rates of current attenuation for monopolar versus bipolar stimulation. The present data are consistent with length constants of 2 mm for bipolar stimulation and 6 mm for monopolar stimulation. These values fall within the broad range reported in previous studies and, similar to earlier studies, indicate that length constants for monopolar stimulation are approximately three times greater than those for bipolar stimulation. Additional within-subjects research is needed to more precisely define the effects of electrode configuration on fmSTC slopes.

(5) Bandwidths of fmSTCs measured 1 dB above the STC tip were not significantly different ($p=0.08$) for bipolar subjects and monopolar subjects, even though the average bandwidth for monopolar subjects (4.7 mm) was substantially larger than the average bandwidth for bipolar subjects (2.6 mm). A within-subjects investigation of electrode configuration could eliminate between-subjects variance and potentially reveal a significant effect.
(6) Some of the individual variability in fmSTC slopes and bandwidths among cochlear implant users with the same device may reflect differences in the radial position of the electrode array within scala tympani.
(7) Irregularities near the tips of fmSTCs were more common in bipolar subjects than in monopolar subjects. Such irregularities may be due to the peaks and nulls of current fields associated with the bipolar maskers.
(8) fmSTCs converted to frequency coordinates, by calculating the spatial frequency corresponding to each electrode using a tonotopic place-frequency function (Greenwood, 1961), yield Apical and Basal slopes of fmSTCs that are similar to the Tail slopes of tuning curves measured in acoustic listeners. Consistent with this, bandwidths of fmSTCs are similar to those measured in acoustic listeners at high stimulus levels: The range of Q values in electric hearing overlapped the range of Q values seen at high levels in normal-hearing and hearing-impaired acoustic listeners.
(9) Taken together, the above-noted findings indicate that cochlear implant users have roughly the same broad spatial frequency resolution at low to moderate stimulus levels that normal-hearing and hearing-impaired acoustic listeners have at high stimulus levels. This relationship is coincidental, since spatial resolution in cochlear implant listeners is governed by different factors (current attenuation rate, radial electrode position, neural survival, and current pathways) than the factors that govern sharpness of tuning in acoustic hearing (active and passive cochlear mechanics). However, it does indicate that the typical cochlear implant user can perform spectral resolution tasks under direct electrical stimulation as well as normal-hearing and hearing-impaired acoustic listeners can perform similar tasks at high sound levels. Thus, constraints imposed by the speech processor and the electrode array, rather than the underlying spatial tuning characteristics, may be the primary factors limiting speech recognition for many cochlear implant users.
(10) fmSTC obtained from multiple electrodes in an individual cochlear implant user may identify regions of poor or absent neural survival. This knowledge may prove useful for developing remapping strategies to improve implant performance.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. R01-DC006699 and by the Lions 5M International Hearing Foundation. John Van Essen converted Robert Shannon's computer software into the C language and made modifications to that software for testing Nucleus subjects. Cochlear Corporation provided Nucleus subjects' calibration tables. Advanced Bionics Corporation provided the research interface used for testing Clarion subjects, and Eric Javel developed the experimental software to control that interface. The authors would like to extend special thanks to the subjects who participated in this work, and to Andrew Oxenham, Bob Shannon, and an anonymous reviewer, who made valuable suggestions during the review of this manuscript.

[1]It should be noted here that a between-subjects design, using subjects with two different types of processors, is not optimal for demonstrating differences due to electrode configuration. However, at the time this research was carried out, it was not possible to implement both monopolar and bipolar stimulation on the same subjects, using the subjects that were available to us. Therefore, we report here the results from a group of Clarion C-I and C-II subjects stimulated in monopolar mode and a group of Nucleus N-22 subjects stimulated in bipolar mode, with the caveat that differences across subjects and processors could influence our group comparisons of monopolar and bipolar electrode configuration.

[2]Although BP+1 and BP+2 maskers may result in broader tuning than BP maskers, some subjects could not achieve adequate loudness growth with BP probes or adequate forward masking with BP maskers to permit measurement of fmSTCs using a BP/BP configuration. Thus, the present results reflect the spatial tuning available to each subject using the same electrode configuration used in their clinical map.

[3]The split tip is particularly obvious for N09. It seems unlikely that this is related to the slower pulse rates used for this subject; however, it may be related to the use of a wider bipolar mode (BP+1) for the probe stimulus. Note that the slopes and bandwidths for N09 are not substantially different from the other Nucleus subjects, despite the differences in stimulation rate and electrode configuration.

[4]The term spatial frequency is used here to indicate the acoustic frequency associated with a particular spatial distance along the cochlear duct. It is intended only as an *approximation* of the acoustic frequency previously associated with a particular place. Individual differences in cochlear length or errors in our estimates of insertion depth could have significant effects on our place-to-frequency estimates.

[5]The Q values obtained from normal-hearing listeners at the lowest probe levels compare favorably with the average Q_{ERB} value of 9 obtained using two forward maskers, one above and one below a 1000 Hz probe (Oxenham and Shera, 2003).

Baskent, D., and Shannon, R. V. (**2003**). "Speech recognition under conditions of frequency-place compression and expansion," J. Acoust. Soc. Am. **113**, 2064–2076.
Baskent, D., and Shannon, R. V. (**2005**). "Interactions between cochlear implant electrode insertion depth and frequency-place mapping," J. Acoust. Soc. Am. **117**, 1405–1416.
Bierer, J. A. (**2007**). "Threshold and channel interaction in cochlear implant users: evaluation of the tripolar electrode configuration," J. Acoust. Soc. Am. **121**, 1642–1653.
Black, R. C., and Clark, G. M. (**1980**). "Differential electrical excitation of the auditory nerve," J. Acoust. Soc. Am. **67**, 868–874.
Black, R. C., Clark, G. M., and Patrick, J. F. (**1981**). "Current distribution measurement within the human cochlea," IEEE Trans. Biomed. Eng. **28**, 721–724.
Black, R. C., Clark, G. M., Tong, Y. C., and Patrick, J. F. (**1983**). "Current distribution in cochlear stimulation," Ann. N.Y. Acad. Sci. **405**, 137–145.
Boex, C., Kos, M. I., and Pelizzone, M. (**2003**). "Forward masking in different cochlear implant systems," J. Acoust. Soc. Am. **114**, 2058–2065.
Briaire, J. J., and Frijns, J. H. M. (**2000**). "Field patterns in a 3D tapered spiral model of the electrically stimulated cochlea," Hear. Res. **148**, 18–30.
Busby, P. A., and Plant, K. L. (**2005**). "Dual electode stimulation using the Nucleus CI24RE cochlear implant: electrode impedance and pitch ranking studies," Ear Hear. **26**, 504–511.
Chatterjee, M. (**1999**). "Temporal mechanisms underlying recovery from forward masking in multielectrode implant listeners," J. Acoust. Soc. Am. **105**, 1853–1863.
Chatterjee, M., Galvin, J. J., Fu, Q. J., and Shannon, R. V. (**2006**). "Effects of stimulation mode, level and location on forward-masked excitation pat-

terns in cochlear implant patients," J. Assoc. Res. Otolaryngol. **7**, 15–26.
Chatterjee, M., and Shannon, R. V. (**1998**). "Forward masked excitation patterns in multielectrode electrical stimulation," J. Acoust. Soc. Am. **103**, 2565–2572.
Cohen, L. T., Richardson, L. M., Saunders, E., and Cowan, R. (**2003**). "Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking," Hear. Res. **179**, 72–87.
Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psychophysics of a peri-modiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.
Cohen, L. T., Saunders, E., Knight, M. R., and Cowan, R. S. C. (**2006**). "Psychophysical measures in patients fitted with Contour (TM) and straight nucleus electrode arrays," Hear. Res. **212**, 160–175.
Collins, L. M., and Throckmorton, C. S. (**1998**). "Investigation of the relationship between psychophysical measures of channel interaction and perceptual structure in subjects with cochlear implants," Assoc. Res. Otolaryngol. Abstracts, #872, St. Petersburg, FL.
Dingemanse, J. G., Frijns, J. H. M., and Briaire, J. J. (**2006**). "Psychophysical assessment of spatial spread of excitation in electrical hearing and single and dual electrode contact maskers," Ear Hear. **27**, 645–657.
Donaldson, G. S., Kreft, H. A., and Litvak, L. (**2005**). "Place-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users," J. Acoust. Soc. Am. **118**, 623–626.
Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.
Donaldson, G. S., Viemeister, N. F., and Nelson, D. A. (**1997**). "Psychometric functions and temporal integration in electrical hearing," J. Acoust. Soc. Am. **101**, 3706–3721.
Dorman, M. F., Dankowski, K., McCandless, G., Parkin, J. L., and Smith, L. (**1990a**). "Longitudinal changes in word recognition by patients who use the Ineraid cochlear implant," Ear Hear. **11**, 455–459.
Dorman, M. F., Smith, L., McCandless, G., Dunnavant, G., Parkin, J., and Dankowski, K. (**1990b**). "Pitch scaling and speech understanding by patients who use the Ineraid cochlear implant," Ear Hear. **11**, 310–315.
Dorman, M. F., Smith, L. M., Smith, M., 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.
Firszt, J. B., Koch, D. B., Downing, M., and Litvak, L. (**2007**). "Current steering creates additional pitch percepts in adult cochlear implant recipients," Otol. Neurotol. **28**, 629–636.
Frijns, J. H., Briaire, J. J., de Laat, J. A., and Grote, J. J. (**2002**). "Initial evaluation of the Clarion CII cochlear implant: speech perception and neural response imaging," Ear Hear. **23**, 184–197.
Frijns, J. H. M., Briaire, J. J., and Grote, J. J. (**2001**). "The importance of human cochlear anatomy for the results of modiolus-hugging multichannel cochlear implants," Otol. Neurotol. **22**, 340–349.
Fu, Q. J., Chatterjee, M., Shannon, R. V., and Zeng, F. G. (**1997**). "Comparison of electrode interaction measures in multichannel cochlear implants," Assoc. Res. Otolaryngol. Abstracts, #305, St. Petersburg, FL.
Gani, M., Valentini, G., Sigrist, A., Kos, M. I., and Boex, C. (**2007**). "Implications of deep electrode insertion on cochlear implant fitting," J. Assoc. Res. Otolaryngol. **8**, 69–83.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.
Hartmann, R., and Klinke, R. (**1990**). "Impulse patterns of auditory nerve fibres to extra and intracochlear electrical stimulation," Acta Oto-Laryngol., Suppl. **469**, 128–134.
Henry, B. A., McKay, C. M., McDermott, H. J., and Clark, G. M. (**2000**). "The relationship between speech perception and electrode discrimination in cochlear implantees," J. Acoust. Soc. Am. **108**, 1269–1280.
Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.
Hinojosa, R. and Lindsay, J. R. (**1980**). "Profound deafness: Associated sensory and neural degeneration," Arch. Otolaryngol. **106**, 193–209.
Hughes, M., Vander Werff, K. R., Brown, C. J., Abbas, P. J., Kelsay, M. R., Teagle, H. F. B., and Lowder, M. (**2001**). "A longitudinal study of electrode impedance, the electrically evoked compound action potential, and behavioral measures in Nucleus 24 cochlear implant users," Ear Hear. **22**, 471–486.
Jolly, C. N. (**1998**). "Breakthrough in perimodiolar concepts," presented at the Seventh Symposium on Cochlear Implants in Children, Iowa City (unpublished).
Jolly, C. N., Spelman, F. A., and Clopton, B. M. (**1996**). "Quadrupolar stimulation for cochlear prostheses: Modeling and experimental data," IEEE Trans. Biomed. Eng. **43**, 857–865.
Kawano, A., Seldon, H. L., Clark, G. M., Ramsden, R. T., and Raine, C. H. (**1998**). "Intracochlear factors contributing to psychophysical percepts following cochlear implantation," Acta Oto-Laryngol. **118**, 313–326.
Kessler, D. K. (**1999**). "The CLARION multi-strategy cochlear implant," Ann. Otol. Rhinol. Laryngol. Suppl. **177**, 8–16.
Koch, D. B., Downing, M., Osberger, M. J., and Litvak, L. (**2007**). "Using current steering to increase spectral resolution in CII and HiRes 90K users," Ear Hear. **28**, 38S–41S.
Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (**1998**). "Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents," Hear. Res. **121**, 11–28.
Kwon, B. J., and van den Honert, C. (**2006**). "Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners," J. Acoust. Soc. Am. **119**, 2994–3002.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liang, D. M., Lusted, H. S., and White, R. L. (**1999**). "The nerve-electrode interface of the cochlear implant: Current spread," IEEE Trans. Biomed. Eng. **46**, 35–43.
Lim, H. H., Tong, Y. C., and Clark, G. M. (**1989**). "Forward masking patterns produced by intracochlear electical stimulation of one and two electrode pairs in the human coclea," J. Acoust. Soc. Am. **86**, 971–979.
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.
Moore, B. C. J. (**2001**). "Dead regions in the cochlea: diagnosis, perceptual consequences, and implications for the fitting of hearing aids," Trends Amplif. **5**, 1–34.
Moore, B. C. J., and Alcantara, J. I. (**2001**). "The use of psychophysical tuning curves to explore dead regions in the cochlea," Ear Hear. **22**, 268–278.
Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Braz. J. Phys. **34**, 205–224.
Morris, D. J., and Pfingst, B. E. (**2000**). "Effects of electrode configuration and stimulus level on rate and level discrimination with cochlear implants," J. Assoc. Res. Otolaryngol. **1**, 211–223.
Nadol, J. B., Shia, J. Y., Burgess, B. J., Ketten, D. R., Eddington, D. K., Gantz, B. J., Kos, I., Montandon, P., Coker, N. J., Roland, J. T., and Shallop, J. K. (**2001**). "Histopathology of cochlear implants in the human," Ann. Otol. Rhinol. Laryngol. **110**, 883–891.
Nelson, D. A. (**1991**). "High-level psychophysical tuning curves: Forward masking in normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **34**, 1233–1249.
Nelson, D. A., and Donaldson, G. S. (**2002**). "Psychophysical recovery from pulse-train forward masking in electric hearing," J. Acoust. Soc. Am. **112**, 2932–2947.
Nelson, D. A., and Freyman, R. L. (**1984**). "Broadened forward-masked tuning curves from intense masking tones: delay-time and probe-level manipulations," J. Acoust. Soc. Am. **75**, 1570–1577.
Nelson, D. A., and Schroder, A. C. (**2004**). "Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2221–2233.
Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.
Nelson, D. A., Van Tasell, D. J., 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.
O'Leary, S. J., Black, R. C., and Clark, G. M. (**1985**). "Current distribution in the cat cochlea: A modelling and electrophysiological study," Hear. Res. **18**, 273–281.
Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.

Patrick, J. F., and Clark, G. M. (**1991**). "The nucleus 22-channel cochlear implant system," Ear Hear. **12**, Suppl., 3S–9S.
Pfingst, B. E., and Xu, L. (**2004**). "Across-site variation in detection thresholds and maximum comfortable loudness levels for cochlear implants," J. Assoc. Res. Otolaryngol. **5**, 11–24.
Pfingst, B. E., Xu, L., and Thompson, C. S. (**2004**). "Across-site threshold variation in cochlear implants: Relation to speech recognition," Audiol. Neuro-Otol. **9**, 341–352.
Saunders, E., Cohen, L., Aschendorff, A., Shapiro, W., Knight, M., Stecker, M., Richter, B., Waltzman, S., Tykocinski, M., Roland, T., Laszig, R., and Cowan, R. (**2002**). "Threshold, comfortable level and impedance changes as a function of electrode-modiolar distance," Ear Hear. **23**, 28S–40S.
Schindler, R. A., and Kessler, D. K. (**1993**). "Clarion cochlear implant: phase I investigation results," Am. J. Otol. **14**, 263–272.
Shannon, R. V. (**1983a**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.
Shannon, R. V. (**1983b**). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.
Shannon, R. V. (**1986**). "Temporal processing in cochlear implants," in *Sensorineural Hearing Loss: Mechanisms, Diagnosis and Treatment—The Reger Conference*, edited by M. J. Collins, T. J. Glattke, and L. A. Harker, pp. 349–367.
Shannon, R. V. (**1990a**). "Forward masking in patients with cochlear implants," J. Acoust. Soc. Am. **88**, 741–744.
Shannon, R. V., Adams, D. D., Ferrel, R. L., Palumbo, R. L., and Grandgenett, M. (**1990**). "A computer interface for psychophysical and speech research with the Nucleus cochlear implant," J. Acoust. Soc. Am. **87**, 905–907.
Shannon, R. V., Galvin, J. J., and Baskent, D. (**2001**). "Holes in hearing," J. Assoc. Res. Otolaryngol. **3**, 185–199.
Shepherd, R. K., Matsushima, J., Martin, R. L., and Clark, G. M. (**1994**). "Cochlear pathology following chronic electrical stimulation of the auditory nerve: II Deafened kittens," Hear. Res. **81**, 150–166.
Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., Neely, J. G., Kletzker, G. R., Brunsden, B., and Blocker, B. (**2002**). "CT-Derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients," J. Assoc. Res. Otolaryngol. **3**, 332–350.
Spelman, F. A., Clopton, B. M., and Pfingst, B. E. (**1982**). "Tissue impedance and current flow in the implanted ear; implications for the cochlear prothesis," Ann. Otol. Rhinol. Laryngol. **91**, 1–8.
Spoendlin, H. and Schrott, A. (**1988**). "The spiral ganglion and the innervation of the human Organ of Corti," Acta Oto-Laryngol. **105**, 403–410.
Whiten, D. M., and Eddington, D. K. (**2007**). "Validating patient-specific electro-anatomical models: Empirically measured and model-predicted evoked potentials and EFI data from the same individual," presented at the 2007 Conference on Implantable Auditory Prostheses, Lake Tahoe, CA (unpublished).
Wilson, B. S., Lawson, D. T., Zerbi, M., and Finley, C. C. (**1994**). "Recent developments with the CIS Strategies," in *Advances in Cochlear Implants, Proceedings Third International Cochlear Implant Conference (Innsburck, Austria, April 1993).* edited by I. J. Hochmair-Desoyer and E. S. Hochmair (Manz, Wien), pp. 103–112.
Won, J. H., Drennan, W. R., and Rubinstein, J. T. (**2007**). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," J. Assoc. Res. Otolaryngol. **8**, 384–392.
Zappia, J. J., Niparko, J. K., Oviatt, D. L., Kemink, J. L., and Altschuler, R. A. (**1991**). "Evaluation of the temporal bones of a multichannel cochlear implant patient," Ann. Otol. Rhinol. Laryngol. **100**, 914–921.

A variant temporal-masking-curve method for inferring peripheral auditory compression[a)]

Enrique A. Lopez-Poveda[b)] and Ana Alves-Pinto
Unidad de Audición Computacional y Psicoacústica, Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, 37007 Salamanca, Spain

(Received 14 September 2007; revised 14 December 2007; accepted 18 December 2007)

Recent studies have suggested that the degree of on-frequency peripheral auditory compression is similar for apical and basal cochlear sites and that compression extends to a wider range of frequencies in apical than in basal sites. These conclusions were drawn from the analysis of the slopes of temporal masking curves (TMCs) on the assumption that forward masking decays at the same rate for all probe and masker frequencies. The aim here was to verify this conclusion using a different assumption. TMCs for normal hearing listeners were measured for probe frequencies (f_P) of 500 and 4000 Hz and for masker frequencies (f_M) of 0.4, 0.55, and 1.0 times the probe frequency. TMCs were measured for probes of 9 and 15 dB sensation level. The assumption was that given a 6 dB increase in probe level, linear cochlear responses to the maskers should lead to a 6 dB vertical shift of the corresponding TMCs, while compressive responses should lead to bigger shifts. Results were consistent with the conclusions from earlier studies. It is argued that this supports the assumptions of the standard TMC method for inferring compression, at least in normal-hearing listeners. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2835418]

PACS number(s): 43.66.Ba, 43.66.Dc, 43.66.Mk [MW] Pages: 1544–1554

I. INTRODUCTION

It is now believed that our ability to perceive sounds over a wide range of sound pressure levels (SPLs) is accomplished via a form of compression that takes place in the peripheral auditory system and is almost certainly related to basilar membrane (BM) compression (Bacon *et al.*, 2004). There exist several psychoacoustical methods to estimate the amount of peripheral compression in humans: the growth-of-masking (Oxenham and Plack, 1997), the additivity of masking (Plack *et al.*, 2006), the pulsation threshold (Plack and Oxenham, 2000), and the temporal-masking-curve (TMC) (Nelson *et al.*, 2001) methods are the most significant. Of these, the TMC method of Nelson *et al.* (2001) is perhaps the most accurate because it minimizes off-frequency listening. However, some of its assumptions and the results obtained with this method have been recently questioned (see the following text). This report describes a variant TMC method that is based on different assumptions than the standard method and yet yields similar results.

The method of Nelson *et al.* (2001) consists of measuring the level of a tonal forward masker required to just mask a fixed tonal probe as a function of the time interval between the masker and the probe. A TMC is a graphical representation of the resulting masker levels against the corresponding masker–probe intervals. Because the probe level is fixed, the masker level increases with increasing the masker–probe time interval and hence TMCs have positive slopes. Nelson *et al.* (2001) argued that the slope of any given TMC depends *simultaneously* on two factors: (a) the amount of BM compression affecting the masker at a cochlear place whose characteristic frequency (CF) equals approximately the probe frequency and (b) the rate of decay of the internal (postcochlear) masker effect. [A full justification of these assumptions can be found in Nelson *et al.* (2001)]. By *assuming* that the time course of decay of the postcochlear masker effect is identical for *all* masker frequencies (and levels), the degree of BM compression can be estimated by comparing the slope of a TMC for any masker frequency against the slope of a TMC for a masker that is processed linearly by the BM (usually referred to as the linear reference TMC). Clearly, the validity of the compression estimates obtained with this method depends on (1) the selection of a suitable linear reference TMC and (2) the accuracy of the assumption that the time course of decay of the postcochlear masker effect is the same for all masker frequencies. Note, however, that this method does not require making assumptions about the actual form of the decay function.

The selection of the linear reference TMC has been justified based on physiological evidence. Nelson *et al.* (2001) reasoned that the TMC for a masker frequency approximately an octave below the probe frequency would be a good linear reference because the BM responds linearly to stimuli well below CF (Robles and Ruggero, 2001). Lopez-Poveda *et al.* (2003) pointed out, however, that there is physiological evidence that compression extends to very low frequencies (relative to CF) for apical cochlear sites (Rhode and Cooper, 1996) and proposed using an off-frequency masker (2 kHz) for a *high*-frequency probe (4 kHz) as the linear reference throughout. This suggestion stands on the

[a)]Portions of this work were presented at the British Society of Audiology Short Papers Meeting on Experimental Studies on Hearing and Deafness, London, 16–17 September 2004; and at the 151st Meeting of the Acoustical Society of America, Providence, Rhode Island, 5–9 June 2006.

[b)]Author to whom correspondence should be addressed. Electronic mail: ealopezpoveda@usal.es

 0001-4966/2008/123(3)/1544/11/$23.00

additional (third) assumption that the time course of decay of the postcochlear masker is the same for all *probe* frequencies.

Based on these three assumptions, Lopez-Poveda *et al.* (2003) concluded that the degree of on-CF BM compression is similar across cochlear places with CFs from 500 Hz to 8 kHz, and that BM compression extends to a wider range of frequencies (relative to CF) in apical than in basal cochlear sites. These results have been confirmed independently by other authors using similar (e.g., Plack and Drga, 2003; Nelson and Schroder, 2004) and also different assumptions (e.g., Plack and Drga, 2003).

The assumptions of the TMC method, however, have been recently challenged by the characteristics of TMCs of hearing-impaired listeners. First, off-frequency TMCs at 4 kHz appear on average twice as steep for normal-hearing listeners as for listeners with mild-to-moderate sensorineural hearing loss (Plack *et al.*, 2004; Lopez-Poveda *et al.*, 2005). This suggests that either the postcochlear effect of the masker decays more slowly for hearing impaired listeners *or* that the off-frequency TMCs of normal hearing listeners are subject to some degree of cochlear compression even at 4 kHz. If the latter were true, it might invalidate the "standard" linear reference TMC, as defined by Lopez-Poveda *et al.* (2003). Second, the TMCs of *some* listeners with uniform moderate-to-severe sensorineural hearing loss are steeper for low than for high probe frequencies (Stainsby and Moore, 2006). The lack of measurable distortion-product otoacoustic emissions (DPOAEs) in these listeners was taken as indicative of a passive, thus linear, cochlea. Therefore, the result was interpreted as though the postcochlear masker effect decays more rapidly with time for lower than for higher probe frequencies (Stainsby and Moore, 2006), which undermines the conclusion that the degree of BM compression is similar across cochlear sites (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; Nelson and Schroder 2004).

The present study aimed to reassess on- and off-CF compression at 0.5 and 4 kHz in normal-hearing listeners using a variant TMC method that does not require assuming identical time courses of decay across frequencies or a linear reference TMC. The results support the conclusions of earlier studies based on the standard TMC method. The results also suggest that the TMC for a 4 kHz probe and a 2.2 kHz masker may still be subject to compression and hence that those studies that have employed this TMC as a linear reference may have underestimated the degree of compression. It will be argued that the TMC for a 4 kH probe and a 1.6 kHz masker is an appropriate linear reference and that, contrary to what has been suggested by studies in hearing-impaired listeners, the standard TMC method is valid to estimate BM compression in *normal-hearing* listeners, even at 500 Hz, when this linear reference is used.

II. METHODS

A. Rationale and assumptions

The standard TMC method allows for the inference of compression for any given frequency by comparing the slopes of the TMC for that frequency with that of a linear reference TMC. The method proposed here, by contrast, allows estimating compression from two TMCs that are identical except that they are measured with slightly different probes levels. The *assumption* is that given a small increase in probe level (ΔL_P dB), linear cochlear responses to the maskers should lead to a similar increase in masker level for any masker–probe time interval and thus to a ΔL_P-dB vertical shift of the corresponding TMCs, while compressive responses should lead to bigger shifts. For a fixed masker–probe interval, the ratio of the masker-level increment to the probe-level increment provides a direct compression-ratio estimate and its inverse an estimate of the slope of the BM input/output (I/O) function (or the compression exponent). For example, a 24 dB increase in masker level for 6 dB probe-level increase would imply a 4:1 compression ratio or a 0.25 compression exponent. The masker level increases as the masker–probe time interval increases. Therefore, measuring the masker-level increment for different masker-probe intervals allows for the inference of the degree of compression for different masker levels.

This new method has the same advantages as the standard TMC method. As with the standard TMC method, the negative effects of off-frequency listening (O'Loughlin and Moore, 1981) are minimized by using two *low* probe levels (say 9 and 15 dB SL). The two probe levels are similar, low and fixed for each TMC, hence it is also reasonable to assume that the compression estimates correspond to a BM place with a CF approximately equal to the probe frequency. Furthermore, like the standard TMC method, the new method does not require making assumptions about the actual form of the function describing the decay of the postcochlear masker effect. The proposed method, however, has the advantage over the standard method that it circumvents the need for a linear reference TMC or for assuming that the decay function is the same across frequencies. Hence, this new method also allows testing the linearity of the linear reference TMCs used in earlier studies (see the following text).

The proposed method, however, assumes that the BM responds linearly over the range of *probe* levels considered, which is based on the fact that BM responses to CF tones grow linearly at low levels [reviewed by Cooper (2004) and Robles and Ruggero (2001)]. It also assumes that the time course of decay of the postcochlear masker excitation is the same for the two probe levels considered here. This is supported by the fact that it is almost certain that auditory nerve adaptation contributes significantly to forward masking [although it is probably not the only factor; Meddis and O'Mard (2005), Oxenham (2001)] and the time course of recovery of auditory nerve firing from previous stimulation is independent of probe level (Smith, 1977). Failure to meet these assumptions could result in inaccurate compression estimates. These issues are fully discussed in Sec. IV.

B. Stimuli

TMCs were measured for probe frequencies (f_P) of 0.5 and 4.0 kHz, for masker frequencies (f_M) of 0.4, 0.55, and $1.0f_P$, and for probe levels (L_P) of 9 and 15 dB above the

TABLE I. Absolute threshold (in dB SPL) for each probe frequency (f_P) and listener.

	Listener					
f_P (Hz)	S1	S2	S3	S4	S5	S6
500	32.6	42.8	35.8	34.0		
4000	23.8	27.9	24.8		34.4	35.0

listeners' absolute thresholds for the probes (SL). For any given condition (f_P,f_M,L_P), the masker levels required to just detect the probe were measured for increasing masker–probe intervals (defined from masker offset to probe onset) ranging from 2 to 100 ms, with 5 ms interval increments within the range 5–100 ms. The sinusoidal maskers and probes had total durations of 110 and 10 ms, respectively, and were both gated with 5 ms onset/offset raised-cosine ramps.

Stimuli were generated with a Tucker Davis Technologies psychoacoustics workstation (System III) at a sampling rate of 48.8 kHz and 24 bit resolution. All stimuli were played monaurally via the workstation's headphone connections through the same pair of Etymotic ER2 headphones. Listeners sat in a double-walled sound-attenuating room. The SPLs reported in the following are nominal electrical levels.

C. Procedure

Masker levels at threshold were measured with a two-interval, two-alternative forced-choice (2IFC) adaptive tracking procedure with an interstimulus interval of 500 ms. The masker and probe frequencies were fixed in any given block of trials. The probe level was fixed at 9 or 15 dB SL. The initial masker level was typically set 6 dB below the probe level and was varied adaptively using a two-up, one-down rule to obtain the value needed to achieve the 70.7% correct on the psychometric function (Levitt, 1971). The masker level was increased and decreased by 4 dB for the first 2 reversals and by 2 dB thereafter. Fourteen reversals were recorded in each experimental block and the threshold estimate was taken as the mean of the last 12 reversals. The estimate was discarded when the corresponding standard deviation (s.d.) exceeded 6 dB. The estimate was also discarded, and attempted later if possible, when the masker reached the maximum allowed SPL output of the system (104 dB SPL) in two consecutive trials. Three threshold estimates were obtained in this way and their values averaged to obtain a mean threshold. If the corresponding s.d. exceeded 6 dB, a fourth threshold estimate was measured and included in the mean.

For every listener, all TMCs for the 15 dB SL probe were measured before the corresponding TMCs for the 9 dB SL. For every TMC, masker levels were measured for increasing masker–probe time intervals.

D. Listeners

TMCs were measured for six normal-hearing listeners, although not all conditions were attempted for all listeners (see Table I). Prior to measuring TMCs, absolute thresholds were measured for the probe tones using a 2IFC adaptive procedure. Each probe threshold was measured at least three times and the results were averaged (Table I). All listeners had previous experience or were given at least 4 h of training in the forward-masking task.

E. TMC fits

An *ad hoc* function was fitted to the measured TMCs and the fitted masker levels were used instead of the measured masker levels to infer BM I/O functions (see the following text). The details of the fitting procedure are given elsewhere (Lopez-Poveda *et al.*, 2005). The R^2 between the original and the fitted masker levels was typically $\geqslant$0.98.

III. RESULTS

A. TMCs for a 4 kHz probe frequency

The left panels of Fig. 1 illustrate the TMCs for a probe frequency of 4 kHz. Closed and open symbols illustrate the measured TMCs for probe levels of 9 and 15 dB SL, respectively. Different symbol shapes represent TMCs for different masker frequencies: circles, $1.0f_P$; squares, $0.55f_P$; triangles, $0.4f_P$. Lines illustrate function fits to the measured TMCs.

Except for listener S5, the on-frequency TMCs for the 9 dB SL probe (closed circles) show a segment with a shallow slope for short masker–probe intervals followed by a steeper segment at longer intervals. For listener S5, masker level also increased rapidly for masker–probe intervals $\geqslant$35 ms, but the actual values could not be measured because they exceeded the maximum system output. The off-frequency TMCs for the 9 dB SL probe (closed triangles and squares) can be described, by contrast, as single segments with a shallow slope. Based on the fact that the BM responds linearly to frequencies an octave or so below CF (Robles and Ruggero, 2001), the steeper segment of on-frequency TMCs has been interpreted to reflect BM compression (Nelson *et al.*, 2001). In summary, the overall shapes of these TMCs are consistent with those reported in earlier studies (Lopez-Poveda *et al.*, 2003; Nelson *et al.*, 2001; Nelson and Schroder, 2004; Plack and Drga, 2003).

The threshold masker levels required to mask the 15 dB SL probe (open symbols) are consistently higher than those required for the 9 dB SL probe. Following the rationale of the present method, the degree of BM compression may be estimated as the ratio of the masker-level increment to the probe-level increment, and the compression exponent as the inverse of this ratio. The increments in masker level for the 6 dB increment in probe level are shown in the right panels of Fig. 1. Masker-level increments of approximately 6 dB occur for the lowest masker frequency (triangles) only. The level increments for the on-frequency masker (circles) range from 6 to 48 dB and those for the $0.55f_P$ masker (squares)

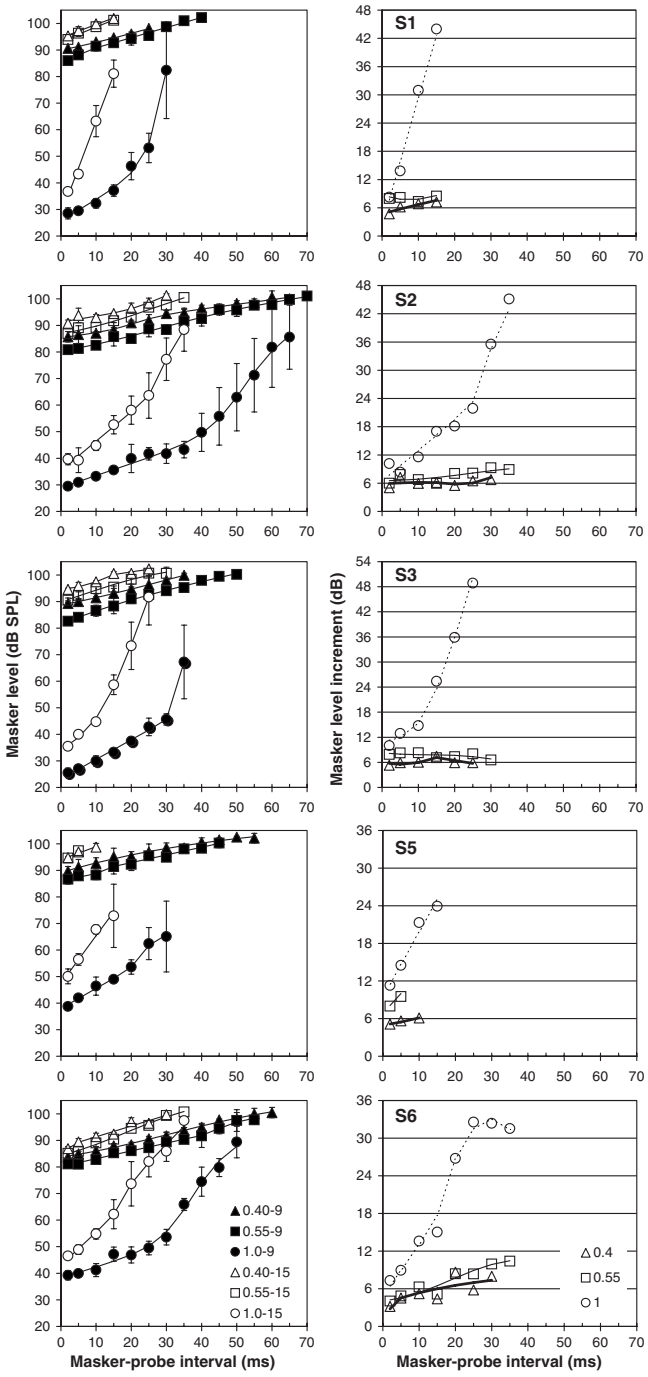


FIG. 1. *Left panels*: TMCs for a probe frequency of 4000 Hz. Each panel illustrates the results for a different listener (as indicated in the top left corner of the right panels). The legend (bottom panel) informs of the masker frequency (relative to CF) followed by the probe level (in dB SL). Symbols illustrate experimental data points. Error bars illustrate 1 s.d. of the mean. Lines illustrate function fits to the measured TMCs. *Right panels*: Corresponding masker-level increments for a 6 dB increment in probe level. Different symbols illustrate the results for different masker frequencies (legend in the bottom panel). Lines illustrate the corresponding differences based on function fits to the measured TMCs.

can be as large 10 dB (S6). This suggests that linear BM responses occur for the $0.4f_P$ masker only, and that the response for the $0.55f_P$ masker is slightly compressed for some listeners (e.g., squares for S6). This also suggests that the compression ratio for on-frequency maskers ranges from 4:1 (S5) to as much as 8:1 (S2 and S3) (see the following text).

It is noteworthy that masker-level increments generally increase with increasing masker–probe interval, except perhaps for the $0.4f_P$ masker (triangles) (see also Sec. IV). Based on the current interpretation of TMCs, this almost certainly reflects that masker-level increments increase with increasing masker level, which in turn increases with increasing masker–probe time interval.

B. TMCs for a 500 Hz probe frequency

Figure 2 illustrates the results for a 500 Hz probe and four listeners. The format is the same as in Fig. 1. Unlike for 4 kHz, the TMCs (left panels) for the 9 dB SL probe (closed circles) do not show two clear segments with different slopes. Instead their slope varies gradually with increasing masker–probe interval from shallow to steep to shallow (S1) or from steep to shallow (S2, S4). Further, the TMCs for the off-frequency maskers (closed squares and triangles) are overall steeper than the corresponding TMCs for the 4 kHz probe (Fig. 1). Overall this pattern is consistent with earlier reports on TMCs for low-frequency probes (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; Nelson and Schroder, 2004).

The right panels of Fig. 2 illustrate the masker-level increment for the 6 dB increment in probe level for the different masker frequencies. The results were broadly similar for listeners S1 to S3, but different for S4 and so they will be described separately. For S1–S3, masker-level increments were always $\geqslant$6 dB for all masker frequencies tested. For the on-frequency masker (circles), masker-level increments were as large as 30 dB (S1, S3), which suggests 5:1 compression. Masker-level increments were as large as 18 dB for the $0.55f_P$ masker (squares), which suggests 3:1 compression. Masker-level increments suggests less but still significant ($\sim$2:1) compression for the $0.4f_P$ masker (triangles).

The masker-level increments for listener S4 (right-bottom panel of Fig. 2) were between 6 and 12 dB for all masker frequencies and conditions. This suggests less overall BM compression than for the other listeners. This is also supported by the comparatively shallower slope of her corresponding off-frequency TMCs. The reason for this is uncertain, but it is unlikely that it relates to some form of cochlear hearing loss because her absolute thresholds for the probe (Table I) were comparable to those of other listeners (S1, S3).

C. BM I/O functions inferred with the standard and the new method

BM I/O functions were inferred with the standard method by plotting the masker levels for a linear reference TMC against the masker levels for any other TMC, paired according to masker–probe time interval (Nelson *et al.*, 2001). While earlier studies have used the TMCs for masker frequencies of 0.5 or $0.55f_P$ for a 4 kHz probe as linear references (e.g., Lopez-Poveda *et al.*, 2003, 2005; Nelson and Schroder, 2004; Plack *et al.*, 2004; Rosengard *et al.*, 2005), the results shown in the preceding sections (right

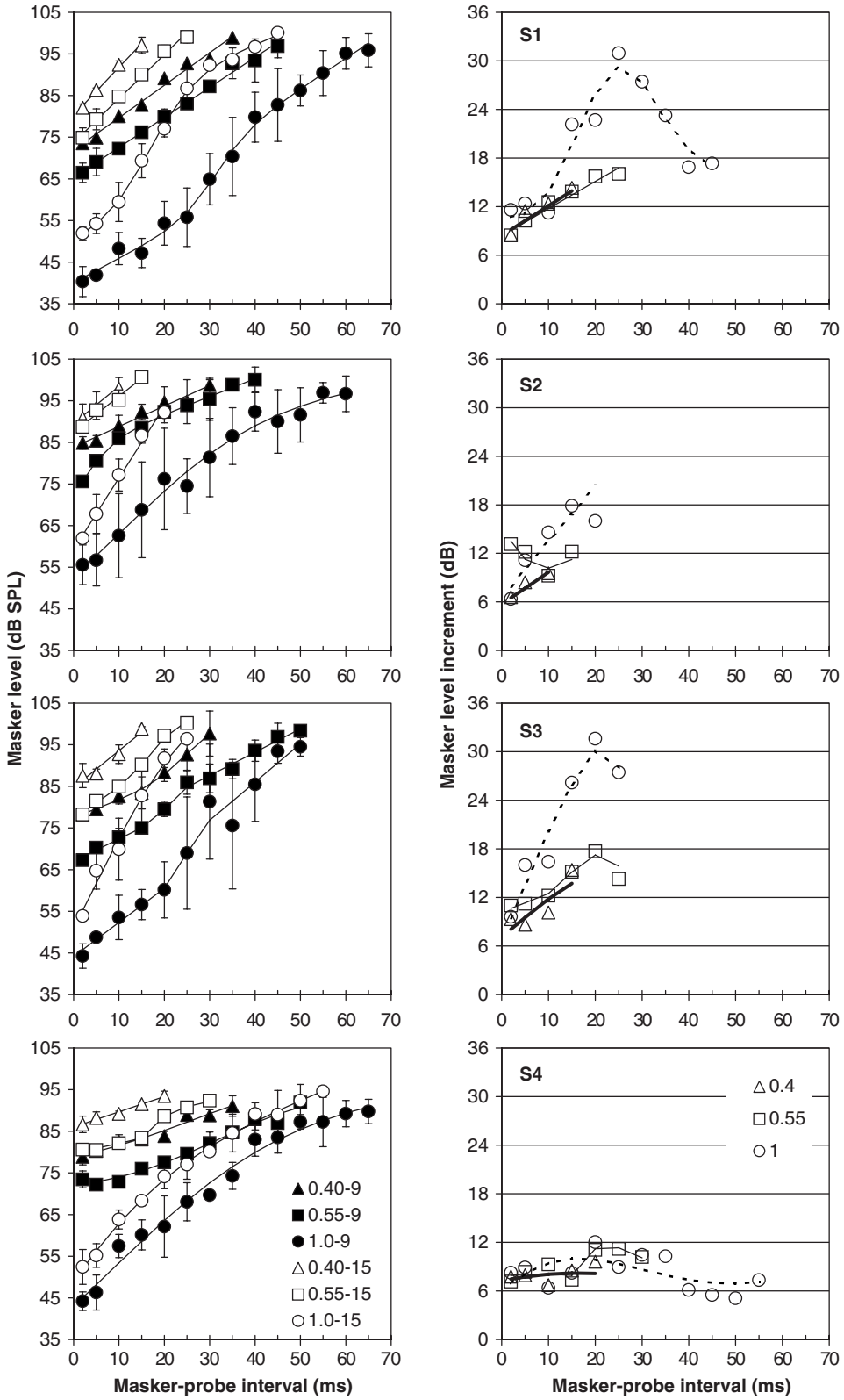


FIG. 2. Same as for Fig. 1 but for a probe frequency of 500 Hz.

panels of Fig. 1) suggest that these maskers may still be subject to compression and that the TMC for a 1600 Hz masker and a 4 kHz probe is a more appropriate linear reference. The closed symbols of Fig. 3 illustrate the resulting BM I/O curves using this linear reference for the three listeners (S1–S3) for whom TMCs were measured for probe frequencies of 500 Hz (left panels) and 4 kHz (right panels). Different symbol shapes illustrate different stimulus frequencies (inset in the middle-left panel). These curves illustrate that significant compression occurs for low, off-frequency tones at 500 Hz, which is consistent with earlier studies (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; Nelson and Schroder, 2004). They also illustrate that the functions for the $0.55f_P$ maskers in the right panels of Fig. 3 are

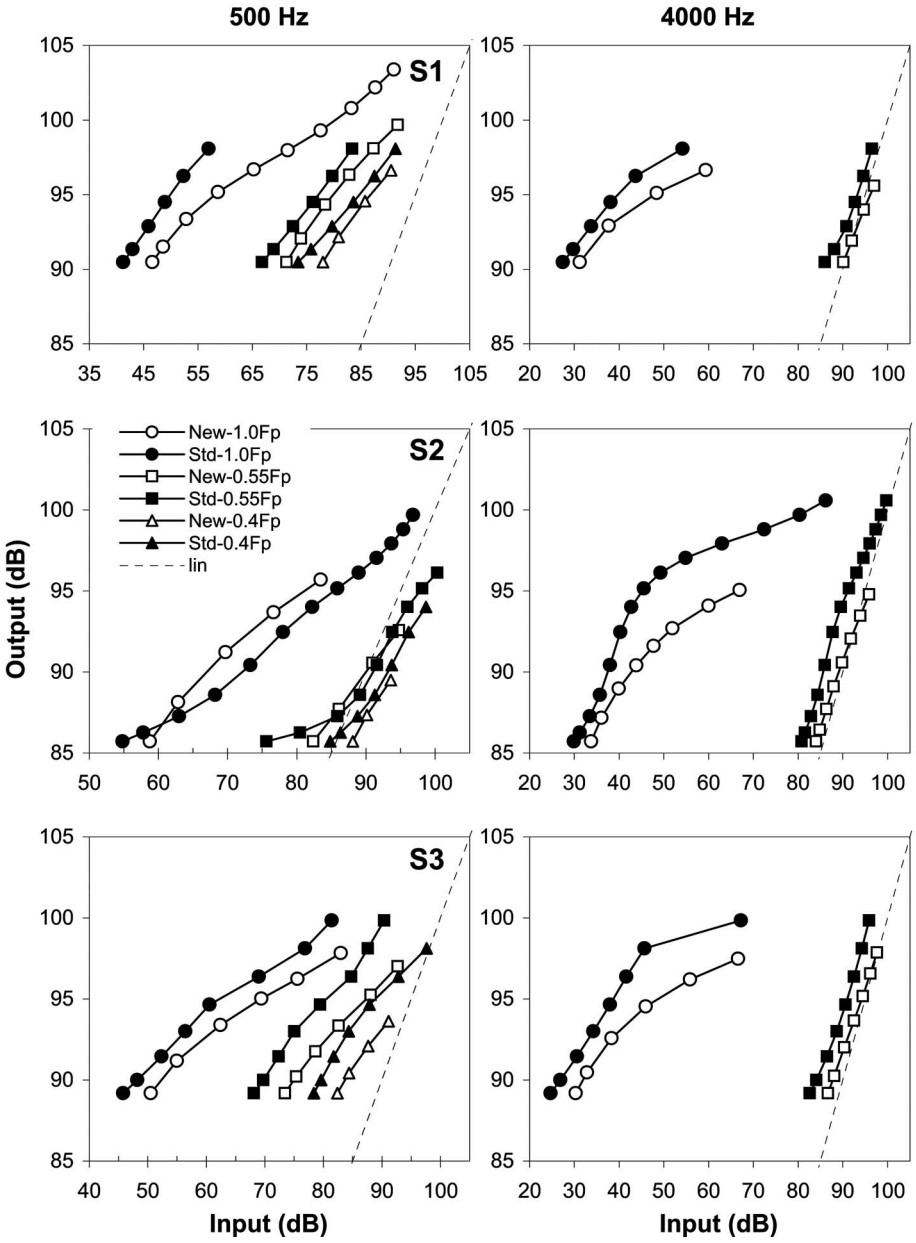


FIG. 3. BM I/O functions at 500 (left panels) and 4000 Hz (right panels) inferred with the standard TMC method (closed symbols) and with the new method (open symbols). Different symbol shapes illustrate the I/O functions for different masker frequencies (as expressed relative to the probe frequency in the inset of the middle-left panel), but the same symbol shapes are used to represent the corresponding functions obtained with the two methods. Note that both sets of curves are based on function fits to the experimental TMCs shown in Figs. 2 and 4. Dashed lines illustrate linear responses with zero gain.

slightly shallower than the dashed lines and hence that slight compression affects the response of the 0.55CF tone at 4 kHz over a narrow range of input levels (see also the following text).

It is also possible and informative to infer BM I/O functions based on the assumptions of the present method. The procedure is described in the Appendix. The resulting I/O curves are illustrated by open symbols in Fig. 3. The shift of these curves with respect to those inferred with the standard TMC method comes from the choice of a constant involved in deriving the new functions, as discussed in the Appendix. With some exceptions (e.g., S2 at 500 Hz), their shapes closely resemble those of the I/O curves inferred with the standard TMC method. Most important, the I/O curves inferred with the new method confirm the two main characteristics of those inferred with the standard TMC method: (1) significant compression affects BM responses to low, off-frequency tones at 500 Hz and (2) slight compression affects the $0.55f_P$ tones at 4 kHz. These coincidences are remarkable considering that the two sets of BM I/O functions were inferred using fundamentally different assumptions.

BM I/O curves inferred with both methods sometimes extend over a different range of input levels (Fig. 3). The actual range of input levels for the I/O curves depends on the number of data points available. For the I/O curves obtained with the new method, fewer data points were available for the TMC measured with the higher-level probe than for the TMC measured with the lower-level probe. As for I/O curves obtained with the standard TMC method, the range of input levels was typically limited by the linear reference TMC, because it contained fewer data points than some of the other TMCs (Figs. 1 and 2). To infer BM I/O functions over a wider range of input levels with the standard TMC method, some researchers (e.g., Lopez-Poveda *et al.*, 2003; Plack *et al.*, 2004) have extrapolated the linear reference TMC to longer masker–probe time intervals using linear regression. This was based on the assumptions that the slope of the linear reference TMC is unaffected by compression and im-

TABLE II. Compression exponent estimates. f_M denotes masker frequency. Two compression estimates are given per listener per condition: The top and bottom values were obtained with the standard and the new TMC methods, respectively. p indicates the probability of the means across listeners being equal for the two methods (two-tailed, paired, Student's t-test).

		Listener						
f_M (Hz)	Method	S1	S2	S3	S5	S6	Mean ± s.d.	p
				4000 Hz				
2400	Standard	0.40	0.57	0.53	0.59	0.58	0.53 ± 0.08	0.12
	New	0.69	0.66	0.73	0.59	0.57	0.65 ± 0.07	
4000	Standard	0.17	0.09	0.08	**0.16**	0.15	0.13 ± 0.04	0.13
	New	0.14	0.14	0.12	**0.24**	0.18	0.16 ± 0.05	
				500 Hz				
200	Standard	0.37	0.37	0.34			0.36 ± 0.02	0.14
	New	0.43	0.62	0.44			0.50 ± 0.11	
275	Standard	0.40	**0.11**	0.32			0.28 ± 0.15	0.44
	New	0.36	**0.44**	0.35			0.38 ± 0.05	
500	Standard	**0.39**	0.18	0.20			0.26 ± 0.12	0.72
	New	**0.20**	0.29	0.19			0.23 ± 0.06	

plies that the postcochlear rate of decay is independent of the masker–probe time interval. However, it will be discussed in the following text that the rate of decay almost certainly varies with time and therefore that linear extrapolation of the linear reference TMC is inaccurate and misleading.

D. Compression-ratio estimates

The main aim of the study was to compare compression estimates for apical (CF ~500 Hz) and basal (CF ~4 kHz) BM regions obtained with the standard TMC method and with the present method. The idea was that because the two methods are based on different assumptions, such comparison allows assessment of the validity of the assumptions of the standard TMC method, particularly that the rate of decay of forward masking is comparable for all masker frequencies and throughout the cochlear CF range from 0.5 to 4 kHz (e.g., Lopez-Poveda *et al.*, 2003). The similarity between the sets of I/O curves inferred with the two methods (Fig. 3) already provides qualitative support for these assumptions. Quantitative support was sought by statistically comparing compression-ratio estimates as inferred with the two methods.

As for the standard TMC method, the compression ratio was calculated as the *minimum* slope of every I/O curve of Fig. 3 (closed symbols) across input levels. The results for each listener and condition are given as compression exponents in Table II (denoted *standard*), where they are compared with the corresponding estimates obtained with the present method (denoted *new*). The latter were obtained as the ratio of the probe-level increment (6 dB) to the *maximum* masker-level increment across masker–probe intervals shown in the right panels of Figs. 1 and 3.

The two methods yield similar compression estimates across listeners and conditions with rare exceptions (denoted in bold in Table II). Indeed, the *mean* compression estimates across listeners are statistically similar for both methods ($p > 0.05$) without exceptions. Furthermore, of the three most significant individual differences (denoted in bold in Table II), two (S5 at 4 kHz and S1 at 500 Hz) occur because the I/O functions obtained with both methods extend over a different range of input levels (as shown for S1 in Fig. 3).

IV. DISCUSSION

A new method for assessing BM compression has been presented that is based on measuring TMCs for two low, slightly different probe levels. Compression was assumed to occur whenever the increase in masker level exceeds the corresponding increase in probe level. Unlike the standard TMC method, this new method does *not* require assuming that the time course of the decay of the postcochlear masker effect be the same across masker and probe frequencies. The compression ratio has been estimated as the ratio of the fixed probe-level increment to the maximum masker-level increment across masker–probe intervals. Compression ratio estimates obtained with the new method were slightly lower but still similar to those obtained with the standard TMC (Table II) method. The measurements obtained with this new method confirm that compression extends to frequencies as low as 0.4CF in the 500 Hz region of the human cochlea. They also suggest that slightly compressive responses (maximum 2:1) occur over a narrow range of levels for frequencies as low as 0.55CF for CF ~4 kHz (right panels of Fig. 1).

A. The time course of decay of the postcochlear masker effect and the assumption that it is independent of probe level

Like the standard TMC method, the present method does not require making assumptions on the actual form of the function describing the time course of the decay of the postcochlear masker effect. For the standard TMC method, it is sufficient to assume that the function is the same across masker frequencies. For the present method, it is sufficient to assume that it is the same for the two probe levels. The latter is supported by the fact that auditory nerve adaptation contributes significantly to forward masking [although probably it is not the only factor; Meddis and O'Mard (2005), Oxen-

ham (2001)] and the poststimulatory rate of recovery of auditory nerve firing appears to be independent of probe level (Smith, 1977).

Some details of the present TMCs, however, undermine this assumption. For example, on-frequency TMCs for the higher-level probe appear sometimes steeper than those for the lower-level probe over the same range of masker levels (e.g., S2 in Figs. 1 and 3). This should not be the case because BM compression must be the same over the same range of levels. One possible explanation for this result is that the masker effect decays faster at short delays and more slowly at long delays. This would agree with earlier modeling studies that suggest that a double exponential recovery process with a fast and a slow time constant gives account of psychophysical (e.g., Oxenham and Moore, 1994; Plack and Oxenham, 1998; Meddis and O'Mard, 2005) and physiological (e.g., Meddis and O'Mard, 2005) forward masking data. This explanation would be consistent also with the fact that some of the linear-reference TMCs become shallower with increasing masker–probe time interval (e.g., closed triangles for S2 and S5 in Fig. 1).

On the other hand, the two on-frequency TMCs of listener S6 in Fig. 1 follow a different trend. They have different slopes for masker levels below around 55 dB SPL and run approximately parallel to each other above that level. Based on the above-presented explanation, one would expect the steeper portion of the on-frequency TMC to be shallower for the lower- than for the higher-level probe because it occurs over longer masker–probe intervals. Furthermore, the slope of the corresponding off-frequency TMCs seems constant across masker–probe time intervals.

Maybe these inconsistencies are only apparent and the effect of a time-dependent decay rate on the TMCs is obscured by its interaction with the effects of BM compression. Or maybe the values of the two time constants and/or their relative contributions to the net decay function vary significantly across listeners. For instance, the TMCs for S6 in Fig. 1 could be consistent with a compression threshold of ~50 dB SPL and a fast time constant of less than 10 ms. This would explain why the slope of the corresponding linear reference TMC hardly changes with increasing masker–probe interval and that the on-frequency TMCs for the two probe levels run parallel to each other. The TMCs for S2, however, would be consistent with a compression threshold of ~45 dB SPL and a fast time constant of ~35 ms because this is the time interval at which the corresponding linear reference TMCs become shallower.

There is, nevertheless, another aspect of the present TMCs that undermine the assumption of identical decay time courses for the two probe levels. For *some* listeners the linear reference TMC is slightly steeper for the higher- than for the lower-level probe (closed versus open triangles in Fig. 1). This subtle difference is more clearly reflected in Fig. 4(a), which shows the increment in the level of the $0.4f_P$ masker for a 6 dB increment in probe level as a function of masker–probe time interval. If the slope of the two linear reference TMCs were identical, then the masker level increment should be approximately equal to 6 dB and constant across masker–probe time intervals. This is the case for some listeners (e.g., S2 or S3). For others (e.g., S1 or S6), however, the increment increases slightly with increasing time interval. This cannot be explained based on a time-dependent decay rate.

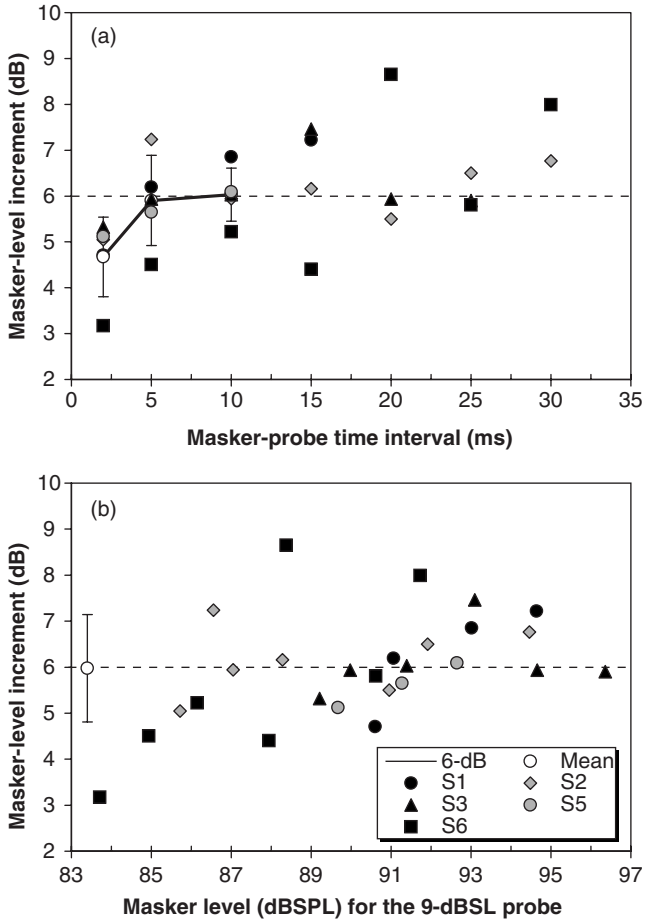


FIG. 4. Increments in the level of a 1600 Hz masker for a 6 dB increment (9–15 dB SL) in the level of a 4000 Hz probe: (a) As a function of masker–probe time interval and (b) as a function of the level for the 9 dB SL masker. Different symbols illustrate the results for different listeners as indicated by the inset in (b). Open circles indicate mean values across the five listeners (a) or across listeners and levels (b). Error bars indicate 1 s.d. The horizontal dashed lines illustrate expected values if the BM response to the probe and the masker were both linear.

Therefore, it is possible that the present TMCs may have been affected idiosyncratically by a factor other than BM compression or the decay rate of the postcochlear masker excitation. The nature of this factor is uncertain. Maybe the TMCs for the 9 dB SL probe were influenced by the difficulty in detecting a very-low level probe. Indeed, we have observed that some untrained listeners find it more difficult to detect very-low level probes for long (>30–40 ms) than for short intervals. This sometimes led to nonmonotonic TMCs (i.e., masker level first increases and then decreases with increasing masker–probe interval). It was as though for moderate-to-long intervals these listeners had lost the cue as to when the probe would occur. The difficulty disappeared when the probe level was raised by a few decibels. It cannot be ruled out that this (or a related factor) may have affected the present on- and off-frequency TMCs, particularly those measured with the 9 dB SL probe. This might explain why

the linear-reference TMCs are sometimes slightly shallower for the lower- than for the higher-level probe, as discussed earlier. Of course, TMCs may be simultaneously affected by BM compression, by a time-dependent decay rate, and by this factor.

An estimate of the significance of this uncertain factor was obtained by comparing BM I/O functions derived using a common linear reference for the TMCs for the two probe levels. The idea was that a small effect should produce overlapping I/O functions for the two probe levels (after allowing for a constant 6 dB difference in output level). The TMC for the 1600 Hz masker and the 9 dB SL probe was used as the common linear reference because it generally contained more data points than the corresponding TMC for the 15 dB SL probe. There resulting I/O functions (not shown) for the two probe levels generally overlapped, but there were subtle differences for some listeners and conditions. Specifically, the compression threshold level appeared to be higher in the I/O curves for the lower-level probe in some cases (S2 at 0.5 and 4 kHz; S1 at 4 kHz; and S3 at 0.5 kHz). In any case, this test showed that *maximum* compression estimates were similar for the two sets of I/O curves and hence that the influence of this factor does not compromise the main conclusions of the present study.

The influence of this confusing factor might be reduced by using a cue indicating when the probe would occur and/or by raising the level of the lower-level probe. The latter would also bring the level of the two probes closer, which would minimize factors related to probe-level differences. Unfortunately, using higher probe levels would make it more difficult to measure the linear portion of BM I/O functions that may occur at low levels.

B. BM linearity near threshold

The new method rests on the assumption that for all CFs the BM response to CF tones is linear over the range of probe levels considered. Direct BM responses in lower mammals suggest that this is a reasonable assumption provided that the probe levels are low [evidence reviewed by Cooper (2004), Fig. 2; and Robles and Ruggero (2001), p. 1308], as the present method requires. The assumption is also supported by physiological evidence in humans (Gorga *et al.*, 2007). Psychophysical evidence is, however, confusing. Some studies have suggested that compressive BM responses may occur for *some* listeners even at levels near threshold, but only at *high* CFs (e.g., Fig. 2 of Plack and Oxenham, 2000). Recent studies (Plack and Skeels, 2007), however, suggest that the BM responds linearly near threshold at CFs of 2 and 4 kHz.

If BM compression affected the response to the probe, then the present method would underestimate the degree of BM compression: a 6 dB increment in probe level would produce a smaller increment in internal excitation, which would lead to masker-level increments smaller than would be required had the probe been processed linearly. Incidentally, it is possible to identify instances where this assumption does not hold at *high* CFs considering that linear BM responses occur for sufficiently low stimulus frequencies (relative to CF). Indeed, if compression affected the probe response, a 6 dB increase in probe level should lead to increases in the level of the $0.4f_P$ masker smaller than 6 dB for all masker–probe time intervals (or masker levels). This, however, was rarely observed in the present data (Fig. 4). Except, perhaps, for listener S6, the level increments for the $0.4f_P$ masker at 4 kHz were on average ~6 dB (Fig. 4). Furthermore, on-frequency TMCs at 4 kHz for the 9 dB SL probe show a shallow segment for short masker–probe time intervals over which the rate of growth of masker level with increasing masker–probe delay is approximately similar to that of the $0.4f_P$ masker (Fig. 1). Since the BM response to the latter is likely linear, this supports the assumption that on-frequency BM responses near threshold must also be linear. In any case, the risk of failing to meet this assumption may be minimized by considering probe levels lower and closer than the ones considered here.

C. Implications of the results

Previous studies have shown that the slopes of off-frequency TMCs (2200 Hz) for hearing-impaired listeners at 4 kHz are approximately half of those for normal-hearing listeners (Plack *et al.*, 2004; see also Fig. 3B of Lopez-Poveda *et al.*, 2005). It has been suggested that this may be interpreted as evidence that the internal effect of the masker decays more slowly at high levels *or* that the 2200 Hz maskers may be compressed (by approximately 2:1) in normal-hearing (but not in hearing-impaired) listeners even at CF ~4 kHz (Plack *et al.*, 2004). The above-presented evidence supports the latter. First, it has been shown with two different methods that compression is likely to affect the response of a 2200 Hz tone at 4 kHz (Fig. 3) and the degree of compression approximately matches 2:1 (Table II). Second, if the rate of decay were slower for higher masker levels, then the masker-level increment for a fixed probe-level increment should be gradually smaller with increasing masker level but this was not the case, as shown in Fig. 4(b). If anything, the opposite was true. An alternative explanation for off-frequency TMCs being shallower for hearing-impaired than for normal-hearing listeners is that age, and not the hearing loss *per se*, seems to reduce the rate of decay of forward masking (Gifford and Bacon, 2005). On the other hand, a similar difference in slope has been reported between off-frequency TMCs measured in the two ears of a 24-year-old listener with unilateral hearing loss (Fig. 3 in Lopez-Poveda *et al.*, 2005).

Several studies have considered the TMC for a probe frequency of 4 kHz and a masker frequency of 2200 Hz ($f_M=0.55f_P$) or higher as the linear reference TMC (Plack and Drga, 2003; Rosengard *et al.*, 2005; Nelson and Schroder, 2004). The data in Table II [and Fig. 1(b)] suggest that the slope of this TMC may be affected by approximately 2:1 compression and thus that it may be inappropriate to use this TMC as the linear reference. Therefore, it is likely that those studies have underestimated the degree of BM compression. The present data also suggest that the TMC for the $0.4f_P$ would be a better linear reference TMC.

The new method provides compression estimates that

are similar to those obtained with the standard TMC method when a linear reference of $0.4f_P$ is used (Fig. 3 and Table II). Given that the two methods are based on fundamentally different assumptions, this provides indirect support for the postulates of the standard TMC method; particularly, for the assumption that the rate of decay of forward masking is similar across probe frequencies (Lopez-Poveda *et al.*, 2003, 2005; Plack and Drga, 2003; Nelson and Schroder, 2004; Rosengard *et al.*, 2005). This, however, appears inconsistent with the conclusion of Stainsby and Moore (2005) that "the rate of decay of forward masking is not the same for all probe frequencies" (or equivalently, for all cochlear sites). The reason for the discrepancy is uncertain, but it may relate to the following.

The conclusion of Stainsby and Moore (2005) was based on TMCs of hearing-impaired rather than normal-hearing listeners. Stainsby and Moore regarded the absence (or the presence of small values) of *measurable* DPOAEs in their listeners as indicative of purely passive and thus linear cochlear responses. It is conceivable, however, that this was not the case. It is technically difficult to measure DPOAEs below 1000 Hz even in normal-hearing listeners because of ambient noise (e.g., Gorga *et al.*, 2007). Therefore, the lack of *measurable* DPOAEs in their hearing-impaired listeners did not necessarily imply that their listeners lacked them or that their cochlea responded linearly. In fact, Stainsby and Moore acknowledged that some of their listeners actually showed some small DPOAEs at some frequencies. This could also explain that their conclusion held for two of their three listeners only.

D. On the merits of different methods for inferring BM compression

Both the present and the standard TMC methods allow inferring BM I/O functions and peripheral auditory compression in normal-hearing listeners and yield similar results (Fig. 3). The present method, however, may be advantageous to infer BM I/O functions in instances when it is not possible to obtain a reliable linear reference TMC. This can happen, for instance, for some (hearing-impaired) listeners for whom the levels of the desired linear reference commonly exceed the maximum system output level.

V. CONCLUSIONS

(1) It is possible to infer BM I/O functions and peripheral auditory compression by measuring corresponding TMCs for two, slightly different probe levels. The compression exponent may be estimated directly as the ratio of the increase in probe level to the *maximum* increase in masker level.
(2) Strong on-frequency compression occurs both at 500 Hz and 4 kHz. The average compression exponents estimated with the new method were 0.23 and 0.16, respectively. The new method confirms that compression extends to a wider frequency range at 500 Hz than at 4 kHz.
(3) At CF ~4 kHz, BM responses to 0.55CF tones show approximately 2:1 compression. Therefore, previous studies that have used the TMC for a masker frequency of 2200 Hz and a probe frequency of 4 kHz as a linear reference may have underestimated the degree of compression. BM responses to 0.4CF tones are linear at 4 kHz and therefore the TMC for a $0.4f_P$ masker is a more appropriate linear reference TMC than that for a $0.55f_P$.
(4) The present method sometimes underestimates the degree of compression but its results generally match those obtained with the standard TMC method. Because the two methods are based on fundamentally different assumptions, this provides circumstantial support to the assumptions of the standard TMC method; particularly that the postcochlear rate of decay of the internal masker excitation is comparable for cochlear sites with CFs of 500 Hz and 4 kHz.

ACKNOWLEDGMENTS

Work supported by MEC (BFU2006-07536/BFI), PROFIT (CIT-390000-2005-4), and IMSERSO (131/06). We thank Chris Plack for useful discussions on the ideas presented in this paper. We are most grateful to Magdalena Wojtczak and an anonymous reviewer for their comments on an earlier version of this paper.

APPENDIX: INFERRING BM I/O FUNCTIONS FROM TMCS FOR TWO PROBE LEVELS

It is possible to infer *approximate* BM I/O functions for any CF and stimulus frequency from a pair of TMCs measured with two slightly different probe levels. The procedure is as follows. Let m_1 be the ratio of the (fixed) probe-level increment (ΔL_P) to the (measured) masker-level increment for a given masker–probe time interval, Δt_1:

$$m_1 = \frac{\Delta L_p}{L'_{M1} - L_{M1}}, \tag{A1}$$

where L_{M1} and L'_{M1} denote the measured masker levels for the lower- and the higher-level probes, respectively, for the masker–probe interval in question. m_1 may be regarded *approximately* as the *local* slope of the BM I/O function affecting the masker for an intermediate masker (input) level between L_{M1} and L'_{M1}; that is for input level $x_1=(L_{M1}+L'_{M1})/2$. By estimating m for all possible masker–probe intervals, one obtains an approximate function describing the slope of the BM I/O curve against the input masker level, $m(x)$. Then, the I/O function, $y=f(x)$, may be obtained by integrating the slope function, $m(x)$, numerically as follows:

$$y_1 = C,$$

$$y_{n+1} = m_n(x_{n+1} - x_n) + y_n \quad \text{for } n > 1, \tag{A2}$$

where $x_{n+1} > x_n$ and C is an undetermined constant. C determines the absolute output values (i.e., it shifts the I/O vertically) but the relative output range and the slope of the resulting I/O function are independent of this constant. Thus, it may be set to any suitable value.

Example I/O curves inferred with this method are shown in Fig. 3, where they are compared with corresponding I/O

functions inferred with the standard TMC method. In this case, C was set so that the curves for the two methods start at the same ordinate value. The new method requires pairs of masker levels for the lower- and the higher-level probes, and it may be sometimes impossible to measure the latter for the longer masker–probe time intervals (Figs. 1 and 3). This sometimes limits the range of input levels of the I/O curves estimated with the present method. One advantage of the present method over the standard TMC method is that it does *not* require assuming that the time course of the decay is similar across masker and probe frequencies.

Bacon, S. P., Fay, R. R., and Popper, A. N., eds. (**2004**). *Compression: From Cochlea to Cochlear Implants* (Springer, New York).

Cooper, N. P. (**2004**). "Compression in the peripheral auditory system," in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 18–61.

Gifford, R. H., and Bacon, S. P. (**2005**). "Psychophysical estimates of nonlinear cochlear processing in younger and older listeners," J. Acoust. Soc. Am. **118**, 3823–3833.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2007**). "Low-frequency and high-frequency cochlear nonlinearity in humans," J. Acoust. Soc. Am. **122**, 1671–1680.

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

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.

Lopez-Poveda, E. A., Plack, C. J., Meddis, R., and Blanco, J. L. (**2005**). "Cochlear compression in listeners with moderate sensorineural hearing loss," Hear. Res. **205**, 172–183.

Meddis, R., and O'Mard, L. P. (**2005**). "A computer model of the auditory-nerve response to forward-masking stimuli," J. Acoust. Soc. Am. **117**, 3787–3798.

Nelson, D. A., and Schroder, A. C. (**2004**). "Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2221–2233.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.

O'Loughlin, B. J., and Moore, B. C. J. (**1981**). "Off-frequency listening-Effects on psychoacoustical tuning curves obtained in simultaneous and forward masking," J. Acoust. Soc. Am. **69**, 1119–1125.

Oxenham, A. J. (**2001**). "Forward masking: Adaptation of integration?," J. Acoust. Soc. Am. **109**, 732–741.

Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Plack, C. J., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low characteristic frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.

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.

Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.

Plack, C. J., and Oxenham, A. J. (**2000**). "Basilar-membrane nonlinearity estimated by pulsation threshold," J. Acoust. Soc. Am. **107**, 501–507.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2006**). "Masking by inaudible sounds and the linearity of temporal summation," J. Neurosci. **26**, 8767–8773.

Plack, C. J., and Skeels, V. (**2007**). "Temporal integration and compression near absolute threshold in normal and impaired ears," J. Acoust. Soc. Am. **122**, 2236–2244.

Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo," Aud. Neurosci. **3**, 101–121.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **117**, 3028–3041.

Smith, R. L. (**1977**). "Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects," J. Neurophysiol. **40**, 1098–1111.

Stainsby, T. H., and Moore, B. C. J. (**2006**). "Temporal masking curves for hearing impaired listeners," Hear. Res. **218**, 98–111.

Role of attention in overshoot: Frequency certainty versus uncertainty

Bertram Scharf,[a] Adam Reeves, and Holly Giovanetti
Department of Psychology, Northeastern University, Boston, Massachusetts 02115

(Received 3 April 2007; revised 15 December 2007; accepted 21 December 2007)

Overshoot, the elevation in the threshold for a brief signal that comes on close to masker onset, was measured with signal frequency certain (same frequency on every trial) or uncertain (randomized over trials). In broadband noise, thresholds were higher 2 ms after masker onset than 200 ms later, by 9 dB with frequency certainty, by 6–7 dB with uncertainty. In narrowband noise centered on the signal frequency, thresholds at 2 ms were not elevated with certainty, but were elevated 4–5 dB with uncertainty. Thus, frequency uncertainty leads to *less* overshoot in broadband noise, to *more* overshoot in narrowband noise. Reduced overshoot in broadband noise may come about because the masker, given its many frequencies, disrupts focusing at onset as much under certainty as uncertainty. Once the initial disruption dissipates, threshold is lower with certainty so overshoot is greater. In contrast, a narrowband noise with frequencies only near the signal does not disrupt focusing when the signal frequency is known beforehand, so overshoot is absent. When frequency is uncertain, the narrowband noise serves to focus attention on the signal frequency; as this requires time, detection near noise onset is poorer than later on, so overshoot is present.
 [DOI: 10.1121/1.2835436]

PACS number(s): 43.66.Dc, 43.66.Ba, 43.66.Mk [AJO] Pages: 1555–1561

I. INTRODUCTION

A brief signal is often more difficult to detect when it comes on close to the onset of a masking noise than when it comes on a few hundred milliseconds after noise onset. This effect is usually called overshoot (e.g., Zwicker, 1965; Miskiewicz *et al.*, 2006) or the temporal effect (e.g., Bacon and Viemeister, 1985b; Strickland, 2004.) We report new experiments that confirm our hypothesis that overshoot is caused, at least in part, by disruption of the focusing on a targeted frequency. A major achievement of our explanation is that it accounts for the absence or near absence of overshoot in the presence of a masker, such as a narrowband noise, with no energy outside the critical band centered on the signal frequency. At the same time, we propose a new approach to the question of why frequency uncertainty leads to smaller threshold increases than predicted by most models of auditory detection. Our primary aim is to enrich the discussion of overshoot and to encourage new avenues of research.

We hypothesize that the sudden onset of a broadband noise with energy at all frequencies makes it difficult for the listener to continue to focus on an expected target frequency whose level must then be increased to permit detection. It usually takes over 100 ms to recover fully from the noise onset and to refocus on the target frequency. Accordingly, threshold in broadband noise is higher for a brief signal near noise onset than for the same signal 200 ms later.

Our disruption hypothesis is close to an explanation by Dai and Buus (1991) of a different phenomenon, the better detection of unexpected signals in gated than in continuous noise. Elaborating on that explanation, Wright and Dai (1994, p. 939) suggested that the onset of a broadband noise "masker initially forces the observer to listen to a broad range of stimulated frequencies even though only a single tone at an expected frequency is to be detected (Dai and Buus, 1991). This explanation fits with the common intuition that novel events, such as the onset of a clearly audible sound, attract attention." Wright and Dai (1994, p. 947) went on to interpret the results of their own measurements as lending "some support to the idea that signal thresholds are higher at masker onset, in part, because masker gating forces the subject to listen to remote frequencies." Similar but much less explicit references have been made to a possible role in overshoot of attention (Elliott, 1965) or of distraction and centrally mediated effects (Carlyon and Moore, 1986; Bacon and Moore, 1986).

As noted earlier, we extend the disruption hypothesis to clarify why overshoot is usually absent or small when the masker is a narrowband noise (or pure tone) and the signal is a brief tone at a frequency centered in the noise (or at the same frequency as the masking tone) (e.g., Zwicker, 1965; Fastl, 1977; Bacon and Viemeister, 1985a, b; Bacon and Smith, 1991). Our disruption hypothesis can be tested directly both in broadband noise and in narrowband noise by measuring overshoot when the signal frequency changes randomly from trial to trial. Such measurements of overshoot, with *frequency uncertainty*, have not been reported; all past measurements of overshoot appear to have been made with *frequency certainty.* Our disruption hypothesis leads to opposite and testable predictions of how frequency uncertainty affects overshoot—less overshoot in broadband noise, more overshoot in narrowband noise.

Our results fully confirm these predictions and lend strong credence to the underlying attentional disruption hypothesis. However, neither the results nor the hypothesis ex-

[a]Author to whom correspondence should be addressed. Electronic mail: scharf@neu.edu

plain readily the many other ways in which overshoot depends on stimulus parameters besides masker bandwidth. Other mechanisms may be involved to account for why overshoot is markedly smaller at low rather than at high frequencies (e.g. Zwicker, 1965; Fastl, 1976), why it is evoked by a masker at frequencies above the signal but seldom by frequencies below (e.g., Schmidt and Zwicker, 1991), why it is reduced in cochlear hearing loss (Bacon and Takahashi, 1992), after aspirin intake (McFadden and Champlin, 1990), or after exposure to an intense tone (Champlin and McFadden, 1989), and so forth. Among the many hypotheses offered to explain overshoot, an involvement of active processing, which probably reflects the action of the outer hair cells, on the basilar membrane has received considerable attention (e.g., Von Klitzing and Kohlrausch, 1994; Strickland, 2001; Bacon and Savel, 2004). Strickland (2001) succeeded in relating the characteristics of the nonlinear input–output function on the basilar membrane to the dependence of overshoot on masker level and signal frequency. To explain overshoot she had to assume that the active processing changes during noise input. As the basis for that change, Strickland (2001) and others (e.g., Schmidt and Zwicker, 1991; Bacon and Liu, 2000) have suggested a role for efferent input to the cochlea via the olivary–cochlear bundle (OCB). The temporal properties of the OCB would seem to match reasonably well those of overshoot. Efferent activity begins to be effective 100–200 ms after signal (or noise) onset and "is best suited to operate on acoustic changes that persist for 100's of milliseconds" (Backus and Guinan, 2006). On the assumption that efferent activity changes the slope of the input–output function so as to increase detection of a signal in noise, the reduced threshold 200 ms after noise onset would reflect the positive role of that activity. However, the assumption that efferent input improves signal detection has not been supported psychophysically, either in animals (e.g. Igarashi *et al.*, 1972) or in humans (Scharf *et al.*, 1997). Moreover, some measurements of overshoot in patients with normal audiometric thresholds have shown as much overshoot in ears to which the OCB had been severed as in ears with normal efferent input.[1] Nonetheless, enough uncertainty remains about efferent activity in hearing (e.g., Guinan, 2006) that its involvement in overshoot remains an attractive possibility.

Returning to the current experiments, we first consider the usual experimental paradigm in studies of overshoot, i.e., with frequency certainty. According to our hypothesis, with frequency known beforehand, the listener can prepare for the signal frequency before any stimulation. The onset of a broadband noise temporarily diverts attention to a wide range of frequencies, thereby raising threshold for a signal close to noise onset. However, after 200 ms, focusing on the (known) signal frequency is reestablished, so threshold is no longer elevated. This difference in thresholds is overshoot. In contrast, the onset of a narrowband noise that is centered on the same critical band as the signal does not displace focusing to frequencies away from the target. Consequently, detection is as good right after the onset of a narrowband noise as a few hundred milliseconds later. No overshoot is present.

We now consider the new situation, frequency uncertainty. In the case of broadband noise, with frequency uncertain, focusing on the unknown signal frequency first begins upon the arrival of the signal, but if signal onset is too close to noise onset, the focusing is prevented or disrupted and threshold is elevated. Presented 200 ms or so later, the same signal is able to initiate focusing with little or no disruption from the temporally "remote" noise onset. Hence, the threshold for the tone is higher near noise onset than after 200 ms; overshoot is present. However, overshoot with frequency uncertainty should be smaller than with frequency certainty because (1) signals delayed 200 ms are nearly equivalent to signals in continuous noise and so would be more difficult to detect with frequency uncertainty (e.g. Green, 1961; Scharf *et al.*, 2007) and (2) for signals near noise onset, threshold will be no better with frequency certainty than with uncertainty owing to the disruption of focusing in both cases. The situation with narrowband noise under frequency uncertainty is quite different. The onset of a narrowband noise, with all its energy in the same critical band as the signal, directs focusing toward the frequency region of the signal. However, the focusing requires time so that the threshold for a signal coming on close to noise onset is higher than for the same signal coming on later. Accordingly, overshoot, which is normally absent in narrowband noise with frequency certainty, should now be present with frequency uncertainty.

To test these predictions, we measured overshoot both under frequency certainty and uncertainty, in narrowband noise and in broadband noise.

II. EXPERIMENT 1: OVERSHOOT IN NARROWBAND NOISE

A. Method

A single-interval yes/no procedure with feedback was used to measure d' for tone bursts presented either 2 or 200 ms after the onset of a brief narrowband noise masker. (We did not use two observation intervals, because the noise in the first interval would eliminate frequency uncertainty in the second interval.) Signals had a duration of 10 ms including raised cosine fall/rise times of 2 ms. Signal delay (2 or 200 ms) was calculated as the time between zero amplitude at noise onset and zero amplitude at signal onset. Masker noise bands were approximately one critical band wide, with the following frequency limits: 630–770, 920–1080, 1265–1475, 1710–1990, 2310–2690, 3125–3675, 4350–5250, and 5250–6350 Hz. Maskers were synthesized so that at 100 Hz below and above these frequency limits, the spectrum level was 40 dB down from its maximum value. The masker was centered geometrically on the signal frequency on every signal trial. All maskers had a duration of 398 ms including raised cosine fall/rise times of 2 ms. The noise spectrum level was always 31.14 dB.

In the certainty condition, owing to time constraints, measurements were made at only two signal frequencies, 2500 and 4800 Hz. Throughout a block of 64 trials the same noise was presented on every trial; the signal was presented at the same frequency and with the same delay on every signal trial.

In the uncertainty condition, the frequency was chosen on each trial from eight possible values: 700, 1000, 1370, 1850, 2500, 3400, 4800, and 5800 Hz. The choice was such that each frequency was presented four times during a block of 64 trials. Signals were set at levels expected to yield a d' between 1.0 and 2.0. Each block contained 64 trials with a signal presented on 32 of them. On noise-only trials, one of the eight noise bands was chosen such that in each block of 64 trials, each noise was presented on a total of eight trials, four with a signal and four without. A trial began with a visual cue on a computer terminal followed after 500 ms by a second visual cue together with the masking noise. On signal trials, the tone burst came on after the appropriate delay (2 or 200 ms depending on the ongoing block) from noise onset. After termination of the noise, the listener used a computer mouse to indicate whether or not the signal was heard. The correct response was immediately shown on the monitor. No time limit was imposed on the response. The next trial began 500 ms later. The first block of the session was repeated so that signal levels could be adjusted as necessary to achieve the target values of d' at as many frequencies as possible. After one or two such repetitions, a full set of eight blocks was run with all signal levels unchanged; only results from those blocks were used in the final calculations.

Twenty-two Northeastern undergraduates aged 18–23 years participated in these experiments, about half in partial fulfillment of a course requirement, the other half for remuneration or as laboratory members. All listeners reported normal hearing. The experiment was completed in two sessions at least one day apart. Half the listeners were run in the first session with the signal delay set to 200 ms and the other half with it set to 2 ms. (As average thresholds were independent of the order in which the sessions were run, data were pooled over order.) In each session, measurements were made first with signal frequency uncertain in eight to ten blocks, and then with frequency certain in two final blocks.

Listeners sat in a small sound-isolated booth (Eckel Industries). Stimuli were generated on the basis of custom software (Matlab) and a Tucker-Davis (TDT) system III signal processor (RP2.1) under the control of a microcomputer (Dell Optiplex GX270). Tones were generated at a rate of 48.83 kHz. Noise was generated digitally at the start of each trial, with a TDT driver simulating an analogue bi-quad band-pass filter. Tone and noise were mixed digitally and sent through a headphone driver (HB7) to the left ear via a Sony MDR-V6 headphone. Waveforms, frequency content, and distortion were checked with a wave-analyzer (GRC 1900) and an oscilloscope (Tektronix TAS220). The Optiplex computer was programmed to conduct the experiment and record responses.

B. Results

For each of the 22 listeners, from the fixed SPL and the d' measured by the yes/no procedure, the SPL at which d' would equal 1.0 was calculated. [The value d' was calculated as $z(P_{hit})-z(P_{fa})$, where $z(P_{hit})$ is the z-score (normal inverse) of the hit rate, $z(P_{fa})$ is the z-score of the false-alarm rate. Calculations were made on the assumption that a change of 3 dB in SPL yielded a change of 1.0 in d'. See Scharf *et al.* (2007) for a discussion of this assumption.] The means of these values together with their associated standard errors are given on the left-hand side of Table I under all the conditions of experiment 1. (The right-hand side gives the results of experiment 2 with a broadband noise masker.) With certainty, the thresholds measured at 2500 and 4800 Hz at a delay of 200 ms are close to the values calculated on the basis of critical-band theory and an assumed integration time of 200 ms (e.g., Scharf and Buus, 1986). With uncertainty, threshold is significantly higher ($F_{1,21}=19.44$, $p=0.0002$) at both frequencies.

TABLE I. Thresholds to achieve a d' of 1.0 (narrowband noise, left side) or 76% correct (broadband noise, right side) at two signal frequencies and at two signal delays re noise onset. (Mean SPL in dB for $d'=1$ or 76% correct, and *standard error*.)

	Narrowband noise				Broadband noise			
	Certainty		Uncertainty		Certainty		Uncertainty	
Delay	2 ms	200 ms	2 ms	200 ms	2 ms	200 ms	2 ms	200 ms
2500 Hz	**64.4**	**64.2**	**69.5**	**65.5**	**75.6**	**66.6**	**75.4**	**68.4**
	0.51	*0.67*	*0.82*	*0.72*	*1.77*	*0.79*	*1.69*	*0.58*
4800 Hz	**67.8**	**67.6**	**74.5**	**69.8**	**76.6**	**68.7**	**76.2**	**70.2**
	0.69	*0.72*	*0.80*	*0.68*	*1.17*	*0.56*	*0.88*	*0.35*

From the values in Table I, overshoot is taken as the 2-ms threshold minus the 200-ms threshold. Since with certainty, at both 2500 and 4800 Hz, threshold is essentially the same at the two delays, overshoot is near 0 dB as shown by the white bars in Fig. 1. With uncertainty, overshoot is markedly greater than with certainty (by t-test, $p<0.0001$), as shown by the black bars. (Of the 22 listeners, all but 5 showed overshoot at 2500 Hz and all but 4 at 4800 Hz.) Vertical lines through the bars represent one standard error.

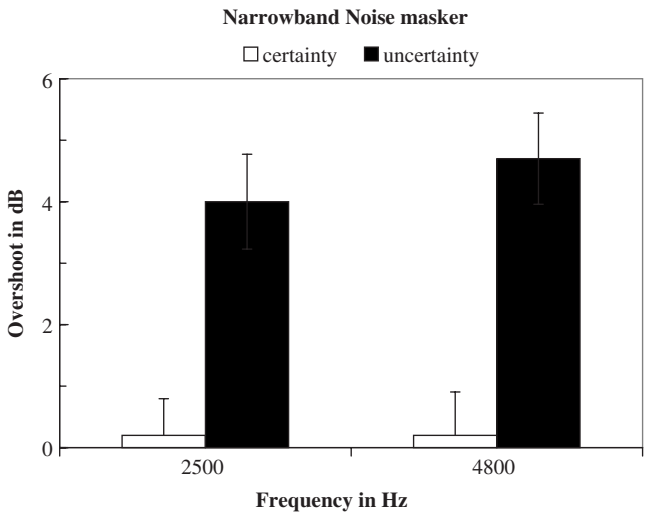


FIG. 1. Overshoot in narrowband noise.

III. EXPERIMENT 2: OVERSHOOT IN BROADBAND NOISE

A. Method

All measurements were made with one of two two-alternative forced choice (2AFC) paradigms. In the adaptive version, signal level changed according to a three-down/one-up rule to yield the SPL required for 79% correct (cf. Levitt, 1971). In the constant version, the signal level was kept invariant throughout a block of trials, and the measured variable was the percentage correct. The adaptive version was used only with frequency certainty, the constant version with both certainty and uncertainty. (Mean levels for 79% correct as measured with the adaptive and constant versions were within 0.5 dB of each other.) A yes/no procedure like the one in experiment 1 was not used because the 2AFC procedure has been the usual method in most recent studies of overshoot and because it is thought to provide somewhat more precise data.

Stimulus durations and temporal relations were the same as in experiment 1, i.e., signals lasted 10 ms, noise 398 ms, delay from noise to signal onset was either 2 or 200 ms, etc. Signal frequencies were 700, 2500, 4800, and 5800 Hz. The masking noise was broad band, from 300 to 8000 Hz with a spectrum level of 31.14 dB (overall SPL of 70 dB).

In the first of three experimental sessions, 79% thresholds were measured by the adaptive procedure at each of the four frequencies, first with a signal delay of 200 ms and then with a delay of 2 ms. The timing of each trial was like that in experiment 1. This first session provided training for the listeners as well as estimates of the signal levels yielding performance in the vicinity of 80% correct, which were then used as the initial fixed levels in subsequent sessions. In the second session, measurements were made first with frequency uncertainty. For approximately half the listeners, the 2-ms delay was tested in this session, and for the remainder, the 200-ms delay was tested. A series of eight blocks of 48 trials each were run with the appropriate signal delay. On each trial, the signal frequency was chosen from among the same eight values as in experiment 1: 700, 1000, 1370, 1850, 2500, 3400, 4800, and 5800 Hz. Every frequency occurred six times in a block of trials, in random order, for a total of 48 presentations of each frequency over the eight blocks. The first block was run with signal levels set on the basis of the results of the first session. If performance was not in the vicinity of 80% correct over most of the signal frequencies, the block was repeated one or two times, with signal level adjusted 1 or 2 dB as necessary. Eight blocks were then run with the signal levels unchanged. Finally, threshold was measured with frequency certainty at two or three frequencies (depending on time) at which overshoot had been strongest for the listener in the first session. (As expected, at 700 Hz overshoot was strong for only a few listeners.) The third session was a repeat of the second session with the signal delay not yet tested.

Listeners were students at Northeastern University, four of whom served also in experiment 1. Owing to time constraints, sufficient data became available only at 2500 Hz for six listeners and at 4800 Hz for seven listeners, with five listeners serving at both signal frequencies.

The equipment and signal generation were the same as for experiment 1.

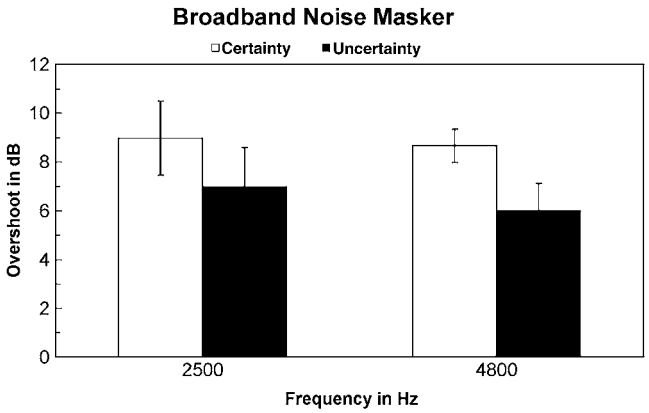


FIG. 2. Overshoot in broadband noise.

B. Results

The right-hand side of Table I lists the SPLs expected to yield 76% correct under the various stimulus conditions for broadband noise; 76% correct corresponds approximately to the criterion of d' of 1.0 used for narrowband noise. Levels were calculated on the assumption that a 5% change in performance corresponds to a 1-dB change in level (e.g., Buus *et al.*, 1986). With a signal delay of 200 ms and frequency certainty, thresholds at 4800 Hz in broadband noise are reasonably close to those in narrowband noise—shown on the left-hand side of Table I—as they should be since the spectrum level was the same. However, the 200-ms thresholds with frequency certainty at 2500 Hz are approximately 2.6 dB higher in broadband noise than in narrowband noise. We do not have an explanation for this discrepancy.

The important measure to be gleaned from Table I is the decrease in threshold when the signal delay goes from 2 to 200 ms, which is the measure of overshoot and is plotted in Fig. 2. At both signal frequencies, overshoot is significantly greater with certainty than with uncertainty (by analysis of variation, at 4800 Hz, $F_{1,6}=6.51$, $p=0.043$; at 2500 Hz, $F_{1,6}=6.12$, $p=0.056$).

IV. DISCUSSION

The results confirm our predictions of how frequency uncertainty would differentially affect overshoot in narrowband noise and in broadband noise. We discuss first overshoot in narrowband noise.

A. Narrowband masker

As predicted, overshoot in narrowband noise is absent when the listener knows the signal frequency beforehand, a finding reported often in the literature; what is new is that overshoot is present when the listener does not know the signal frequency beforehand. Our disruption hypothesis ac-

counts for these results in the following way. When the signal frequency is known, i.e., with frequency certainty, the abrupt onset of a narrowband noise centered on the signal frequency does not disrupt focusing at that place by, for example, drawing attention to other frequencies. Consequently, the threshold for a brief tone shortly following noise onset is not raised. In contrast, when the signal frequency is not known beforehand, the noise, centered on the signal frequency, leads the listener to focus on the correct frequency region, but the focusing is not accomplished immediately. Accordingly, a signal coming on only 2 ms after noise onset must be raised in level to be detected. Delaying the signal 200 ms appears to provide almost enough time to focus on the target frequency (cf. Scharf *et al.*, 2007). Almost enough, because uncertainty is still detrimental even after a 200-ms delay. Table I shows that for uncertainty, thresholds were 1.3 and 2.2 dB higher than for certainty; although small, these differences are statistically significant. They suggest that longer than 200 ms may be needed for full focusing.

The disruption hypothesis may also help to explain the presence of overshoot in certainty for signals that are not at the center of a narrowband noise masker (Elliott, 1965) or that are at a frequency different from that of a tonal masker (Bacon and Viemeister, 1985a, b). Elliott (1965) found that with noise bands approximately one critical band wide, overshoot was small at a signal frequency centered in the noise, but was large at frequencies below and above the frequency limits of the noise (e.g. her Table I). Corresponding results have been reported by Bacon and Viemeister (1985b) for a tonal masker. They measured the threshold for a 1000-Hz tone pulse as a function of its delay relative to the onset of an 800-, 1000-, or 1250-Hz sinusoidal masker. Signal duration was 20 ms, and masker duration was 300 or 800 ms. Overshoot was small or absent for the three listeners when the masker was at 800 or 1000 Hz but was at least 10 dB in most conditions when the masker was at 1250 Hz. In another study, Bacon and Viemeister (1985a) measured the threshold for a 20-ms 1000-Hz tone pulse masked by a 50-ms tone that was either gated on 15 ms before the signal or was continuous. Generally, detection was the same whether the masker was gated or continuous, provided signal and masker were at the same frequency. At higher masker levels, detection became markedly poorer as the frequency of the masker was increased relative to that of the signal, less so as it was decreased.

The observed increase of overshoot as a tonal or narrowband noise masker moves away from the signal frequency is explained by the disruption hypothesis in accordance with which attention is drawn exogenously to the masker frequency at masker onset. However, under frequency certainty, attention is focused again on the known signal frequency after 200 ms or so. This explanation for overshoot in frequency certainty with off-frequency maskers parallels that given for overshoot in uncertainty with narrowband noise at the signal frequency. The disruption hypothesis would also explain why narrowband noise is especially disruptive at frequencies higher than the signal frequency, if it were to be shown that off-frequency maskers divert attention more readily at frequencies above the signal than below. Some data in the literature do suggest that listeners attend more to high- than to low-frequency tone bursts (e.g. Richards and Neff, 2004; Scharf *et al.*, 2007). On the whole, these results with off-frequency sinusoidal and narrowband noise maskers lend support to the disruption hypothesis in showing that it is not the bandwidth of the masker but its frequency location that determines the amount of overshoot.

B. Broadband masker

The prediction of the disruption hypothesis for a broadband masker was that overshoot is greater with frequency certainty than with uncertainty. This prediction was based upon the following reasoning. In continuous noise, uncertainty causes an increase of around 3 dB in threshold (e.g. Green, 1961; Scharf *et al.*, 2007). As a signal presented 200 ms after noise onset is usually detected as well or nearly as well as the same signal in continuous noise, it follows that after 200 ms, threshold in our experiments should be higher with uncertainty than with certainty, as it was.[2] In contrast, near masker onset, threshold should be the same whether frequency is known or not because noise onset disrupts attention on every trial, eliminating the advantage of advance frequency focusing.

C. Overshoot as a measure of the "true" effect of frequency uncertainty on detection

At both signal frequencies, overshoot in broadband noise was near 9 dB which is close to the average value reported in the literature at similar frequencies (e.g., Bacon and Savel, 2004). We now suggest that the disturbance produced by the noise onset reveals the true effect of frequency uncertainty on detection. In previous attempts to measure the effect of uncertainty, the signal was presented in continuous noise, at a different frequency on every trial. Undisturbed by noise onset, the onset of the signal was able to set off a process that resulted in attention to the frequency of the ongoing signal. Focusing could then go to the correct critical band to permit detection. In frequency certainty, already prior to signal onset, attention is focused on the correct critical band. To the extent that focusing takes time, signal processing begins sooner with certainty than with uncertainty and the signal can be set a few decibels lower. Tentative support for this hypothesis is provided by Scharf *et al.* (2007) who measured the detection of a 40-ms tone burst in continuous broadband noise under frequency uncertainty. A preceding cue with the same frequency and duration as the signal, but only 4 dB above threshold (as previously measured under certainty), fully overcame the detrimental effect of frequency uncertainty on detection.

The question arises as to why the narrowband noise masker in our experiment 1 needed 200 ms before it could focus attention on its frequency locus. (Recall that we attributed the presence of narrowband-noise overshoot with uncertainty to the delay in focusing at the noise frequency.) Now we postulate that a signal, 3 dB above certainty threshold, focuses attention very rapidly on its spectral locus. Why does not the onset of the narrowband noise enlist focusing just as quickly as the onset of the tone? At least two possibilities

present themselves. One possibility is that rapid focusing at sound onset may require a sinusoidal signal. A test of this possibility would be to measure overshoot with a sinusoidal masker with uncertainty. Another possibility is that the sound must be near threshold to enlist rapid focusing on the correct critical band at onset; more intense sounds, stimulating a broader frequency region, would make focusing initially less precise. This possibility would be more difficult to test directly. Indirect support comes from a report by Sorkin (1965) that a contralateral sinusoidal cue that matches the simultaneous signal of uncertain frequency enhances performance less as the cue level increases.

Thus, it may be that in previous attempts to measure the effect of uncertainty, signals presented 3 dB above certainty threshold had onsets that were in the intensity range that permitted focusing on the signal frequency and thereby overcame the effect of uncertainty. Accordingly, uncertainty appeared to lead to only a 3-dB increase in threshold, well below the 10 dB or so predicted by energy-detection theories (cf. Green, 1961).[3] By presenting the signal close to the onset of a broadband noise, the listener is prevented from using signal onset to focus on the correct critical band and so the signal level must be increased 9 or 10 dB to permit detection. It would seem that signal onset is of preponderant importance in offsetting the lack of information about signal frequency in uncertainty. Once focusing is in the right critical band, the rest of the incoming signal determines threshold. Thus, threshold with uncertainty decreases with increasing signal duration just as it does with frequency certainty (cf. Scharf *et al.*, 2007).

D. Unanswered questions

Although our disruption hypothesis may account for the failure of a narrowband noise masker to elevate threshold for a centered signal that comes on near noise onset, it leaves unanswered many riddles about overshoot, as indicated in the Introduction. Moreover, the disruption hypothesis would seem intuitively at variance with the finding that in broadband noise, overshoot is greatest at moderate masker levels (e.g. Fastl, 1976; Bacon and Savel, 2004). One might expect the disruption to increase further at still higher noise levels. On the other hand, the large individual differences in the amount of overshoot that have been often reported (e.g. Zwicker, 1965; Wright, 1995) would seem more congruent with an attentional mechanism than with a mechanism based on basilar-membrane input–output functions (e.g., Von Klitzing and Kohlrausch, 1994; Strickland, 2001; Bacon and Savel, 2004).

ACKNOWLEDGMENTS

Many of the data were collected by John Suciu, Jennifer Olejarczy, and Kaitlyn Amato. Zhenlan Jin programmed all the experiments. The authors also thank Yves Cazals and Jacques Magnan for providing the facilities in their laboratory in Marseille for preliminary measurements and Thierry Voinier for his earlier programming assistance. This research was supported by AFOSR Grant No. FA9550-04-1-0244.

[1]Two papers report measurements of overshoot in patients whose OCB had been severed, thereby eliminating efferent input to the cochlea. Scharf *et al.* (2007) found in two patients that overshoot was 15 and 11 dB in their operated ears and 14 and 12 dB, respectively, in their normal ears (both patients had normal audiometric thresholds at all frequencies). Masked thresholds for the 5-ms signals were normal and the same in the operated and unoperated ears after a delay of 100 or 200 ms re masker onset; overshoot was due entirely to threshold increases at the 5-ms delay, just as it is in normal listeners. On the other hand, Zeng *et al.* (2000) found that overshoot was greater in the nonoperated ears of five patients than in the operated ears. However, the difference stemmed both from threshold increases in the operated ear after a 300-ms delay re noise onset [corresponding to "the steady-state condition," according to Zeng *et al.* (2000)] and from decreases after a 3-ms delay. (It is to be noted that two of the patients had hearing losses at the test frequencies.) The finding of threshold increases at 300 ms is contrary to the report of Scharf *et al.* (1997) that masked thresholds in continuous noise were normal in 16 patients, including six tested before and after surgery; the finding of threshold increases at 3 ms is contrary to the fundamental assumption of the efferent hypothesis, reiterated by Zeng *et al.* (2000, p. 108), namely that efferent input comes into play only after many milliseconds. Nonetheless, it is clear that efferent input is involved in auditory attentional effects. For example, Delano *et al.* (2007) showed that when attention is focused on a visual task, physiological responses from the chinchilla's auditory nerve are reduced. In humans, Scharf *et al.* (1997) showed that efferent input is needed to avoid, in effect, the detection of weak signals at unattended frequencies. Thus, our own current demonstration of a likely involvement of attention in overshoot makes an efferent explanation highly attractive. Unfortunately, just how input from the OCB would account for overshoot is unclear, and until now almost none of the behavioral evidence supports such an account.

[2]Although the threshold at 200 ms was indeed higher with uncertainty, it was closer to 2 dB than to the expected 3 dB, at both our signal frequencies. It may be that 200 ms did not suffice for full recovery from noise onset in the conditions of our experiment, a possibility already noted in the discussion of the results with a narrowband noise masker. It may also be that more than eight possible signal frequencies are needed to achieve maximum uncertainty.

[3]The 10-dB approximation arises as follows. Given our broadband noise, the ideal listener attends equally to all 17 critical bands in the noise with frequency uncertain, but attends to just the signal band with frequency certain. If energy is pooled across all attended critical bands, the benefit of certainty (relative to uncertainty) is $10 \log_{10} 17 = 12.3$ dB. However, if decisions (not energy) are pooled over critical bands, and the critical bands act as channels with identical and independently distributed noise, then the benefit of certainty is only 4.7 dB (for details, see the Appendix in Scharf *et al.*, 2007). In this latter case, we must assume that noise onset reveals more than just the effects of uncertainty.

Backus, B. C., and Guinan, J. J., Jr., (**2006**). "Time-course of the human medial olivocochlear reflex," J. Acoust. Soc. Am. **119**, 2889–2904.

Bacon, S. P., and Liu, L. (**2000**). "Effects of ipsilateral and contralateral precursors on overshoot," J. Acoust. Soc. Am. **108**, 1811–1818.

Bacon, S. P., and Moore, B. C. J. (**1986**). "Temporal effects in simultaneous pure-tone masking: effects of signal frequency, masker/signal frequency ratio, and masker level," Hear. Res. **23**, 257–266.

Bacon, S. P., and Savel, S. (**2004**). "Temporal effects in simultaneous masking with on and off-frequency noise maskers: Effects of signal frequency and masker level," J. Acoust. Soc. Am. **115**, 1674–1683.

Bacon, S. P., and Smith, M. A. (**1991**). "Spectral, intensive, and temporal factors influencing overshoot," Q. J. Exp. Psychol. A **43**, 373–399.

Bacon, S. P., and Takahashi, G. A. (**1992**). "Overshoot in normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **91**, 2865–2871.

Bacon, S. P., and Viemeister, N. F. (**1985a**). "Simultaneous masking by gated and continuous sinusoidal maskers," J. Acoust. Soc. Am. **78**, 1220–1230.

Bacon, S. P., and Viemeister, N. F. (**1985b**). "The temporal course of simultaneous tone-on-tone masking," J. Acoust. Soc. Am. **78**, 1231–1235.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**). "Decision rules in detection of simple and complex tones," J. Acoust. Soc. Am. **80**, 1646–1657.

Carlyon, R. P., and Moore, B. C. J. (**1986**). "Continuous versus gated pedestals and the 'severe departure' from Weber's law," J. Acoust. Soc. Am. **79**, 453–460.

Champlin, C. A., and McFadden, D. (**1989**). "Reductions in overshoot following intense sound exposures," J. Acoust. Soc. Am. **85**, 2005–2011.
Dai, H., and Buus, S. (**1991**). "Effect of gating the masker on frequency-selective listening (L)," J. Acoust. Soc. Am. **89**, 1816–1818.
Delano, P. H., Elgueda, D., Hamame, C. M., and Robles, L. (**2007**). "Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas," J. Neurosci. **27**, 4146–4153.
Elliott, L. L. (**1965**). "Changes in the simultaneous masked threshold of brief tones," J. Acoust. Soc. Am. **38**, 738–746.
Fastl, H. (**1976**). "Temporal masking effects. I. Broad band noise masker," Acustica **35**, 287–302.
Fastl, H. (**1977**). "Temporal masking effects. II. Critical band noise masker," Acustica **36**, 317–331.
Green, D. M. (**1961**). "Detection of auditory sinusoids of uncertain frequency," J. Acoust. Soc. Am. **33**, 897–903.
Guinan, J. J., Jr. (**2006**). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," Ear Hear. **27**, 589–607.
Igarashi, M., Alford, B. R., Nakai, Y., and Gordon, W. P. (**1972**). "Behavioral auditory function after transection of crossed olivo-cochlear bundle in the cat. I. Pure-tone threshold and perceptual signal-to-noise ratio," Acta Oto-Laryngol. **73**, 455–466.
Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
McFadden, D., and Champlin, C. A. (**1990**). "Reductions in overshoot during aspirin use," J. Acoust. Soc. Am. **87**, 2634–2641.
Miskiewicz, A., Buus, S., and Florentine, M. (**2006**). "Effect of masker-fringe onset asynchrony on overshoot (L)," J. Acoust. Soc. Am. **119**, 1331–1334.
Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.
Scharf, B., and Buus, S. (**1986**). "Audition I," *Handbook of Perception and Human Performance, Vol. 1*, edited by K. Boff, L. Kaufman, and J. Thomas (Wiley, New York).
Scharf, B., Chays, A., and Magnan, J. (**1997**). "The role of the olivocochlear bundle in hearing: Sixteen case studies," Hear. Res. **103**, 101–122.
Scharf, B., Reeves, A., and Suciu, J. (**2007**). "The time required to focus on a cued signal frequency," J. Acoust. Soc. Am. **121**, 2149–2157.
Schmidt, S., and Zwicker, E. (**1991**). "The effect of masker spectral asymmetry in simultaneous masking," J. Acoust. Soc. Am. **89**, 1324–1330.
Strickland, E. A. (**2001**). "The relationship between frequency selectivity and overshoot," J. Acoust. Soc. Am. **109**, 2062–2073.
Sorkin R. D. (**1965**). "Uncertain signal detection with simultaneous contralateral cues," J. Acoust. Soc. Am.," **38**, 207–212.
Strickland, E. A. (**2004**). "The temporal effect with notched-noise maskers: Analysis in terms of input–output functions," J. Acoust. Soc. Am. **115**, 2234–2245.
Von Klitzing, R., and Kohlrausch, A. (**1994**). "Effect of masker level on overshoot in running- and frozen-noise maskers," J. Acoust. Soc. Am. **95**, 2192–2201.
Wright, B. A. (**1995**). "Detectability of simultaneously masked signals as a function of signal bandwidth for different signal delays," J. Acoust. Soc. Am. **98**, 2493–2503.
Wright, B. A., and Dai, H. (**1994**). "Detection of unexpected tones with short and long durations," J. Acoust. Soc. Am. **95**, 931–938.
Zeng, F.-G., Martino, K. M., Linthicum, F. H., and Soli, S. D. (**2000**). "Auditory perception in vestibular neurectomy subjects," Hear. Res. **142**, 102–112.
Zwicker, E. (**1965**). "Temporal effects in simultaneous masking and loudness," J. Acoust. Soc. Am. **38**, 132–141.

Contributions of talker characteristics and spatial location to auditory streaming

Kachina Allen[a] and Simon Carlile
Department of Physiology, University of Sydney, Sydney, NSW 2106, Australia

David Alais
School of Psychology, University of Sydney, Sydney, NSW 2006, Australia

(Received 17 May 2006; revised 7 December 2007; accepted 10 December 2007)

To examine whether auditory streaming contributes to unmasking, intelligibility of target sentences against two competing talkers was measured using the coordinate response measure (CRM) [Bolia *et al.*, J. Acoust. Soc. Am. **107**, 1065–1066 (2007)] corpus. In the control condition, the speech reception threshold (50% correct) was measured when the target and two maskers were collocated straight ahead. Separating maskers from the target by ±30° resulted in spatial release from masking of 12 dB. CRM sentences involve an identifier in the first part and two target words in the second part. In experimental conditions, masking talkers started spatially separated at ±30° but became collocated with the target before the scoring words. In one experiment, one target and two different maskers were randomly selected from a mixed-sex corpus. Significant unmasking of 4 dB remained despite the absence of persistent location cues. When same-sex talkers were used as maskers and target, unmasking was reduced. These data suggest that initial separation may permit confident identification and streaming of the target and masker speech where significant differences between target and masker voice characteristics exist, but where target and masker characteristics are similar, listeners must rely more heavily on continuing spatial cues.
 [DOI: 10.1121/1.2831774]

PACS number(s): 43.66.Dc, 43.71.Bp, 43.71.Gv, 43.66.Qp [AK] Pages: 1562–1570

I. INTRODUCTION

The ability to comprehend speech in a multitalker environment has generally been recognized as a perceptually easy but theoretically difficult task. Cherry (1953) identified a number of cues which could be used to separate out a target in a multiple-talker environment. These include visual cues such as lip-reading (Benoit *et al.*, 1994; Grant and Seitz, 2000; Grant, 2001; Schwartz *et al.*, 2004; Helfer and Freyman, 2005; Wightman *et al.*, 2006) or gestures (e.g., Munhall *et al.*, 2004), difference in voice characteristics due to gender, voice pitch, or accents (sometimes referred to as monaural cues for segregation (Brungart *et al.*, 2006), lexical expectations (Sitler *et al.*, 1983; Boothroyd and Nittrouer, 1988), and the differing spatial location of sound sources (e.g., Bronkhorst and Plomp, 1988; Bregman, 1990; Kidd *et al.*, 1994; Culling and Summerfield, 1995; Kidd *et al.*, 1998; Freyman *et al.*, 2001; Noble and Perrett, 2002; Litovsky, 2005).

Hawley *et al.* (2004) report that when target and masker talkers are collocated, unmasking was increased by increasing the differences in the fundamental frequencies between the target and masker talkers (F0). Similarly, differences in F0 have been used to correctly stream target words with target sentences (Darwin and Hukin, 1999, 2000).

To understand speech in noise requires the identification of aspects of the auditory scene which are attributable to the target source. Yost (1991) argued that the perception of separate auditory objects was facilitated by binaural and spectral cues. The association of spectral components with individual sources probably involves both a short-term grouping process and, separately, a streaming of the grouped elements over time (Bregman, 1990; Shinn-Cunningham *et al.*, 2005; Snyder *et al.*, 2006). Spatial location may play different roles in the grouping and streaming processes. Edmonds and Culling (2005) suggest that spatial location is important in streaming objects, but not necessarily in grouping objects. They showed that interaural time difference (ITD) does not have to be consistent across frequencies for spatial unmasking so long as it is different from the target ITD. This allows target and masker to be streamed as separate objects but not grouped at set locations. Changing the spatial location of objects during streaming and thereby covarying all of the cues to location and not just ITD may help to clarify the role of location as opposed to just ITD differences in grouping and streaming.

The aim of the present study was to explore the interaction of spatial location and voice differences in unmasking in a multitalker listening scenario. Differences in spatial location provide significant unmasking and presumably a clearer identification of features such as voice differences. In this study, spatial location cues to the talkers were removed following identification of the target talker in order to examine any ongoing contribution of nonspatial cues to unmasking.

[a]Electronic mail: kachina-allen@yahoo.com

II. GENERAL METHODS

A. Participants

Four female volunteers and one male volunteer (mean age 32.2, range 30–35 yr) were recruited. Subjects had normal hearing, with pure tone thresholds below 20 dB HL at all standard audiometric frequencies between 250 and 8000 Hz. Four out of five subjects spoke English as a first language, with subject S2 (a native Russian speaker) having learned English at age 18. All currently use English as a main language for communication. Subjects had not previously participated in tests of auditory perception and were not remunerated for participation. All gave written informed consent.

B. Setup

Subjects were seated, facing forward, in a sound attenuated audiometric booth (size=$3.5\times4.6\times2.4$ m), lined with 76 mm acoustic foam to create a semianechoic environment. Three Tannoy active loudspeakers were placed 1.3 m away on the subject's audiovisual horizon at 0° (directly ahead) and ±30° on either side of the midline. Subjects were instructed to remain facing directly ahead at all times, though their heads were not restrained.

C. Stimuli

Stimuli were taken from the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000). The CRM corpus has been used extensively in multitalker testing and has demonstrated release from masking under spatially separated conditions (e.g., Brungart *et al.*, 2001; Ericson *et al.*, 2004; Shinn-Cunningham and Ihlefeld, 2004). It consists of sentences which take the form "Ready [call sign] go to [color][number] now" where the call sign consisted of the target identifying "Baron" or one of the masking call signs ("Charlie," "Ringo," "Eagle," "Arrow," "Hopper," "Tiger," and "Laker"). Colors were chosen from red, white, blue, and green and the numbers from 1 to 8. Complete sentence stimuli were cosine ramped for 15 ms at the beginning and end and where sentences changed location midway, ramping occurred at the beginning and end of the section of a sentence at each location. Preexperimental testing ensured no audible click was present during the location change. For all experiments, masker talkers, call signs, colors, and numbers were different for each masker and the target.

Stimuli were generated and presented using MATLAB (MathWorks Inc.) and a Hammerfall multiface sound card (RME Ltd.). Stimuli were part of a wide-band corpus (80 Hz–8 kHz) and were upsampled to a sampling rate of 44 100 Hz. The target and maskers were played at a combined sound level of 57 dB sound pressure level (SPL).

D. Procedure

Stimuli sentences consisted of an identifier in the first segment (the call sign) and two target words (a color and a number) in the second segment. The target talker was identified by the call sign "Baron" and the subject's task was to identify both of the two target words following the call sign in the presence of two masker talkers with different call signs and scoring words. Subjects responded to the target words by typing answers into a laptop.

The signal to noise ratio (SNR) for each trial was calculated using a maximum 20 dB range per block of 50 trials with 10 trials being presented at each of 5 evenly spaced SNR levels. Maximum and minimum SNR levels were varied between blocks depending on the performance of the subject by altering the level of the target while maintaining a constant overall SPL. This was done to maximize the number of data points around the threshold of interest. All subjects were presented with stimuli between the SNR ranges of −35 and +10 dB. The majority of trials were presented between −20 and 0 dB SNR, to provide maximum data.

A test sentence was scored correct when both the target color and number were correctly identified. A psychometric function (cumulative Gaussian) was fitted to the data using a maximum likelihood procedure (Watson, 1979) and the speech reception threshold (SRT) defined as the point where 50% of the test sentences were correctly identified. The SRT was calculated for each condition and expressed as the level of the target (in decibels) relative to the level of the combined maskers. Release from masking (RFM) was calculated as

$$\mathbf{RFM\ (condition) = SRT(collocated) - SRT(condition)}.$$

A bootstrapping procedure (Efron and Tibshirani, 1993; Davison and Hinkley, 1997) was used to resample each psychometric function 1000 times. Standard deviations for SRTs were estimated by calculating the standard deviation of SRTs obtained from the resampled curves. Post-hoc t-tests (α =0.05) with Bonferroni corrections were carried out using the bootstrapped standard deviations (see Alais and Carlile, 2005).

III. EXPERIMENT 1

A. Design

All subjects carried out 100 unrecorded practice trials of the *separated*, *collocated*, and *start-separated* conditions (see the following) and 150 experimental trials of each condition in blocks of 50 trials.

Four conditions were tested:

(i) ***Collocated***: Target and both maskers played from the central loudspeaker.
(ii) ***Separated***: Target played from central loudspeaker, one masker played from each of symmetrically spaced loudspeakers 30° from the central loudspeaker.
(iii)***Start separated***: Target and maskers start as in *separated* condition but maskers move to central loudspeaker after 700 ms (just after the identifying call sign).
(iv)***Start collocated***: Target and maskers start from central speaker as in *collocated* condition but move to locations as in the *separated* condition after 700 ms.

The CRM corpus was recorded with talkers pacing themselves against a standard rhythm; however, timing was

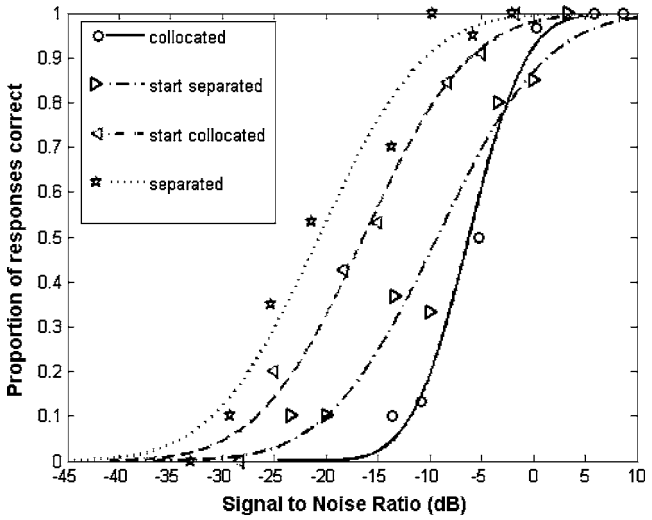


FIG. 1. Representative (binned) data for one subject (S4) showing psychometric functions for each condition in Experiment 1. Binning was carried out for illustrative purposes only. Speech Reception Thresholds (SRTs) calculated at 50% intelligibility demonstrate varying levels of performance in the conditions.

not identical for all sentences. Thus, 700 ms fell somewhere between the end of the call sign and the end of the word "to." The movement was easily noticeable when played as a single masker, but during experimental conditions where subjects concentrated on the stable target and only the maskers moved, subjects rarely noticed the difference between the *start-separated* and *start-collocated* conditions and the *collocated* condition.

For each trial, the masker and target talkers were randomly selected without replacement from the four male and four female talkers from the CRM corpus. This ensured that the sex of the talkers involved in each trial varied randomly throughout the experiment.

B. Results

The psychometric functions for one subject for each of the four conditions are plotted in Fig. 1. For plotting purposes for Fig. 1 only, the proportion correct have been binned into up to eight SNR ranges with the associated cumulative Gaussian curves fitted to the binned data also plotted. All other calculations were carried out on unbinned data.

SRTs for each subject show progressively increasing performance (decreasing SNR) with the *collocated*, *start-separated*, and *separated* conditions. See *collocated* (Fig. 1, solid line and circles; Fig. 2, white columns, range −3.7 to −11.1 dB, mean −7.0 dB), *start-separated* (Fig. 1, right triangles and dot-dash line; Fig. 2, gray columns, range −5.9 to −15.3 dB, mean −10.6 dB), and *separated* (Fig. 1, dotted line and stars; Fig. 2, black columns, range −11.4 to −25.8 dB, mean −19.0 dB) conditions. The *start-collocated* condition (Fig. 1, left triangles and dashed line; Fig. 2, striped columns, range −9.0 to −26.4 dB, mean −18.0 dB) shows slightly better performance than the *separated* condition in two subjects (S1 and S3) and marginally poorer performance in the other three subjects.

When the target (directly ahead) was *separated* from two symmetrical maskers (±30°), significant spatial RFM

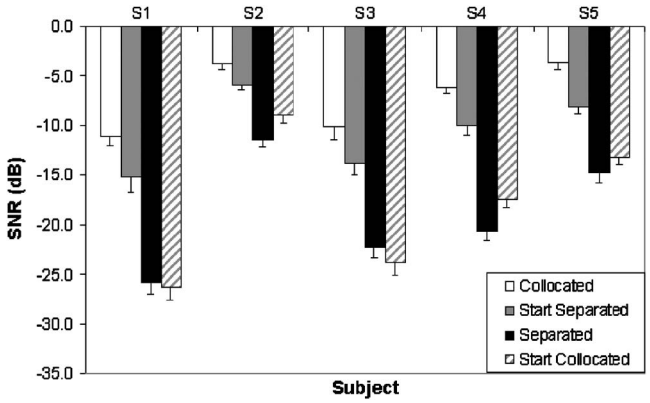


FIG. 2. SRTs for conditions using the full (mixed-sex) CRM corpus in Experiment 1. Error bars are standard deviations calculated from 1000 iterations of a bootstrapping procedure.

was found (Fig. 3, black bars, range 7.7–14.7 dB, mean 12.0 dB). Similarly, all subjects showed significant unmasking in the *start-collocated* condition (Fig. 3, striped, range 5.2–15.3, mean 11.0 dB). In this condition the target and masker started collocated directly ahead but then the maskers were separated ±30° prior to the scoring words.

The *start-separated* condition, where the symmetrical maskers became collocated with the target prior to scoring words, also demonstrated significant RFM for all subjects (Fig. 3: gray bars, range 2.1–4.5 dB, mean 3.6 dB). However, the level of unmasking for the *start-separated* condition was significantly smaller than for both the *separated* and *start-collocated* conditions (Friedman's test with post-hoc t-tests, $p < 0.01$). The level of unmasking between the start-collocated and the separated conditions was not significantly different.

C. Discussion

When target and masking voices are collocated, it is likely that listeners rely entirely on cues such as voice characteristics (gender, tone, accent, etc.) and other primitive grouping cues to isolate the target talker. When separated, the listener can also use the spatial location of the talkers to aid in the identification and streaming of talkers. The overall

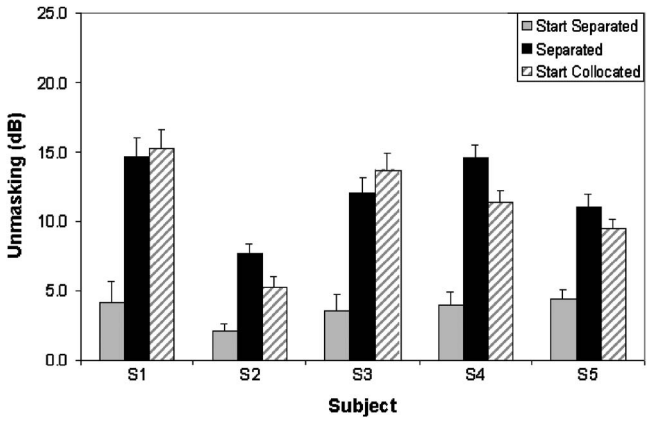


FIG. 3. Release from masking (RFM) as compared to the *collocated* baseline condition for conditions using the full (mixed-sex) CRM corpus in Experiment 1. Error bars are standard deviations calculated from 1000 iterations of a bootstrapping procedure. All conditions show significant levels of unmasking.

location, based on cues such as the binaural (interaural time and level differences) and spectral cues may be exploited, possibly through spatial attention (Arbogast and Kidd, 2000).

The spatial release from masking between the *collocated* and the 30° symmetrically *separated* condition (12.2 dB) is toward the upper limits found in previous studies which range from 2–5 dB (Noble *et al.*, 1997; Noble and Perrett, 2002) to 12 dB (Hawley *et al.*, 2004). The CRM corpus has the same format of carrier phrase for target and maskers, with the same onset time and similar (but not identical) talking rates. This synchronization maximizes both the energetic and informational masking. Informational masking using speech or speech-like stimuli has been shown to provide more spatial release from masking than energetic masking (Kidd *et al.*, 1998; Arbogast and Kidd, 2000; Brungart and Simpson, 2002; Noble and Perrett, 2002; Hawley *et al.*, 2004). Thus the relatively large spatial release from masking demonstrated in this study could reflect the maximization of informational masking.

Previous studies have disagreed over the relative contributions of voice cues and location cues in grouping and streaming (Culling and Summerfield, 1995; Mondor *et al.*, 1998; Edmonds and Culling, 2005). In the current study, all subjects showed a significant level of unmasking in the *start-separated* condition. This may be due to the initial separation allowing the listener to more easily create and identify the target streams and thus detect the voice characteristics of the individual talkers. Identifying the target talker would permit attention to that talker to be maintained (based on voice characteristics), similar to previous priming experiments discussed earlier (e.g., Ericson *et al.*, 2004; Freyman *et al.*, 2004). Potentially it could also allow the listener to actively dis-attend to the masker talkers. When spatial cues are eliminated by collocation of the streams, attentional resources focused on these identifying characteristics could continue to enhance the response to the target talker or possibly filter out the masker talkers.

The differences in unmasking between the *separated* (mean 12.2 dB) and *start-separated* (mean 3.6 dB) conditions may indicate the proportion of unmasking which relates to streaming due to location cues (spatial release from masking) and that which relates to unmasking through identification and attendance to nonspatial attributes of the target.

Reducing talker differences and hence the reliability of these nonspatial cues may increase the reliance on spatial information and thus the benefit gained from initially separating maskers from the target. Voice-characteristic differences between target and maskers can be reduced by using a single-sex corpus. In Experiment 3 (to follow), the importance of these differences is explored further.

The lack of a significant difference in unmasking between the *separated* and *start-collocated* conditions suggests that, in the latter condition, listeners may have adopted the strategy of simply listening to the central location after target and maskers separated, without attempting to identify the target prior to separation. Experiment 2 examines whether this strategy may have been adopted by the subjects.

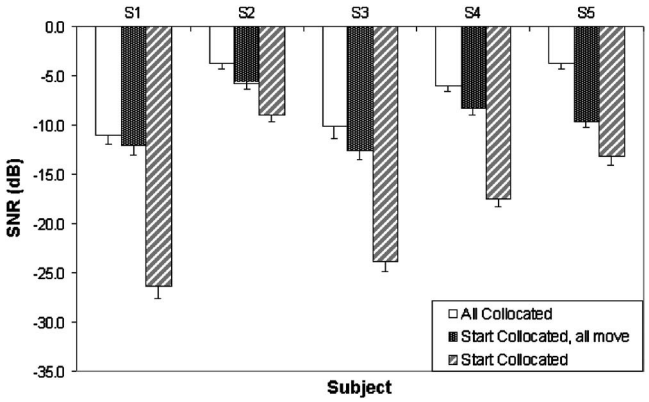


FIG. 4. SRTs for the central loudspeaker for the *start-collocated-all-move* condition in Experiment 2 compared to the *collocated* and *start-collocated* conditions from Experiment 1 using the full (mixed-sex) CRM corpus. Error bars are standard deviations calculated from 1000 iterations of a bootstrapping procedure.

IV. EXPERIMENT 2

A. Design

To test if, in the start collocated condition, subjects were simply listening to the central loudspeaker another condition, "*start-collocated-all-move*," was devised in which subjects were forced to follow the target talker after separation. As before, the target and maskers began collocated at the central loudspeaker, directly ahead, then moved to locations where target and maskers were on different loudspeakers prior to the scoring words, with the target randomly assigned to one of the three loudspeakers, and each masker randomly assigned to one of the remaining loudspeakers. All subjects completed a minimum of 250 trials consisting of one block of 50 trials and two blocks of 100 trials. This ensured that the target sentence moved to each of the speakers for a minimum of 75 trials during the experiment. All subjects were the same as for the previous experiment.

B. Results and discussion

SRTs for the *start-collocated-all-move* condition for Experiment 2 (Fig. 4, stippled columns, SRT range −5.8 to −12.2 dB, mean −9.7 dB) have been compared to the *collocated* and *start-collocated* conditions from Experiment 1 (white and striped columns, respectively). For every subject, SRT results from the *start-collocated-all-move* condition are significantly poorer than those from the *start-collocated* condition. The RFM was also calculated for each subject and plotted in Fig. 5.

There are three main observations relating to the data from Experiment 2. First, subjects demonstrated significant release from masking in the *start-collocated-all-move* condition, at the central loudspeaker (mean 2.7 dB, Fig. 5), however this was significantly lower compared to the *start-collocated* condition in Experiment 1 (mean 11 dB, Fig. 3). Friedman's test also showed differences between these two conditions, with post-hoc t-tests with Bonferroni corrections demonstrating a significant reduction in release from masking for all subjects when the final location of the target was uncertain. That the release from masking was reduced by

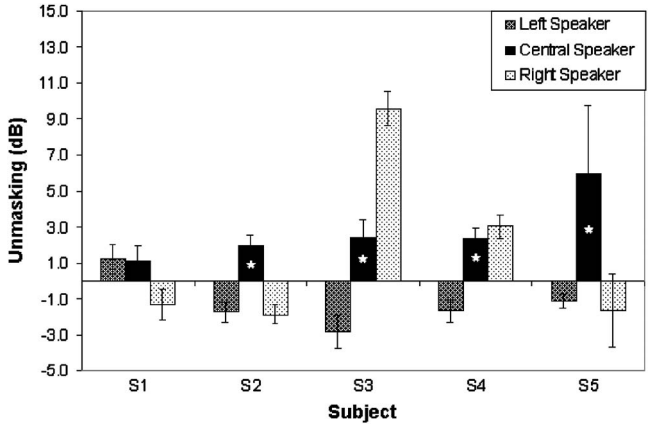


FIG. 5. RFM as compared to the *collocated* baseline condition for thresholds for loudspeakers for the *start-collocated-all-move* condition using the full (mixed-sex) CRM corpus in Experiment 2. Error bars are standard deviations calculated from 1000 iterations of a bootstrapping procedure. Results from the central loudspeaker indicated with a star show significant release from masking.

randomization is consistent with results from Kidd *et al.* (2005), who showed that scores for a known target location compared with a randomly assigned (1 in 3) target location were reduced from 92% to 67% correct.

The second observation arises from the result that performances are slightly better for the central speaker in the *start-collocated-all-move* condition compared to the *collocated* condition and thus there is a moderate release from masking. Notably, in the latter case, post-hoc t-tests with Bonferroni correction indicated that release from masking was only significant in four of the five subjects (S2, S3, S4, S5: "⋆" in Fig. 5). Nonetheless, the subjects' ability to determine who the target talker was should be the same in both conditions as the call sign occurs under identical masking conditions. This would predict that the SRT and release from masking should also be the same. Where the conditions do differ is that during the scoring words in the *start-collocated-all-move* condition the targets and maskers are in different spatial locations. As discussed earlier, the spatial separation of the talkers will lead to an advantage in identifying all three colors and numbers associated with the target talkers and the two maskers. In the absence of a clear cue as to which was the target talker's, the subject may have up to a 1/3 chance at correctly guessing both the target color and number. Importantly, when the test words are collocated and the target words are more heavily masked there are four possible colors and eight possible numbers. Therefore the better performance in the *start-collocated-all-move* condition compared to the *collocated* condition may be related to the difference in probabilities in guessing the correct test words from the limited set available.

Third, the trials in which the target was played from the +30° (right) or −30° (left) loudspeaker were not acoustically identical to the trials where the target was played from the central loudspeaker. In these trials, the SNR should be better in one ear, reducing energetic masking and permitting the listener to use a better ear listening strategy. However, release from masking for test words from the central loudspeaker (mean 2.7 dB, range 1.1–6.0 dB) was in general larger than the release from masking for the scored words from the other loudspeakers (left loudspeaker mean −1.2 dB, range −2.9 to 1.2 dB; right loudspeaker mean 1.6 dB, range −1.9 to 9.6 dB). A t-test (α=0.05) revealed a significant release from masking for the mean results for this central loudspeaker, and Friedman's test showed significant release from masking over individual results. A "better-ear" listening strategy seems to have been adopted only in the case for the right-hand loudspeaker for subject S3. For all ether subjects, release from masking was not significantly better at flanking loudspeakers than at the central loudspeaker. It appears that, despite the acoustic prediction, the randomization procedure with three potential targets may have compromised a better-ear listening strategy. Furthermore, the pattern of results also suggests that subjects generally favored the central loudspeaker either as a consequence of exposure to the previous experiment where the test words were always at the central loudspeaker or due to benefits of the centrally directed gaze (Lie and Branch Coslett, 2006). This is consistent with the interpretation of the *start-collocated* result in Experiment 1, where it is argued that subjects simply attended to the central loudspeaker during the test words without attempting to identify the target talker and stream the sentence from the maskers.

TABLE I. Mean F0 (Hz) calculated from each CRM talker. For group means, standard deviations are given in parentheses.

Male talker	Mean F0	Female talker	Mean F0
Talker 0	100.19	Talker 4	211.06
Talker 1	118.33	Talker 5	178.44
Talker 2	137.96	Talker 6	196.47
Talker 3	112.25	Talker 7	196.98
Mean	117.2 (±15.8)	Mean	195.7 (±13.4)
Overall mean (male and female)	156.5 (±44.1)		

V. EXPERIMENT 3

A. Design

The persistence of unmasking where maskers and target start separated and become collocated (Experiment 1) indicates that the initial separation permits the listener to identify and attend to features other than location in the target stream. In a mixed-sex corpus, there is a great deal of variation between talkers which could provide strong cues to support streaming. In Experiment 3, we used a single-sex corpus by selecting either male or female talkers from the CRM corpus for both target and masker. Table I shows the variation in fundamental frequency of the CRM corpus, both as a mixed-sex and as a single-sex collection. The standard deviation in the fundamental frequency is highest for the mixed-sex corpus (44.1 Hz), lowest for the female talkers (13.4 Hz), and intermediate for the male talkers (15.8 Hz). Fundamental frequencies were measures using Kay Elemetrics Computerized Speech Lab version 4500.

The aim of Experiment 3 is to test whether the persistence in streaming following collocation of the targets and

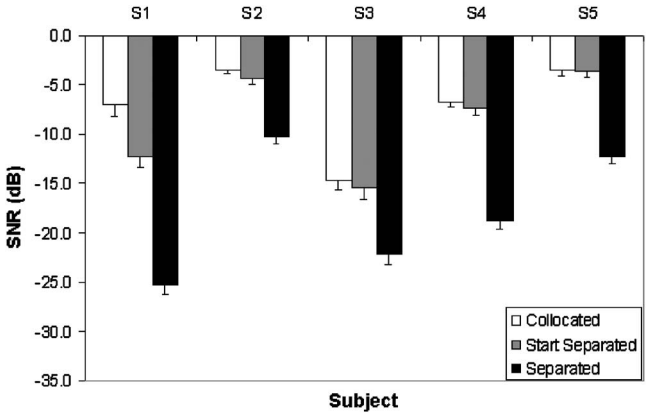


FIG. 6. SRTs for conditions using female talkers from the CRM corpus in Experiment 3. Error bars are standard deviations calculated from 1000 iterations of a bootstrapping procedure.

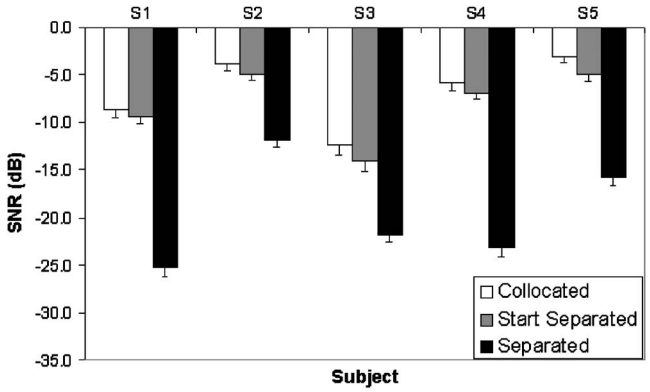


FIG. 7. SRTs for conditions using all *male* talkers from the CRM corpus in Experiment 3. Error bars are standard deviations calculated from 1000 iterations of a bootstrapping procedure.

maskers would be diminished as a consequence of reduced variation in corpus voice cues. Three conditions, which were identical to those used in Experiment 1, were used: *Collocated* (target and two maskers collocated directly ahead), *separated* (target directly ahead and two symmetrical maskers at ±30°), and *start-separated* (maskers start separated ±30° from the target but become collocated prior to the target words). The stimuli used were a same-sex subset from the CRM corpus and consisted of either the four female talkers or the four male talkers. One hundred and fifty trials of each condition were carried out in three blocks of 50 trials. All subjects were the same as for the previous experiment.

B. Results and discussion

When tested with only the female talkers from the corpus, SRTs (see Fig. 6) for each subject show progressively increasing performance (decreasing SNR) with the *collocated* (white columns, SRT range −3.4 to −14.7 dB, mean −7.1 dB), *start-separated* (gray columns, SRT range −3.6 to −15.5 dB, mean −8.6 dB), and *separated* (black columns, SRT range −10.2 to −25.4 dB, mean SRT −17.8) conditions. However, the *collocated* and *start-separated* conditions were only significantly different for one subject (S1).

All subjects showed significant spatial release from masking in the *separated* condition (6.7–18.4 dB range, mean 10.7 dB). in the *start-separated* condition, spatial unmasking was only significant for one subject (S1, 5.3 dB) and a z-test (α=0.05) showed unmasking was not significant for the group (Fig. 8).

When tested with the male-only component of the corpus, SRTs for each subject (see Fig. 7) showed progressively increasing performance for the *collocated* (white columns, SRT range −3.1 to −12.4 dB, mean −6.7 dB), *start-separated* (gray columns, SRT range −5.0 to −14.0 dB, mean −8.1), and *separated* conditions (black columns, SRT range −11.8 to −25.2 dB, mean −19.5 dB).

All subjects showed significant spatial release from masking in the *separated* condition (range 8.0–17.4 dB, mean 12.8 dB). In the *start-separated* condition the spatial unmasking was significantly different from zero in only two subjects (S3 1.6 dB and S5 1.9 dB). As a group, a z-test (α =0.05) showed unmasking was significant in the *start-separated* condition for the male talkers (Fig. 8).

The mean data from Experiments 1 and 3 are plotted in Fig. 8. Spatial release from masking in the *separated* condition was not significantly different between same-sex or mixed-sex talkers. However, the *start-separated* condition showed different levels of masking release depending on the mix of sexes used. When mixed male and female talkers were used, unmasking was greater than when either all male or all female talkers were used. This was true both for individual subjects and for the group. Although the magnitudes of the unmasking is similar (males 1.3 dB, females 1.5 dB), z-tests (α<0.05) showed that unmasking for the start-separated condition with the male corpus was significant, while for the female corpus it was not.

Spatial unmasking might be expected to be greater in a same-sex corpus than in a mixed one as similarities between target and maskers should maximize informational masking. Under listening conditions where there is substantial informational masking, the relative contributions of differences in spatial location to release from masking should be greater. This has been demonstrated in previous studies (e.g., Noble and Perrett, 2002; Festen and Plomp, 1990). However, in this

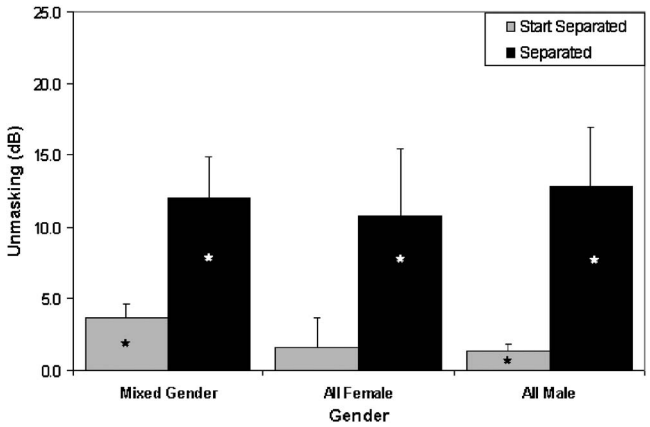


FIG. 8. Mean RFM across subjects in *separated* and *start-separated* conditions from Experiments 1 and 3 (mixed-sex, all female talkers, and all male talkers, respectively). Error bars are standard errors. Significant unmasking is shown by a star.

study unmasking in the separated condition (Fig. 8) did not differ significantly between the mixed-sex, all male, and all female conditions (Friedman's test). This has probably resulted from the quite substantial intersubject variation for each condition. The subject with English as a second language, S2, consistently had the poorest performance on all conditions in all experiments. Individual variation in proportional reliance on spatial versus voice cues may also account for some of the differences, particularly between conditions.

VI. GENERAL DISCUSSION

The principal finding in this study was that to some extent the speech intelligibility advantage provided by spatial separation of target and masker talkers can persist once those cues have been removed. The average masking release of 3.6 dB in the *start-separated* condition, where maskers and targets became collocated prior to the target words, was significantly greater than zero. This indicates that the initial separation provides a significant advantage compared to listening to stimuli which were collocated for the duration of the stimulus. Previous studies on auditory priming using a segment of the target speech before a trial have demonstrated that prior identification of the target can give significant release from masking, even in collocated conditions (Ericson *et al.*, 2004; Freyman *et al.*, 2004). This suggests that initial spatial separation may act as a prime by aiding identification or grouping of maskers and target (depending to some extent on SNR), using nonspatial differences in the talkers. These features could include identification of voice cues based on gender, accent, pitch (F0), speed, temporal fluctuations, and intonation. Once clearly identified, these other factors may provide the basis for an improved streaming of the target even after maskers and target become collocated. That is, the persistence of unmasking could reflect the fact that the initial separation results in a facilitation of attention to salient voice cues.

Where masker and target voices were similar (as with the single-sex corpora), subjects demonstrated reduced unmasking in the *start-separated* condition as compared to the same manipulation using a mixed-gender corpus. This probably reflects in part the reduced differences between the talker characteristics in the single-sex corpora and the probability of salient differences between the target and the masker talker characteristics in any one trial. This may also reflect reliance on different cues varying with task requirements. For instance, Harding and Meyer (2001) have argued that where low-level representations fail to provide source segregation cues, subjects are likely to use higher-level processing based on location, frequency, or other cues used to group and stream auditory sources under the control of attention. The type of higher-level processing would vary with the salience and reliability of various cues. Thus reduced unmasking in the same-sex corpus when the talkers become collocated may also be the result of increased reliance on (or attention to) spatial cues, and a reduction in the reliance on (or attention to) voice cues.

No significant differences were found between the *collocated* conditions of Experiments 1 and 3 (Friedman's test, $\alpha < 0.05$). Previous studies have shown poorer performance where masker and target talkers were the same gender in collocated tests as compared to opposite gender maskers (Brungart, 2001; Darwin *et al.*, 2003; Ericson *et al.*, 2004). However, in previous studies examining same versus different gender maskers, the target gender was identified prior to the trial. In Experiment 1, subjects were not given any indication of target or masker gender and trials consisted of opposite, same, and mixed-gender maskers and target. Thus differences in uncertainty in the talker gender may account for the lack of (expected) differences in collocated SRTs.

Likewise, previous studies have demonstrated that spatial unmasking may increase with higher similarity between target and maskers. This is thought to reflect an added reliance on spatial cues where other cues are ambiguous (Festen and Plomp, 1990; Noble and Perrett, 2002). However, in this study the mean results for the *separated* conditions showed no significant differences in the spatial release from masking for mixed or same-sex talkers. This suggests that spatial cues in the *separated* condition for a same-sex talker trial would have been at least as effective as with mixed-gender trials. On the other hand, the *start-separated* condition did show a difference in release from masking for mixed and same-gender talkers. In this condition, the similarity in the voices in the same-sex conditions may have induced a heavier reliance on spatial cues during streaming—cues which disappeared when maskers and target became collocated. Similar psychophysical experiments exploiting other modalities are consistent with this idea. Where cues such as visual or voice cues are strong, they are utilized, but when these are compromised or confusable, location cues such as binaural cues are used to a greater degree in streaming (Freyman *et al.*, 1999; Noble and Perrett, 2002; Helfer and Freyman, 2005). Alais and Burr (2004) demonstrated that where auditory cues are a less reliable cue for location, visual cues are given preference (leading to the ventriloquist effect), but when visual cues become less reliable, auditory cues are given preference (a reverse ventriloquist effect). Thus the relative reliance on visual or auditory location cues depends on the reliability or salience of each cue.

This interpretation is also consistent with previous work which has demonstrated that differences in voice characteristics can be stronger cues in speech intelligibility trials than spatial cues in streaming sentences (Darwin and Hukin, 2000). In contrast, Kanai *et al.* (2005) demonstrated that a cue to the spatial location of a subsequent pure-tone target improved response times more markedly than a frequency cue for the target. This would suggest that, under certain conditions, spatial cues may be weighted more than frequency cues.

Other studies have argued that location and pitch (and timbre and loudness) are not independent features and listeners follow a combination of cues rather than concentrate on a single feature (Mondor *et al.*, 1998; Hawley *et al.*, 2004). Noble and Perrett (2002) suggested that the reliance on spatial location cues varies depending on the strength of other cues. Shinn-Cunningham and Ihlefeld (2004) showed that, even where frequencies did not overlap between a target and maskers, spatial separation improved performance. Again,

this suggests the importance of both pitch and location in understanding a target in the presence of masker talkers. Culling and Summerfield (1995), using short vowel sounds, showed that interaural delays permitted the grouping of less intense sounds independent of spatial location, arguing that localization of auditory sources is only carried out after the sources have been identified, or possibly that location becomes important over time. That F0 is a strong cue is indicated by Darwin and Hukin (1999) using speech, who showed that a 27% change in F0 frequency improved performance in collocated conditions by 14%. Further increases in F0 differences did not result in additional improvements in performance. Similarly, Darwin *et al.* (2003) demonstrated increasing release from masking with increasing differences in F0 values (between 1 and 12 semitones) between two simultaneous talkers.

The CRM mixed-sex corpus has an F0 standard deviation of 44.1 Hz (Table I). Many of the trials in Experiment 1 would have had significant differences in F0 between maskers and target. This would give strong F0 cues for identification and streaming of maskers and target. In the same-sex case (Experiment 3), the maximum standard deviation of F0 is 15.8 Hz; therefore in most trials it is likely that there would only be small differences in F0 between the maskers and target. This could account for the differences in unmasking in the *start-separated* condition between the mixed and same-sex corpora, though the differences in unmasking in the *start-separated* condition are not significant. In absolute terms there is a greater dispersion in the F0 of the male (15.8 Hz) compared to the female corpora (13.4) (Table I). Thus on a trial to trial basis there is likely to be greater differences between the F0s of target and the maskers for the male corpus than the female corpus. This may have provided the listener with perceptually greater differences in voice characteristics for the male-gender corpus, and thus may explain why the unmasking for the male corpus was significant but for the female corpus it was not.

The amount of unmasking provided by initial separation varied with the masker and target voice differences (Experiments 1 and 3). As discussed earlier, differences in unmasking between the *separated* and *start-separated* conditions could also be due to a number of factors. Furthermore, the level of intersubject variability found in this study also suggests that there may be differences in individual listening strategies in terms of the weighting of different relevant cues. These could include different capacities to capitalize on better ear listening, the possibility of active suppression of masker locations or differences in the ability to support streaming based on location cues. The current data do not permit us to discriminate between these possibilities, though it is possible that all play a role. However, taken together the current results provide support for the hypothesis that the proportion of reliance on voice or location cues depends on the relative robustness of these different cues.

VII. CONCLUSION

Initial spatial separation facilitates the identification of the target and masker voices while the subsequent collocation of target and maskers prior to the key words ensured that a continued streaming advantage due to spatial attributes was impossible. The identification of the target voice could facilitate a subsequent focus of attention to identifying speech cues such as pitch, gender, and accent that continue to support streaming in the absence of location cues. That the unmasking was significantly lower than unmasking with the continuously separated sound sources also indicates that spatial location plays a key role for continued streaming.

A two-stage process may underlie the understanding of speech in a noisy environment: Different sound sources are first grouped, and then streamed (see for instance Shinn-Cunningham, 2005). Improved speech intelligibility due to spatial unmasking may result from either increased ease of grouping, or identification of the target; and is probably also partly due to spatial cues which assist streaming of the target against separated maskers.

In the case where frequency differences between target and masker are sufficient for sustained streaming of the targets from the masker, initial separation facilitates target identification. Frequency differences and spatial cues both support streaming of the target. Where frequency cues are ambiguous, as in a same-sex corpus, these data suggest that spatial cues are assigned a greater perceptual weight throughout the streaming. Thus for similar masker and target voice cues, when target and maskers become collocated, much of the benefit from original identification of the target is lost, because the spatial cues are continuously required for effective streaming.

ACKNOWLEDGMENTS

We thank Karen Froud and Tapasya Wancho for their assistance on the project. Also thanks to Teachers College, Columbia University for the generous loan of equipment and resources. This work was supported by the Australian Research Council.

Alais, D., and Burr, D. (**2004**). "The ventriloquist effect results from near-optimal bimodal integration," Curr. Biol. **14**, 257–262.

Alais, D., and Carlile, S. (**2005**). "Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound," Proc. Natl. Acad. Sci. U.S.A. **102**, 2244–2247.

Arbogast, T., and Kidd, G. J. (**2000**). "Evidence for spatial tuning in informational masking using the probe-signal method," J. Acoust. Soc. Am. **108**, 1803–1810.

Benoit, C. Mohamadi, T., and Kandel, S. (**1994**). "Effects of phonetic context on audio-visual intelligibility of French," J. Speech Hear. Res. **37**, 1195–1203.

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

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

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**, 1508–1516.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Brungart, D. S., and Simpson, B. D. (**2002**). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," J. Acoust. Soc. Am. **112**, 664–676.
Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.
Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.
Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.
Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**, 2913–2922.
Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.
Darwin, C. J., and Hukin, R. W. (**2000**). "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," J. Acoust. Soc. Am. **107**, 970–977.
Davison, A. C., and Hinkley, D. V. (**1997**). *Bootstrap Methods and Their Application* (Cambridge University Press, Cambridge).
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**, 3069–3078.
Efron, B., and Tibshirani, R. (**1993**). *An Introduction to the Bootstrap* (Chapman-Hall, New York).
Ericson, M. A., Brungart, D. S., and Simpson, B. D. (**2004**). "Factors that influence intelligibility in multitalker speech displays," Int. J. Aviat. Psychol. **14**, 313–334.
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.
Freyman, R., Balakrishnan, U., and Heifer, K. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.
Freyman, R., Helfer, K., McCall, D., and Clifton, R. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.
Grant, K. W. (**2001**). "The effect of speech reading on masked detection thresholds for filtered speech," J. Acoust. Soc. Am. **109**, 2272–2275.
Grant, K. W., and Seitz, P. F. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197–1208.
Harding, S., and Meyer, G. (**2001**). "A case for multi-resolution and auditory scene analysis," in *Eurospeech* (Aalborg, Denmark), Vol. **1**, pp. 159–162.
Hawley, M., Litovsky, R., and Culling, J. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.
Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.
Kanai, K., Ikeda, K., and Tayama, T. (**2007**). "The effect of exogenous spatial attention on auditory information processing," Psychol. Res. **71**, 418–426.
Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.
Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.
Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory pattems," J. Acoust. Soc. Am. **104**, 422–431.
Lie, E., and Branch Coslett, H. (**2006**). "The effect of gaze direction on sound localization in brain-injured and normal adults," Exp. Brain Res. **168**, 322–336.
Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.
Mondor, T., Zatorre, R., and Terrio, N. (**1998**). "Constraints on the selection of auditory information," J. Exp. Psychol. Hum. Percept. Perform. **24**, 66–79.
Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and Vatikiotis-Bateson, E. (**2004**). "Visual prosody and speech intelligibility: Head movement improves auditory speech perception," Psychol. Sci. **15**, 133–137.
Noble, W., Byrne, D., and Ter-Horst, K. (**1997**). "Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners," J. Acoust. Soc. Am. **102**, 2343–2352.
Noble, W., and Perrett, S. (**2002**). "Hearing speech against spatially separate competing speech versus competing noise," Percept. Psychophys. **64**, 1325–1336.
Schwartz, J. L., Berthommier, F., and Savariaux, C. (**2004**). "Seeing to hear better: Evidence for early audio-visual interactions in speech identification," Cognition **93**, B69–78.
Shinn-Cunningham, B. (**2005**). "Influences of spatial cues on grouping and understanding sound." Forum Acustica, Budapest.
Shinn-Cunningham, B., and Ihlefeld, A. (**2004**). "Selective and divided attention: Extracting information from simultaneous sound sources," in ICAD 04—Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia.
Shinn-Cunningham, B. G., Kopco, N., and Martin, T. J. (**2005**). "Localizing nearby sound sources in a classroom: Binaural room impulse responses," J. Acoust. Soc. Am. **117**, 3100–3115.
Sitler, R. W., Schiavetti, N., and Metz, D. E. (**1983**). "Contextual effects in the measurement of hearing-impaired speakers' intelligibility," J. Speech Hear. Res. **26**, 30–35.
Snyder, J. S., Alain, C., and Picton, C. W. (**2006**). "Effects of attention on neuroelectric correlates of auditory stream segregation," J. Cogn Neurosci. **18**, 1–13.
Watson, A. B. (**1979**). "Probability summation over time," Vision Res. **19**, 515–522.
Wightman, F., Kistler, D., and Brungart, D. (**2006**). "Informational masking of speech in children: Auditory-visual integration," J. Acoust. Soc. Am. **119**, 3940–3949.
Yost, W. A. (**1991**). "Auditory image perception and analysis: The basis for hearing," Hear. Res. **56**, 8–18.

The mid-difference hump in forward-masked intensity discrimination[a)]

Daniel Oberfeld[b)]
Department of Psychology, Johannes Gutenberg–Universität Mainz, 55099 Mainz, Germany

(Received 6 March 2007; revised 3 January 2008; accepted 3 January 2008)

Forward-masked intensity-difference limens (DLs) for pure-tone standards presented at low, medium, and high levels were obtained for a wide range of masker-standard level differences. At a standard level of 25 dB SPL, the masker had a significant effect on intensity resolution, and the data showed a mid-difference hump: The DL elevation was greater at intermediate than at large masker-standard level differences. These results support the hypothesis that the effect of a forward masker on intensity resolution is modulated by the similarity between the masker and the standard. For a given masker-standard level difference, the effect of the masker on the DL was larger for a 55-dB SPL than for the 25-dB SPL standard, providing new support for a midlevel hump. To examine whether the masker-induced DL elevations are related to masker-induced loudness changes [R. P. Carlyon and H. A. Beveridge, J. Acoust. Soc. Am. **93**, 2886–2895 (1993)], the effect of the masker on target loudness was measured for the same listeners. Loudness enhancement followed a mid-difference hump pattern at both the low and the intermediate target level. The correlation between loudness changes and DL elevations was significant, but several aspects of the data are incompatible with the predicted one-on-one relation between the two effects.
 [DOI: 10.1121/1.2837284]

PACS number(s): 43.66.Fe, 43.66.Dc, 43.66.Cb, 43.66.Ba [AJO] Pages: 1571–1581

I. INTRODUCTION

Nonsimultaneous maskers can strongly affect intensity resolution and produce a rather complex pattern of effects (for a review see Plack and Carlyon, 1995). One of the most prominent findings is the *midlevel hump in intensity discrimination* (Zeng *et al.*, 1991): An intense forward masker (e.g., 90 dB SPL) causes a large elevation in the intensity-difference limen (DL) for a midlevel standard (60 dB SPL), relative to the DL in quiet. The 90-dB SPL masker has only a small effect on the DLs for standards presented at low levels (30 dB SPL) or high levels (90 dB SPL), however.

Three explanations have been proposed for the effects of a nonsimultaneous masker on intensity resolution: the *recovery-rate model* (Zeng *et al.*, 1991), the *referential encoding hypothesis* (Plack and Viemeister, 1992b; Carlyon and Beveridge, 1993; Plack *et al.*, 1995), and the *loudness enhancement hypothesis* (Carlyon and Beveridge, 1993). The former two models attribute the midlevel hump to differences in intensity processing at low and high compared to intermediate standard levels, and the research discussed in this paper is concerned with an alternative or complementary explanation.

According to the recovery-rate model (Zeng *et al.*, 1991), the elevation of the DL at midlevels is a consequence of adaptation of the small population of low spontaneous-rate (SR) auditors nerve neurons, which recover slower than the high-SR population (Relkin and Doucet, 1991). It is assumed that in quiet there is a smooth transition between the operating ranges of the two populations, but that 100 ms after the presentation of an intense forward masker, the high-SR fibers have already recovered, while the threshold of the low-SR fibers is still elevated. The resulting midlevel "coding gap" leads to impairment in intensity resolution at intermediate standard levels. Unfortunately, this model based on peripheral mechanisms is incompatible with the DL elevations caused by contralaterally presented forward maskers (Plack *et al.*, 1995; Zeng and Shannon, 1995; Schlauch *et al.*, 1999), and backward maskers (Plack and Viemeister, 1992b; Plack *et al.*, 1995).

The referential encoding hypothesis (Plack and Viemeister, 1992b; Carlyon and Beveridge, 1993; Plack *et al.*, 1995) assumes that the masker interpolated between the two target tones in a two-interval (2I) intensity-discrimination task degrades the memory trace (Durlach and Braida, 1969) for the target tone presented in the first observation interval (see also Mori and Ward, 1992). Consequently, the listener uses the "context-coding mode" (Durlach and Braida, 1969), in which a temporally stable representation of target intensity is based on a comparison with internal or external references. Referential encoding is assumed to work efficiently at low standard levels, where the internal coding reference detection threshold is available (Carlyon and Beveridge, 1993). At high standard levels, the discomfort level or the level of the intense forward masker may be used as a reference (Braida *et al.*, 1984; Carlyon and Beveridge, 1993). At intermediate standard levels, however, the perceptual distance to these references is large, and discrimination performance will thus

[a)]Portions of the data, which were part of the author's doctoral dissertation, were presented at the 29th Jahrestagung für Akustik – DAGA '03 (Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 29th Jahrestagung für Akustik DAGA '03, Aachen, Germany, March 2003), and at the joint congress CFA/DAGA (Proceedings of the Joint Congress CFA/DAGA '04, Strasbourg, France, March 2004).

[b)]Electronic mail: oberfeld@uni-mainz.de

be inferior (Braida *et al.*, 1984). Therefore, the referential encoding hypothesis can account for the midlevel hump caused by both forward and backward maskers. It can also explain the reduction in the DLs at midlevels found if a notched noise is presented simultaneously with the standard (Plack and Viemeister, 1992a) by assuming that the noise is used as a within-interval coding reference.

At this point however, it is necessary to discuss a methodological problem in previous experiments studying the midlevel hump. Generally, a fixed-level, intense masker was combined with various standard levels. Thus, the *masker-standard level difference* and the standard level were correlated. For a low-level standard, the level difference was always larger than for a medium-level standard. Therefore, the different DL elevations caused by an intense forward masker at different standard levels could in principle be due to the variation in the masker-standard level difference rather than to the variation in standard level. The effect of a forward masker on intensity resolution has been shown to depend on the perceptual similarity between masker and standard (cf. Schlauch *et al.*, 1997, 1999). An alternative explanation for the midlevel hump could thus be that for, e.g., a 30-dB SPL standard combined with a 90-dB SPL masker, the perceptual distance between the masker loudness and the standard loudness is so large that the masker has only a small effect on intensity resolution, while a 60-dB SPL standard and the same masker are sufficiently similar in loudness for the masker to have a significant effect.

The third explanation proposed for the midlevel hump, the loudness enhancement hypothesis (Carlyon and Beveridge, 1993), provides a basis for explaining the influence of the masker-standard similarity. The model assumes that the effect of a forward masker on intensity resolution is related to the effect of the masker on the loudness of the target tones. This idea was based on the observation that *loudness enhancement*, that is, an increase in the loudness of a proximal target caused by a forward or backward masker (e.g., Galambos *et al.*, 1972; Zwislocki and Sokolich, 1974; Elmasian *et al.*, 1980; Oberfeld, 2007), and the masker-induced DL elevation depend in a similar manner on various stimulus parameters. For instance, both loudness enhancement and the DL elevation caused by an intense forward masker are most pronounced at intermediate standard levels (Zeng, 1994; Plack, 1996a), and the two phenomena show a similar dependence on the inter-stimulus interval (ISI) between the masker and the target (Zwislocki and Sokolich, 1974; Zeng and Turner, 1992), on the relation between masker frequency and target frequency (Zwislocki and Sokolich, 1974; Zeng and Turner, 1992), and on the laterality of the masker relative to the target (Elmasian *et al.*, 1980; Plack *et al.*, 1995). Carlyon and Beveridge (1993) suggested that a forward masker impairs intensity resolution because loudness enhancement introduces variability in the loudness of the target tones. Data by Zeng (1994) supported their hypothesis. He presented a 90-dB SPL forward masker and reported loudness enhancement, loudness variability (estimated in an adaptive matching procedure; cf. Jesteadt, 1980; Schlauch and Wier, 1987), and intensity DLs for the same listeners. The effect of the masker on all of the three measures was maximal at intermediate levels (midlevel hump). Plack (1996a) reported that with a 90-dB SPL forward or backward masker, loudness enhancement and loudness variabilitywere significantly correlated for three of the four listeners (Spearman rank correlation coefficient r_s=0.5–0.76).

Oberfeld (2007) proposed an explanation for loudness enhancement which is based on an idea by Elmasian *et al.* (1980), who suggested that the loudness representations of the masker and the target are merged automatically. Applied to the three-tone matching task used in most experiments (cf. the lower row of Fig. 1), it follows that the initial value of the target loudness is no longer available at the presentation of the comparison, but that the listener will instead compare a weighted average of the masker loudness and the target loudness with the loudness of the comparison. This explains why the loudness of the target always seems to be shifted toward masker loudness (Zwislocki and Sokolich, 1974; Elmasian *et al.*, 1980). The mergence hypothesis alone cannot account for the midlevel hump in loudness enhancement (Zeng, 1994; Plack, 1996a) because with the masker level fixed at, e.g., 90 dB SPL, loudness enhancement should increase with decreasing target level if a simple weighted average between masker loudness and target loudness was used in the loudness match. This limitation can be overcome, Oberfeld (2007) suggested, by assuming that the effect of the masker depends on the *perceptual similarity* between masker and target, that is, that the masker loudness will receive a smaller weight in the computation of the weighted average if the masker and the target differ strongly in, e.g., spectral content, duration, or loudness. An influence of similarity on the operation of a memory system would not be surprising because effects of the target-distractor similarity are one of the best-established findings in cognitive psychology, for example in experiments studying visual search (e.g., Duncan and Humphreys, 1989), or recognition memory (e.g., Baddeley, 1966). The most prominent evidence for a similarity effect in intensity discrimination is the observation of Schlauch *et al.* (1997, 1999) that adding a 4.133-kHz component to a 1-kHz forward masker strongly reduced the size of the midlevel hump for a 1-kHz standard. Closely related is the reduction in the midlevel DLs if a 10-ms standard was combined with a 250-ms rather than with a 10-ms forward masker (Schlauch *et al.*, 1997).

The *similarity hypothesis* (i.e., loudness enhancement hypothesis extended by the assumption that a reduction in the masker-standard similarity reduces the effect of the masker) predicts that that the effect of a forward masker on intensity resolution is modulated by the similarity between the masker and the standard. To test this prediction, intensity DLs were obtained in Experiment 1 at low, intermediate, and high standard levels (25, 55, and 85 dB SPL) with a wide range of masker-standard level differences (–30 to 60 dB), avoiding the confound between standard level and the masker-standard level difference. According to the similarity hypothesis, it is conceivable that the perceptual distance between the loudness of an 85-dB SPL masker and the loudness of a 25-dB SPL standard is too large for the masker to have a strong effect on intensity resolution. In contrast, a 55-dB SPL masker and the 25-dB SPL standard should be

sufficiently similar in loudness for the masker to cause a significant effect. Therefore, a significant elevation of the DLs for a low level standard was expected at intermediate masker-standard level differences, resulting in a *mid-difference hump* pattern.

In Experiment 2, the effect of the forward masker on target loudness was measured for the same listeners and the same conditions as in Experiment 1, to test for the correlation between masker-induced loudness changes and DL elevations predicted by the loudness enhancement hypothesis.

II. EXPERIMENT 1: INTENSITY-DIFFERENCE LIMENS AS A FUNCTION OF THE MASKER-STANDARD LEVEL DIFFERENCE

Intensity DLs were measured for standards presented at 25, 55, and 85 dB SPL, as a function of the level difference between the masker and the standard ($L_M - L_S$). A two-interval, two-alternative forced-choice (21, 2AFC), adaptive procedure was used. $L_M - L_S$ was varied between −30 and +60 dB in 15-dB steps. The lowest masker level was 10 dB SPL. To keep stimulus levels within safe limits, the highest masker level was 100 dB SPL and the maximum sound pressure level was restricted to 105 dB SPL. DLs in quiet were also obtained.

Unfortunately, due to a programming error, listener AL did not receive the 10-dB SPL masker/25-dB SPL standard and the 25-dB SPL masker / 55-dB SPL standard combination, while listener BS did not receive the 70-dB SPL masker combined with the 85-dB SPL standard.

A. Method

Six volunteers participated in the experiment (three female, three male; age 20–32 years). One of them (DO) was the author, the remaining participants were paid an hourly wage. For the ear tested, all had hearing levels better than 11 dB HL in the frequency range between 125 and 8000 Hz, with one exception (listener AL: HL = 17.6 dB at 8 kHz). All listeners except BS received stimulation to their right ear. For listener BS, the left ear was used because in the right ear, the HL was 19.6 dB at 8000 Hz, while the HLs in the left ear were better than 6.6 dB at all frequencies tested. The listeners were fully informed about the course of the experiment. All except the author were naïve with respect to the hypotheses under test.

The standard and the masker were 1-kHz pure tones with a steady-state duration of 20 ms, gated on and off with 5-ms $\cos^2$ ramps. As the upper row in Fig. 1 shows, there were two observation intervals. In one of the intervals (selected with an equal *a priori* probability), a level increment was added to the standard. The interval between the offset of the first target tone and the onset of the second target tone was 650 ms. In the forward masking conditions, a masker was presented in both intervals. The silent interval between masker offset and standard onset was 100 ms. The target tones were marked by visual signals.

The stimuli were generated digitally and played back via an M-Audio Delta 44 PCI audio-card (sampling rate 44.1 kHz, 24-bit resolution). One channel was used for the masker, a separate channel for the standard/standard-plus-increment. The increment was produced digitally. The masker and the standard/standard-plus-increment were fed into two separate channels of a custom-made programmable attenuator, summed in an inverting summing amplifier, amplified by a headphone amplifier, and fed into one channel of Sennheiser HDA 200 headphones calibrated according to IEC 318 (1970). The experiment was conducted in a single-walled sound-insulated chamber. Listeners were tested individually.

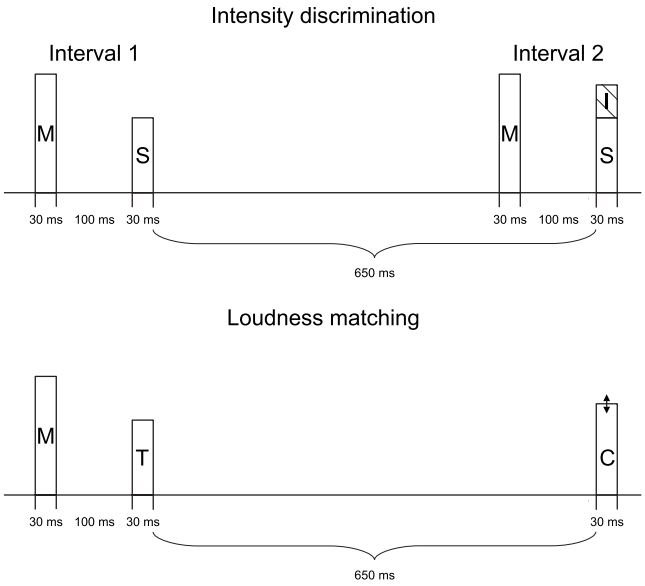


FIG. 1. Upper row: Trial configuration used in the 2I, 2AFC intensity discrimination task in Experiment 1. In each of the two observation intervals, a forward masker *M* and the standard *S* were presented. All stimuli were 30-ms, 1-kHz tone bursts. The silent interval between masker and standard was 100 ms. The level increment *I* was presented in interval 1 or interval 2 with an identical *a priori* probability. Listeners responded whether the increment (i.e., the louder target tone) had occurred in the first or in the second interval. The level of the increment was adjusted by an adaptive procedure with a two-down, one-up rule. Lower row: Trial configuration used in the loudness matching procedure in Experiment 2. Listeners responded whether the target (*T*) or the comparison (*C*) had been louder. The level of the target was fixed. The level of the comparison was adjusted by an interleaved-staircase, adaptive procedure (Jesteadt, 1980).

A 2I, 2AFC, adaptive procedure with a two-down, one-up tracking rule (Levitt, 1971) was used to measure intensity DLs corresponding to 70.7% correct. The listeners selected the interval containing the louder target tone. After two consecutive correct responses, the increment was reduced. After each incorrect response, the increment was increased. Up to the fourth reversal, the step size was 5 dB. For the remaining eight reversals, the step size was 2 dB. The difference limen ΔL_{DL} was computed as the arithmetic mean of $10 \log_{10}(1 + \Delta I/I)$ at the final eight reversals. A track was discarded if the standard deviation was greater than 5 dB. At least three runs were obtained for each data point. Time permitting, additional tracks were run if the standard deviation of the DLs measured in the first three runs exceeded 5 dB, so that each data point is based on three to nine runs. Visual false/correct feedback was provided after each trial. The experiment was self-paced.

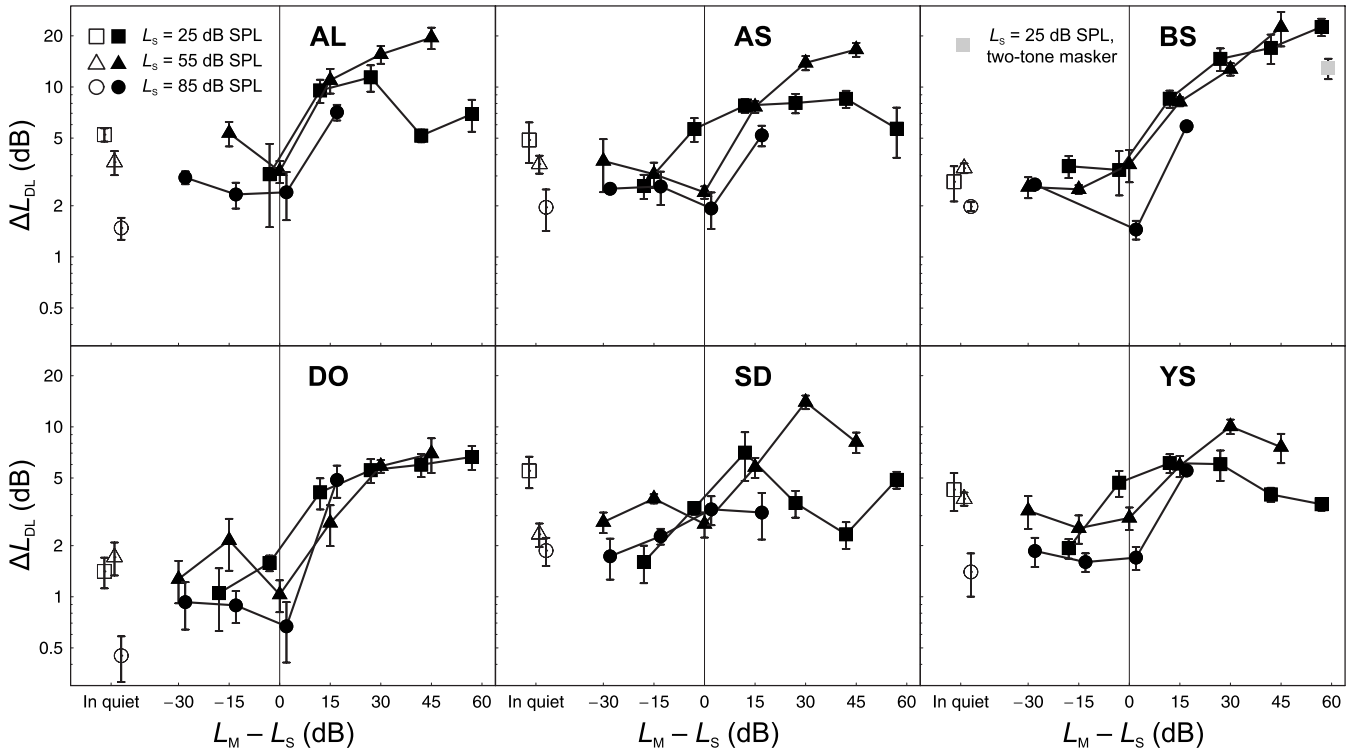


FIG. 2. Experiment 1. Individual intensity-difference limens $[\Delta L_{DL}=10\log_{10}(1+\Delta I_{DL}/ID)]$ for 1-kHz, 30-ms tones as a function of the masker-standard level difference L_M-L_S, and standard level L_S. Panels represent listeners. Open symbols: in quiet. Closed symbols: in forward masking. Squares: 25-dB SPL standard. Triangles: 55-dB SPL standard. Circles: 85-dB SPL standard. Filled gray square: Two-tone masker presented to listener BS in the 85-dB SPL masker/25-dB SPL standard condition. Lines are shifted by 2 dB on the *x* axis. Error bars show plus and minus one standard error of the mean (SEM) for the three or more measurements obtained for each data point.

The listeners were instructed to ignore the maskers. In each block, only one masker-standard level combination was presented. The conditions occurred in pseudorandom order, with the exception that blocks presenting the 100-dB SPL masker were always run at the end of a session.

Listeners received at least 2 h of practice. If necessary, further practice was allowed until performance stabilized. A testing session lasted approximately 1 h with one or two short breaks.

B. Detection thresholds

Detection thresholds were obtained for 30-ms (including 5-ms $\cos^2$ ramps), 1-kHz tones in quiet and under forward masking. A 2I, 2AFC, adaptive procedure (two-down, one-up rule; Levitt, 1971) was used. In one interval (selected randomly), the signal was presented, while no tone was presented in the other interval. In the forward masking conditions, a masker was presented in both intervals. The silent interval between masker offset and signal onset was 100 ms. Masker levels were 25, 55, 85, and 100 dB SPL, except for listener BS, who did not receive the 100-dB SPL masker.

Initially, the signal level was 20 dB SPL. Step size was 5 dB until the fourth reversal, and 2 dB for the remaining eight reversals. Visual trial-by-trial feedback was provided. The threshold level was computed as the arithmetic mean of the signal levels at the last eight reversals. For each condition, at least three measurements were obtained. If the standard deviation was larger than 5 dB within a track, the track was rerun. Time permitting, additional tracks were obtained if the standard deviation of the three threshold estimates exceeded 5 dB.

Thresholds in quiet ranged between 4.5 and 9.2 dB SPL [M=7.0 dB SPL standard deviation (SD) = 1.8 dB]. Thresholds in forward masking were virtually identical to the threshold in quiet for listener BS and elevated by only about 2.5 dB for DO. This finding is compatible with results by Zeng *et al.* (1991). For the remaining listeners, the masker caused an elevation of the detection thresholds that tended to increase with the masker level. The maximum individual elevation of 9.5 dB is comparable to data by Carlyon and Beveridge (1993). Mean thresholds (with *SDs* in parentheses) with the 25-, 55-, and 85-dB SPL masker were 7.9 dB SPL (1.7 dB), 9.9 dB SPL (2.9 dB), and 11.2 dB SPL (4.3 dB), respectively. Five listeners were tested with a 100-dB SPL masker, mean threshold in this condition was 12.9 dB SPL (SD=2.2 dB).

C. Results and discussion

Individual results from the intensity discrimination experiment are displayed in Fig. 2, where ΔL_{DL} is plotted on a logarithmic axis because this measure is compressive at small values. At the 25-dB SPL standard level (squares in Fig. 2), ΔL_{DL} was largest at intermediate masker-standard level differences of 15–45 dB for listeners AL, AS, SD, and YS, resulting in a *mid-difference hump* compatible with the predictions of the similarity hypothesis. The remaining two

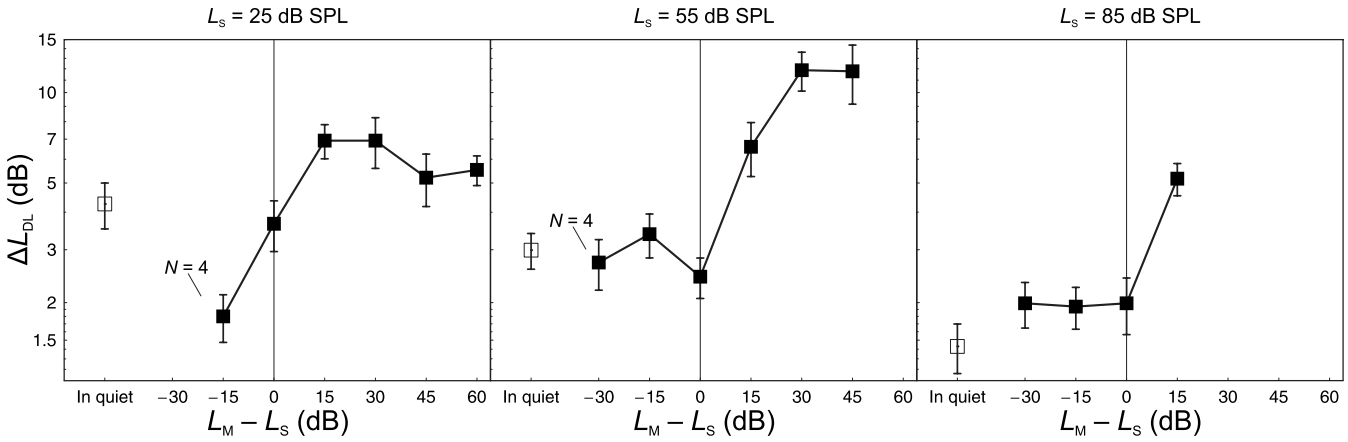


FIG. 3. Experiment 1. Mean intensity-difference limens for five of the six listeners, as a function of the masker-standard level difference $L_M - L_S$, and standard level L_S. The data from listener BS, whose forward-masked DLs for the 25-dB SPL standard strongly deviated from the DLs of the remaining listeners, were excluded. Panels represent standard levels. Open symbols: in quiet. Closed symbols: in forward masking. Error bars show ±1 SEM of the individual values. For the two data points with missing values (see the text), the labels show the number of listeners contributing to the respective data point.

listeners deviated from this pattern. For listener DO (the author), the DL showed a rather small monotonic increase with the masker-standard level difference. For listener BS, the DL increased dramatically with $L_M - LS$. His data are in some aspects similar to the pattern Schlauch *et al.* (1997) reported for one subject in their study, for whom the DLs were even larger (up to 50 dB) for low-level standards combined with a 90-dB SPL masker. Schlauch *et al.* hypothesized that the listener integrated the intensities of the masker and the standard, presumably due to the perceptual similarity between the two tones. Large DLs for a low-level standard combined with an intense masker were also reported by Zeng *et al.* (1991; listener RB), Carlyon and Beveridge (1993; listener LW), and Schlauch *et al.* (1999; listener 2). Schlauch *et al.* (1997) reported that adding a 4.133-kHz component to the masker greatly reduced the DLs in the critical conditions, presumably because the additional component served as a cue helping the listener to differentiate between the masker and the standard. To examine whether the large DLs produced by BS could also be attributed to the perceptual similarity between the masker and the standard, he was tested with a two-tone masker in the 85-dB SPL masker/25-dB SPL standard condition. The two-tone masker consisted of a 1- and a 4.133-kHz tone burst sharing the same temporal envelope. The sound pressure level of the two components was identical so that the overall level was 3 dB higher than the level of the 1-kHz tone. As the gray square in Fig. 2 shows, the two-tone masker resulted in a substantially lower DL than the 1-kHz masker, although the DL was still 10 dB higher than in quiet. Because the data for listener BS deviated strongly from the DLs for the other listeners, his data were excluded from all following analyses.

Mean data for all listeners except BS are displayed in Fig. 3. On average, maskers 15 or 30 dB higher in level than the 25-dB SPL standard caused the largest DLs. Repeated-measures analyses of variance (ANOVAs) were conducted using a maximum-likelihood approach (SAS PROC MIXED; Littell *et al.*, 1996) because there were two missing data points due to a programming error. The Satterthwaite method was used for computing the denominator degrees of freedom for approximate F tests of fixed effects. The "heterogeneous compound symmetry" (CSH) was selected to model the covariance structure. An ANOVA for the data obtained at the 25-dB SPL standard level showed a significant effect of $L_M - L_S [F(6,7.45)=9.71, p=0.0034]$, providing evidence against the recovery-rate model, which predicts the forward masker to have no effect at low standard levels. Paired-samples t-tests indicated that the DL in forward masking was significantly larger than the DL in quiet at the +15-dB masker-standard level difference $[t(4)=5.51, p=0.003$, one-tailed], and marginally significantly elevated at $L_M - L_S = 30$ dB $[t(4)=1.96, p=0.06$, one-tailed]. At the two highest masker-standard level differences, there was no significant DL elevation. Additional evidence for the expected nonmonotonic relation between the DL elevation and the masker-standard level difference was provided by a post-hoc contrast which showed that the DLs obtained at the two intermediate level differences (15 and 30 dB) were marginally significantly higher than the DLs at the two largest level differences (45 and 60 dB) $[F(1,12.8)=3.64, p=0.079$, two-tailed].

For the 55-dB SPL standard (triangles in Fig. 2), the DLs at the largest masker-standard level difference (45 dB) were smaller than at the 30-dB level difference for two listeners (SD and YS), resulting in a mid-difference hump. The remaining listeners produced no mid-difference hump at this standard level. It remains unclear whether for them the maximum masker-standard level difference of 45 dB was not large enough for the DL to decrease again. Zeng and Turner (1992) also reported a monotonic increase of the DLs for a midlevel standard as the masker-standard level difference was increased from 0 to 50 dB. In the mean data (center panel in Fig. 3), the DL showed no further increase, but also no decrease, as $L_M - L_S$ was increased from 30 to 45 dB. A one-factorial ANOVA conducted for the data at the 55-dB SPL standard level indicated a significant effect of $L_M - L_S [F(6,6.76)=7.16, p=0.011]$. Paired-samples t-tests showed that all maskers higher in level than the standard had caused a significant DL elevation $[p<0.05$ (one-tailed)].

On average, the effects of the forward maskers higher in level than the standard were more pronounced at the intermediate than at the low standard level. The DL elevations (i.e., the DL in forward masking minus the DL in quiet) were analyzed via an ANOVA with the within-subjects factors L_S (25, 55 dB SPL) and $L_M - L_S$ (15, 30, and 45 dB). The significant effect of L_S [$F(1,12.7)=26.59, p=0.001$] confirmed the observation of larger DL elevations at the intermediate standard level. The $L_S \times (L_M - L_S)$ interaction was also significant [$F(2,12.8)=7.65, p=0.007$]. These results indicate that forward masking has the strongest effect at intermediate standard levels, even if the masker-standard level difference is not correlated with the standard level as in previous experiments.

Except for listener SD, the 100-dB SPL masker caused an elevation in the DL for the 85-dB SPL standard (circles in Fig. 2). An ANOVA conducted at this standard level indicated a significant effect of $L_M - L_S$ [$F(4,4.13)=7.23, p=0.038$]. The DL obtained with the 100-dB SPL masker was significantly larger than the DL in quiet [$t(4)=5.15$, $p=0.004$, one-tailed]. This effect is not compatible with the recovery-rate model, as a 100-dB SPL masker cannot be expected to shift the threshold of the low-SR fibers to values even near 85 dB SPL. In terms of the referential encoding hypothesis, it seems likely that the DL elevation at the 15-dB level difference should be smaller for an 85-dB SPL standard than for a 55-dB SPL standard, as the distance to the internal reference discomfort level is smaller at high intensities. However, the DL elevation was virtually identical at the two standard levels (L_S=85 dB SPL: M=3.74 dB, SD=1.62 dB; L_S=55 dB SPL: M=3.62 dB, SD=2.41 dB; $t(4)=0.109$, n.s.).

For masker levels lower than or equal to the standard level, the DLs were generally close to those in quiet. For three listeners, the 10-dB SPL masker even caused a reduction in the DL for the 25-dB SPL standard. A potential explanation for this finding is a cueing effect (Moore and Glasberg, 1982).

For comparison with previous studies where a fixed-level, intense masker was used and only the level of the standard was varied, consider the data obtained with the 85-dB SPL masker (data points at $L_M - L_S$=60, 30, and 0 dB in the left, center, and right panel of Fig. 3, respectively). The DL elevation at the intermediate standard level was approximately 9 dB, which is comparable to the value of about 13 dB reported by Plack *et al.* (1995).

To summarize the results, the significant DL elevations observed for 25 and 85-dB SPL standards are incompatible with the recovery rate model.

The mid-difference hump observed at the low standard level is directly compatible with the similarity hypothesis. In terms of the referential encoding hypothesis, the effects of the masker-standard level difference are somewhat difficult to predict as the masker can be assumed to have two effects. First, the amount of trace degradation caused by the masker could be a function of the masker level, although the model contains no explicit assumptions concerning this parameter. It seems reasonable, however, to assume that for example a 60-dB SPL masker should have a stronger detrimental effect on the trace of a 30-dB SPL standard than on the trace of a 90-dB SPL standard. Second, the masker could serve as a within-interval coding reference (Plack, 1996b). Taken together, there should be virtually no trace degradation at masker levels well below standard level. If the masker level is increased to values above the standard level, the trace degradation increases and the system has to rely increasingly on context coding. At the same time, the effectiveness of the masker as a coding reference is reduced according to the perceptual anchor model (Braida *et al.*, 1984). A slightly different prediction can be derived from data by Plack (1996b, 1998), who suggested that the level of the standard is compared to the level of the decaying "temporal excitation pattern" of the masker at the time of occurrence of the standard, so that the optimum level of the masker for referential encoding purposes would be somewhat higher than the standard level. Both variants suggest a maximal effect of the masker at larger masker-standard level differences, but similarity effects are not in principle incompatible with the referential encoding hypothesis. For example, as a reviewer noted, if the masker is grouped into a separate perceptual stream by pairing it with a higher tone (Schlauch *et al.*, 1997), it might cause less trace degradation. Consequently, the assumption that a perceptual difference between masker and standard results in less memory trace interference could be integrated into the referential encoding hypothesis, so that the mid-difference hump could be accounted for.

The data also indicate that there is a midlevel hump not only in the sense that with a fixed-level, intense forward masker the largest DL elevation occurs at medium standard levels, but also in the sense that for a given masker-standard level difference, the effect of the masker on the DL is larger for a medium-level than for a low-level standard. Thus, the masker-standard level difference does not completely determine the effect of the masker. This finding is compatible with the referential encoding hypothesis. On the other hand, at a 15-dB level difference between masker and standard, the DL elevation was not larger for the 55- than for the 85-dB SPL standard, contrary to the predictions of the latter model. It seems possible to integrate the influence of the standard level into the similarity hypothesis by assuming that the effect is due to the compressive behavior of the cochlea that is more pronounced at intermediate than at low levels (for a recent review see Oxenham and Bacon, 2003), and which results in a steeper slope of the loudness function at low levels (e.g., Hellman and Zwislocki, 1964; Yates, 1990). Therefore, it could be argued that the masker-standard level difference at which the difference between the loudness of the masker and the loudness of the target becomes large enough for the effect of the masker to decrease again is smaller at low than at intermediate standard levels. In this line of reasoning, probably the maximum masker-standard level difference of 45 dB presented at L_s=55 dB SPL was not large enough for the DL to decrease again, which would explain the absence of a mid-difference hump at this standard level for most listeners.

III. EXPERIMENT 2: LOUDNESS MATCHES AS A FUNCTION OF THE MASKER-TARGET LEVEL DIFFERENCE

A. Rationale

The data from Experiment 1 were in large part compatible with the explanation for loudness enhancement proposed by Oberfeld (2007) combined with the loudness enhancement hypothesis (Carlyon and Beveridge, 1993). If this model is valid, then the masker-target level difference should have a similar effect on target loudness as on the intensity DL: The loudness change caused by the masker should be maximal at intermediate level differences, resulting in a mid-difference hump. Results compatible with the predicted pattern were previously reported by Oberfeld (2007). Additionally, there should be a correlation between the masker-induced loudness change and the masker-induced DL elevation. Experiment 2 was designed to test these hypotheses. A loudness matching task was used to measure the effect of the forward masker on the loudness of a proximal target, for the same listeners as in Experiment 1. The temporal configuration of a trial is displayed in the lower row of Fig. 1. Except for the omission of the masker in the second interval, the temporal structure was the same as in Experiment 1 (upper row in Fig. 1), and the same stimuli and masker-target level combinations were presented.

Note that recent data by Scharf *et al.* (2002), Arieh and Marks (2003), and Oberfeld (2007) suggest that a masker has two effects on a target following it by less than about 400 ms. First, the masker causes a shift in the target loudness toward the masker loudness. At the same time, if the masker level is higher than the target level, the masker causes a reduction in target loudness (loudness recalibration; Marks, 1994). If the comparison is presented at the same frequency as the masker, the masker also induces loudness recalibration in the comparison if the comparison level is lower than the masker level (Scharf *et al.*, 2002; Oberfeld, 2007). According to this two-process model (Arieh and Marks (2003); Oberfeld, 2007), the effect of the masker on the loudness level of the target in a three-tone matching procedure with all tones sharing the same frequency is an estimate of the effect of the process causing loudness enhancement or loudness decrement, because the effect of the process resulting in loudness recalibration is assumed to be constant for at least several seconds following the presentation of the masker (Arieh and Marks, 2003; Oberfeld, 2007), so that its effect on the target loudness and its effect on the comparison loudness cancel.

B. Method

The same listeners as in Experiment 1 took part, except for listener AS, who chose not to participate due to lack of time. For listener BS, the 25-dB SPL target combined with the 85-dB SPL masker again presented a problem. The listener requested comparison levels exceeding the maximum level difference between comparison and target the test equipment could deliver (45 dB), unlike the remaining listeners, for whom loudness enhancement was rather small in this situation. As BS could not be tested in all conditions, his data were excluded from the analyses.

The same stimuli and apparatus as in Experiment 1 were used. Due to a programming error, the 10-dB SPL masker/25-dB SPL target combination was not presented to listener SD. As can be seen in the lower row of Fig. 1, the target was followed by a comparison after a silent interval of 650 ms. The level of the target was fixed. The level of the comparison was adjusted by an adaptive procedure. In forward-masked trials, the silent interval between masker offset and target onset was 100 ms. Listeners responded whether the target or the comparison had been louder. They were instructed to ignore the masker. No feedback was provided.

A 2I, 2AFC, interleaved-staircase procedure (Jesteadt, 1980) was used. Each block comprised two randomly interleaved tracks. The upper track converged on the comparison level corresponding to the 70.7% *Comparison louder* point on the psychometric function. If the listener indicated on two consecutive trials that he or she had perceived the comparison as being louder than the target, the level of the comparison was reduced. After each response indicating that the target had been perceived as being louder, the level of the comparison was increased. In the lower track, a one-down, two-up rule was used to track the 29.3% *Comparison louder* point on the psychometric function. The upper track and the lower track started with a comparison level 11 dB above or below the target level, respectively. The step size was 5 dB until the fourth reversal, and 2 dB for the remaining eight reversals. If in one of the tracks 12 reversals had already occurred before the other track had also reached 12 reversals, trials from the former track were still presented with an *a priori* probability of 0.2.

For each block, the arithmetic mean of the level differences between comparison and target ($L_C - L_T$) at all but the first four reversals was computed separately for the upper and for the lower track, with the restriction that for each track an even number of reversals entered the computation (e.g., if 13 reversals had occurred in one of the tracks, reversal 13 was excluded). The arithmetic mean of these two values was taken as the loudness match, corresponding to the comparison level at the point of subjective equality (PSE) minus the target level. A run was discarded if the standard deviation of $L_C - L_T$ at the counting reversals was greater than 5 dB in either the upper or the lower track. Three runs were obtained in each condition. Time permitting, additional runs were presented if the standard deviation of the loudness matches exceeded 5 dB.

Only one masker-target level combination was presented in each block. Listeners received the conditions in pseudo-random order with the exception that the 100-dB SPL masker was always presented at the end of a session.

C. Results and discussion

Individual loudness matches are displayed in Fig. 4 in terms of the level difference $L_C - L_T$ required to make the comparison sound equally loud as the target. Positive values of $L_C - L_T$ correspond to loudness enhancement of the target,

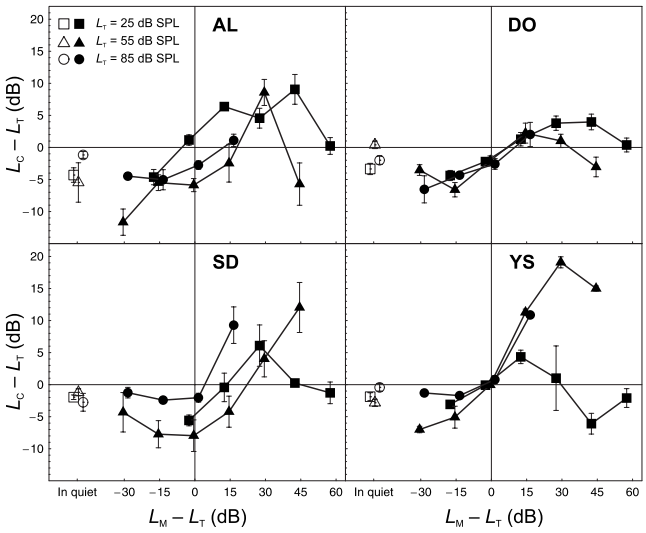


FIG. 4. Experiment 2. Individual loudness matches (comparison level L_C at the PSE minus target level L_T) as a function of the masker-target level difference ($L_M - L_T$) and target level. Same format as Fig. 2.

negative values to loudness decrement. At a target level of 25 dB SPL (squares in Fig. 4), loudness enhancement was a nonmonotonic function of the masker-target level difference for all listeners, consistent with the predicted mid-difference hump. The maximum change in loudness level relative to the match in quiet was found at values of $L_M - L_T$ between 15 and 45 dB, with individual maxima of 6.2–13.4 dB. The 10-dB SPL masker caused loudness decrement. Mean results are shown in the left panel of Fig. 5. On average, loudness enhancement was maximal at intermediate masker-target level differences. A one-factorial repeated-measures ANOVA using a univariate approach with a Huynh–Feldt correction for the degrees of freedom was conducted for the data obtained at the 25-dB SPL target level because PROC MIXED did not converge in this case. The level difference of −15 dB was excluded because of one missing value. There was a marginally significant effect of $L_M - L_T$ [$F(5,15)=2.92, p=0.051$]. Post-hoc paired-samples t-tests indicated a significant difference between the loudness match under forward masking and the baseline match (i.e., the match in quiet) at $L_M - L_T = 30$ dB [$t(3)=6.06$, $p=0.015$ (two-tailed)], and marginally significant differences at the −15- and the +15-dB masker-target level difference [$t(3)=3.17$, $p=0.087$ (two-tailed), and $t(3)=3.02$, $p=0.057$ (two-tailed), respectively]. The observation of substantial loudness enhancement for a 25-dB SPL target (on average 6.7 dB relative to the baseline match at a $L_M - L_S = 30$ dB) indicates that the absence of loudness enhancement at low target levels reported in previous studies presenting only an intense masker (Zeng, 1994; Plack, 1996a) cannot be attributed to the low target level, but rather to the large masker-target loudness difference. At the two largest values of $L_M - L_T$, the loudness matches did not differ significantly from the baseline match ($p > 0.15$). The effect of the masker on the loudness match was significantly smaller at the largest masker-target level difference (60 dB) than at the intermediate difference of 30 dB [$t(3)=4.67$, $p=0.019$ (two-tailed)]. This pattern of results is compatible with the mid-difference hump predicted by the similarity hypothesis.

Mid-difference humps were also present at the 55-dB SPL target level for all listeners except SD (triangles in Fig. 4). At masker-target level differences between 15 and 30 dB, the maximum amounts of enhancement were observed, which showed considerable interindividual variation and ranged between 1.8 and 21.9 dB relative to the match in quiet. An ANOVA conducted at the 55-dB SPL target level showed a significant effect of $L_M - L_T$ [$F(6,18)=4.61, p=0.005$]. On average (center panel in Fig. 5), loudness enhancement was stronger at a masker-target level difference of 30 rather than 45 dB, but this difference was not significant [$t(3)=0.79$]. The average loudness changes induced by the masker (i.e., the match under masking minus the baseline match) were larger at the 55 than at the 25-dB SPL target level (Fig. 5). Yet, in an L_T (25,55 dB SPL) × $L_M - L_T$ (−15 to 45 dB) ANOVA, neither the effect of L_T [$F(1,7.21)=0.03, p=0.86$], nor the $L_T \times (L_M - L_T)$ interaction was significant [$F(4,6.46)=0.72, p=0.61$].

At the 85-dB SPL target level (circles in Fig. 4), loudness decrement was observed at masker-target level differences of −30 and −15 dB, except for listener SD. The 100-dB SPL masker caused loudness enhancement. An

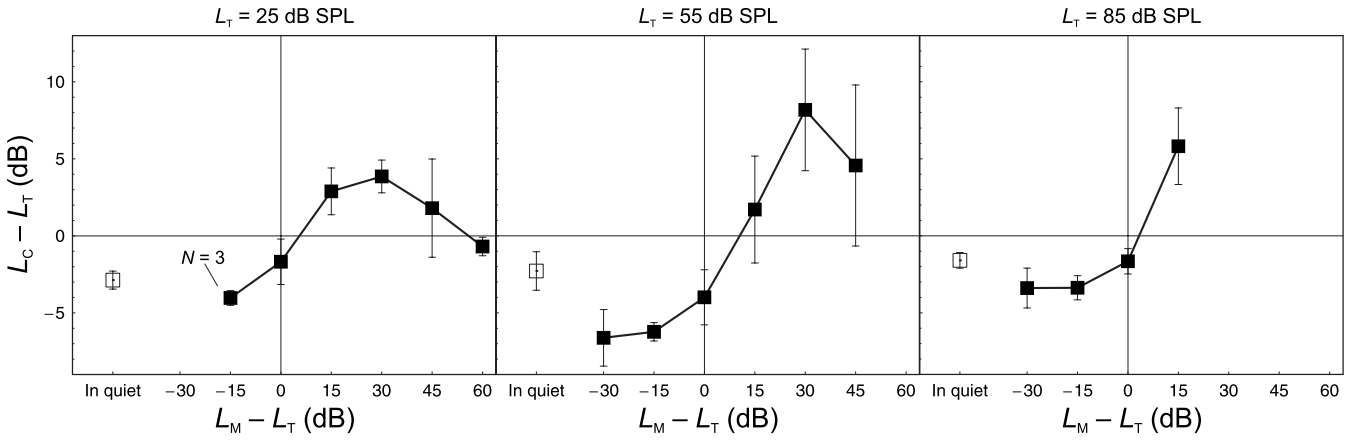


FIG. 5. Experiment 2. Mean loudness matches (comparison level at the PSE minus target level) as a function of the masker-target level difference ($L_M - L_T$) and target level. Same format as Fig. 3.

ANOVA showed that the effect of $L_M - L_T$ was significant at this target level [$F(4, 4.14) = 9.28, p = 0.024$]. The match obtained with the 100-dB SPL masker was marginally significantly higher than the baseline match [$t(3) = 2.98, p = 0.059$].

For a level difference of 0 dB ($L_M = L_T$), the mergence hypothesis (Elmasian *et al.*, 1980) predicts no loudness change relative to the condition in quiet. Compatible with this hypothesis, paired-samples *t*-tests showed that at none of the three target levels did the loudness match at $L_M = L_T$ differ significantly from the match in quiet ($p > 0.4$).

Note that the data obtained with the 85-dB SPL masker followed a midlevel hump pattern (right panel of Fig. 5), compatible with the results by Zeng (1994) and Plack (1996a).

IV. GENERAL DISCUSSION

A. Relation between loudness enhancement and intensity DLs

The loudness enhancement hypothesis (Carlyon and Beveridge, 1993) predicts loudness enhancement and the DL elevation caused by the forward masker to be correlated. Two previous studies reported evidence for such a relation (Zeng, 1994; Plack, 1996a), but only for the case of an intense masker, and consequently only for conditions producing loudness enhancement. What can be expected for the case of maskers lower in level than the target, in which loudness *decrement* is observed? The most parsimonious assumption would be that any change in the target loudness induced by the masker, regardless whether enhancement or reduction, increases the variability of the intensity representation. The data from Experiment 1 do not support this hypothesis, however, as only small DL elevations or in some cases even DL reductions were observed if the masker level was lower than the target level. In an experiment by Zeng and Turner (1992), DLs for standards presented at a level between 40 and 60 dB SPL were also unaffected by maskers lower in level than the standard.

Correlational analyses were used to test the prediction of the loudness enhancement hypothesis. In a first step, the parsimonious assumption that both loudness enhancement and loudness decrement should increase the loudness variability was adopted. Consequently, correlations between the DL elevation and the *absolute value* of the masker-induced change in loudness were computed. For each for the four listeners who had participated in both experiments, and each masker-target level combination, the average DL elevation (i.e., ΔL_{DL} in forward masking minus ΔL_{DL} in quiet; Experiment 1) and the average masker-induced change in loudness level (i.e., L_C at the PSE in the presence of the masker minus L_C at the PSE in quiet; Experiment 2) was computed. The Pearson product-moment correlation coefficient for the relation between the DL elevation and the absolute value of the loudness change was significantly greater than zero [$r = 0.312$, $n = 62$, $p = 0.014$ (two-tailed)], compatible with the loudness enhancement hypothesis. Note that the proportion of variance accounted for by the linear regression was small ($R^2 = 0.097$).

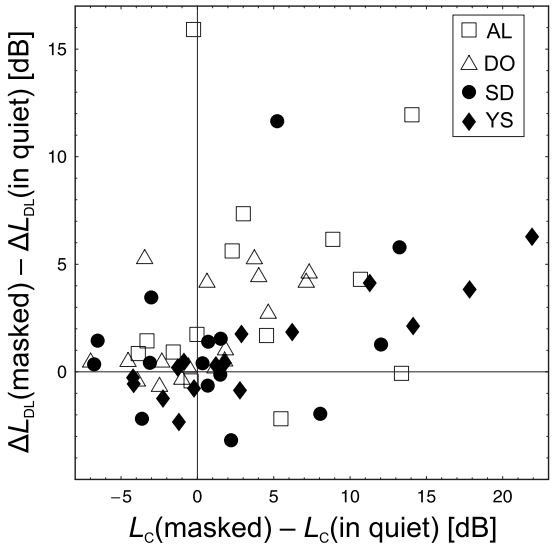


FIG. 6. Scatter plot of the individual masker-induced change in the loudness level of the target measured in Experiment 2 (horizontal axis) and the masker-induced change in ΔL_{DL} measured in Experiment 1 (vertical axis). Each data point represents one listener and one masker-target level combination. Symbols denote listeners.

In a second step, the conditions associated with loudness decrement (i.e., $L_M < L_T$) were excluded. The correlation between the DL elevation and the original rather than the absolute value of the masker-induced loudness change was computed, it was only marginally significant [$r = 0.294$, $n = 44$, $p = 0.053$ (two-tailed)]. The data are displayed in Fig. 6, where it can be seen that loudness decrement was associated with small DL elevations, while the DLs tended to increase with the amount of loudness enhancement.

A potential explanation for the correlations being smaller than in the experiment by Plack (1996a) would be that in the latter study the masker level was fixed at 90 dB SPL. In fact, for the data obtained with the 85-dB SPL masker in Experiments 1 and 2 of the present study, the correlation between the DL elevation and the masker-induced loudness change was $r = 0.636$ ($n = 12$, $p = 0.026$).

If one compares the individual patterns of ΔL_{DL} and loudness enhancement (Figs. 2 and 4), several instances are obvious that speak against a simple one-on-one relation between loudness enhancement and intensity resolution. To give an example, the loudness matches produced by listeners AL and DO at the 55-dB SPL target level show a pronounced mid-difference hump, while their DLs increased monotonically with the masker-standard level difference in this condition.

As noted earlier, a masker higher in level than the target can be assumed to cause not only loudness enhancement, but also loudness recalibration (Arieh and Marks (2003); Oberfeld, 2007). Could loudness recalibration rather than loudness enhancement be the mechanism resulting in impairment in intensity resolution? As loudness recalibration is not observed for masker levels lower than the target level, this would explain the absence of a DL elevation in these conditions. Above that, Mapes-Riordan and Yost (1999) reported only 4 dB of loudness recalibration for a 40-dB SPL, 500-Hz

target combined with an 80-dB SPL masker, but about 11 dB for the target presented at 60 or 70 dB SPL. In other words, a midlevel hump pattern was observed. Oberfeld (2007) also reported that the amount of loudness recalibration caused by a 90-dB SPL forward masker was slightly stronger for a 60- than for a 30-dB SPL target, but the difference was only about 2 dB, and the data were from two different experiments involving different listeners. Evidence against a relation between loudness recalibration and the DL elevation comes from the observation that the reduction in the loudness of a 30-dB SPL tone increases monotonically with the level of the masker (Oberfeld, 2007), showing no evidence for a mid-difference hump. Still, it would be interesting to measure not only loudness enhancement, but also loudness recalibration in future experiments aimed at examining the relation between masker-induced loudness changes and DL elevations.

B. Implications for models of forward-masked intensity discrimination

In Experiment 1, intensity DLs at different standard levels were for the first time obtained for a wide range of masker-standard level differences, avoiding the confound between the latter difference and the standard level present in previous studies. Three important findings emerged. First, significant DL elevations were observed at low and high standard levels if there was an intermediate level difference between masker and standard. Second, a mid-difference hump was found at the 25-dB SPL standard level where the maximal DL elevations occurred at intermediate masker-standard level differences. Third, the data confirmed that there is a midlevel hump because for a given masker-standard level difference, the effect of the masker on the DL was larger for a medium-level than for a low-level standard. The data are incompatible with the recovery-rate model (Zeng *et al.*, 1991).

The loudness enhancement hypothesis by Carlyon and Beveridge (1993) extended by the assumption that a reduction in the masker-standard similarity reduces the effect of the masker (Oberfeld, 2007) can account for the masker-induced DL elevations at low and high standard levels as well as for the mid-difference hump. The fact that the latter was not observed for all listeners weakens the case for perceptual similarity playing a dominant role, but as discussed in Sec. I, several previous experiments also found pronounced intersubject differences for high-level maskers combined with low-level standards. It is tempting to assume that this finding is related to the "central" nature of the effect of the forward masker assumed by both the referential encoding hypothesis and the similarity hypothesis. Studies on informational masking (which clearly represents a central/cognitive rather than a peripheral effect) also reported a large amount of interindividual variability (for a recent discussion see Durlach *et al.*, 2005). The evidence for a mid-difference hump was more uniform in Experiment 2, where the loudness change caused by the forward masker was a nonmonotonic function of the masker-target level difference for all listeners at the 25-dB SPL target level, and for all but one listener at the 55-dB SPL target level.

The effect of the standard level (midlevel hump) in the intensity discrimination experiment is compatible with the referential encoding hypothesis, which could also account for the mid-difference hump if the assumption is made that a perceptual difference between masker and standard results in less memory trace interference.

To summarize, the data do not provide a basis for definitively deciding between the similarity hypothesis and the referential encoding hypothesis. Therefore, some general remarks concerning the two alternative explanations are in order.

A unique feature of the similarity hypothesis is that it addresses the relation between the effects of a nonsimultaneous masker on loudness and intensity resolution. In line with the predictions, there was a significant correlation between the masker-induced loudness changes (Experiment 2) and the DL elevations (Experiment 1). At the same time, due to several discrepancies between the two effects, the predicted one-on-one relation between loudness enhancement and intensity resolution received only partial support. It should also be noted [as Plack (1996a) pointed out] that even if there had been a perfect correlation between loudness enhancement and the DL elevation, it would still remain to find an explanation of why a masker-induced loudness change should increase the variability of the loudness representations used for example in an intensity discrimination task.

The referential encoding hypothesis is very flexible and thus can account for a broad range of findings as, for example, the effects a tone interpolated between the masker and the target (cf. Plack, 1996b). Plack (1996a) noted that the latter results are difficult to explain in terms of the loudness enhancement hypothesis. On the other hand, the flexibility of the referential encoding hypothesis comes at the cost of many degrees of freedom. For example, to predict the effect of varying the masker-standard level difference, it would be necessary to specify at least qualitatively a functional relation not only between the level difference and the amount of trace degradation, but also between the level difference and the effectiveness of the masker as a within-interval coding reference. Another issue is that if the effect of a forward masker was due to the use of the context coding mode, a midlevel hump should be observed in quiet in a one-interval (absolute identification) paradigm, where listeners are also assumed to use context coding (Durlach and Braida, 1969). If a wide range of different stimulus levels is presented within a block ("roving level"), it is known that intensity resolution is superior at the edges of the intensity range (cf. Berliner *et al.*, 1977). However, in one-interval experiments presenting only a *small* range of intensities, just as in forward-masked intensity discrimination experiments, Braida and Durlach (1972, Experiment 5) found no evidence for a midlevel hump, while in the study by McGill and Goldberg (1968), the DLs for standards presented at 5–15 dB SL were somewhat smaller than in the region 25–35 dB SL. The individual DL differences ranged between only 0.5 and 2.5 dB, however, and were thus considerably smaller than the midlevel DL elevations of up to 13 dB found under forward masking (Plack *et al.*, 1995; Schlauch *et al.*, 1997). Now it could be argued that the maskers create a roving-level

situation even if the standard level is fixed. Within a block, a listener encounters, e.g., a 60-dB SPL standard, rather similar standard-plus-increment levels, but also a 90-dB SPL masker, corresponding to a level range of 30 dB. It remains unclear, however, why in this situation the listeners should not be able to use the lower edge of the intensity range (i.e., the intensity of the standard) as an efficient coding reference (Braida *et al.*, 1984), but should have to rely on the detection threshold or the discomfort level.

To conclude, as far as intensity resolution is concerned, the significant masker-induced DL elevations at all standard levels as well as the observed mid-difference hump support the similarity hypothesis, but can also be accounted for by the referential encoding hypothesis.

ACKNOWLEDGMENTS

I am grateful to two anonymous reviewers for helpful comments concerning earlier versions of this manuscript.

Arieh, Y., and Marks, L. E. (**2003**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.

Baddeley, A. D. (**1966**). "Short-term memory for word sequences as a function of acoustic, semantic and formal similarity," Q. J. Exp. Psychol. **18**, 362–365.

Berliner, J. E., Durlach, N. I., and Braida, L. D. (**1977**). "Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects," J. Acoust. Soc. Am. **61**, 1577–1585.

Braida, L. D., and Durlach, N. I. (**1972**). "Intensity perception. II. Resolution in one-interval paradigms," J. Acoust. Soc. Am. **51**, 483–502.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context coding," J. Acoust. Soc. Am. **76**, 722–731.

Carlyon, R. P., and Beveridge, H. A. (**1993**). "Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise," J. Acoust. Soc. Am. **93**, 2886–2895.

Duncan, J., and Humphreys, G. (**1989**). "Visual search and stimulus similarity," Psychol. Rev. **96**, 433–458.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd, G. (**2005**). "Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers," J. Acoust. Soc. Am. **118**, 2482–2497.

Elmasian, R., Galambos, R., and Bernheim, A. (**1980**). "Loudness enhancement and decrement in four paradigms," J. Acoust. Soc. Am. **67**, 601–607.

Galambos, R., Bauer, J., Picton, T., Squires, K., and Squires, N. (**1972**). "Loudness enhancement following contralateral stimulation," J. Acoust. Soc. Am. **52**, 1127–1130.

Hellman, R. P., and Zwislocki, J. J. (**1964**). "Loudness function of a 1000 cps tone in the presence of a masking noise," J. Acoust. Soc. Am. **36**, 1618–1627.

International Electrotechnical Commission (IEC). (**1970**). *An IEC Artificial Ear, of the Wide Band Type, for the Calibration of Earphones Used in Audiometry* (IEC 318, Geneva).

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

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

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (**1996**). *SAS System for Mixed Models* (SAS Institute Inc., Cary, NC).

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.

McGill, W. J., and Goldberg, J. P. (**1968**). "Pure-tone intensity discrimination and energy detection," J. Acoust. Soc. Am. **44**, 576–581.

Moore, B. C. J., and Glasberg, B. R. (**1982**). "Contralateral and ipsilateral cueing in forward masking," J. Acoust. Soc. Am. **71**, 942–945.

Mori, S., and Ward, L. M. (**1992**). "Intensity and frequency resolution: Masking of absolute identification and fixed and roving discrimination," J. Acoust. Soc. Am. **91**, 246–255.

Oberfeld, D. (**2007**). "Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both?," J. Acoust. Soc. Am. **121**, 2137–2148.

Oxenham, A. J., and Bacon, S. P. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hear. **24**, 352–366.

Plack, C. J. (**1996a**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024–1030.

Plack, C. J. (**1996b**). "Temporal factors in referential intensity coding," J. Acoust. Soc. Am. **100**, 1031–1042.

Plack, C. J. (**1998**). "Beneficial effects of notched noise on intensity discrimination in the region of the 'severe departure'," J. Acoust. Soc. Am. **103**, 2530–2538.

Plack, C. J., and Carlyon, R. P. (**1995**). "Loudness perception and intensity coding," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 123–160.

Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). "Intensity discrimination under forward and backward masking: Role of referential encoding," J. Acoust. Soc. Am. **97**, 1141–1149.

Plack, C. J., and Viemeister, N. F. (**1992a**). "The effects of notched noise on intensity discrimination under forward masking," J. Acoust. Soc. Am. **92**, 1902–1910.

Plack, C. J., and Viemeister, N. F. (**1992b**). "Intensity discrimination under backward masking," J. Acoust. Soc. Am. **92**, 3097–3101.

Relkin, E. M., and Doucet, J. R. (**1991**). "Recovery from prior stimulation. I. Relationship to spontaneous firing rates of primary auditory neurons," Hear. Res. **55**, 215–222.

Scharf, B., Buus, S., and Nieder, B. (**2002**). "Loudness enhancement: Induced loudness reduction in disguise? (L)," J. Acoust. Soc. Am. **112**, 807–810.

Schlauch, R. S., Clement, B. R., Ries, D. T., and DiGiovanni, J. J. (**1999**). "Masker laterally and cueing in forward-masked intensity discrimination," J. Acoust. Soc. Am. **105**, 822–828.

Schlauch, R. S., Lanthier, N., and Neve, J. (**1997**). "Forward-masked intensity discrimination: Duration effects and spectral effects," J. Acoust. Soc. Am. **102**, 461–467.

Schlauch, R. S., and Wier, C. C. (**1987**). "A method for relating loudness-matching and intensity-discrimination data," J. Speech Hear. Res. **30**, 13–20.

Yates, G. K. (**1990**). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions," Hear. Res. **50**, 145–162.

Zeng, F.-G. (**1994**). "Loudness growth in forward masking: Relation to intensity discrimination," J. Acoust. Soc. Am. **96**, 2127–2132.

Zeng, F.-G., and Shannon, R. V. (**1995**). "Possible origins of the nonmonotonic intensity discrimination function in forward masking," Hear. Res. **82**, 216–224.

Zeng, F.-G., and Turner, C. W. (**1992**). "Intensity discrimination in forward masking," J. Acoust. Soc. Am. **92**, 782–787.

Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). "Recovery from prior stimulation. II. Effects upon intensity discrimination," Hear. Res. **55**, 223–230.

Zwislocki, J. J., and Sokolich, W. G. (**1974**). "On loudness enhancement of a tone burst by a preceding tone burst by a preceding tone burst," Percept. Psychophys. **16**, 87–90.

Acoustic model adaptation for ortolan bunting (*Emberiza hortulana* L.) song-type classification

Jidong Tao[a] and Michael T. Johnson
Speech and Signal Processing Laboratory, Marquette University, P.O. Box 1881, Milwaukee, Wisconsin 53233-1881

Tomasz S. Osiejuk
Department of Behavioural Ecology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

(Received 11 June 2007; revised 3 January 2008; accepted 4 January 2008)

Automatic systems for vocalization classification often require fairly large amounts of data on which to train models. However, animal vocalization data collection and transcription is a difficult and time-consuming task, so that it is expensive to create large data sets. One natural solution to this problem is the use of acoustic adaptation methods. Such methods, common in human speech recognition systems, create initial models trained on speaker independent data, then use small amounts of adaptation data to build individual-specific models. Since, as in human speech, individual vocal variability is a significant source of variation in bioacoustic data, acoustic model adaptation is naturally suited to classification in this domain as well. To demonstrate and evaluate the effectiveness of this approach, this paper presents the application of maximum likelihood linear regression adaptation to ortolan bunting (*Emberiza hortulana* L.) song-type classification. Classification accuracies for the adapted system are computed as a function of the amount of adaptation data and compared to caller-independent and caller-dependent systems. The experimental results indicate that given the same amount of data, supervised adaptation significantly outperforms both caller-independent and caller-dependent systems.
 [DOI: 10.1121/1.2837487]

PACS number(s): 43.66.Gf, 43.80.Ka, 43.72.Fx, 43.60.Uv [DOS] Pages: 1582–1590

I. INTRODUCTION

Hidden Markov models (HMMs) have been successfully applied to animal vocalization classification and detection in a number of species. Kogan and Margoliash (1998) and Anderson (1999) have shown that HMM-based classification is more robust to noise and more effective for highly confusable vocalizations than a dynamic time warping approach applied to the indigo bunting (*Passerina cyanea*) and zebra finch (*Taeniopygia guttata*). Other species in which HMM-based classification has been investigated include African elephants (*Loxodonta africana*) (Clemins *et al.*, 2005), beluga whale (*Delphinapterus leucas*) (Clemins and Johnson, 2005), ortolan bunting (*Emberiza hortulana* L.) (Trawicki *et al.*, 2005), red deer (*Cervus elaphus*) (Reby *et al.*, 2006), and rhesus macaques (*Macaca mulatto*) (Li *et al.*, 2007). HMM systems have been widely used to examine vocal repertoire, identify individuals, and classify vocalizations according to social context or behavior.

Typically, such classification systems are caller-independent (CI), meaning that the examples used for training the classifier come from a different set of individuals than those used for testing. In contrast to this, systems for human speech recognition are often speaker-dependent (SD), i.e., trained on the same individual who will be using the system, since given sufficient individual-specific training data SD systems have better performance than speaker-independent (SI) systems. When individual-specific training data are limited, an alternative is to use a speaker-adapted (SA) system. In this case a SI system is trained first and then the classification models are adapted with some individual-specific data, called adaptation data, to better account for individual variability in speech and pronunciation patterns. SA systems will typically have better overall accuracy than either SI or SD systems for small or moderate amounts of adaptation data. The error rate of a SD system may be as low as one-third that of a comparable SI speech recognition system tested on the same data (Hazen, 1998; Lee *et al.*, 1991), because individual speech differences are minimized in the SD system. The goal of using adaptation is to achieve performance approaching that of an ideal SD system using only limited amounts of speaker-specific data (Kuhn *et al.*, 2000).

Similarly, it is possible to develop analogous classification systems for animal vocalizations that are caller-dependent (CD) or caller-adapted (CA). The goal of this approach is to maximize the accuracy of the classifier while minimizing the amount of labor required to analyze and transcribe the collected data. Previous studies in animal vocalization analysis have found that individual vocal variability is one of the most important cues impacting vocalization related behavior study in bioacoustics (Reby *et al.*, 2006). Individual variability in acoustic structure has been described in many species such as bottlenose dolphins (*Tursiops truncates*) (Parijs *et al.*, 2002; Janik *et al.*, 2006), zebra finches (*Taeniopygia guttata*) (Vignal *et al.*, 2004), and Belding's ground squirrels (*Spermophilus beldingi*) (McCowan and

[a]Electronic mail: vjdtao@hotmail.com

Hooper, 2002). In ortolan buntings, song vocalization has been found to differ significantly between individuals in terms of repertoire content (Osiejuk *et al.*, 2003) and tonality (Osiejuk *et al.*, 2005). These differences have strong influence on species biology as ortolan bunting males were recently shown to discriminate between neighbors and strangers by song (Skierczyński *et al.*, 2007) and to differentiate response to songs composed of syllables originating from local or foreign population (Osiejuk *et al.*, 2007). This would imply that a CA system for animal vocalization analysis and classification should yield measurable improvements in overall accuracy and performance. Because both the data collection and analysis/transcription processes are much more difficult and time-consuming for most animal species than for human speech, utilizing a CA system to reduce the overall data requirements for developing automated classification systems may result in significant cost-savings. Additionally, cross comparisons of CD, CI, and CA recognition models have the potential to yield significant insight into the source of individual vocal variability.

The aim of this study is to demonstrate the use of adaptation for animal vocalization classification and examine the data requirements and degree of improvement provided by a CA system over comparable CI and CD systems. The CA system implemented for this task is based on the maximum likelihood linear regression (MLLR) technique (Leggetter and Woodland, 1995). The MLLR method works by clustering the states in an HMM into groups using a regression tree, then learning a maximum likelihood (ML) linear transformation for each group. The regression-based transformations tune the HMM mean and covariance parameters to each new individual represented by the adaptation data. To ensure all parameters can be adapted, a global transformation can be used for all HMMs in the system if only a small amount of adaptation data is presented, so that MLLR adaptation can improve recognition performance even with very limited adaptation data (Leggetter and Woodland, 1995). Results indicate that CA does in fact provide substantial performance improvement over both CI and limited-data CD systems.

II. DATA

A. Species under study

Ortolan buntings (*Emberiza hortulana* L.) are the focus of the current study. The species has declined steadily the last 50 years in Western Europe, and is currently listed in Norway as critically endangered on the Norwegian red-list. The population size is now only about 100 singing males and declines an average of 8% annually (Dale, 2001; Steifetten and Dale, 2006). The initial decline of the Norwegian population was probably due to the habitat loss related to changes in agriculture practices (Dale, 2001). However, 10 years of intensive study revealed that the main reason for the continuous decrease is female-biased dispersal pattern, which in isolated and patchy population seriously affects sex ratio, behavior of males, and breeding success measured at the population level (Dale *et al.*, 2005, 2006; Steifetten and Dale, 2006). It is hoped that increasing our understanding of male ortolan bunting vocalizations will enable us to better understand breeding behavior and reduce the risk of extinction.

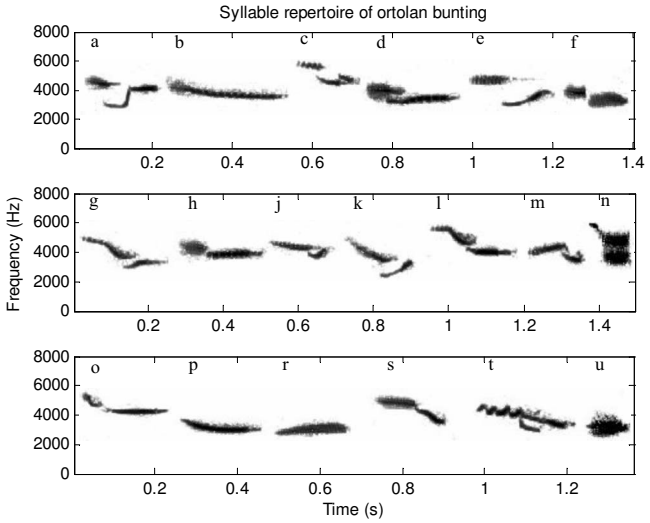


FIG. 1. Complete set of the 19-syllable repertoire of ortolan bunting.

Norwegian ortolan bunting vocalizations were collected from County Hedmark, Norway in May of 2001 and 2002 (Osiejuk *et al.*, 2003). The birds covered an area of approximately 500 km^2 on 25 sites, and males were recorded on 11 of those sites. A team of one to three research members who recognized and labeled the individual male buntings visited the sites. Overall, the entire sample population in 2001 and 2002 contains 150 males, 115 of which were color-ringed for individual identification. Because there are no known acoustic differences between the ringed and nonringed males, all data were grouped together for experimental use.

Ortolan buntings communicate through fundamental acoustical units called syllables (Osiejuk *et al.*, 2003). Figure 1 depicts the 19-syllable vocal repertoire used in this data set. Individual songs are grouped into song-type categories, e.g., *ab*, *cb*, that indicate the sequence of syllable types present. Each song type has many specific song variants, e.g., *aaaab*, *aaabb*, which indicate the exact repetition pattern. Figure 2 shows spectrograms of three specific type *ab* songs, song variants *aaaab*, *aaabb*, and *aaaabb*. The waveforms in Figs. 1 and 2 are low background noise exemplars, taken from different individuals to illustrate the repertoire.

B. Data collection

Vocalizations were recorded in the morning hours between 04:00 and 11:00 in each site, using a HHB PDR 1000 Professional DAT recorder with a Telinga V Pro Science parabola, a Sony TCD-D8 DAT recorder with a Sennheiser ME 67 shotgun microphone or an Aiwa HS-200 DAT recorder with a Sennheiser ME 67 shotgun microphone. All recordings were digitally transferred from Technics SV-DA 10 recorder via a SPDIF cable to a PC workstation with SoundBlaster Live! 5.1 at a sampling rate of 48 kHz with 16-bit quantization. For a more detailed description of the methods used to record the vocalizations, see Osiejuk *et al.* (2003, 2005).

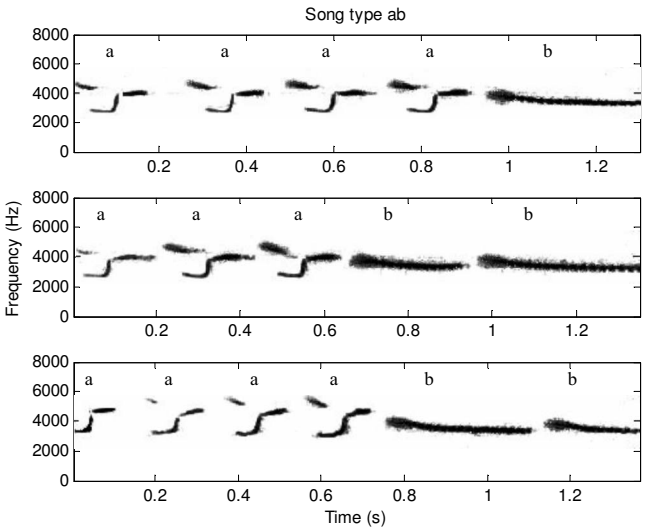


FIG. 2. *ab*-type song variation in ortolan bunting.

C. Data organization

The data set used here is a subset of the Osiejuk *et al.* data (Osiejuk *et al.*, 2003, 2005) including 60 song types and 19 syllables from 105 individuals. In selecting data for this study, calls containing syllables which were identified in only a single individual or a single song type were not included. Different individuals were selected for the training and testing/adaptation sets, balanced to get full coverage of all syllables in each set.

The protocol used to separate the data into training, test, and adaptation sets is as follows:

(1) Remove calls containing syllables identified in only a single individual or a single song type. This gives a resulting data set of 105 individuals, 60 call types, and 19 syllables.
(2) Select individuals for testing/adaptation.
  (a) Sort song types in increasing order according to number of examples.
  (b) Starting with the least common song type, select the individual with the highest number of examples in that song type (minimum two examples).
  (c) Repeat this process for each song type until the individuals selected for testing cover all 60 types.
  This results in a set of 30 individuals for testing/adaptation.
(3) Create explicit test and adaptation data sets by randomly dividing the data into test and adaptation sets for each selected individual, subject to a maximum of 30 vocalizations in each set for any one individual and song type.
(4) Group the remaining individuals into a training data set, again reducing the number of examples to a maximum of 30 for any one individual and song type.

Descriptive statistics of the resulting training, test, and adaptation sets are shown in Table I. From the above-detailed process it is clear that the 75 individuals in the training set are disjoint from the 30 individuals in the test/adaptation data, while the test and adaptation sets share the same group of individuals. All three sets have a full representation of syllables. Note that the training set does not cover the full range of 60 song types, but is still sufficient for training syllable-level HMMs for classification, as discussed in Sec. III. The size of the adaptation set is the same as that of the test set to allow the data to be used for training caller-dependent models as well as to allow a large range of variation for examining the impact of adaptation data quantity on performance.

TABLE I. Distribution of the number of individuals, song types, and vocalizations, and vocalizations with associated frequencies on individual, song type and syllable for training, test, and adaptation sets.

	Training set	Test set	Adaptation set
Number of individuals	75	30	30
Number of song types	53	60	60
Number of syllables	19	19	19
Number of vocalizations	2039	864	886
Mean vocalizations/caller	27.2	28.8	29.5
Mean vocalizations/type	38.5	14.4	14.8
Mean vocalizations/syllable	107.3	45.5	46.6

III. METHODS

A. Feature extraction

The primary features used in this HMM classification system are Greenwood function cepstral coefficients (GFCCs) (Clemins *et al.*, 2006; Clemins and Johnson, 2006). GFCCs are a species-specific generalization of mel frequency cepstral coefficients (MFCCs) (Huang *et al.*, 2001), one of the most common feature sets used in human speech recognition. The process for computing cepstral coefficients begins with segmenting vocalizations into evenly spaced appropriately sized windows (based on the frequency range and vocalization patterns of the species). For each window, a log magnitude fast Fourier transform (FFT) is computed and grouped into frequency bins. A discrete cosine transform is then taken to transform the log magnitude spectrum into cepstral values. For GFCCs, the frequency scale of the FFT is warped according to the Greenwood function (Greenwood, 1961) to provide a perceptually scaled axis. To do this, the parameters of the Greenwood function are estimated from the upper and lower bounds of the species' hearing range along with a warping constant of $k=0.88$ (LePage, 2003). Details of the warping equations and GFCC feature extraction process can be found in Clemins *et al.* (2006) and Clemins and Johnson (2006). Given basic information about a species frequency range, GFCCs provide an accurate and robust set of features to describe spectral characteristics over time.

In addition to the base set of GFCC features, energy is computed on the original time-domain data, and velocity and acceleration coefficients representing the first- and second-order rates of change are added. For the experiments described here, the vocalizations are segmented using 5 ms Hamming windows, with a 2.5 ms overlap. Twelve GFCCs plus normalized log energy along with velocity and acceleration coefficients are calculated, for a total of 39 features.

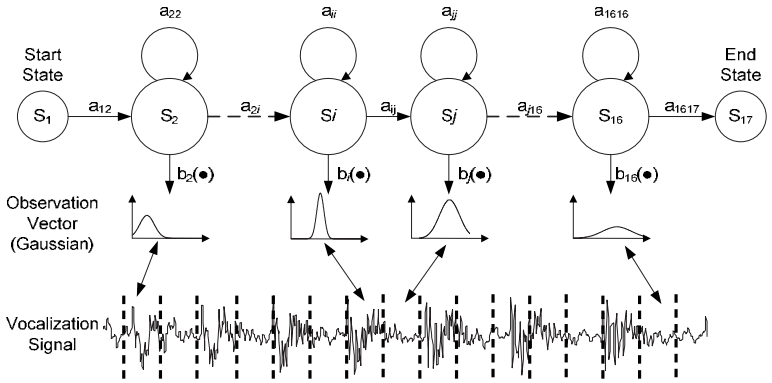


FIG. 3. A 15-state left-to-right hidden Markov model (HMM). Each state emits an observable vector of 39 GFCCs that are characterized by a single Gaussian model.

Frequency warping is done using a given hearing range from 400 to 7200 Hz, with 26 triangular frequency bins spaced across that range. Velocity and acceleration coefficients are computed using a five-window linear regression.

B. Acoustic models

HMMs (Rabiner and Juang, 1993) are the state-of-the-art approach for continuous speech recognition tasks. HMMs are statistical finite-state machines, where states represent spectrally stationary portions of the vocalization and transitions between states represent spectral transitions. This results in the ability to model spectral and temporal differences between an example vocalization and a trained HMM, with an implicit nonlinear time alignment.

In this work, each of the 19 ortolan bunting syllables is modeled with a 15-state left-to-right HMM, as illustrated in Fig. 3. Each state S_j is entered according to a transition probability a_{ij} from the previous state S_i. An observation feature vector o_t at time t is generated from the current state S_j based on a probability distribution $b_j(o_t)$, which in this work is a diagonal covariance Gaussian model.

During the training process, the Baum–Welch algorithm for expectation maximization (EM) (Baum *et al.*, 1970; Moon, 1996) is used to estimate the HMM parameters that maximize the joint likelihood of all training observation sequences. For classification, the Viterbi algorithm (Forney, 1973) is used to find the model sequence having the highest likelihood match to the sequence of test features.

C. Maximum likelihood linear regression adaptation

Once an HMM has been trained, the model parameters can be adapted to tailor the model to more domain-specific data. The key parameters for adaptation are the means and variances corresponding to each state distribution $b_j(o_t)$. In the MLLR adaptation approach, two linear transformation matrices are estimated for each state, one for the mean vector and one for the covariance matrix, under a maximum likelihood criteria function (Leggetter and Woodland, 1995). The underlying principle is to provide a reestimation approach that is consistent with maximizing the HMM likelihood while keeping the number of free parameters under control, thus requiring a smaller amount of adaptation data and allowing for rapid adaptation. MLLR has been widely used to obtain adapted models for both new speakers and new environmental conditions (Huang *et al.*, 2001).

In order to maximize the use of adaptation data, the required linear transformation matrices for each state are grouped into broad acoustic/syllable classes so that the overall number of free parameters is significantly less than the number of mean vectors. This is accomplished by building a regression class tree to cluster states with similar distributions into regression classes, the members of which share the same linear transformation.

The regression class tree is constructed so as to cluster together components that are close acoustically, using the original CI model set (independent of any new data). A centroid splitting algorithm using a Euclidean distance measure is applied to construct the tree (Young *et al.*, 2002). The terminal nodes or leaves of the tree specify the finest possible resolution groupings for transformation, and are termed the base (regression) classes. Each Gaussian component from the CI model set belongs to one specific base class.

The amount and type of adaptation data that is available determines exactly which transformations are applied to the original model. This makes it possible to adapt all models, even those for which there were no observations in the adaptation data, because the regression tree representation allows for adaptation to be done based on similar models that are present in the data. When more adaptation data are available, a larger number of unique transformations are applied, in accordance with the structure of the regression tree.

Specifically, the mean vector μ_i for each state can be transformed using

$$\hat{\mu}_i = A_c m_i + b_c = W_c \mu_i, \tag{1}$$

where m_i is the original mean vector for state i, $\mu_i = [1 \quad m_i^T]^T$ is the extended mean vector incorporating a bias vector b_c, A_c is the transformation matrix for regression class C, and W_c is the corresponding extended transformation matrix $[b_c \quad A_c]$.

While the regression tree itself is built from the caller-independent models, the number of regression classes C actually implemented for a particular set of adaptation data is variable, depending on the data's coverage of the classes. A tiny amount of adaptation data would result in only a single transformation matrix being used across all classes, or even no adaptation at all.

The required transformation matrix W_c for adapting the mean vector μ_i as indicated in Eq. (1) is obtained using the EM technique. The resulting reestimation formula W_c is given by

$$w_q = \left[\sum_t \sum_{i\in C} \xi_t(i)\Sigma_i^{-1} x_t \mu_i^T\right]_q \left[\sum_{i\in C}\left[\sum_t \xi_t(i)\Sigma_i^{-1}\right]_{qq} (\mu_i \mu_i^T)_q\right]_q^{-1}, \tag{2}$$

where w_q is the qth row vector of W_c being estimated, $\xi_t(i)$ is the occupancy likelihood of state i, Σ_i and μ_i are the corresponding diagonal covariance matrix and the extended mean vectors, and x_t is the adaptation data feature vector at time t. The subscripts q and qq in this equation are used to indicate the corresponding row and diagonal element of a matrix, respectively, for compactness of representation.

The Gaussian covariance Σ_i for state i is transformed using

$$\hat{\Sigma}_i = (\Sigma_i^{1/2})^T H_i \Sigma_i^{1/2}, \tag{3}$$

where the diagonal linear transformation matrix H_i is estimated via

$$H_i = \frac{\sum_{j\in C} (\Sigma_j^{-1/2})^T \left[\sum_t \xi_t(j)(x_t-\mu_j)(x_t-\mu_j)^T\right]\Sigma_j^{-1/2}}{\sum_{j\in C}\sum_t \xi_t(j)}. \tag{4}$$

Typically, transformation matrices converge in just a few iterations. At each iteration, all matrices are initialized to the identity transformation, and recognition likelihood statistics are accumulated over the data using the current model. Means alone or both means and variances are then updated using Eqs. (2) and (4). Typically the impact of variance adaptation is much less significant than that of mean adaptation. Transforming the variances can still be significant, however, because by nature variances in a CI system, which come from many individuals, are higher than those of the corresponding CD systems.

To implement the adaptation process, transformation matrices are initialized to the identity matrix. Using the original CI model and the prebuilt regression tree, state occupancies are calculated for all possible states, and the occupation counts are grouped for each class in the tree and compared to a threshold to determine exactly which transformations are to be applied. Following this, several iterations of Eqs. (2) and (4) are run to estimate and apply the mean and variance transformation matrices and create a new adapted model.

MLLR adaptation can be used in various different modes. Supervised adaptation refers to adaptation done using data accompanied by expert transcriptions, so that the process is applied to known model components. It is also possible to implement unsupervised adaptation, where before each adaptation iteration a recognition pass is performed to determine which models to adapt. Clearly, if the initial CI models are not a good match to the new domain, unsupervised adaptation could potentially fail to improve or even degrade the overall system by adapting an incorrect selection of models using the new data. It is also possible to apply adaptation methods either statically, where the entire amount of adaptation data is used together, or incrementally, where adaptation is done repeatedly as the amount of adaptation data increases.

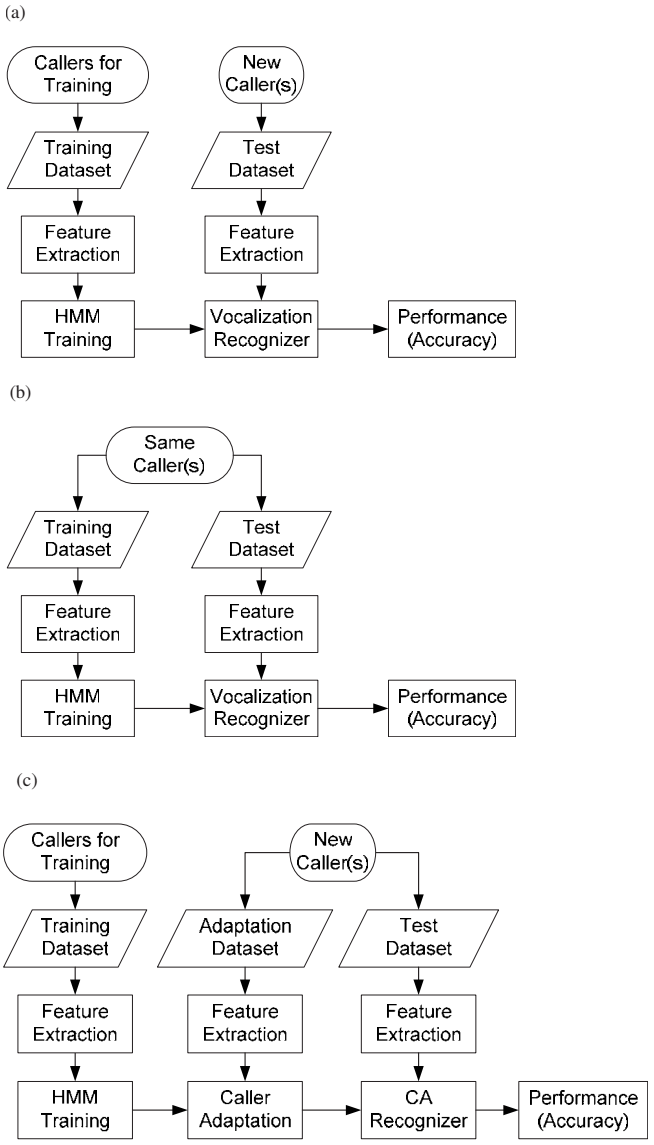


FIG. 4. Vocalization recognition systems. (a) Caller-independent, with separate individuals for the training and testing data. (b) Caller-dependent, with training and testing data coming from the same group of individuals. (c) Caller-adapted, with separate training and testing data, but with a portion of the testing data pulled out and used for adaptation.

D. Song-type recognition experiments

Song-type recognition experiments were implemented on the ortolan bunting data set as previously described. The goal of these experiments is to compare how well a CA HMM system performs compared to a baseline CI system. For reference, a fully CD system was also implemented.

The recognition models used for all experiments were 15 state single Gaussian HMMs with diagonal covariance matrices. The feature vector used for classification, as described previously, was a 39-element vector that included 12 GFCCs plus normalized log energy, accompanied by delta and delta–delta coefficients. The software toolkit HTK ver-

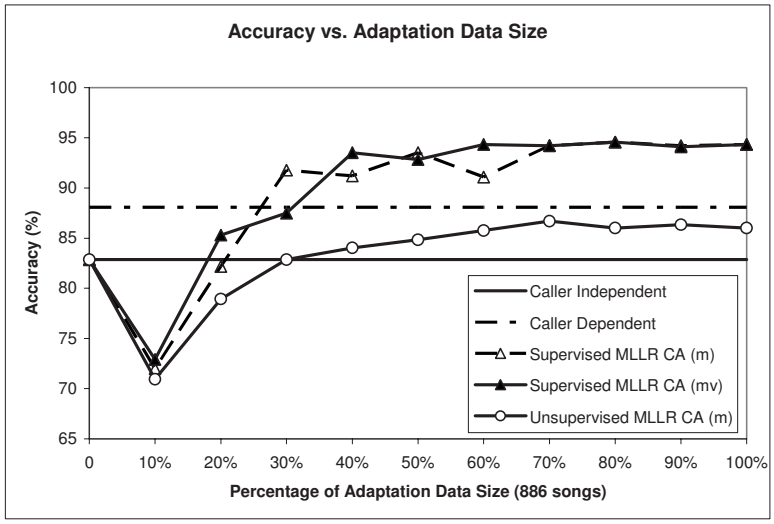


FIG. 5. The CI, CD, and CA system performance with varying amounts of adaptation data. MLLR adaptation was run in both supervised and unsupervised modes on the same data. Two types of adaptation in supervised mode are used: mean (m) only, and both mean and variance (mv).

sion 3.2 (Young *et al.*, 2002) was used to implement the HMMs, perform adaptation, and analyze classification performance. There were 19 different HMMs trained for each system, one for each syllable.

The following song-type recognition systems were implemented for comparison:

CI: the baseline caller-independent models. The system diagram for the CI system is shown in Fig. 4(a). There was no overlap between the training individuals and test individuals, with 75 and 30 individuals in the two data sets, respectively.

CD: the caller-dependent models. The system diagram for the CD system is shown in Fig. 4(b). The training and testing data were separate but came from the same individuals. The training data used for the CD experiments was the same as the adaptation data used for the CA experiments.

CA: the caller-adapted models. The system diagram for the MLLR adaptation systems is shown in Fig. 4(c). The training and testing data were from separate individuals, and the test data were further split into adaptation data and final test data. Three different CA experiments were implemented: supervised mean adaptation, supervised mean and variance adaptation, and unsupervised mean adaptation.

In order to see how the amount of adaptation data affected the results, each adaptation method was implemented multiple times, using increasing amounts of adaptation data. This was done in 10% increments, starting with 0% (no adaptation, equivalent to the initial CI system), then 10%, 20%, and so on up to 100% (full adaptation set in use).

IV. RESULTS

Overall results of the adaptation process can be seen in Fig. 5. The baseline CI system has an 82.9% accuracy, while the CD system has an 88.1% accuracy. Unsupervised adaptation of the means has a peak accuracy of 86.7% and a final accuracy of 86.0% using the full data set. Supervised adaptation yields the highest accuracy, 94.3% overall, representing a net gain of 11.4 percentage points (66% reduction in error) over CI and 6.2 percentage points (52% reduction in error) over CD.

The supervised adaptation using means (m) and that using both means and variances (mv) show a different pattern for lower amounts of data, but reach exactly the same accuracy, 94.3%, as the adaptation data increases. The supervised methods significantly outperform both the CI and the CD systems, reaching the performance level of the CI system at about 20% data and that of the CD at about 30% data. The unsupervised adaptation results, as expected, trail those of the supervised system, but are still able to significantly outperform the baseline CI system.

A. Detailed recognition results across specific individuals

Table II displays the comparison of CI, CD, and CA (full adaptation set) for each individual in the test set, along with the distributions of song types and syllables for each.

Note that in a few cases, 9 out of 30 individuals, the CD system actually gives a lower accuracy than the original CI system. In two of these cases, even the CA system still has a lower accuracy than the CI system. Comparing the CD to the CA systems, only one individual has an accuracy that is lower in the adapted system.

B. Detailed recognition results as a function of song-type frequency

In order to examine the recognition accuracy as a function of how often each song occurs (i.e., the amount of data in the training and test sets for each song), an additional analysis is done by rank-ordering the songs according to frequency of occurrence and plotting the accuracy.

The overall system recognition cumulative accuracies by classified song types are shown in Fig. 6. The CI results drop from 95.5% for the most common *ab* song down to 82.9%

TABLE II. Vocalization recognition comparison among CI, CD, and CA for each new individual, with the distributions of adapted songs, song types, and syllables of each individual bird. Overall accuracies with variances are CI 82.9 ± 16.4%, CD 88.1 ± 10.1%, and CA 94.3 ± 7.3%.

Caller ID	Adapt songs	song types	Syllables	CI (%)	CD (%)	CA (%)
2044	30	ab, cb, c, a	a, b, c	53.3	83.3	100
2049	30	huf, h, jufb, juf, hu	b, f, h, j, u	60	90	100
347	30	ab, gb, hufb, ghuf	a, b, f, g, h, u	63.3	93.3	96.7
2004	30	cb, eb, huf, h, jufb, hufb, c, cufb	b, c, e, f, h, j, u	56.7	90	90
2046	8	cb, er	c, b, e, r	66.7	83.3	100
2026	22	h, jufb, juf, hu, ju	b, f, h, j, u	68.2	72.7	100
2029	30	cb, gb, guf, gufb, gcb, gluf	b, c, f, g, l, u	46.7	70	76.7
385	30	h, ef, e	e, f, h	83.3	96.7	100
502	30	ab, cb, cufb, cf, cfb, tb, sfb	a, b, c, f, s, t, u	83.3	93.3	100
2022	30	huf, jufb, juf, j, ju	b, f, h, j, u	80	90	96.7
2010	30	ab, eb, ef	a, b, e, f	86.2	96.6	100
1303	40	cb, gb, guf, c, g, gh, gu, ch	b, c, f, g, h, u	70	86.7	83.3
205	30	ab, p, pb	a, b, p	73.3	60	83.3
165	30	cd, eb, cdb, suf, tb, sb, tuf	c, d, e, f, s, t, u	90	83.3	96.7
384	30	eb, cufb, cuf	b, c, e, f, u	93.3	93.3	100
430	30	ab, huf, h, hd, a	a, b, d, f, h, u	93.3	93.3	100
176	60	eb, huf, guf, hufb, luf, gufb, lufb	c, d, e, f, s, t, u	69	72.4	74.1
1201	30	gb, h, hb, gh, ghb, hgb	b, g, h	83.3	80	86.7
2038	30	ab, cb, cd, cdb	a, b, c, d	93.3	96.7	96.7
39	30	hd, gd	h, g, d	100	100	100
106	30	ab, kb, a, k	a, b, k	100	100	100
413	25	jufb, jb	b, f, j, u	100	100	100
1030	30	cd, od	c, d, o	93.3	83.3	93.3
1903	30	gb, nu, nuf, n	b, f, g, n, u	100	90	100
2011	30	cb, ghuf, gh, ghu	b, c, g, h, u	100	100	100
2021	30	gb, huf, h, hufb, ghuf, gh, ghufb, hu	b, f, g, h, u	96.7	80	96.7
2025	13	h, gh, hr	g, h, r	91.7	83.3	91.7
2030	30	ab, cb, kb, kab	a, b, c, k	100	100	100
314	30	gb, h, jd, hr	b, d, g, h, j, r	100	93.3	96.7
239	28	ab, kb, luf, mluf, mufl	a, b, f, k, l, m, n	100	95.7	95.7
Total	886	**60 song types**	**19 syllables**	**82.9**	**88.1**	**94.3**

for the least frequent song type, *sfb*. It is interesting to note that the impact of song-type frequency is less pronounced for the CA systems as compared to the CI and CD systems, with smoother accuracy curves across song type.

V. CONCLUSIONS

This work demonstrates the advantages of using an acoustic model adaptation system in classifying ortolan bunting vocalizations. There are two key advantages illustrated by these experiments. The first is that a caller-adapted recognition system typically gives significant performance improvement over either caller-independent or caller-dependent systems. The second is that adaptation provides for extremely efficient data utilization, which is very important for bioacoustics tasks where data collection and labeling is difficult.

The results given here suggest that the classification accuracy of many systems could be improved using adaptation, since individual vocal variation is typically one of the most important factors affecting performance. In essence, caller adaptation works like a flexible interpolation between an independent and a fully dependent system. An adapted system starts from a baseline independent system, surpasses both independent and dependent systems, and approaches an ideal (well trained with unlimited size of individual identity labeled data) dependent system. As the amount of data increases, the specificity of the adaptation is improved through the creation of a larger regression tree.

Although a caller-dependent system with unlimited data is theoretically ideal, it is often impractical because of the large amount of data required for each individual to build well trained models. The performance of such a system will be low if the number of individual vocalizations used to train the HMMs is limited, and recognition accuracy for any new individuals who are not present in the training set will be especially low. In contrast, an adaptation-based system overcomes these limitations by taking advantage of existing well trained CI models. With a moderate amount of labeled adaptation data, an adapted system generally achieves better performance than a CI or even a CD system. Even in unsupervised mode where *no* transcriptions at all are used, an adapted system may still approach the accuracy of a caller-dependent system as the size of adaptation data increases, as illustrated in the experimental results in Fig. 5.

The underlying reason for the high accuracy of adaptation systems is data utilization efficiency. Vocalization data

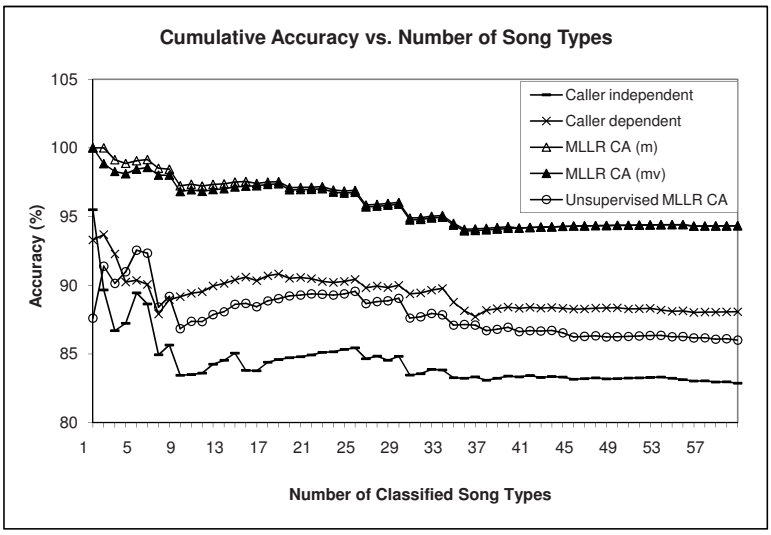


FIG. 6. Vocalization recognition cumulative accuracy vs number of classified song types for CI, CD, and CA systems.

collection is often difficult due to environmental constraints and inconsistent vocal repertoire across individuals, and data transcription or labeling is a time-consuming task requiring great expertise, so it is expensive to develop large data sets. Acoustic model adaptation is a natural solution to this problem, making the most effective use of the data regardless of how much is available. A CA system is initialized by starting from a CI system that is relative cheaper in effort required for data labeling because it does not need individual identity, then adapted using a much smaller amount of identity transcribed adaptation data to customize the system to the new individual vocal models. This enables a controlled trade-off between data labeling effort and system performance. In other words, maximum system performance is obtained with the minimum effort on data labeling.

The acoustic model adaptation methods presented here are applicable to a wide variety of species. Although each species has different vocal characteristics, individual vocal variability is nearly always present. Applying adaptation allows us to achieve high performance in classifying animal vocalizations with a small amount of available data.

ACKNOWLEDGMENTS

This material is based on work supported by National Science Foundation under Grant No. IIS-0326395. The study of ortolan buntings in Norway was supported by the Polish State Committee for Scientific Research, Grant Nos. 6 P04C 038 17 and 3 P04C 083 25 and Adam Mickiewicz University grant no. PBWB-301/2001.

Anderson, S. E. (**1999**). "Speech recognition meets bird song: A comparison of statistics-based and template-based techniques," J. Acoust. Soc. Am. **106**, 2130.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (**1970**). "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," Ann. Math. Stat. **41**, 164–171.

Clemins, P. J., and Johnson, M. T. (**2005**). "Unsupervised classification of beluga whale vocalizations," J. Acoust. Soc. Am. **117**, 2470(A).

Clemins, P. J., and Johnson, M. T. (**2006**). "Generalized perceptual linear prediction features for animal vocalization analysis," J. Acoust. Soc. Am. **120**, 527–534.

Clemins, P. J., Johnson, M. T., Leong, K. M., and Savage, A. (**2005**). "Automatic classification and speaker identification of African elephant (*Loxodonta africana*) vocalizations," J. Acoust. Soc. Am. **117**, 956–963.

Clemins, P. J., Trawicki, M. B., Adi, K., Tao, J., and Johnson, M. T. (**2006**). "Generalized perceptual features for vocalization analysis across multiple species," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Paris, France, Vol. **33**, pp. 1253–256.

Dale, S. (**2001**). "Causes of population decline in ortolan bunting in Norway," *Proceedings of the Third International Ortolan Symposium*, Poznan, Poland, pp. 33–41.

Dale, S., Lunde, A., and Steifetten, Ø. (**2005**). "Longer breeding dispersal than natal dispersal in the ortolan bunting," Behav. Ecol. Sociobiol. **16**, 20–24.

Dale, S., Steifetten, Ø., Osiejuk, T. S., Losak, K., and Cygan, J. P. (**2006**). "How do birds search for breeding areas at the landscape level? Interpatch movements of ortolan buntings," Ecography **29**, 886–898.

Forney, G. D. (**1973**). "The Viterbi Algorithm," Proc. IEEE **61**, 268–278.

Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

Hazen, T. J. (**1998**). "The use of speaker correlation information for automatic speech recognition," Ph.D. dissertation, MIT, Cambridge.

Huang, X., Acero, A., and Hon, H.-W. (**2001**). *Spoken Language Processing* (Prentice Hall, Upper Saddle River, NJ).

Janik, V. M., Sayigh, L. S., and Wells, R. S. (**2006**). "Signature whistle shape conveys identity information to bottlenose dolphins," Proc. Natl. Acad. Sci. U.S.A. **103**, 8293–8297.

Kogan, J. A., and Margoliash, D. (**1998**). "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," J. Acoust. Soc. Am. **103**, 2185–2196.

Kuhn, R., Junqua, J. C., Nguyen, P., and Niedzielski, N. (**2000**). "Rapid speaker adaptation in eigenvoice space," IEEE Trans. Speech Audio Process. **8**, 695–707.

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.

Leggetter, C. J., and Woodland, P. C. (**1995**). "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models," Comput. Speech Lang. **9**, 171–185.

LePage, E. L. (**2003**). "The mammalian cochlear map is optimally warped," J. Acoust. Soc. Am. **114**, 896–906.

Li, X., Tao, J., Johnson, M. T., Soltis, J., Savage, A., Leong, K. M., and Newman, J. D. (**2007**). "Stress and emotion classification using jitter and shimmer features," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, HI, pp. IV1081–1084.

McCowan, B., and Hooper, S. L. (**2002**). "Individual acoustic variation in Belding's ground squirrel alarm chirps in the High Sierra Nevada," J. Acoust. Soc. Am. **111**, 1157–1160.

Moon, T. K. (**1996**). "The expectation-maximization algorithm," IEEE Signal Process. Mag. **13**, 47–60.

Osiejuk, T. S., Ratyńska, K., and Cygan, J. P. (**2007**). "What makes a 'local song' in a population of ortolan buntings without common dialect?," Anim. Behav. **74**, No. 1, pp. 121–130.
Osiejuk, T. S., Ratyńska, K., Cygan, J. P., and Dale, S. (**2003**). "Song structure and repertoire variation in ortolan bunting (*Emberiza hortulana* L.) from isolated Norweigian population," Ann. Zool. Fenn. **40**, 3–16.
Osiejuk, T. S., Ratyńska, K., Cygan, J. P., and Dale, S. (**2005**). "Frequency shift in homologue syllables of the Ortolan Bunting *Emberiza hortulana*," Behav. Processes **68**, 69–83.
Parijs, S. M. V., Smith, J., and Corkeron, P. J. (**2002**). "Using calls to estimate the abundance of inshore Dolphins: A case study with Pacific humpback dolphins (Sousa Chinensis)," J. Appl. Ecol. **39**, 853–864.
Rabiner, L. R., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).
Reby, D., André-Obrecht, R., Galinier, A., Farinas, J., and Gargnelutti, B. (**2006**). "Cepstral coefficients and hidden Markov models reveal idiosyncratic voice characteristics in red deer (*Cervus elaphus*) stags," J. Acoust. Soc. Am. **120**, 4080–4089.
Skierczyński, M., Czarnecka, K. M., and Osiejuk, T. S. (**2007**). "Neighbour-stranger song discrimination in territorial ortolan bunting Emberiza hortulana males," J. Avian Biol. **38**, No. 4, pp. 415–420
Steifetten, Ø., and Dale, S. (**2006**). "Viability of an endangered population of ortolan buntings: The effect of a skewed operational sex ratio," Biol. Conserv. **132**, 88–97.
Trawicki, M. B., Johnson, M. T., and Osiejuk, T. S. (**2005**). "Automatic song-type classification and speaker identification of Norwegian ortolan bunting (*Emberiza hortulana*) vocalizations," IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT, pp. 277–282.
Vignal, C., Mathevon, N., and Mottin, S. (**2004**). "Audience drives male songbird response to partner's voice," Nature (London) **430**, 448–451.
Young, S., Evermann, G., Hain, T., Kershaw, D., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2002**). *The HTK Book (for HTK Version 3.2.1)* (Cambridge University Engineering Department, Cambridge, UK).

Investigation of perceptual constancy in the temporal-envelope domain

Marine Ardoint[a)]
Laboratoire de Psychologie de la Perception (CNRS—Université Paris 5 Descartes), Departement d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

Christian Lorenzi
Laboratoire de Psychologie de la Perception (CNRS—Université Paris 5 Descartes), Departement d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

Daniel Pressnitzer
Laboratoire de Psychologie de la Perception (CNRS—Université Paris 5 Descartes), Departement d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

Andreï Gorea
Laboratoire de Psychologie de la Perception (CNRS—Université Paris 5) Descartes, Université René Descartes, UFR Biomédical des Saints Pères. 45 rue des Saints Pères, 75006 Paris, France

(Received 2 April 2007; revised 21 December 2007; accepted 28 December 2007)

The ability to discriminate complex temporal envelope patterns submitted to temporal compression or expansion was assessed in normal-hearing listeners. An XAB, matching-to-sample-procedure was used. X, the reference stimulus, is obtained by applying the sum of two, inharmonically related, sinusoids to a broadband noise carrier. A and B are obtained by multiplying the frequency of each modulation component of X by the same time expansion/compression factor, α ($\alpha \in [0.35-2.83]$). For each trial, A or B is a time-reversed rendering of X, and the listeners' task is to choose which of the two is matched by X. Overall, the results indicate that discrimination performance degrades for increasing amounts of time expansion/compression (i.e., when α departs from 1), regardless of the frequency spacing of modulation components and the peak-to-trough ratio of the complex envelopes. An auditory model based on envelope extraction followed by a memory-limited, template-matching process accounted for results obtained without time scaling of stimuli, but generally underestimated discrimination ability with either time expansion or compression, especially with the longer stimulus durations. This result is consistent with partial or incomplete perceptual normalization of envelope patterns.
 [DOI: 10.1121/1.2836782]

PACS number(s): 43.66.Mk, 43.66.Ba [JHG] Pages: 1591–1601

I. INTRODUCTION

Normal-hearing listeners understand each other even when the rate of production of their spoken words is increased up to a factor of roughly 3 (e.g., Fairbanks and Kodman, 1957; Fu *et al.*, 2001; Versfeld and Dreschler, 2002). This form of *perceptual constancy* [which may be defined as the ability to listen to critical global aspects of speech and other complex nonspeech sounds, in contrast to the ability to listen to acoustic details (Li and Pastore, 1995)] seems to be based on the speech temporal envelope, as it is relatively independent of the audio carrier. Indeed, Fu *et al.* (2001) have shown that the deterioration of the spectral and temporal fine structure content of speech stimuli does not preclude their recognition after temporal compression or expansion. In addition, Ahissar *et al.* (2001) have shown that speech comprehension of time-compressed signals is correlated with the representation of the speech envelope in MEG (magnetoencephalography) signals. Hence, a possible explanation for the robustness of speech intelligibility to variation in presentation rate is that perceptual normalization is applied to the amplitude envelope of sounds, whether they are speech signals or not.

The general question asked in this paper was whether or not normal-hearing listeners show perceptual constancy for nonlinguistic amplitude envelopes presented at various time scales or, in other words, robust recognition of complex envelope patterns that are temporally compressed or expanded. Here, the nonlinguistic amplitude envelopes were obtained by summing two inharmonic, sinusoidal amplitude modulations. Temporal compression or expansion (i.e., temporal transposition) was achieved by mutiplying the frequency of each modulation component by a given index. Discrimination of the temporally transposed patterns was assessed as a function of their compression/expansion index. A similar approach was taken by Gockel and Colonius (1997) to study perceptual constancy following transposition of spectral patterns. In addition, discrimination of the temporally transposed patterns was assessed here for (i) two frequency ratios

a)Author to whom correspondence should be addressed. Electronic mail: ardoint@ens.fr

(i.e., two frequency spacing) of the two modulation components and (ii) three levels of amplitude compression/expansion applied to the complex envelopes, because both factors should be important determinants of envelope discrimination.

Manipulation of frequency ratio was intended to test the extent to which putative processing either within or across temporal modulation channels (Dau *et al.*, 1997a, b) affects a listener's resistance to temporal transposition. In other words, this manipulation attempted to assess the effect of the resolvability of the modulation components on the discrimination of the transposed envelopes.

The manipulation of the amplitude compression was intended to test the effect of the temporal envelope peak-to-trough ratio on performance. Previous experiments have revealed that this ratio is also an important determinant of speech identification (e.g., Fu and Shannon, 1999; Lorenzi *et al.*, 1999; Apoux *et al.*, 2001). In these experiments, the peak-to-trough ratio was modified by applying a power-law transform to the stimulus envelope. Overall, these experiments showed that increasing the peak-to-trough ratio yields significant improvements in phoneme identification performance in noise. It should be noted that all the speech perception studies investigating the effects of temporal compression/expansion (e.g., Fu *et al.*, 2001) have tested temporal compression/expansion constancy against a ~100% correct recognition performance for nontransformed (control) stimuli. The possibility remains that the observed constancy reflected in fact a ceiling effect. Accordingly discrimination of the temporally transposed patterns is assessed here for three peak-to-trough ratios (obtained by means of a compression/expansion of the envelope amplitude as in the studies cited previously) yielding different levels of discrimination performance with the highest still below perfect performance.

Current models of temporal-envelope processing in the auditory system do not include temporal normalization. One such model proposes that temporal-envelope detection or discrimination is achieved by cross correlating the outputs of amplitude-modulation channels with memory-stored templates according to an "optimal detector" scheme (Dau *et al.*, 1997a, b). This model accounts successfully for a variety of envelope detection data collected in masked and unmasked conditions. However, some form of a *normalization* process in the time domain might be required to account for the discrimination of complex temporal envelopes submitted to various levels of temporal compression or expansion. In the modeling part of the present study, we used a simplified front-end to the optimal detector approach to investigate whether listeners' performances can be predicted with a model that does not include a normalization stage.

II. EXPERIMENTS

A. Method

1. Listeners

Four listeners ranging in age between 20 and 33 years were tested. One of them was one of the authors (M.A.) and the other three were students. All listeners had absolute thresholds of less than 20 dB HL (Hearing Level) at audiometric frequencies between 0.125 and 8 kHz, and no history of hearing difficulty. Practice was given to each listener prior to data collection (see the following). All listeners were fully informed about the goal of the present study and provided written consent before their participation. The present experimental protocol is in accordance with the Helsinki declaration in 2004.

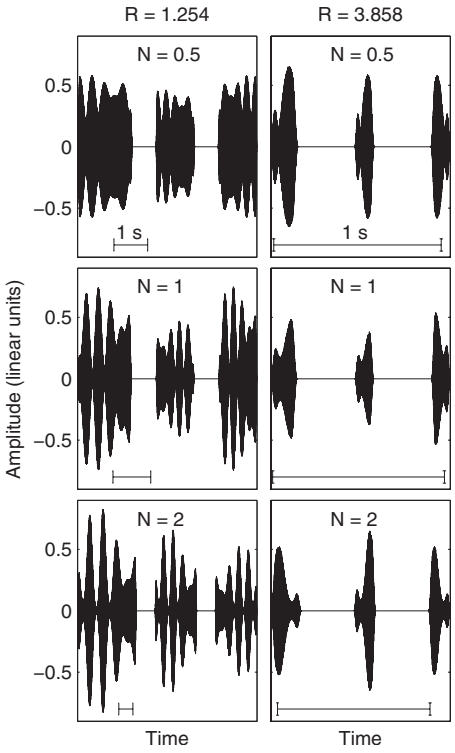


FIG. 1. Examples of wave form for stimuli in six typical trials, obtained with $R=1.254$ (left column) and $R=3.878$ (right column), $\alpha=1.414$, and $N=0.5$ (top panels), 1 (middle panels), and 2 (bottom panels). The center frequency and the modulation depths of the modulation components were varied across trials, whereas the global amplitude was varied independently for each stimulus within a trial. The 1-s time bars are different for different values of N because the center frequency of the modulation components is roved.

2. Stimuli

Examples of stimulus wave forms are illustrated in Fig. 1 for six typical trials. All stimuli were broadband noise audio carriers modulated by a complex temporal envelope equal to the sum of two temporal modulations:

$$S(t) = [1 + m_1 \sin(2\pi\alpha f_{m1} t + \varphi_1) + m_2 \sin(2\pi\alpha f_{m2} t + \varphi_2)]b(t), \quad (1)$$

with t being time, f_{m1} and f_{m2}, m_1 and m_2, and φ_1 and φ_2, respectively, are the frequencies (with $f_{m1}<f_{m2}$), depths and starting phases of the two components and with $b(t)$, the broadband noise carrier. Parameter α is a frequency-multiplication factor explained in the following. The stimuli were generated with a 16-bit digital/andlog converter (44.1 kHz sampling rate) under the control of a PC and delivered binaurally via a Sennheiser HD 600 headphone at a level of 65 dB SPL in a soundproof booth. The broadband

noises were non-Gaussian. They were generated in the time domain using a uniform distribution of amplitudes and were physically different within (i.e., across test and comparison stimuli) and across trials. The bandwidth of the noise was set to half the sampling rate.

Two f_{m2}/f_{m1} inharmonic ratios, R, were used so as to tap, presumably, the same ($R=1.254$, "unresolved components" condition), or two distinct temporal modulation channels [$R=3.879$, "resolved components" condition (e.g., Ewert and Dau, 2000; Lorenzi *et al.*, 2001)]. The two modulating frequencies, f_{m1} and f_{m2}, were symmetric (on a log scale) about a nominal central frequency f_c of 3 Hz [chosen because it corresponds to the most salient and critical frequency in the production and understanding of continuous speech (Houtgast and Steeneken, 1985)]. In order to prevent listeners from building over time a template of the stimuli and storing it in long-term memory, f_c was randomized across trials within a range of ± 0.5 octaves (i.e., 2.12–4.24 Hz). For the same reason, the phases φ_1 and φ_2 of the two modulation components were also independently randomized from trial to trial in a range of $0-2\pi$.

Both within and across trials, the modulation amplitudes m_1 and m_2 were each randomly varied between 0.25 and 0.5 so that their sum never exceeded 1.0 (i.e., overmodulation). The global amplitudes of the modulated noises of all stimuli (within and between trials) were independently randomized in a range of ± 3 dB SPL (with a 1-dB step) about the average 65 dB SPL. Two additional experimental conditions were obtained by elevating the envelope amplitudes to powers $N=0.5$ and 2 (amplitude compression and expansion, respectively). Although amplitude compression minimizes the peak-to-trough contrasts, amplitude expansion exaggerates them.

Manipulation of the factor α was central to the present study. In the *test* condition, it was used to produce the temporally compressed ($\alpha>1$) and expanded ($\alpha<1$) versions of the "reference" stimulus ($\alpha=1$). Factor α ($\alpha\neq 1$) was thus applied only to the two modulation frequencies, f_{m1} and f_{m2}, of the target and comparison stimuli. In the *control* condition, factor α ($\alpha\neq 1$) was also applied to the two modulation frequencies of the reference stimulus, so that in effect all stimuli had the same duration and the reference and target stimuli had identical envelopes. The following 7 α-values were used in all experimental conditions:.35,.5,.7, 1, 1.41, 2, and 2.82.

However, when N was equal to 1.0, 6 extra α-values (.42,.59,.84, 1.18, 1.68, and 2.37) were also used. To facilitate the analysis of the data and their comparison with previous studies, the α-values were converted to a compression/expansion index, CE, computed as $100|1-1/\alpha|$.

Stimuli duration, D, was equal to the period of the modulated envelope, i.e., $D=1/(\alpha f_{m2}-\alpha f_{m1})$, so as to prevent listeners from using more than one temporal envelope beat for their judgments. Obviously, D varies with both the compression-expansion factor, α, and with the frequency ratio, R. As can be seen in Table I, these durations (displayed for each α and R) range from as short a period as 81 ms ($\alpha=2.83$, $R=3.878$) to as long a period as 4199 ms ($\alpha=0.35$, $R=1.254$). Stimuli were ramped on and off with a cosine envelope whose temporal extent was equal to 50 ms for $\alpha=1$, and was proportional to α (ramp duration $=50/\alpha$ ms) when α departed from 1.

TABLE I. Stimulus duration for each value of α and R.

	Duration, D (ms)	
α	$R=1.254$	$R=3.878$
0.35	4199	652
0.42	3499	543
0.5	2939	456
0.59	2491	387
0.71	2070	321
0.84	1750	272
1	1470	228
1.19	1235	192
1.41	1042	162
1.68	875	136
2	735	114
2.38	617	96
2.83	519	81

3. Procedure

In the test condition, envelope discrimination performance (% correct) was measured by means of an XAB matching-to-sample procedure (see MacMillan and Creelman, 2005, Chap. 9) whereby X stands for the reference stimulus (with $\alpha=1$, i.e., a CE index of 0%), whereas A and B are its compressed or expanded temporal versions ($\alpha\lessgtr 1$, $\mathrm{CE}\neq 0\%$), one of which (randomized over trials) is a time-reversed (temporal mirror) rendering of X. The listener's task was to determine whether A or B matched the reference stimulus, X. The temporal interval between the three stimulus versions was 500 ms and the minimum interval between two successive trials was 2 s. Performance was measured in separated blocks for each combination of temporal compression/expansion (α), components frequency ratio (R), and amplitude compression/expansion (N). The control condition involved temporally noncompressed/expanded A and B versions of the reference X (see the following). Listeners completed the control and test conditions in random order. In both cases, one experimental block consisted of 50 trials and was repeated three times in a different random order for each listener. Hence, for each experimental condition, percent correct was computed out of 150 trials.

Before starting the main experiments, listeners passed 3–5 training sessions (i.e., 150–250 trials) with $\alpha=1$, $N=1$, and $R=1.254$ and 3.878. The training sessions were terminated once listeners reached a performance level of at least 75% correct (i.e., $d'=2$) which was achieved over a period of 2–6 h.

Listeners were provided with visual feedback in all sessions (training and testing) and experimental conditions (test and control conditions). Listeners' performance is presented as sensitivity (d') scores obtained from the assessed percentages correct (Macmillan and Creelman, 2005; Table A5.3: Differencing model).

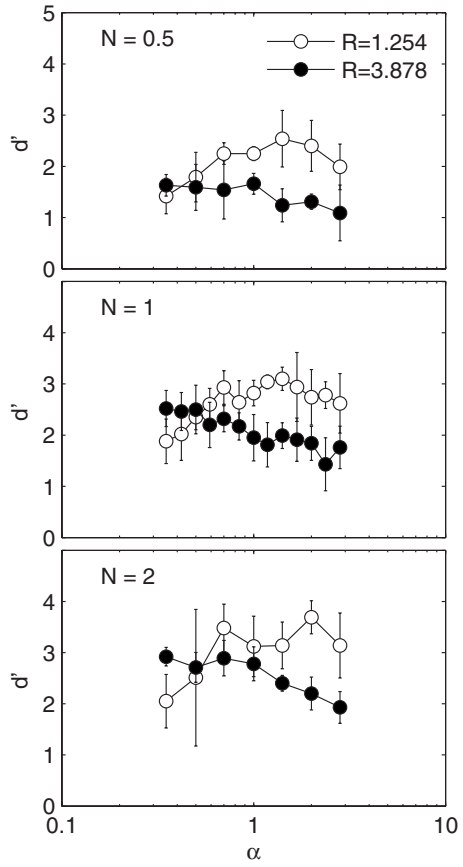


FIG. 2. Mean discrimination sensivity (d') for four listeners obtained in the control condition. Discrimination performance is plotted as a function of the time compression/expansion factor, α. Here, the time compression/expansion factor is applied to all envelopes (i.e., *X*, *A*, and *B*). Error bars represent ± 1 standard deviation across listeners. In each panel, open and filled circles correspond to cases where the frequency ratio, *R*, of the two modulation components of the complex envelopes is 1.254 and 3.878, respectively. The top, middle, and bottom panels show the data obtained with $N=0.5$ (all envelopes are compressed in amplitude), 1 (all envelopes are left intact), and 2 (all envelopes are expanded in amplitude), respectively.

B. Results: Control performance

In the control experiment, the expansion/compression factor α was applied to *all three* envelopes of the XAB sequence (with A or B being the time-reversed version of X) so that its manipulation was only meant to assess the dependence of envelope discrimination performance on the center frequency f_c of the envelopes, or, equivalently, on the duration of the stimuli. As a reminder, an $\alpha=1$ is equivalent to a nominal $f_c=3$ Hz with the two extreme α-values for all listeners ($\alpha=0.35$ and 2.83) yielding nominal f_c-values of 1.1 and 8.5 Hz.

All four listeners behaved similarly in this task. Therefore, for each experimental condition, discrimination sensitivity (d') was averaged across listeners. Figure 2 displays these average data as a function of α with envelope frequency-component ratio (*R*) the parameter [$R=1.254$ (open circles); $R=3.878$ (closed circles)]. The average data are shown for each of the three envelope amplitude compression/expansion indices, *N* (top panel: $N=0.5$; middle panel: $N=1$, bottom panel: $N=2$).

Overall, the discrimination of identical, time-reversed envelopes yields the following main characteristics: (1) it generally peaks for $\alpha=1.41-2$ (i.e., for $f_c=4-6$ Hz) when the modulation frequencies are close ($R=1.254$) and decreases monotonically as a function of α when the modulation frequencies are spaced apart ($R=3.878$); (2) it is globally better for the proximal rather than distal spacing of modulation components, particularly so within the medium-to-high f_c-range and independently of the amplitude expansion/compression index, *N*; (3) it increases with the amplitude expansion index, *N*; (4) it yields a maximum d' of about 3.69 (i.e., 93% correct). Overall, the present discrimination scores are within the range of those obtained for the discrimination of noise modulated envelopes (Takeuchi and Braida, 1995; 78–99% correct with a similar XAB method).

The above-mentioned qualitative account is confirmed by a three-way (α[7], R[2], N[3]) repeated-measures analysis of variance (ANOVA). Each of the main factors yields a significant effect [α: $F(6,18)=3.29$, $p<0.05$; *R*: $F(1,3)=17.84$, $p<0.05$; *N*: $F(2,6)=115.3$, $p<0.0001$]. Of the three second-order interactions, only $\alpha\times R$ is significant [$\alpha\times R$: $F(6,18)=14.34$, $p<0.00001$; $\alpha\times N$: $F(12,36)=1$, NS; $R\times N$: $F(2,6)<1$, NS]. Finally, the third-order interaction is not significant [$F(12,36)=1.94$, NS].

In other words, the present experiment and statistical analysis point to the fact that the discrimination of nontransposed temporal envelopes depends on their central modulation frequency (f_c), is better for a small frequency spacing of modulation components (*R*), and increases with amplitude expansion factor (*N*). Moreover, f_c and *R* interact in such a way that discrimination as a function of f_c has roughly an inverted U-shape for low values of *R* and a monotonically decreasing function for higher values.

C. Results: Test performance (control versus temporally expanded/compressed envelopes)

Again, as all four listeners behaved similarly for each experimental condition, discrimination sensitivity (d') was averaged across listeners. Figure 3 displays the average discrimination scores, d'_{Test}, in the same format as Fig. 2. In order to isolate the effect of the temporal transposition factor (α) from that of the envelope's central frequency (f_c; assessed in the first experiment), the d'_{Test} scores were normalized with respect to those obtained in the "control" experiment (d'_{Control}) and are expressed as $d'_{\text{Test}}/d'_{\text{Control}}$ ratios in Fig. 4.

With this format, a perfect perceptual invariance to temporal transposition will translate into flat functions relating $d'_{\text{Test}}/d'_{\text{Control}}$ to α. It should be also noted that if the effects of the two additional factors studied (*R* and *N*) were the same in the control and "test" experiments, computing $d'_{\text{Test}}/d'_{\text{Control}}$ ratios should cancel them out. Deviations from these predicted null effects of *R* and *N* would therefore indicate contributions of these factors different from those observed in the nontransposed case.

Based on the d' ratios shown in Fig. 4, the discrimination of temporally tranposed envelopes can be characterized as follows. It is an inverted U-shaped function of the transposition factor (with a peak at or just below $\alpha=1$) whatever

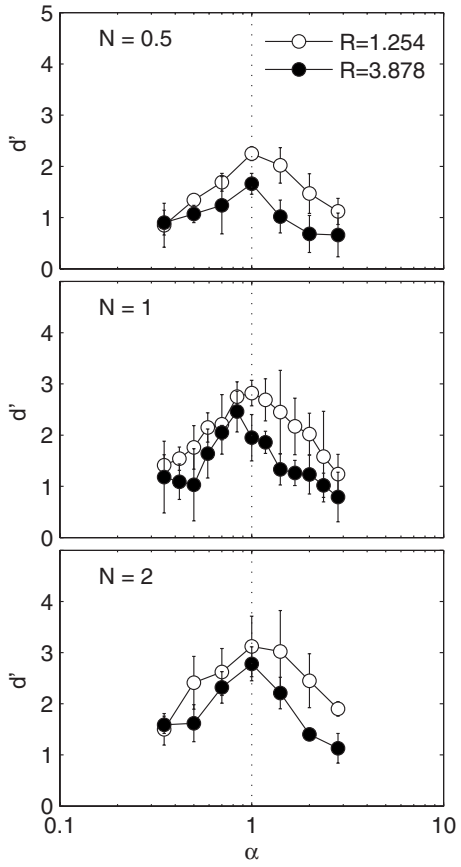


FIG. 3. Mean discrimination sensitivity (d') for four listeners obtained in the test condition. Discrimination performance is plotted as a function of the time compression/expansion factor, α. Here, the time compression/expansion factor is applied to the envelopes of A and B, only. For each value of R, vertical dotted lines indicate $\alpha=1$ Otherwise as in Fig. 2.

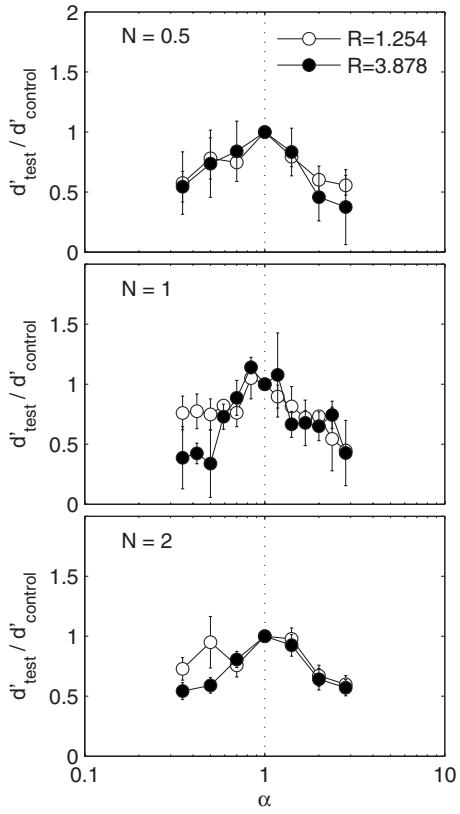


FIG. 4. Mean ratio of discrimination scores (d'_{Test} presented in Fig. 2) normalized with respect to those obtained in the "control" experiment (d'_{Control} presented in Fig. 1). The $d'_{\text{Test}}/d'_{\text{Control}}$ ratios are plotted as a function of the time compression/expansion factor, α. in each panel, the vertical dotted line indicates $\alpha=1$ Otherwise as in Fig. 2.

R or N. Given that perceptual constancy predicts that $d'_{\text{Test}}/d'_{\text{Control}}$ should be independent of α, ratios smaller than 1 indicate a sensitivity reduction due to the temporal transposition *per se*. For the extreme temporal expansion ($\alpha=0.35$) and compression ($\alpha=2.83$) values used, sensitivity drops by a factor of 1.32–2.7. The U-shaped functions of α are symmetrical for $R=3.878$ but temporal compression appears to be more detrimental than temporal expansion for $R=1.254$ (at least for $N=1$ and 2). With the exception of a limited temporal expansion range ($0.35<\alpha<0.5$), $d'_{\text{Test}}/d'_{\text{Control}}$ ratios are relatively independent of R, indicating that this factor contributes equally to the recognition of temporally transposed and non-transposed envelopes. $d'_{\text{Test}}/d'_{\text{Control}}$ ratios are also independent of N suggesting that this factor is also equally involved in the discrimination of temporally transposed envelopes and in the discrimination of nontransposed envelopes.

The previous observations are partially supported by a 3-way (α[7], R[2], N[3]) repeated measures ANOVA performed on the $d'_{\text{Test}}/d'_{\text{Control}}$ ratios. The effect of temporal compression/expansion factor α is significant [$F(6,18)=26.07$, $p<0.000001$], confirming the fact that temporally transposed envelopes are less well discriminated than nontransposed ones. Hence, contrary to previous studies that demonstrated a resistance of word or sentence identification to their temporally compressed/expanded versions (i.e., perceptual constancy; Fairbanks and Kodman, 1957; Fu *et al.*, 2001; Versfeld and Dreschler, 2002), the present data show a lack of temporal transposition constancy for nonlinguistic stimuli. For instance, Fu *et al.* (2001) showed that when a 32-channel vocoder was used to remove temporal fine-structure cues, time-expanded and time-compressed speech remained perfectly intelligible even at half (CE=100%) or two times (CE=50%) the normal speaking rate (equivalent to $\alpha=0.5$ and 2 in the present study, respectively). For such changes in α values in the present discrimination task, d' dropped by a factor of 1.3–2.4. The effect of the R-factor (presumably related to the resolvability of the envelopes' components) is also significant [$F(1,3)=20$, $p<0.05$]. This inference is qualified by the partial comparisons over the two R-levels showing a significant R-effect only for the largest temporal expansion ($\alpha=0.35$) and for the amplitude expanded ($N=2$) envelopes [$F(1,3)=12.22$, $p<0.05$]. These partial comparisons are in line with the significant $\alpha\times R$ interaction [$F(6,18)=3.15$, $p<0.05$]. The effect of the amplitude compression/expansion factor, N, is not significant [$F(2,6)=3.48$, $p=0.1$] and neither is the $\alpha\times N$ interaction [$F(12,36)=1.28$, $p=0.27$] or the $R\times N$ interaction [$F(2,6)=1.52$, $p=0.29$]. Finally, the triple interaction $\alpha\times R\times N$ is not significant either [$F(12,36)=1.02$ $p=0.45$]. Overall, the statistical analysis shows that perfect perceptual constancy is

not maintained for temporally transposed, nonlinguistic envelopes.

III. INTERIM DISCUSSION

The main results of the present study can be summarized as follows. The discrimination of two-component temporal envelopes *equally* compressed/expanded in time is maximized when their two modulation components are close in frequency and centered around 4–6 Hz, but is a monotonically decreasing function of the frequency of their modulation components when the latter are spaced apart (along the modulation frequency axis). Overall, discrimination scores are enhanced when the frequency spacing between the two modulation components is decreased and when the envelopes are expanded in amplitude. Discrimination of temporally transposed envelopes appears to preserve globally these characteristics while displaying a significant drop related to the amount of transposition (whether compression or expansion). Hence, at odds with previous studies that used linguistic stimuli, the present data suggest an absence of perfect perceptual constancy over temporal transpositions.

Effects of resolvability (R) *and duration* (D). The dependence of envelope discrimination on the frequency spacing of modulation components, R, is consistent with the existence of distinct temporal modulation filters. Indeed, the temporal-reversal discrimination task requires the encoding of the phase of the modulation components. On the assumption that temporally modulated signals are discriminated via a comparison (or cross correlation) of their temporal profiles, discrimination based on the phase of their components is possible as long as they feed into the same modulation filter but not otherwise. The R values used in the present experiment were chosen so that the two envelope components tap one ($R=1.254$) or two distinct ($R=3.878$) modulation filter(s) as they have been inferred from modulation masking experiments (e.g., Ewert and Dau, 2000). For these conditions, the modulation filterbank model hence predicts that discrimination of phase-reversed envelopes should be better for the smaller R, just as presently found. Sek and Moore (2003) who have measured the discrimination of two envelopes that differed only in the phase of one of their three components found a similar dependence on the frequency ratio of these components.

Inasmuch as the hypothetical modulation filters have a constant quality ratio, Q, the observed R-effect should be independent of the envelopes' central frequencies, f_c. The present data, however, show a significant $R \times f_c$ interaction, with the disappearance of the R-effect for the lower f_c values ($\alpha < 0.5$ that is $f_c < 1.5$ Hz: cf. first experiment and Fig. 2). To this we offer one possible interpretation relating to the duration of the stimuli. As noted in Sec. II A, in order to prevent listeners from using more than one envelope beat for their judgments, all envelopes were temporally windowed so that they included only one envelope beat period [$D = 1/\alpha(f_{m2}\text{-}f_{m1}) = 1/\alpha f_{m1}(R-1)$]. Hence, stimulus duration D was inversely proportional to both α and R (cf. Table I). Figures 5 and 6 replot the mean control and test data shown in Figs. 2 and 3, respectively, as a function of D (instead of α) in order to show the confounded effect of changes in duration on discrimination performance. In Fig. 5, the replotted control data (open and filled circles) reveal that discrimination performance is a nonmonotonic function of stimulus duration. More precisely, discrimination performance peaks at intermediate durations ranging from 735 to 1042 ms (corresponding to $\alpha = 1.41-2$, or $f_c = 4.2-6$ Hz). This seems consistent with the notion that, in the first (i.e., control) experiment, changes in stimulus duration are—at least partly—responsible for the effect of α or f_c (temporal compression/expansion). For instance, an increase in a listener's memory load or a decay of the sensory trace stored in auditory short-term memory could explain why envelope discrimination deteriorates for the longest duration. In Fig. 6, the replotted test data (open and filled circles) indicate that discrimination performance is a nonmonotonic function of stimulus duration. Discrimination performance peaks when all three stimuli of the XAB sequence have the same duration (as shown by vertical dotted lines), and degrades as a function of the departure in duration betwen the reference and comparison (target and standard) stimuli. Can changes in duration also account for the effect of R? For comparable duration intervals—that is for D between 326 and 521, 456 and 735, and 625 and 1042 ms—post-hoc comparisons (LSD–Least Significant Difference test) indicate that discrimination

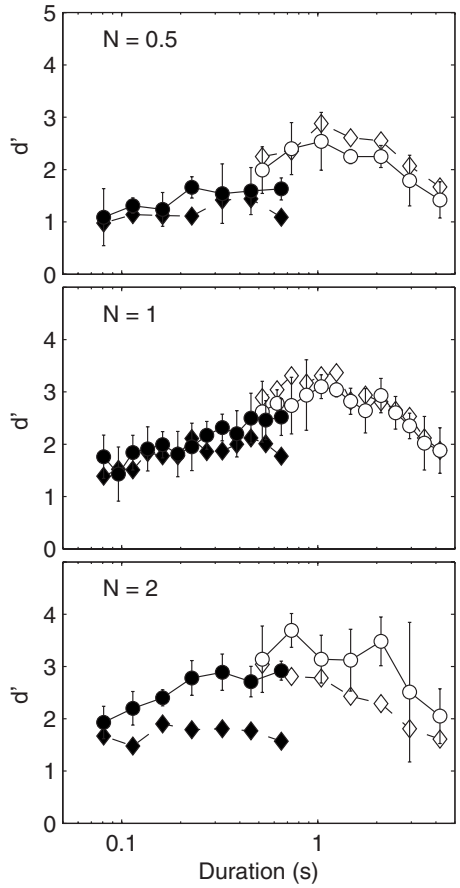


FIG. 5. Mean discrimination sensitivity (d'_{Control}) for the four listeners obtained in the control condition (open and filled circles). Discrimination performance is plotted as a function of stimulus duration, D (ms). Otherwise as in Fig. 2. Open and filled diamonds correspond to simulation data obtained with $R=1.254$ and 3.878, respectively.

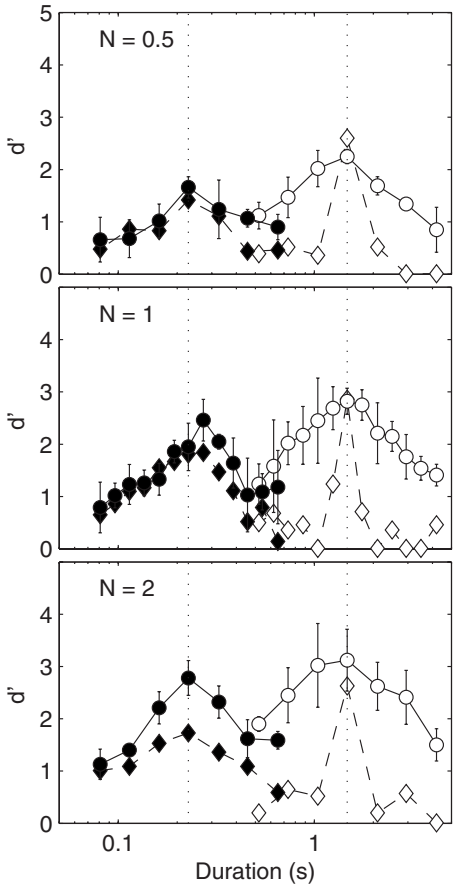


FIG. 6. Mean discrimination sensitivity (d'_{Test}) for the four listeners obtained in the test condition (open and filled circles). Discrimination performance is plotted as a function of stimulus duration, D (ms). Otherwise as in Fig. 2. Open and filled diamonds correspond to simulation data obtained with $R=1.254$ and 3.878, respectively. For each value of R, vertical dotted lines indicate $\alpha=1$.

scores obtained with $R=1.254$ are still significantly greater than those obtained with $R=3.878$ [$p<0.05$] for $N=0.5$, 1, and 2, except when $N=1$ and D is between 465 and 735 ms [$p=0.15$] and when $N=2$ and D is between 326 and 521 ms [$p=0.17$]. These tests hence sustain a genuine effect of components' resolvability. However, this effect is more apparent when the magnitude of envelope components is small (i.e., when $N=0.5$) and tends to disappear when envelope components are presented at levels (i.e., depths) well above detection threshold (i.e., when $N=1$ or 2). In addition, the effect of resolvability, when observed here, is relatively small in magnitude. Thus, in the present experiment, envelope discrimination performance seems to be more constrainted by stimulus duration (and presumably memory factors) than by envelope resolvability per se.

Effects of amplitude compression/expansion (N). The beneficial effect of envelope amplitude expansion indicates that complex envelopes discrimination depends on their overall peak-to-trough ratio. It can also be related to the notion that envelope discrimination is at least partly based on listeners using local features of the envelope, particularly their local peaks, as suggested by a speech-perception study conducted by Drullman (1995). Indeed, the effect of raising the envelope amplitude by a power larger than 1 is equivalent to reinforcing its peaks (relative to troughs). Effects of amplitude expansion are also found for speech signals presented in noise (e.g., Fu and Shannon, 1999; Lorenzi *et al.*, 1999; Apoux *et al.*, 2001). Moreover, amplitude expansion is "naturally" observed in hearing-impaired listeners as a consequence of the loss of fast-acting cochlear compression. On a more audiological side, this suggests that peripheral amplitude compression (and its loss in the case of coch-le-ar lesions) affects not only detection (as shown previously for hearing-impaired listeners, e.g., Moore *et al.*, 1992) but also discrimination. The current results predict therefore that hearing-impaired listeners with loudness recruitment should show better-than-normal ability to discriminate between complex temporal envelopes of linguistic and nonlinguistic stimuli.

Perceptual constancy for envelope discrimination? The present study demonstrates a strong limitation in the discrimination of temporally compressed or expanded nonlinguistic envelopes regardless of their amplitude expansion. In fact, the data (Figs. 3 and 4) show a discrimination deterioration even for the smallest temporal compression/expansion used (CE: 16% and 18%; $\alpha=0.84$ or 1.18).

This lack of constancy for temporally transposed nonlinguistic temporal envelopes appears to be at odds with the constancy reported for both linguistic and musical signals. Indeed, identification of temporally transposed linguistic signals remains unaffected by transposition up to a compression/expansion index (CE) of 50% (e.g. Fairbanks and Kodman, 1957; Daniloff *et al.*, 1968; Vaughan and Letowski, 1997; Gordon-Salant and Fitzgibbons, 2001; Versfeld and Dreschler, 2002). Some studies on categorical perception of phonemes also seem to provide evidence for the existence of some form of temporal normalization (Summerfield, 1981; Miller and Volaitis, 1989). It may then be argued that the current discrepancy is related to the fact that linguistic signals are coded by a speech-specific system (Liberman and Mattingly, 1985) that may well be designed so as to resist temporal alterations. A resistance to temporal alterations has also been reported for musical sequences [as long as the duration of their component notes is within a 160–1280 ms interval (Warren *et al.*, 1991)] hence rejecting the singularity of the speech coding system.

The alternative, more plausible account of this discrepancy is that previous studies have compared categorization performance for reference and transposed signals under conditions where the former were always discriminable (e.g., Fairbanks and Kodman, 1957; Daniloff *et al.*, 1968; Fu *et al.*, 2001). It may then well be that, though degraded, performance for the transposed signal did not show a measurable drop due to a ceiling effect. This putative methodological concern was circumvented in the present study by utilizing a reference task in which performance was below 100% correct (i.e., a d' not larger than 4; see Fig. 2).

IV. MODEL PREDICTIONS

The present data show that the discrimination of nonlinguistic temporal envelopes is degraded by temporal transpositions. Hence, perfect perceptual constancy is not achieved

for time-stretched or time-compressed random envelopes. It is unclear, however, whether the observed degradation is consistent with the total absence of perceptual constancy in the envelope domain, or whether it still requires some sort of normalization mechanism. To investigate this issue, we now present a qualitative modeling study in which we compare listeners' performances with the predictions of an envelope cross correlator after auditory filtering. The cross correlator did not include any normalization stage. We could obtain a good fit to the control data, which indicates that envelope cross correlation was sufficient to explain behavioral performance when comparing stimuli with the same duration. The model failed however in the test conditions, strongly suggesting the need for an additional normalization stage when stimuli have different durations.

Model structure. The model was an envelope extractor with a limited memory store followed by a cross-correlation decision stage. The first stage was a single linear gammatone filter that simulated the band pass filtering at one locus on the basilar membrane (Patterson *et al.*, 1987). In a second stage, the temporal envelope of the band pass-filtered signal was extracted using half-wave rectification followed by lowpass filtering [cutoff=64 Hz, rolloff=6 dB/oct (see Viemeister 1979)]. The envelope obtained was then temporally windowed with an exponential function in order to simulate a decay of the memory trace. A similar approach to account for memory constraint was taken by Sheft and Yost (2005).

The decision stage was realized as a cross correlation between the windowed envelopes. On each trial, the output of the model for the three stimuli (X, A, and B) was computed. The windowed envelope of the reference stimulus, X, served as a "template" that was cross correlated with the windowed envelopes of A and B. The response was determined by the largest cross-correlation coefficient (X better correlated to A or X better correlated to B). Note that this differed from a simple Pearson product-moment correlation in two important ways. First, the correlation was applied on the envelopes including the direct current component. The measure was thus sensitive to modulation depth to some extent. Second, the whole cross-correlation function was computed so envelopes were effectively time shifted to find the highest correlation. This approach was very similar to that used by van de Par and Kohlrausch (1998) to model monaural and binaural envelope correlation detection, and it could be viewed as simplified version of the optimal detector described in Dau *et al.* (1997a, b).

Stimuli were generated as in the behavioral experiment, except that no level rove was applied. Center frequency f_c (or, equivalently, duration) was roved across trials just as in the behavioral experiment. Six hundred trials were simulated for each condition. To restrict the numbers of degrees of freedom in the model, no internal noise was added. The noise carrier, refreshed for each interval, was thus the sole source of variability in the predictions for a given set of stimulus parameters. The randomization of modulation depth, phase, and f_c are other sources of variability across trials. Percent correct was transformed into d'. The half-life of the exponential window and the gammatone center frequency were varied to fit the data in the control condition, where stimuli had identical durations within each trial.

Model results, control condition. Fits were obtained by minimizing the root mean square (rms) error between experimental data and model predictions for the two R values and for $N=1$. The best fit was obtained for a half-life of 1.2 s and a filter center frequency of 5 kHz (Fig. 5). Model predictions for these parameters (open and filled diamonds) and empirical data (open and filled circles) for the control conditions are shown in Fig. 5. The results have been replotted as a function of the duration of the stimuli. As indicated earlier, this duration covaried with R, except for a few values where the same duration could be obtained with two different R's. Most predicted values fell within the variability range of the empirical data for $N=1$. There was also a relatively good fit for $N=0.5$, even though the parameters were not optimized for this condition. The fit was poorer for $N=2$, where the model consistently underestimated performance.

The discrimination performance peaked at intermediate stimulus durations. In the model, this was because performance first increases with stimulus duration and then decreases because of the exponential weighting window, which limits the maximum stimulus duration that can be accurately stored. Performance would increase indefinitely with stimulus duration without such a window, because d' increases by the square root of duration for a correlation receiver. For any given duration, the model also predicted poorer discrimination for high R compared to low R. The poorer discrimination for high R was also observed in the listeners' discrimination scores, although the model overestimates the effect. It is noteworthy that the model predicted an effect of R without any modulation filterbank and thus without any notion of modulation frequency resolvability. We hypothesize that the model's behavior for these points is related to the complexity of the envelope pattern. For low R, there are more distinct features in the envelope, as illustrated in Fig. 1 (left versus right column). The decision stage of the model has to extract a signal template from the noisy stimulus, so having many peaks in the envelope will make this stage more resistant to noise. This is less the case for high R where the envelopes are broadly similar, without sharp features, and hence more susceptible to noise. Such an interpretation of the model's behavior remains speculative and should be verified by further testing.

Overall, the simulations show that the effects of α, f_c (or duration) and R on complex envelope discrimination observed in the control experiment, where all stimuli within a trial have the same duration, can be accounted for reasonably well by a simple model of envelope cross-correlation with a memory limit.

Model results, test condition. Figure 6 shows empirical data (open and filled circles) and model predictions (with the half-life time parameter used to simulate control data) for the test condition, again plotted as a function of duration. Model parameters were kept identical to the ones used for the control condition. The model predictions (open and filled diamonds) always peaked at a value corresponding to $\alpha=1$ (indicated by vertical dotted lines for each value of R). Such a

value represents the case where reference and comparison stimuli have the same duration. It is not surprising that the model should perform well in these conditions as these are similar to control conditions. For other values of α, predicted performance decreased, for each value of R. The same trend was observed in listeners' performances. For $N=0.5$ and $N=1$ and for small durations (high R), the model accurately predicted the rate of decrease in performance due to the mismatch in durations. Crucially, however, for long durations (low R) the rate of decrease was much faster in the model's predictions than in the listeners' data. This suggests that, for long durations, envelope cross-correlation underestimates listeners' performance. A different mechanism or an additional, as yet unspecified normalization stage is thus needed to account for listeners' performance.

Model discussion. The aim of the present model was to illustrate the prediction of an envelope-correlation approach when comparing two random envelopes. Such an approach has been used before in the context of envelope perception (Dau *et al.*, 1997a, b, van de Par and Kohlrausch, 1998; Sheft and Yost, 2005). The main finding of the present study is that envelope correlation predicts well behavioral performance when stimuli durations are equal, but fails when durations are unequal. In order to keep the focus on the predictions of envelope correlation in the context of perceptual constancy, we tried to keep the model as simple as possible. For instance, no adaptation or compression front-end was used, even though such processing would affect model behavior (Derleth *et al.*, 2001). We now examine briefly this and other choices made in the modeling and show that they do not bear on our general conclusion.

No attempt was made to model the influence of N or of the level rove imposed on the stimulus. Adaptation is important to account for these parameters in at least two ways. First, static compression would change the effective peak to trough ratio, as well as the effect of the level rove. Second, dynamic changes in the adaptation characteristics would result in different behavior for forward and reversed envelopes. Accurately capturing these effects in a model would require adding a realistic front-end with respect to absolute and dynamical changes in level. Although this would be of interest for future modeling studies to better account for performance in the control conditions, it is unlikely that such a front-end would change anything regarding the failure to predict performance in the test conditions where stimuli durations are unequal.

The choice of the auditory filter considered was based on the optimization of the fit between model and experimental data, and it was found that a center frequency of 5 kHz provided the best fit. The relatively wide bandwidth (ERB $=564$ Hz) of the 5 kHz filter minimizes two disruptive effects on envelope perception resulting from band pass filtering, that is, envelope filtering, and masking produced by the intrinsic envelope fluctuations of the noise carrier. Figure 7 illustrates the quality of the fit between model and data when the half-life of the exponential memory window is varied, with filter center frequency as a parameter. Low-frequency filters produced a worse fit to the data (rms error, upper panel) and a lower performance overall (mean error, lower panel). We hypothesize that listeners would ignore the low-frequency filters and listen to frequency regions providing more reliable cues. The fit also got worse for filters with frequencies above 5 kHz, but for a different reason. In this case, the model predicted higher performance than observed behaviorally. Note however that the model did not include any source of internal noise. Results similar to what we are presenting could be obtained by choosing the most accurate envelope representation, that is the highest available auditory filter, and then adding a source of internal noise. Again, this would introduce additional parameters to the model without affecting the general conclusion.

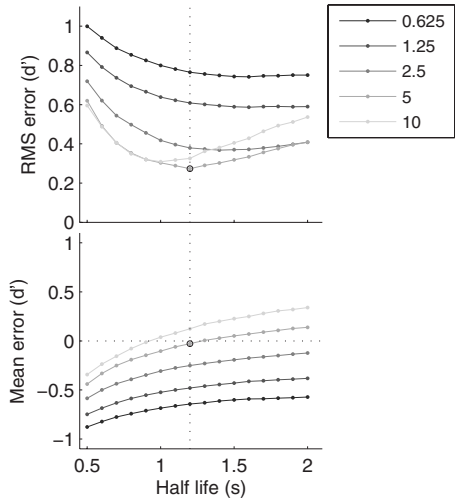


FIG. 7. Influence of model parameters on the quality of fit to the behavioral data, for $R=1.254$ and 3.878 and $N=1$. rms error (top) and mean error (bottom) are plotted as a function of the half-life of the exponential window applied to the envelopes. Each shade of gray indicates a different auditory filter frequency (0.625, 1.25, 2.5, 5, or 10 kHz). The best fit, half-life $=1.2$ s and center frequency$=5$ kHz, is indicated by circles.

A model applicable to the comparison of temporal patterns of different length has been proposed by Sorkin and colleagues (Sorkin and Montgomery, 1991; Sorkin *et al.*, 1994). Sorkin and colleagues used tone sequences rather than envelopes and, in the case of equal stimulus durations, listeners' performances could be predicted by cross correlating the sequences of onset times after introduction of an internal noise. In order to account for performance with stretched or compressed sequences, Sorkin and Montgomery (1991) assumed a normalization of all sequences to the same duration, but the internal noise was proportional to the amount of normalization required. Such a model is based on correlations of time-of-occurrence between salient features in the sequence, the tone onsets in the case of Sorkin and Montgomery (1991). Applying it to the comparison of temporal envelopes would require extracting 'features' from the continuous envelope function, as was in fact proposed by Sheft and Yost (2005). An interesting future direction for modeling the test data of the present experiment would thus be to apply a noisy normalization mechanism, similar to Sorkin and colleagues, to salient features of the internal envelope.

V. GENERAL DISCUSSION

The discrimination of nontransposed temporal envelopes (that is envelopes of identical duration) can be accurately

accounted for by an auditory model using an envelope extraction stage similar to that proposed by Viemeister (1979) followed by a template-matching process similar to that proposed by Dau *et al.* (1997a, b). Although beyond the goal of the present study, the empirical and simulated results obtained in these control discrimination experiments suggest that envelope filtering via selective modulation filters such as those described initially by Dau *et al.* (1997a, b) is not a necessary prerequisite for the discrimination of equal-duration time-reversed envelopes.

The poor discrimination of temporally transposed, nonlinguistic envelopes suggested—at a first sight—a total absence of perceptual constancy to be contrasted with previous work on speech recognition. The comparison of the current psychoacoustical and modeling data argues nevertheless in favor of the existence of some form of (incomplete) normalization, whose effects are mainly visible for envelopes of long duration and highly contrasted envelopes (as produced by amplitude expansion). Detailed inspection of the longest envelopes indicates that they display more "local" features (i.e., primary and secondary peaks and troughs) than the shorter ones. This suggests that within each trial, perceptual constancy may be achieved, although imperfectly, by comparing the temporal sequences of envelope peaks and troughs across stimuli when these local features are numerous and salient enough. These conjectures warrant further experimental investigation.

In light of the present results, the resistance to temporal alterations of speech and music signals reported in previous studies may result from the operation of these normalization and template-matching processes. Compared to the stimuli of the current study, the higher level of redundancy of speech and music may account for the improvement in perceptual constancy with these stimuli.. In addition or alternatively, the possibility still remains that the reported constancy reflected partly—as suggested in Secs. I and V—a ceiling effect artifact.

VI. CONCLUSIONS

The current research investigated perceptual constancy in the temporal-envelope domain using nonlinguistic stimuli. Taken together, the psychophysical results indicate that the discrimination of temporally transposed envelopes degrades continuously as a function of the degree of temporal transposition. At least for moderate temporal expansion/compression rates, this deterioration is only slightly modulated by manipulations of stimulus parameters (frequency spacing between envelope components, peak-to-trough ratio of the envelopes) shown to influence the discrimination of (nontransposed) complex temporal envelope patterns.

A quantitative model of temporal envelope processing using a memory-limited envelope extraction stage followed by a template-matching process accounts for the discrimination of equal-duration envelopes, but generally underestimates listeners' discrimination of temporally transposed envelopes for the longest stimuli. This suggests that the auditory system applies some form of incomplete normalization to the temporal envelopes of incoming sounds, whether linguistic or nonlinguistic in nature.

ACKNOWLEDGMENTS

This research was supported by a MENRT grant to M. Ardoint, a grant from the Institut Universitaire de France to C. Lorenzi, an ANR grant (ANR-06-NEURO-022-01) to D. Pressnitzer, and an ANR grant (ANR-06-NEURO-042-01) to A. Gorea. The authors thank two anonymous reviewers for helpful comments on an earlier version of this manuscript.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (**2001**). "Speech comprehension is correlated with temporal response patterns recorded from auditory cortex," Proc. Nat. Acad. Soc. **98**, 13367–13372.

Apoux, F., Crouzet, O., and Lorenzi, C. (**2001**). "Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: Effects on identification performance and response times," Hear. Res. **153**, 123–131.

Daniloff, R., Shriner, T. H., and Zemlin, W. R. (**1968**). "Intelligibility of vowels altered in duration and frequency," J. Acoust. Soc. Am. **44**, 700–707.

Dau, T., Kollmieier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Dectection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2893–2905.

Dau, T., Kollmieier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Derleth, R. P., Dau, T., and Kollmeier, B. (**2001**). "Modeling temporal and compressive properties of the normal and impaired auditory system," Hear. Res. **159**, 132–149.

Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Fairbanks, G., and Kodman, F. (**1957**). "Word intelligibility as a function of time compression," J. Acoust. Soc. Am. **29**, 636–641.

Fu, Q.-J., Galvin, J. J., and Wang, X. (**2001**). "Recognition of time-distorted sentences by normal-hearing and cochlear-implant listeners," J. Acoust. Soc. Am. **109**, 379–384.

Fu, Q. J., and Shannon, R. V. (**1999**). "Recognition of spectrally-degraded speech in noise with nonlinear amplitude-mapping," Proceedings of the 1999 IEEE (Institute of Electrical and Electronics Engineers) ICASSP (International Conference on Acoustics, Speech, and Signal Processing) , Vol. **1**, pp. 369–372.

Gockel, H., and Colonius, H. (**1997**). "Auditory profile analysis: Is there perceptual constancy for spectral shape for stimuli roved in frequency?" J. Acoust. Soc. Am. **102**, 2311–2315.

Gordon-Salant, S., and Fitzgibbons, P. F. (**2001**). "Sources of age-related recognition difficulty for time-compressed speech," J. Speech Lang. Hear. Res. **44**, 709–719.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Li, X., and Pastore, R. E. (**1995**). "Perceptual constancy of a global spectral property: Spectral slope discrimination," J. Acoust. Soc. Am. **98**, 1956–1968.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Lorenzi, C., Berthommier, F., Apoux, F., and Bacri, N. (**1999**). "Effects of envelope expansion on speech recognition," Hear. Res. **136**, 131–138.

Lorenzi, C., Soares, C., and Vonner, T. (**2001**). "Second order temporal modulation transfer functions," J. Acoust. Soc. Am. **110**, 1030–1038.

MacMillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

Miller, J. L., and Volaitis, L. E. (**1989**). "Effect of speaking rate on the perceptual structure of a phonetic category," Percept. Psychophys. **46**, 505–512.

Moore, B. C. J., Shailer, M. J., and Schooneveldt, G. P. (**1992**). "Temporal modulation transfer functions for band-limited noise in subjects with co-

chlear hearing loss," Br. J. Audiol. **26**, 229–237.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," presented at the Meeting of the IOC Speech Group on Auditory Modeling at RSRE (Royal Signals and Radar Establishment), 14–15 December.
Sek, A., and Moore, B. C. (**2003**). "Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators," J. Acoust. Soc. Am. **113**, 2801–2811.
Sheft, S., and Yost, W. A. (**2005**). "Minimum integration times for processing of amplitude modulation," *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 244–250.
Sorkin, R. D., and Montgomery, D. A. (**1991**). "Effect of time compression and expansion on the discrimination of tonal patterns," J. Acoust. Soc. Am. **90**, 846–857.
Sorkin, R. D., Montgomery, D. A., and Sadralodabai, T. (**1994**). "Effect of sequence delay on the discrimination of temporal patterns," J. Acoust. Soc. Am. **96**, 2148–2155.
Summerfield, Q. (**1981**). "Articulatory Rate and Perceptual Constancy in Phonetic Perception," J. Exp. Psychol. Hum. Percept. Perform. **5**, 1074–1095.
Takeuchi, A. H., and Braida, L. D. (**1995**). "Effect of frequency transposition on the discrimination of amplitude envelope patterns," J. Acoust. Soc. Am. **97**, 453–460.
van de Par, S., and Kohlrausch, A. (**1998**). "Analytical expressions for the envelope correlation of narrow-band stimuli used in CMR and BMLD research," J. Acoust. Soc. Am. **103**, 3605–3620.
Vaughan, N. E., and Letowski, T. (**1997**). "Effects of age, speech rate, and type of test on temporal auditory processing," J. Speech Lang. Hear. Res. **40**, 1192–1200.
Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am. **111**, 401–408.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Warren, R. M., Gardner, D. A., Brubaker, B. S., and Bashford, J. A. (**1991**). "Melodic and nonmelodic sequences of tones: effects of duration on perception," Music Percept. **8**, 277–290.

Detection of combined changes in interaural time and intensity differences: Segregated mechanisms in cue type and in operating frequency range?[a)]

Shigeto Furukawa[b)]
Human and Information Science Laboratory, NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan

(Received 14 May 2007; revised 13 November 2007; accepted 26 December 2007)

Although physiological studies have revealed segregated binaural pathways, namely the medial and lateral superior olives, it is unclear whether the human auditory system has separate mechanisms for different cue types (interaural time and intensity differences; ITD and IID, respectively) and for operating frequency ranges. This study hypothesized "channels" for ITD and IID processing, and examined channel interaction at low and high frequencies based on the signal detection theory. The stimuli were a 125- or 500-Hz tone and a 4-kHz tone amplitude-modulated with a half-wave-rectified 125-Hz sinusoid, presented dichotically with various baseline ITDs and IIDs. The detectability indices, d', for ITD and IID changes, imposed individually or simultaneously in the same direction, were derived from the results of a forced-choice task. The degree of channel interaction was estimated by comparing d' for combined cues with those for individual cues. The estimated interaction showed little effect of baseline ITD or IID. The results generally exhibited nonzero interaction, indicating that the cue processes are not completely independent. The interaction was stronger for high frequencies than for low frequencies. The results can be interpreted as indicating the involvement of different binaural mechanisms for different frequency regions.
 [DOI: 10.1121/1.2835226]

PACS number(s): 43.66.Pn, 43.66.Qp [RLF] Pages: 1602–1617

I. INTRODUCTION

Traditionally, the auditory system has often been regarded as having two separate pathways. The "duplex theory" of sound localization, which dates back to Lord Rayleigh (1907), states that the primary cues to the lateral position of a sound source differ for low and high frequency signals. For low frequency signals (<1000 Hz), human listeners are sensitive to ongoing phase differences between ears, and thus the interaural time differences (ITDs) could be the primary cue for localization. For high frequency signals, acoustic shadowing by the head makes the interaural intensity differences (IIDs) the primary cue. Results of physiological studies are broadly consistent with the duplex theory [see Irvine (1992) for a review]. Low frequencies are overrepresented in the medial superior olivary nucleus (MSO), in which primarily ITD is coded by cells predominantly with an excitatory–excitatory type of binaural response. In contrast, high frequencies are overrepresented in the lateral superior olivary nucleus (LSO), in which cells are sensitive to IID, mediated by the excitatory–inhibitory mechanism.

Detailed physiological studies, however, indicate that this view oversimplifies the actual binaural processes. For example, the MSO has an appreciable amount of inhibitory input from the contralateral ear (Yin, 2002), and has neurons sensitive to high frequency ITD (Spitzer and Semple, 1995; Batra *et al.*, 1997a, b). The LSO also has a population of low frequency neurons that are binaurally sensitive (Tollin, 2003). It should be noted that the function and relative contribution of the two nuclei can vary between species (Grothe, 2000). Therefore, the existing physiological data provide limited information as to the degree of functional segregation in the *human* auditory system. Although human neurophysiological studies (e.g., Schröger, 1996; Pratt *et al.*, 1997; Tardif *et al.*, 2006) and a study of a patient with a focal brain lesion (Griffiths *et al.*, 1998) have shown some indications of separate pathways for ITD and IID processing, they were not designed to reveal frequency-dependent processes; they used either low frequency tones or broadband complex signals.

This study was driven by an interest in the segregation of cue processing mechanisms in the binaural pathway. A particular focus was placed on whether or not the binaural mechanism depends on the spectral regions of the stimuli. Existing psychophysical data have generally indicated that the between-frequency differences in sensitivity to binaural cues are more quantitative than qualitative. For example, the detection performance as regards IID changes, when measured by using headphone stimulations, is equally good for low and high frequency signals, although the performance worsens somewhat for frequencies around 1000 Hz (Grantham, 1984; Yost and Dye, 1988). Also, even for high frequency signals, listeners exhibit sensitivity to ongoing ITD when it is represented by the stimulus envelope (e.g., Henning, 1974; Nuetzel and Hafter, 1981; Hafter and Dye, 1983; Bernstein and Trahiotis, 2002). This leads to the argument that the mechanisms underlying ITD processing at low

[a)]Portions of this work were presented at the 30th midwinter meeting of the Association for Research in Otolaryngology, Denver, CO, 10–17 February 2007.
[b)]Electronic mail: shig@avg.brl.ntt.co.jp

and high frequencies are essentially the same, and that differences in ITD sensitivity between frequency ranges can be attributed to the peripheral efficiency with which timing information is encoded (Colburn and Esquissaud, 1976; van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002; Zhang and Wright, 2007).

Existing binaural models have also been unable to provide a definitive answer regarding the segregation of low and high frequency mechanisms. Currently, the main models that incorporate human sensitivities in ITD and IID can be classified into two broad categories in terms of their underlying mechanisms. One is based on the interaural cross-correlation mechanism (Stern and Colburn, 1978; Lindemann, 1986; Gaik, 1993). It can be considered that the computation of a cross-correlation function is performed by the excitatory–excitatory type of binaural interaction with delay lines. The cross-correlation mechanism would process primarily ITD information. In the model, the IID modulates the cross-correlation function via an additional mechanism, thereby accounting for human sensitivity to IID. The other class is the equalization-cancellation mechanism, in which the excitation by an input from one ear is inhibited by an input from the other ear (the "E–I" type interaction) (Breebaart *et al.*, 2001). This type of mechanism could be sensitive to both ITD and IID by incorporating the transmission delay of excitatory and inhibitory inputs. Either class of model can account for various binaural phenomena almost equally well for low and high frequency signals, and thus the modeling studies have not provided information as to whether different mechanisms operate for different frequency regions.

To examine the frequency dependence of binaural mechanism, the present study focused on the interaction between ITD and IID. Cue interaction was previously examined by using various paradigms. A well-known approach is to measure a listener's lateralization or localization judgments as a joint function of ITD and IID (e.g., Harris, 1960; Domnitz and Colburn, 1977; Macpherson and Middlebrooks, 2002). The studies adopting this approach have shown that ITD-based and IID-based lateralization or localization effects are to some extent tradable or additive, with the relative weights on the cues depending on the stimulus condition. This indicates that the two binaural cues are processed by a common mechanism at some stage(s) in the central auditory system. Supporting this notion, studies on the binaural masking level difference (e.g., Hafter and Carrier, 1970; Yost, 1972) and on just-noticeable differences in ITD or IID at various combinations of baseline ITD and IID (Domnitz and Colburn, 1977) revealed marked ITD and IID interactions. There is evidence, however, that those two cues are not entirely processed by a common mechanism. Hafter and Carrier (1972) and Ruotolo *et al.* (1979) showed that a signal with a certain ITD and an opposing IID that would centralize the intracranial image could be distinguished from a diotic signal, i.e., ITD cannot be completely traded by an opposing IID. This indicates the existence of a mechanism in which the ITD and IID information is treated differently. There has also been a report indicating that thresholds for detecting IID changes are somewhat independent of interaural correlation (Hartmann and Constan, 2002), although the threshold IIDs are markedly elevated when the stimulus ITD roved stimulus by stimulus (Bernstein, 2004). Wright and Fitzgerald (2001) and Zhang and Wright (2007) showed that the patterns of discrimination learning along multihour training were different for the two cues. In summary, the combined results of these earlier studies imply that there is an appreciable interaction between the ITD and IID in the central auditory system, but they are not entirely processed by common mechanism(s). However, those studies were generally not designed to examine the extent or the manner of cue interaction in such a way that would enable comparisons between frequency regions. In addition, the majority of studies used disproportionately low frequency signals (typically about 500 Hz), and thus there is little information about the cue interaction at high frequencies.

The present study used low frequency (typically 500-Hz) tonal signals and high frequency "transposed stimuli" (van de Par and Kohlrausch, 1997; Bernstein, 2001) and measured the detectabilities of changes in ITD alone, IID alone, and their combination. The relationship between the detectabilities of individual and combined cue changes was considered to reflect certain properties of the binaural mechanisms that cause cue interaction. Since the peripheral representations of the tonal signal and the transposed stimuli are considered to be comparable, any across-frequency differences in the results could be regarded as indicating differences in the binaural mechanism. In order to evaluate the effects of cue combination, the present study adopted a framework that incorporates the signal detection theory (Green, 1958; Green and Swets, 1974) with a hypothesis of the linear additivity of the cues. This framework was chosen because it deals with (hypothetical) internal noise sources attributed to both individual processes for the two cues, and common processes, and thus is sensitive to the relative contribution of independent mechanisms as regards individual cues, if any. In addition, the present framework is convenient because, as described in the following, it provides a single measure, ρ, which summarizes the effect of cue combination on the psychophysical performance, and thus enables us to compare signal frequencies.

In the framework, hypothetical "channels" for ITD and IID processing were considered, and the processes in the channels were characterized by the *signal* (strength of channel activity related to cue strength) and by the internal *noise*. It was assumed that the information from the two channels is combined linearly, at least for a small, near-threshold range of ITD or IID changes. As working hypotheses, let us consider two rather extreme cases where the listener detects ITD and IID changes in the *same* direction. One case assumes that there are separate channels for ITD and IID with independent noise sources and that the decision device combines information from those channels linearly and optimally [Fig. 1(a); the *independent channel hypothesis*]. In this case, the detectability index, d', for simultaneous changes of ITD and IID would be the root mean square sum of the d' values for individual changes. The other extreme case assumes a single channel (thus, a single noise source) that processes both ITD and IID, in which the ITD and IID information is optimally weighted and combined linearly [Fig. 1(c); the *common*

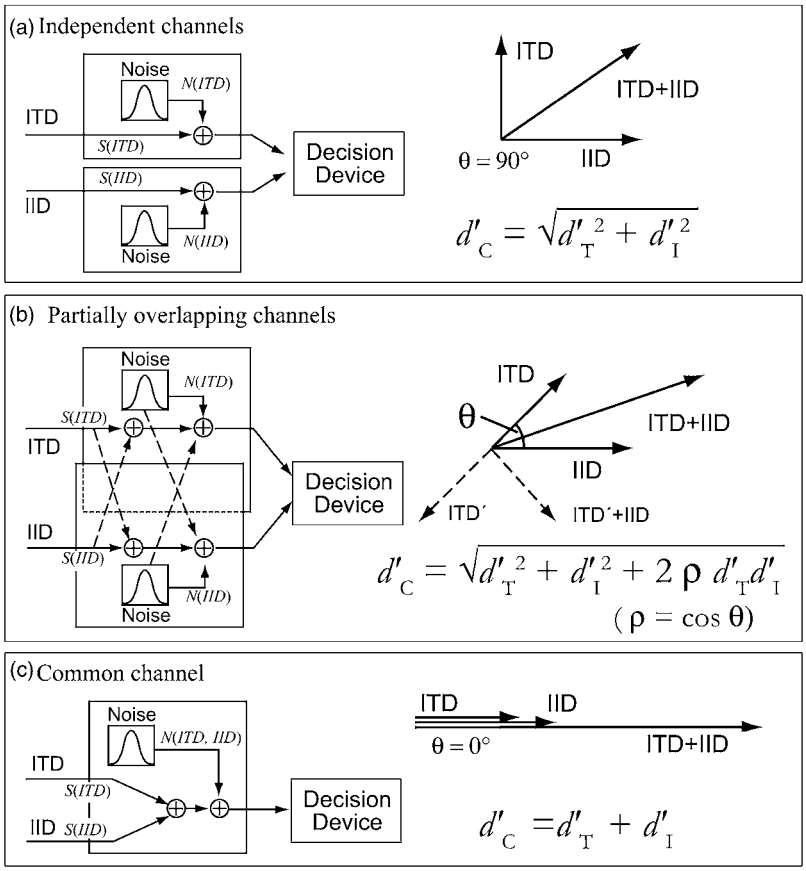


FIG. 1. Illustration of the working model used for evaluating cue interaction. The model assumes channels for ITD and IID processing, which overlap each other to various degrees. Each channel has internal *noise*, which is added to the *signal* that is related to cue strength. The magnitudes of the internal noise and the signal in the ITD and IID channels depend on the input ITD and IID, respectively, and on the stimulus frequency. The degree of channel overlap also depends on the stimulus frequency. The outputs from the channels are added linearly. In the present task of detecting ITD and/or IID change, the decision device was assumed to base its judgments on changes in the weighted linear sum. It was also assumed that the weights on the channel outputs are chosen to optimize the performance. (a)–(c) Examples arranged in increasing order of the degree of channel overlap. Each panel shows a somewhat conceptual illustration of the model (left), its vector expression (right), and a formula expressing the expected relationship between d' for combined cues (d'_C) and d' for individual cues (d'_T,d'_I). See the text for details. (a) An extreme case in which the two channels are completely independent (*independent channel hypothesis*). Each of the channels has independent noise sources. In this case, the dimensions for ITD and IID are thought of as orthogonal, and the d' value for simultaneous and consonant changes of ITD and IID would be the root mean square sum of the d' values for individual changes. (b) The two channels partially overlap. Fractions of the signal and the noise in one channel are added mutually to the other channel. The degree of channel overlap can be expressed as the angle of the ITD and IID dimensions, θ (see the vector diagram of solid-line arrows). The cos θ value is equal to the correlation coefficient ρ for the noise at the outputs of the two channels. In this working model, the ρ value is expected to range between 0 [channels are independent; (a)] and 1 [ITD and IID are processed in a common channel; (c)]. The vectors with the dashed-line arrows represent a case where the "effective" direction of an ITD is reversed in relation to the "cue-reversal point" (see Sec. IV C). (c) ITD and IID information is combined linearly in a single channel, in which the common internal noise is added to the signal (*common channel hypothesis*). Equivalently, ITD and IID are represented by a single dimension. The d' value for combined ITD and IID changes is the simple sum of the d' values for individual changes.

channel hypothesis]. Here, the d' for combined ITD and IID changes is simply the sum of the d''s for individual changes.

A general case that is intermediate between the two hypotheses is that the two channels partially overlap. That is, fractions of the signal and the noise in one channel are added to those in the other channel, and thus the noises at the two channel outputs are correlated to some degree [Fig. 1(b)]. In this general case, the d' values for combined and individual changes of ITD and IID are as follows:

$$d'_C = \sqrt{d'^2_T + d'^2_I + 2\rho d'_I d'_I}, \tag{1}$$

where d'_C is the detectability index for the combined changes of ITD and IID, and d'_T and d'_I, respectively, are the detectability indices when the ITD and IID changes are presented alone. The parameter ρ is the "correlation coefficient" that indicates the degree of channel interaction: ρ is the correlation coefficient for the internal noises that limit ITD and IID performance, respectively (Hafter *et al.*, 1990), and is also the proportion of signal strength in one channel that is added to the signal in the other channel. As illustrated by the vector expressions in Fig. 1(b) (see the vector diagram with solid-line arrows), the parameter ρ can be viewed as representing the orthogonality of the ITD and IID dimensions, cos θ, where θ is the angle between the perceptual dimensions for ITD and IID (Cohen, 1997). ρ=0 (θ=90°) corresponds to the independent channel hypothesis, and ρ=1 (θ=0°) corresponds to the common channel hypothesis.[1] Thus, ρ is expected to range between 0 and 1 (it is assumed that consonant ITD and IID changes have an additive rather than a subtractive effect). It should be noted that the prediction by Eq. (1) indicates the upper bound of the d'_C for given d'_T, d'_I, and ρ, i.e., the weights for the two channels are chosen to

maximize d'_C. Thus, $\rho > 0$ would invalidate the hypothesis of completely independent processes, implying the existence of a common process in which ITD and IID information interacts. The empirical estimates of ρ based on the results of detection experiments can fall outside the range between 0 and 1, e.g., $\rho > 1$, which corresponds to $d'_C > d'_T + d'_I$. If this were to be the case, it would indicate the failure of the assumption of linear additivity: The cue interactions in the binaural pathway could not be approximated by the linear addition of ITD and IID information. Unless the nature of the nonlinearity is specified, the absolute value of ρ would provide insufficient information as to whether the channels are completely overlapping (i.e., a common channel). Nonetheless, the ρ value is a measure that relates the d' value for the combined cue to those for individual cues: A larger ρ indicates a larger increase in d' when two cues are combined relative to d' for an individual cue. Thus, the ρ value remains a convenient quantity that summarizes the effect of combining cues on detectability.

Two studies have employed comparable frameworks for evaluating channel interaction. Gilliom and Sorkin (1972) performed an experiment on the discrimination between stimuli differing along ITD and/or IID dimensions using 500-Hz tones. They evaluated the results on the basis of a theory of recognition (Tanner, 1956), and derived a correlation coefficient (comparable to ρ in the present study) of 0.88 for the ITD and IID dimensions. This value of the correlation coefficient ($0 < \rho < 1$) implies a mechanism that is intermediate between the independent and common channel mechanisms. Hafter *et al.* (1990) used essentially the same approach as in the present study and examined the cue interaction for high frequency stimuli (bandpass-filtered pulse trains; passband around 4 kHz). They derived $\rho = 1.0$, which is favorable for the common channel mechanism. Unfortunately, the studies by Gilliom and Sorkin (1972) and by Hafter *et al.* (1990) used different experimental paradigms (discrimination between stimuli differing along a certain dimension compared with the detection of consonant changes in ITD and IID) and examined different frequency regions (500 Hz compared with 4 kHz). Thus, it is unclear to which factor we should attribute the difference between the ρ values in the two studies—the experimental paradigm, stimulus frequency, or interlistener variability.

The present study, using the detection paradigm with a within-listener design, examined specifically (1) the extent to which the ITD and IID information interacts in central mechanisms, (2) whether the degree of interaction is different for low and high frequency regions, and (3) to what extent the interaction is sensitive to baseline ITD and IID. The third point was included because it was expected that a central nervous mechanism could combine ITD and IID information most efficiently when the signal was ecologically naturalistic. Gaik (1993), analyzing head-related transfer functions for various sound-source directions, derived ecologically natural combinations of ITD and IID (i.e., combinations that normally occur in a free field) at individual critical bands. Gaik (1993) also used bandpass-filtered (passband ~0.3 oct) noises centered around 500 Hz and 4 kHz, and showed that listeners reported a higher rate of perceived "unnaturalness" (i.e., the intracranial image was hard to lateralize, broad, or split into multiple components) as the stimulus ITD and IID departed from the ecologically natural combination. The present study examined baseline ITD and IID that constitute a subset of the conditions tested by Gaik.

II. EXPERIMENT 1: EFFECTS OF FREQUENCY AND BASELINE ITD/IID

A. Rationale

Two types of stimuli, a 500-Hz sinusoid (hereafter referred to as the "500-Hz tone") and a "transposed stimulus" (van de Par and Kohlrausch, 1997; Bernstein, 2001), which was a 4-kHz tone amplitude-modulated with a half-wave-rectified 125-Hz sinusoid (hereafter referred to as the "4-kHz transposed tone"), were used as representative low and high frequency stimuli, respectively. These stimuli were chosen because a frequency around 500 Hz and a combination consisting of a 4-kHz carrier and a 125-Hz modulator are conditions for which human listeners generally show the highest ITD sensitivity for sinusoidal and transposed stimuli, respectively (e.g., Bernstein and Trahiotis, 2002). In addition, for frequencies around 500 Hz and 4 kHz, data also are available regarding the ecological and perceptual "naturalness" of various combinations of baseline ITD and IID (Gaik, 1993). The present experiment used stimulus frequencies and baseline ITD and IID combinations that were fairly similar to those used by Gaik (1993), and examined the sensitivity of the cue interaction to baseline ITD and IID combinations.

B. Methods

1. Stimuli

The stimuli were synthesized digitally with a sampling rate of 24 414.0625 Hz. For the 4-kHz transposed tone, the modulator was low-pass filtered with a cutoff frequency of 2 kHz before the modulation, so that the energies of the resulting stimulus were restricted within the ±2-kHz range around the carrier frequency (Bernstein and Trahiotis, 2002).

For each type of stimulus, two baseline ITDs (0 and +600 μs) and three baseline IIDs (−12, 0, and +12 dB), hence six combinations, were tested (a positive ITD or IID indicates that the signal in the right ear is advanced in time or greater in intensity, respectively). Ecologically, a positive ITD would usually accompany a positive IID. A combination of positive ITD and negative IID (and to a lesser degree, combinations of zero ITD and nonzero IID, and vice versa) would be unnatural both ecologically and perceptually (Gaik, 1993).

One experimental trial consisted of two listening intervals with each interval containing bursts of three tones (see Sec. II B 2). Each tone burst had a duration of 200 ms, and was gated on and off synchronously for the left and right ears with 50-ms long raised-cosine ramps (included in the tone duration). This relatively long ramp duration was used to reduce the saliency of the stimulus onset, thereby making the ongoing ITD and IID valid cues for listener's judgments. The starting phases of the carrier and the modulator were randomized burst by burst. The IID was implemented by increasing the level in one ear and decreasing it in the other

relative to the center level, so that the average for the two ears (on a decibel scale) was the center level. The center level was varied randomly burst by burst, so that it had a uniform distribution within 65 ± 3 dB SPL. This way of implementing IID by symmetrical level change, rather than by increasing or decreasing the level in one ear only, was expected to make the binaural loudness change an ineffective cue for IID detection, at least for small baseline IIDs (less than about 10 dB, Keen, 1972; Domnitz and Colburn, 1977; Bernstein, 2004). The binaural loudness cue that might remain for large baseline IIDs (12 dB) was reduced, if not completely removed, by the burst-by-burst randomization of the center level. There were 200-ms silent gaps between tone bursts in an interval, and there was a 600-ms gap between the listening intervals.

Stimuli were generated by using a digital-to-analog converter with a resolution of 24 bits (TDT RP2.1), attenuated as necessary (TDT PA5), and presented dichotically with Sennheiser HDA200 headphones through a headphone amplifier. The high frequency signal was mixed with a continuous, low-pass filtered Gaussian noise (cutoff frequency 1300 Hz; spectrum level 20 dB SPL), which was interaurally uncorrelated. This was done to prevent the listener from using any information at low spectral frequencies (Bernstein and Trahiotis, 2002).

2. Procedure

A two-interval two-alternative forced-choice (2I-2AFC) method was used to measure the detectability of the ITD/IID change. The *nonsignal interval* contained three tone bursts with identical ITDs and IIDs (i.e., baseline ITD and IID). In the signal interval, the second of the three bursts departed in terms of ITD and/or IID from the baseline values, while the remaining bursts had baseline ITDs and IIDs. The listener's task was to indicate, by pressing a button, the signal interval that was randomly chosen trial by trial from the two listening intervals. Correct answer feedback was provided after each response by lights on the response box. The listener was instructed to use any attributes in laterality-related sensations (i.e., not only the centroid, but also the dispersion of the intracranial image elicited by a stimulus) as cues for performing the detection task.

The use of this specific, rather nonstandard, two-interval task in the present study requires an explanation. A more straightforward alternative might be a lateralization task that involves judgments of the change direction of the intracranial image. An example is a 2I-2AFC task in which each interval would contain one tone burst, the interaural differences in one interval favoring the left, and those in the other interval favoring the right (e.g., Hafter *et al.*, 1990). Listeners would be instructed to indicate the direction of the lateral movement of the image centroid produced in the two intervals. There were two reasons for not choosing this type of lateralization task in the present study. First, for combinations of nonzero baseline ITD or IID, the intracranial image of a stimulus is often strongly biased toward left or right, or is dispersed, which would make it difficult for listeners to indicate the *direction* of a laterality change, even in the presence of a laterality-related sensation. In such a case, a task requiring change direction judgments was thought to underestimate listeners' binaural sensitivities. Second, the author assumed that outputs from the channels and their linear combination as the decision variables could be represented by various forms in the intracranial image elicited by the stimuli. Previous studies (Hafter and Carrier, 1972; Ruotolo *et al.*, 1979) demonstrated that even when the image centroid is centralized, stimuli with ITD and IID in opposing directions can be discriminated from a diotic stimulus, thereby supporting the existence of independent processes for ITD and IID. In those studies, listeners reported that they used the dispersion and/or multiple peaks of the internal image as cues for the task. Given those reports, the author was concerned that the presence of independent channels might not be detected if the image centroid were the only available cue in the present experiments.

Another alternative and more standard task might be a simple three-interval three-alternative forced-choice task, in which the listener chooses the tone burst with interaural differences departing from the baseline values from three tone bursts with a randomized order. However, in pilot experiments that used this paradigm, listeners often showed confusion when performing the task, reporting that *all* of the three tone bursts differed in intracranial images even for relatively large changes in interaural differences. The exact reason for this confusion is unknown, but may be related to the phenomenon of auditory saltation (Bremer *et al.*, 1977). The specific task adopted in the present study was chosen to avoid this type of confusion by temporally surrounding the tone burst of interest with bursts at baseline ITD/IID.

One block of trials consisted of 112 trials that tested one baseline ITD and IID combination. The set of 112 trials comprised 14 repetitions of 8 combinations of ITD and IID changes. The directions of the ITD and IID changes, when combined, were always consonant. The 8 combinations included two sizes of ITD and IID changes, presented individually or in combination; namely $(\Delta ITD_1, 0)$, $(\Delta ITD_2, 0)$, $(0, \Delta IID_1)$, $(0, \Delta IID_2)$, $(\Delta ITD_1, \Delta IID_1)$, $(\Delta ITD_1, \Delta IID_2)$, $(\Delta ITD_2, \Delta IID_1)$, $(\Delta ITD_2, \Delta IID_2)$, where each set of parentheses represents one combination of ITD and IID changes, and the first and second terms, respectively, in parentheses indicate the sizes of the ITD and IID changes. Within a block of trials, the 8 combinations were tested in pseudorandom order such that all 8 combinations were tested once before all the conditions were repeated in a different random order. The values of ΔITD_* and ΔIID_* were determined for each listener, stimulus type, and baseline ITD/IID on the basis of pilot experiments, so that the resulting proportions of correct responses were designed to range from about 0.60 to 0.95 for all 8 conditions. For each stimulus type and each combination of baseline ITD and IID, trial blocks were repeated 15 times in pseudorandom order, yielding a total of 210 trials per condition. Initially, all the conditions with negative ITD and IID changes (i.e., the signal in the right ear was delayed in time and reduced in level) were completed, and then the positive change directions were tested.

3. Listeners

Initially, four listeners with normal hearing were employed. However, one was excluded from the formal experiments, because she failed to perform tasks reliably for the high frequency stimulus even for very large (>600 μs) ITD changes at zero baseline-ITD and IID despite extensive training. Thus, the present experiment represents the data provided by three listeners, ranging from 31 to 37 years of age. One listener was the author (SF). The other two listeners (and the one who was excluded from the formal experiments) had experience in psychoacoustic tasks, but not in sound lateralization judgment. They received extensive training until their performance appeared to be stable before starting the formal data collection. The experiments were conducted in a sound-attenuating chamber.

4. Data analysis

The analyses were performed with MATLAB (MathWorks).

The proportion of correct responses, P_c, obtained with the 2AFC method for each condition was converted to the detectability index, d', by the equation, $d' = \sqrt{2}\,\text{z-score}(P_c)$ (Macmillan and Creelman, 2005).

The value of the "correlation coefficient," ρ, expressed in Eq. (1) was estimated for each subject and each baseline ITD and IID by the least-square fitting of Eq. (1) to measured values of d'_C, d'_T, and d'_I. Specifically, the ρ value was varied between −1 and 5 in 0.01 steps, and the ρ estimate was defined as the ρ value that minimized the sum of the squared difference (SSD) between the measured d'_C and the d'_C predicted from d'_T, d'_I, and ρ on the basis of Eq. (1). Every SSD vs ρ function was convex with a minimum within the range of the ρ variation (−1 to 5). One estimation involved 8 pairs of d'_T and d'_I (4 combinations of ΔITD and ΔIID × 2 change directions). The ρ value was also estimated for pooled data for the three listeners, and here 24 pairs of d'_T and d'_I (4 combinations of ΔITD and ΔIID × 2 change directions × 3 listeners) were involved. The reliability of the ρ estimate was evaluated by the bootstrap test. A ρ estimate was obtained for a bootstrap set of 8 d'_T-d'_I pairs (or 24 pairs for the pooled data) that were randomly selected from the original 8 pairs (or 24 pairs) with replacement. This was repeated to obtain 100 bootstrap estimates of ρ. The 95% confidence interval for a ρ estimate was defined as the range between the 2.5th and 97.5th percentile values of the bootstrap estimates.

C. Results

The overall sensitivity to individual ITD or IID changes (ΔITD or ΔIID, respectively) generally varied depending on stimulus type, baseline ITD/IID, and listener. For the purpose of inspecting the effects, the threshold ΔITD or ΔIID values were estimated from the data for ITD or IID change alone, i.e., (ΔITD, 0) or (0, ΔIID), and are summarized in Fig. 2. For threshold estimation, first, a straight line passing the origin was fitted to the d' vs ΔITD or d' vs ΔIID plot of the data with the least-squares method. Then, the ΔITD or ΔIID value that would produce $d' = 1$ was defined as the threshold. Since, with a few exceptions, the ITD or IID sensitivities

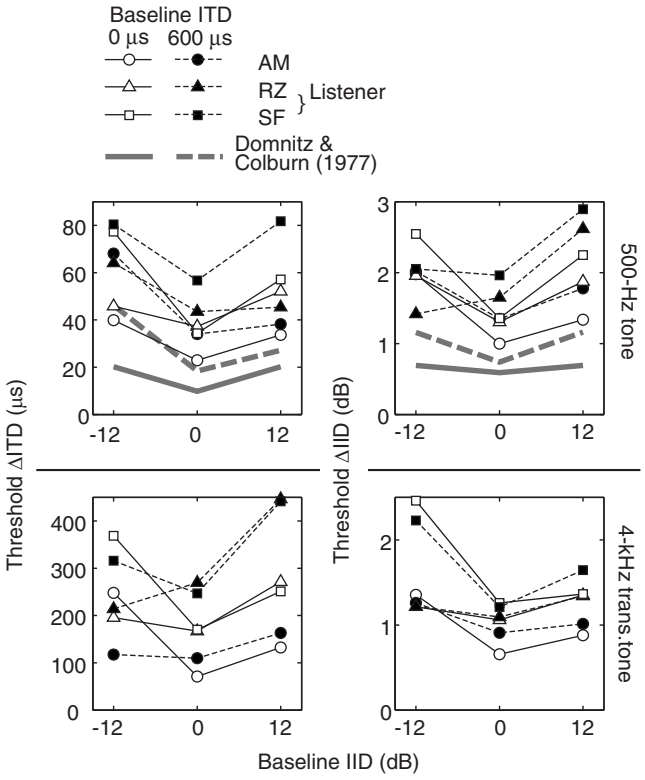


FIG. 2. Threshold for detecting ITD or IID change alone. For each listener, stimulus type, and baseline ITD/IID, a straight line passing the origin was fitted to the d' vs ΔITD or ΔIID plot of the data with the least-squares method. Then, the threshold was defined as the ΔITD or ΔIID value that would produce $d' = 1$. The obtained thresholds were plotted against baseline IID; baseline ITD being the parameter (indicated by the open and closed symbols). The threshold ITDs and IIDs are shown in the left and right columns of the panels, respectively. The upper and the lower panels represent the 500-Hz and the 4-kHz transposed tones, respectively. The thick gray lines in the upper panels indicate the thresholds measured by Domnitz and Colburn (1977, average data in their Figs. 2 and 3); their data were linearly interpolated, where necessary, to make the baseline ITD/IID comparable.

were similar for the negative and positive ITD or IID changes, the average thresholds are shown. For the 500-Hz tone, Fig. 2 also shows the thresholds derived (by linear interpolation) from the data published by Domnitz and Colburn (1977, average data in their Figs. 2 and 3). The thresholds obtained in the present experiment (lines with symbols) were consistently greater than those reported by Domnitz and Colburn (1977, thick gray lines) by a factor of about 2 (ΔITD) or by about 1 dB (ΔIIID). This difference was possibly due to the procedural differences between the studies (elaborated in Sec. IV). Nonetheless, the effects of baseline ITD and IID showed certain tendencies that were common across studies and listeners. That is, the threshold was lowest at zero baseline ITD and IID, and generally increased when the baseline ITD and/or IID was a nonzero value. Exceptions were the IID thresholds of listeners RZ and SF (triangles and squares, respectively, in the right panel), where the thresholds for the incompatible baseline (i.e., 600 μs/−12 dB) were lower than for the 0- μs/0-dB baseline. The above-mentioned trends were generally true for the 4-kHz

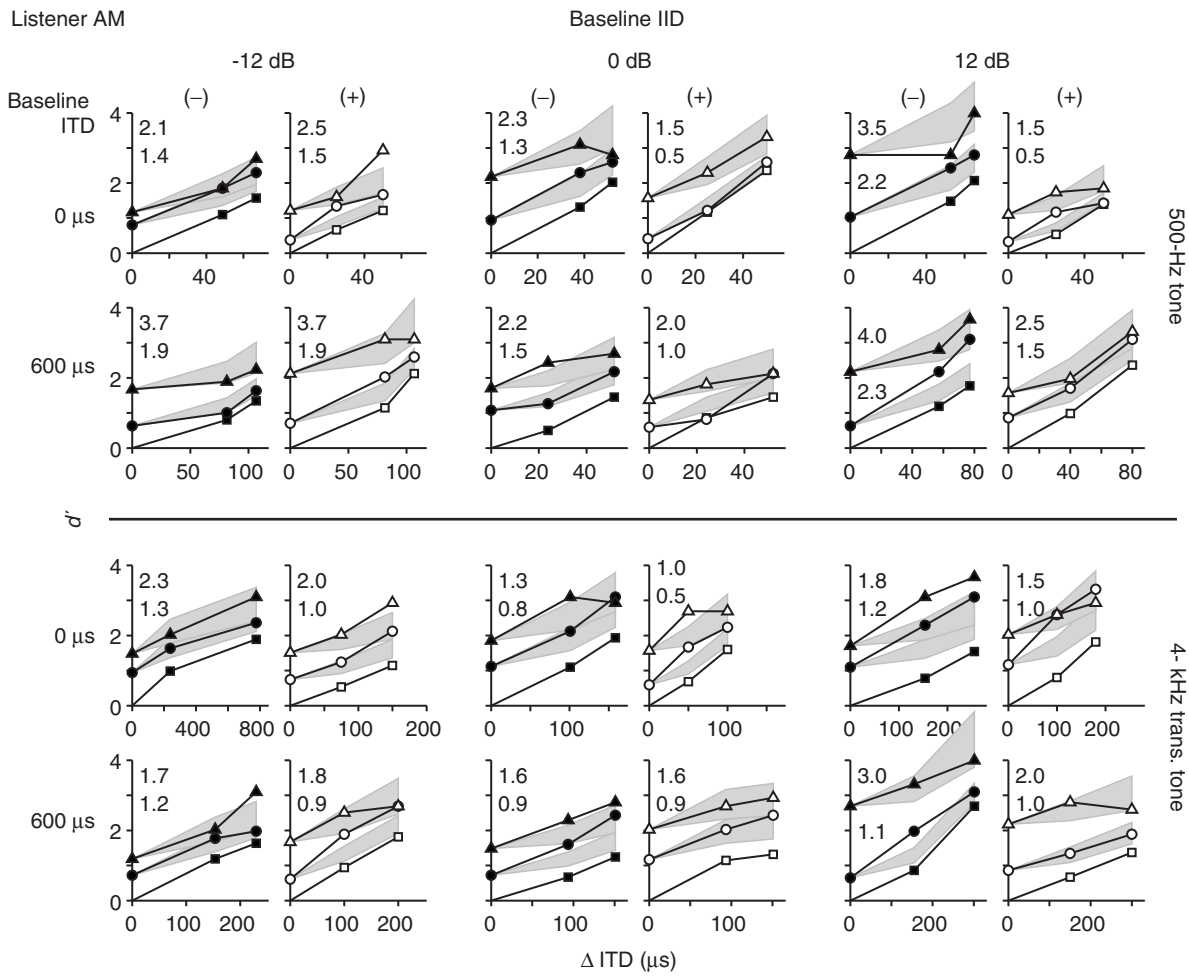


FIG. 3. d' values plotted as a function of the size of ITD change (ΔITD) for listener AM. Each small panel represents one stimulus type (indicated on the extreme right), one combination of baseline ITD and IID (indicated on the extreme left and at the top, respectively), and one direction of ITD and IID change (indicated by the ± signs above the top panel, and by the closed/open symbols; (+) and open symbols indicate that the ITD and IID changed in such a way that the stimulus in the right ear was further advanced in time and increased in intensity). In each small plot, the three lines connecting symbols correspond to three ΔIID values. The data with a line locked to the origin (square symbols) represent ΔIID=0, i.e., a psychometric function with regard to ITD change alone. The data with the other two lines indicate cases when IID was changed together with ITD (except for the y intercepts, when only IID was changed). The ΔIID size varied depending on condition, and is shown by the two numbers (in dB) in each plot. The top and bottom numbers correspond to ΔIIDs for the lines with triangles and circles, respectively. The scales of the abscissa are the same for positive (+; closed symbols) and negative (−; open symbols) change directions. One exception was for listener AM; a baseline ITD/IID of 0 μs/−12 dB; the 4-kHz transposed tone, where the tested ΔITDs were markedly different for different change directions.

transposed tone, except that the thresholds were higher for ΔITD and lower for ΔIID, and that the ITD thresholds, not the IID thresholds, showed lower values for the incompatible baseline than for the 0-μs/0-dB baseline (left panel).

Complete data, including the results when ITD and IID were changed simultaneously, are presented in Figs. 3–5 for the three subjects. Each small panel in Figs. 3–5 is a psychometric function, which is the plot of d' as a function of ITD change (ΔITD), ΔIID being a parameter. Of the three lines in a panel, the lowest represents the results when ITD alone was changed (marked by squares). The other two indicate cases where ITD and IID changes were combined, except for the y intercepts, for which IID alone was changed. There, the sizes of the IID changes (i.e., ΔIID) are indicated by the two numbers corresponding to (from lower to upper) circles and triangles, respectively, in the panels. Each group of two panels represents one combination of baseline ITD (sorted in rows) and IID (sorted in columns) for each stimulus type. In each group of two panels, the left and right panels represent the positive and negative directions of the ITD and IID changes, respectively. The gray area indicates the range of d'_C predicted for the ρ range of $0 \leqslant \rho \leqslant 1$ in Eq. (1). The lower and upper boundaries correspond to predictions based on $\rho = 0$ (independent channel hypothesis) and $\rho = 1$ (common channel hypothesis), respectively. Thus, in the plots, the degree of cue interaction can be evaluated by comparing the measured d'_C with reference to the gray area.

For the 500-Hz tone (upper half of the panels), the measured d'_C values often fell inside the gray area, indicating that the degree of cue interaction was intermediate between that expected from the independent and common channel hypotheses (see, for example, the baseline ITD/IID of 0 μs/0 dB). A marked exception was for the baseline ITD/IID of 600 μs/−12 dB where the measured d'_C was generally at or below the lower boundary, often even lower than the d' for an ITD or IID change alone. In contrast, for the 4-kHz trans-

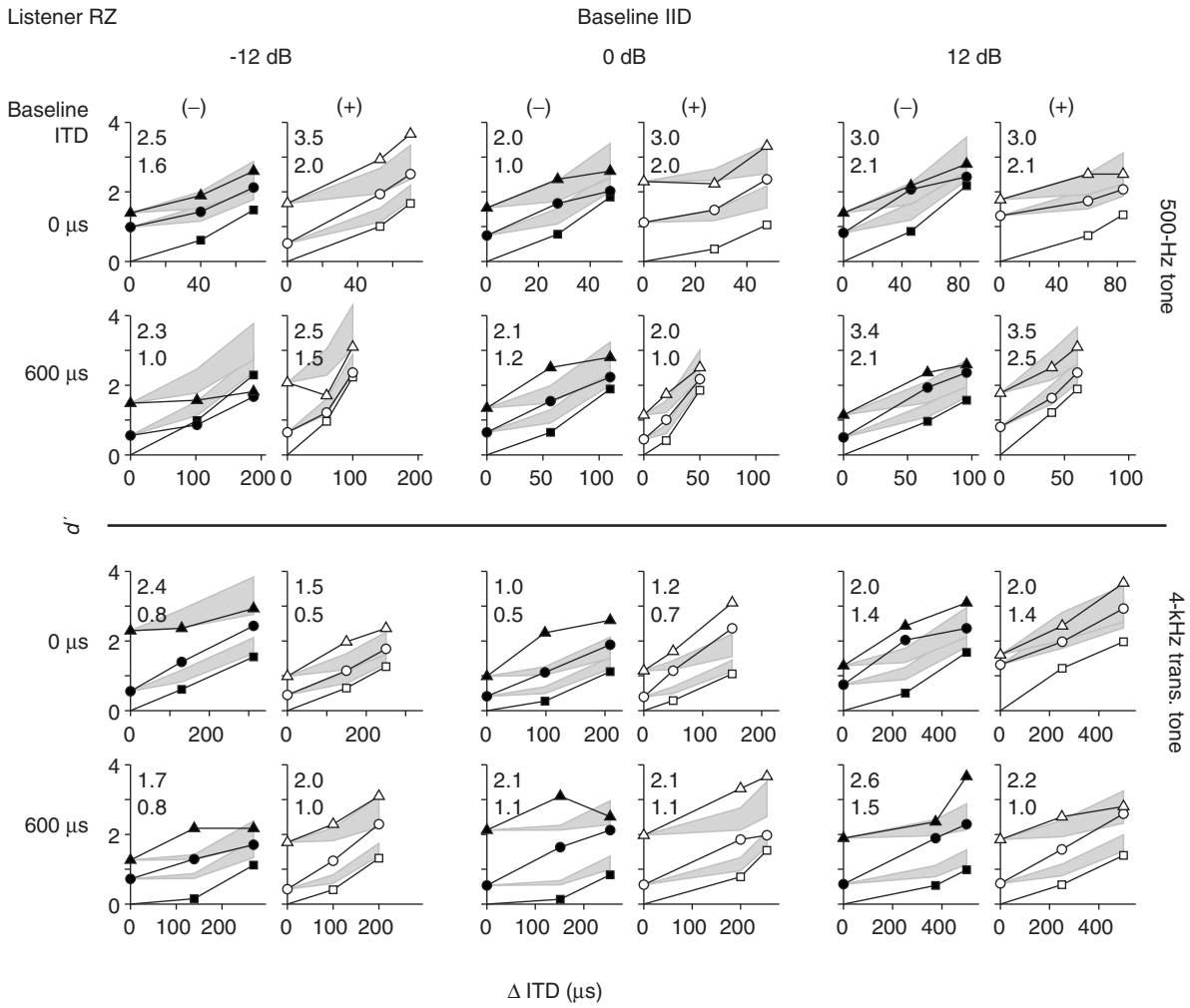


FIG. 4. Same as Fig. 3. but for listener RZ.

posed tone, the measured d'_C values were generally at the upper boundaries. This is consistent with the prediction by the common channel hypothesis. There were even cases where the measured values were well above the prediction. This was particularly noticeable for listener RZ, with a baseline ITD/IID of 0 μs/0 dB.

The data presented here were used to test the independent channel and common channel hypotheses. For each ΔITD and ΔIID pair, the d'_C value was predicted on the basis of those hypotheses using the measured d'_T and d'_I values. The results of a comparison of the measured and predicted d'_C values are summarized in Table I. The results support the above-noted observations. For a low frequency stimulus (500-Hz tone), both independent and common channel hypotheses failed to account for the measured d'_C for most of the baseline ITD/IIDs; differences between the measured and the predicted d' were statistically significant. The measured d'_C was generally greater than the value predicted by the independent channel hypothesis (i.e., positive mean difference in the Table I) and below the value predicted by the common channel hypothesis (i.e., negative difference). This implies that the underlying binaural processes fall between the processes assumed in those two hypotheses. For a 4-kHz transposed tone, the independent channel hypothesis again generally failed to predict d'_C, but the common channel hypothesis was not rejected except in one case, namely for a baseline ITD/IID of 0 μs/0 dB. It should be noted that in this exceptional case, the measured d'_C was *greater* than the predicted value.

To examine the extent of channel interaction, the "correlation coefficient," ρ, expressed in Eq. (1) was estimated by the least-squares fitting of the equation to the measured d'_C, d'_T, and d'_I values (see Methods Sec. II B 4). Results for individual subjects and for the pooled data from all three listeners are shown in Fig. 6. The stimulus types, namely the 500-Hz tone and the 4-kHz transposed tone, are represented by closed circles and open triangles, respectively. The baseline ITD and IID are represented by different line types (0 μs—solid lines; 600 μs—dashed lines) and by the abscissa, respectively. The vertical bars indicate 95% confidence intervals. Again, the results shown here were consistent with the above-described observations. The ρ estimates for the 500-Hz tone generally fell in the 0 to 1 range (gray area), except for the incompatible combination of baseline ITD and IID (600 μs/−12 dB). The ρ estimates for the 4-kHz transposed tone were around 1, and this is consistent

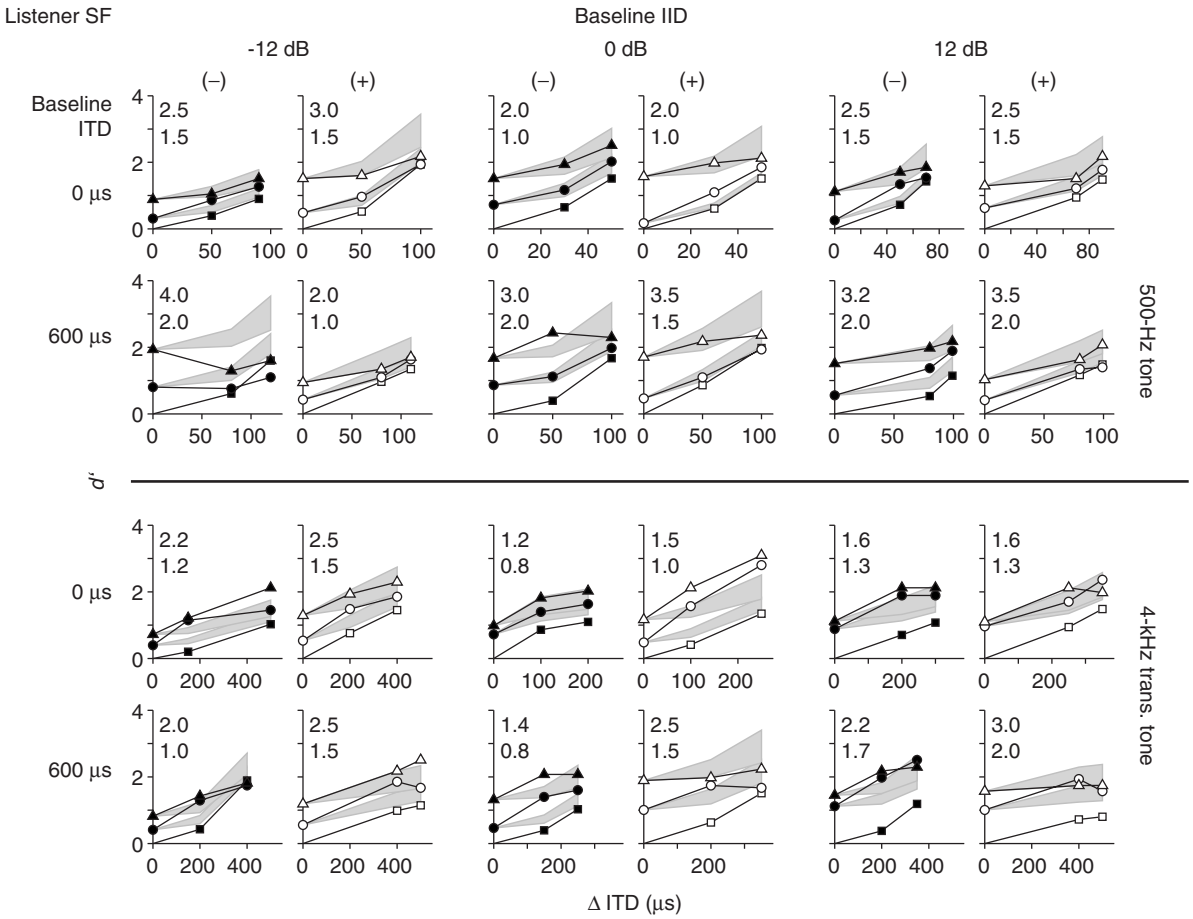


FIG. 5. Same as Fig. 3. but for listener SF.

with the common channel hypothesis. It should be noted, however, that the estimate for listener RZ at 0-μs/0-dB baseline ITD/IID was significantly above 1; its 95th confidence interval did not include 1.

TABLE I. Summary of statistical tests on the independent channel and common channel hypotheses. Predicted d'_C values were derived using measured d' values for individual ITD and IID changes, i.e., d'_T and d'_I for each hypothesis (ρ=0 or 1). Mean differences between the predicted and measured d'_C (a positive value indicates that the measured d'_C was greater than the predicted d'_C) are shown. Paired t-tests (two tailed) were performed, and significant differences are indicated by **bold** letters and asterisks ($^{*}p<0.05$, $^{**}p<0.01$). Each test was based on 24 d'_C pairs (3 listeners ×2 change directions ×4 ΔITD and ΔIID combinations).

		Baseline IID (dB)		
	Baseline ITD (μs)	−12	0	12
500-Hz tone	Independent channel hypothesis (ρ=0)			
	0	**0.45****	**0.33****	**0.29****
	600	−0.11	**0.29****	**0.48****
	Common channel hypothesis (ρ=1)			
	0	−0.09	**−0.29****	**−0.35****
	600	**−0.77****	**−0.28****	**−0.16****
4-kHz transposed tone	Independent channel hypothesis (ρ=0)			
	0	**0.50****	**0.79****	**0.75****
	600	**0.54****	**0.61****	**0.58****
	Common channel hypothesis (ρ=1)			
	0	−0.03	**0.28****	0.07
	600	0.03	0.05	−0.02

When we compare stimulus types, it is clear that the ρ values were generally greater for the 4-kHz transposed tone than for the 500-Hz tone ($p<0.001$ paired t-test; DoF=17). As regards the effect of the baseline ITD and IID, it was difficult to find consistent tendencies among listeners in Fig. 6. The only exception was for a baseline ITD/IID of 600 μs/−12 dB in the 500-Hz tone, for which the ρ value was consistently smaller than any other baseline ITD and IID combinations. A within-subject analysis of variance with baseline ITD/IID as a factor was performed for each stimulus type. For the 500-Hz tone, the effect of the baseline ITD/IID was significant [p=0.029; $F(5,10)$=4.05]. A post hoc, multiple comparison of pairs of baseline ITD/IIDs revealed that ρ for a baseline ITD/IID of 600 μs/−12 dB was significantly different from that for any other combination except 0 μs/12 dB (least significant difference test, $p<0.05$).

III. EXPERIMENT 2: COMPARISON BETWEEN FREQUENCIES

A. Rationale

The results of Experiment 1 suggest that the interaction between ITD and IID, indicated by the ρ value, was generally greater for a high frequency stimulus (4-kHz transposed tone) than for a low frequency stimulus (500-Hz tone). It is, however, possible that this relatively small interaction with the lower frequency stimulus was specific to 500 Hz, and would not apply generally to other low frequencies. Thus,

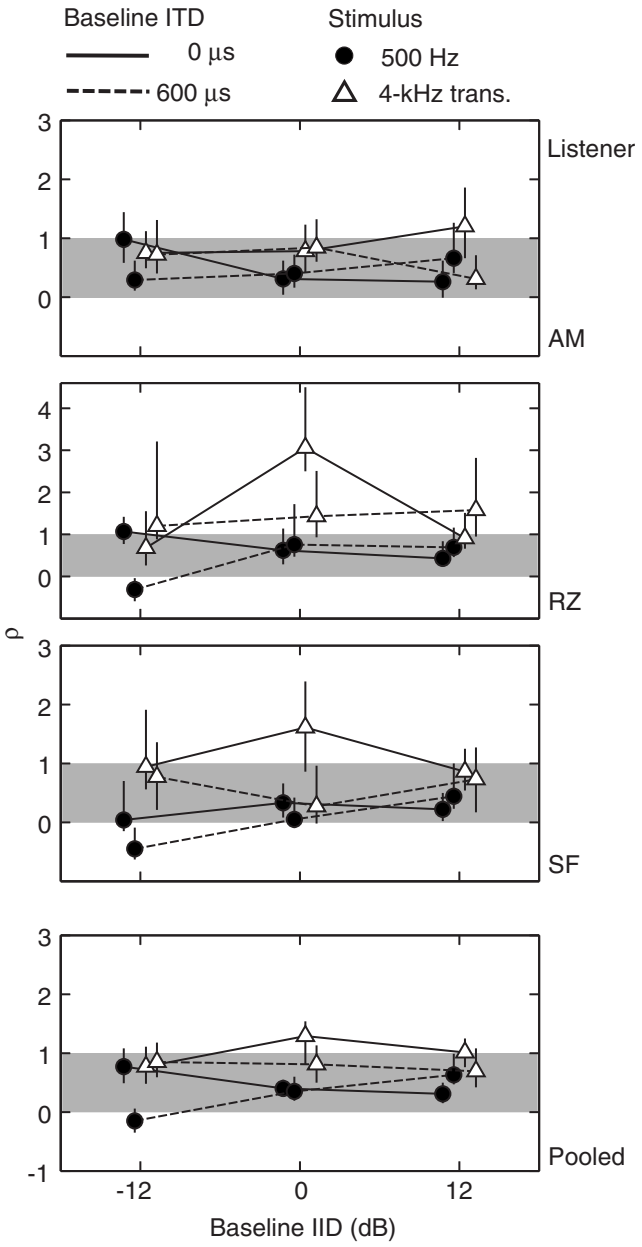


FIG. 6. Estimated ρ values. For each listener, stimulus type, and baseline ITD/IID, the value of ρ was estimated by the least-squares fitting of Eq. (1) to the measured d'_C, d'_T, and d'_I values (see the Methods II B 4). Results for individual subjects and for the pooled data from all the three listeners are shown in the panels. The ρ estimates were plotted against baseline IID, and the baseline ITD was indicated by line type (solid and dashed). Stimulus types, 500-Hz tone and 4-kHz transposed tone, are represented by closed circles and open triangles, respectively. The vertical bars indicate 95% confidence intervals, estimated by the bootstrap method. The gray area indicates a range of $0<\rho<1$.

Experiment 2 was conducted to test a 125-Hz tone, in addition to the 500-Hz tone and 4-kHz transposed tone that were used in Sec. II. The 125-Hz tone is arguably more comparable to the 4-kHz transposed tone, since the tone frequency matches the modulator frequency of the 4-kHz transposed tone. In addition, this experiment employed two additional listeners. This related mainly to the results obtained from listener RZ, who exhibited an unexpectedly large ρ estimate ($\geqslant 1$) for the 4-kHz transposed particularly at a baseline ITD/IID of 0 μs/0 dB. Experiment 2 examined whether listener RZ could be regarded as an outlier, or whether such a large ρ can be observed widely among listeners.

B. Methods

The stimuli and the experimental procedure were identical to those used in Experiment 1, except that a 125-Hz tone was also tested, and only a baseline ITD/IID of 0 μs/0 dB was tested.

The three listeners tested in Experiment 1 and two additional listeners with normal hearing participated in the experiment. The ages of the five listeners ranged between 22 and 37 years. The new listeners had experience of psychoacoustic tasks, but not of sound lateralization judgment. They received extensive training before starting the formal data collection. For those listeners who participated in Experiment 1, only the additional stimulus, the 125-Hz tone, was newly tested, and for the other stimulus types, the data from the previous experiment are reproduced in Sec. III.

C. Results

As with the earlier analysis, the independent channel and common channel hypotheses were tested individually for the stimulus types. The independent channel hypothesis was rejected for all three stimulus types: The measured d'_C values were significantly greater than the predicted values (two-tailed, paired t-test: $p<0.001$ for all the stimulus types, DoF d.f.=39). The test results for the common channel hypothesis varied depending on stimulus type: The hypothesis was not rejected for the 125-Hz tone ($p=0.173$). For the 500-Hz tone, the measured d'_C was significantly smaller than the predicted value ($p<0.001$), whereas for the 4-kHz transposed tone, the measured values were significantly greater than the predicted value ($p=0.015$). The results for the latter two stimulus types were consistent with those described in Sec. II.

Figure 7 summarizes the estimated ρ values. For each listener, the symbols from left to right represent estimated ρ values for the 125-Hz, 500-Hz, and 4-kHz transposed tones, respectively. The symbols and the vertical lines indicate the ρ estimates and the 95% confidence intervals obtained by the bootstrap method. The ρ estimates for the pooled data across the five listeners are also shown (labeled "All" in Fig. 7).

The ρ value was consistently the largest among listeners for the 4-kHz transposed tone, and the smallest for the 500-Hz tone. The value for the 125-Hz tone was intermediate. A statistical analysis supported this observation. A within-subject analysis of variance with stimulus type as a factor showed that the main effect of the stimulus type was significant [$p=0.018$; $F(2,8)=6.88$]. A post hoc, multiple comparison of pairs of stimulus types revealed significant differences between the ρ values of the 500-Hz and 4-kHz transposed tones (the least significant difference test, $p<0.01$).

Comparing ρ among listeners, it is noticeable that the value for listener RZ was exceptionally high. This was particularly apparent for the 4-kHz transposed tone; ρ for listener RZ was 3.06, while ρ for all the other listeners could be

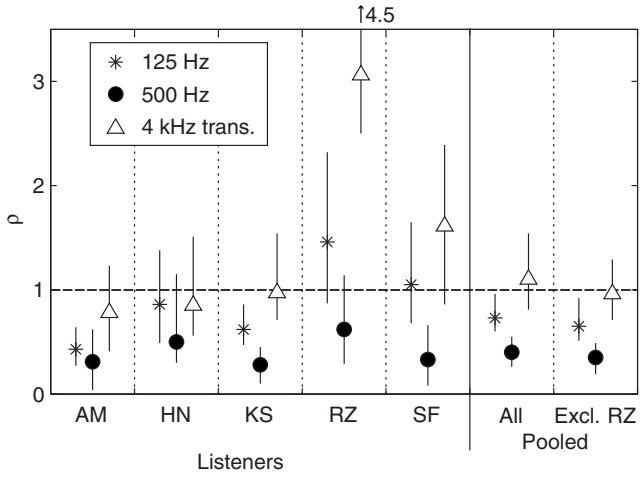


FIG. 7. The ρ values estimated for the three stimulus types, namely the 125-Hz, 500-Hz, and 4-kHz transposed tones. The data for the five listeners are represented by the five groups of three symbols on the left, indicating stimulus types (see the key). The vertical lines indicate the 95% confidence intervals estimated by the bootstrap method. The number above the data for listener RZ indicates the upper limit of the confidence interval for the corresponding ρ estimate. The data in the two groups of the symbols on the right represent ρ estimates for pooled data for all five listeners (labeled with "All") and of all but listener RZ (labeled with "Excl. RZ").

regarded as being about 1 (i.e., the 95% confidence intervals included 1). This value of ρ for RZ deviated by 5.29σ from the mean across the other four listeners (σ: standard deviation across the four listeners). An examination of measured d' values for the 4-kHz transposed tone (data not shown, except for listeners AM, RZ, and SF in Figs. 3–5) revealed that d'_C values for listener RZ were consistently greater than the expectation by the hypothesis of $\rho=1$ for both change directions, whereas no such consistent difference for the other listeners, except for listener SF with positive ITD/IID changes. Thus, listener RZ was tentatively treated as an outlier.

When listener RZ was excluded, the analyses performed earlier led to slightly different conclusions. The test of the common channel hypothesis that was undertaken by comparing the measured d'_C values with a prediction assuming $\rho=1$ indicated that for both low frequency tones (i.e., 125 and 500 Hz), the measured d'_C values were significantly smaller than the predicted values ($p=0.011$ and $p<0.001$, respectively, DoF=31). This suggests a mechanism that is intermediate between the independent and the common channel mechanisms (the null hypothesis of $\rho=0$ was also rejected). On the other hand, the difference between the measured and predicted values was not significant for the 4-kHz transposed tone ($p=0.327$), suggesting the common channel mechanism. These results are consistent with the observation of $\rho \cong 1$, which was estimated based on the pooled data excluding listener RZ (the rightmost three symbols in Fig. 7, labeled with "Excl. RZ").

One might be concerned that the difference in ρ among the listeners and the conditions reflected differences in stimulus parameters, such as ΔITD and ΔIID rather than the degree of cue interaction *per se*. The stimulus ΔITD was generally the smallest and largest for 500-Hz and 4-kHz transposed tones, respectively (cf., Figs. 3–5), and large ΔITD values might artifactually lead to large ρ values for certain reasons. A multiple regression analysis was used to test this possibility. For each listener and stimulus type (125-Hz tone, 500-Hz tone, or 4-kHz transposed tone), the least-squares fit of a linear function of ΔITD and ΔIID to corresponding eight ρ values was found. Those ρ values were computed by solving Eq. (1) for individual pairs of ΔITD and ΔIID, which yielded 8 values=4 combinations ×2 change directions. If ΔITD and, to some degree, ΔIID were the true factors for determining ρ, one would expect the ρ variance to be well accounted for by the linear fit. For all (3 stimulus types ×5 listeners) but in two cases the regression failed to account for the variance of ρ ($F(2,5)$; criterion $p=0.05$). The results indicate that the observed ρ variation could not be accounted for simply by ΔITD and/or ΔIID alone.

IV. DISCUSSION

A. Comparison with previous studies

Thresholds for detecting changes in ITD or IID generally increased as the baseline ITD and/or IID departed from the midline, i.e., 0 μs/0 dB (Fig. 2). This tendency was consistent with previous studies [Domnitz and Colburn, (1977) for 500-Hz tone; see also a review by Yost and Hafter, (1987) for broad range of frequencies but at limited baseline ITD/IID conditions]. However, it is notable that the thresholds in the present study were generally higher. For instance, Domnitz and Colburn (1977) showed that thresholds for the 500-Hz tone were, at best, about 10 μs (ITD change) and about 0.5 dB (IID change), in contrast to the present results of >20 μs and >1 dB, respectively (as can be seen in Fig. 2). For the 4-kHz stimuli, the thresholds reported previously were about 70 μs for an ITD change (Bernstein and Trahiotis, 2002) and about 0.5 dB for an IID change (reviewed by Yost and Hafter, 1987), and the thresholds in the present study were again higher. These discrepancies are possibly for three reasons, which would not critically affect the interpretation of the present data as regards cue interaction. The first reason relates to the detailed differences in the stimulus parameters. For instance, the stimuli used in the present study had durations of 200 ms, whereas previous studies, for example those undertaken by Domnitz and Colburn (1977) and Bernstein and Trahiotis (2002), used 300-ms long stimuli. Also, the onset and offset ramps in Bernstein and Trahiotis (2002) were shorter (20 ms) than those in the present study (50 ms). The second reason is the roving of the center level used in the present study. In fact, previous studies showed that IID thresholds measured with a roving level were about 1 dB higher than those measured without one (Yost and Hafter, 1987). The third possible reason is related to listeners' strategies under the present stimulus configuration. The formula computing d' from the proportion of correct responses, $d'=\sqrt{2}$ z-score(P_c), adopted in this study is derived assuming the listener directly compares two internal variables, one being elicited by the tone burst at baseline ITD and IID ("noise only" in the terminology of the signal detection theory) and the other by the tone burst with ΔITD and/or

ΔIID ("signal plus noise"). However, the actual listener might look at the difference between the second burst and the baseline burst within an interval, and compare the size of the difference between the two intervals. If this were the case, the above-presented formula would underestimate d' by some factor.[2] Nonetheless, this factor should be common to all the estimates of d'_C, d'_T and d'_I, and thus should not affect the estimation of ρ.

The present estimates of the ρ values for 500-Hz tones fell between 0 and 1, which is generally consistent with the results reported by Gilliom and Sorkin (1972), who estimated ρ values with a discrimination experiment for 500-Hz stimuli that differed along ITD and/or IID dimensions. However, the magnitudes of ρ were different, namely it was about 0.88 as reported by Gilliom and Sorkin compared with 0.35–0.4 in the present study (pooled data in Fig. 7). It is unclear what the possible source(s) of the difference was. It may be simply due to experimental errors or to some nonsensory factors. The smaller ρ values obtained in the present study may indicate that the listeners relied to a greater extent on aspects of internal images for which the cue interaction was relatively small, e.g., a "time image," which is sensitive to ITD but invariant with IID (Whitworth and Jeffress, 1961). However, it is puzzling, because not only the present detection task but also Gilliom and Sorkin's discrimination task were designed to avoid limiting the listener's use of any ITD- or IID-related cues represented in the intracranial image.

The ρ values for the 4-kHz transposed tone were consistent with those reported by Hafter *et al.* (1990), who adopted the same theoretical framework. To make the ρ estimates in the two studies more comparable, the author recalculated the ρ values from the data provided by Hafter *et al.* (1990, Fig. 3) by allowing $\rho > 1$.[3] It was found that the ρ estimates for three of their four listeners were about 1 (ranging from 0.9 to 1.3), while one listener (RF) exhibited an exceptionally high value of 3.5 (cf., listener RZ in the present study). This consistency between the studies, despite differences in the experimental task, implies that the listeners in both studies used effectively the same cues to perform the task. More specifically, the task in the experiment described by Hafter *et al.* involved left–right judgment of the intracranial image, which would lead listeners to focus on the image centroid, as opposed to the task in the present study, which allowed other ITD- or IID-related aspects of the image to be taken into account, e.g., image width, and multiple peaks. It can be inferred that the ρ values obtained in the present study reflected the cue interaction mainly in the image centroid.

It is interesting to see the extent to which an existing binaural model can account for the ρ value observed in the present study. This section predicted a ρ value of the 0-μs/0-dB baseline for a 500-Hz tone by using the model proposed by Stern and Colburn (1978). Their model is based on an interaural coincidence-detection process (thereby reflecting ITD) that is modulated by a weighting function sensitive to IID. Its internal noise is attributed to the stochastic processes at the auditory nerve level. The paper by Stern and Colburn (1978) presents the predicted mean and standard deviations of the "pointer position" (a value associated with the centroid of an intracranial image) for various combinations of ITD and IID. These mean and standard deviations (hereafter referred to as $E[P]$ and σ_P) can be regarded, respectively, as corresponding to the signal and the noise in the present study [cf., Fig. 1(c)]. Here, the d' value for a given ITD and/or IID change can be predicted as the ratio of the shift size of $E[P]$ to σ_P. In fact, using this method, Stern and Colburn (1985) were generally able to account well for psychophysical performance on detecting changes in ITD and IID. In the present analysis, first, the $E[P]$ and σ_P values presented in their Figs. 3 and 5 were spline-interpolated (in 1-μs and 0.1-dB steps, respectively, for ITD and IID), and the individual ITD and IID changes needed to achieve $d'=1$ ($\Delta\mathrm{ITD}_1$ and $\Delta\mathrm{IID}_1$, respectively) were determined (i.e., $d'_T=1$ and $d'_I=1$ for $\Delta\mathrm{ITD}_1$ and $\Delta\mathrm{IID}_1$). Then, the d'_C value for the combined changes by $\Delta\mathrm{ITD}_1$ and $\Delta\mathrm{IID}_1$ was estimated based on the interpolated $E[P]$ and σ_P values. By solving Eq. (1) with the estimated d'_T, $d'_I(=1)$ and d'_C values, the ρ value was predicted to be 0.93. This value is close to one, as expected from the common channel hypothesis in the present linear model, although the Stern and Colburn model is a nonlinear model. The model behaved like a common-channel mechanism because the shift size of $E[P]$ for combined ITD and IID changes was approximately the linear sum of those for individual ITD and IID changes within a small range around the baseline ITD and IID, and because the σ_P was essentially constant for individual and combined ITD and IID changes. It should be noted that the measured $\rho(0.35-0.4)$ was markedly smaller than the predicted ρ. This discrepancy may be simply due to some problems with the present analysis (e.g., inaccurate interpolation of $E[P]$ and σ_P) or with the Stern and Colburn model. However, it is also possible that the Stern and Colburn mechanism is valid but that there is an additional binaural mechanism with an independent noise source. If the decision device combines the outputs from those two mechanisms, the whole system can behave as predicted by the case where the channels partially overlap [Fig. 1(b)], resulting in a measured ρ value that is intermediate between 0 and 1.

B. Interpretations of ρ

For all the stimulus types (125 and 500-Hz tones, and 4-kHz transposed tones), the independent channel hypothesis ($\rho=0$) was rejected for all but one of the tested conditions (Table I); the ρ values being significantly greater than 0. Under the present framework, this suggests that the two channels are not completely independent [Fig. 1(b)]. This is consistent with the fact that the sensitivity to changes of ITD (or IID) alone depends on the baseline IID (or ITD) (Domnitz, 1973; Domnitz and Colburn, 1977, present study, Fig. 2). It is possible that the baseline-IID (or -ITD) dependent internal noise in the IID (or ITD) channel is added to the ITD (or IID) channel, thereby affecting the ITD (or IID) change detection performance.

For the 500-Hz tone, the measured d'_C values were generally significantly smaller than the values predicted from the common channel hypothesis. The ρ value was between 0 and 1, namely $\rho=0.35-0.4$ for baseline ITD/IID=0 μs/0 dB

(Fig. 7). These results imply a mechanism that is intermediate between the independent and common channel mechanisms [Fig. 1(b)]. This is consistent with the notion that ITD and IID are not completely tradable (Hafter and Carrier, 1972; Ruotolo *et al.*, 1979). Moreover, although the evidence was weaker, the results for the 125-Hz tone also indicate that the ITD and IID are not entirely processed by a common channel. The measured d'_C values (when listener RZ was excluded) were significantly smaller than those predicted by the common channel hypothesis. The ρ estimate for the pooled data was about 0.7 (Fig. 7). For the 4-kHz transposed tone, the degree of ITD and IID interaction was generally as great as that predicted by the common channel hypothesis ($\rho=1$). This suggests that ITD and IID information is processed or summed in a common mechanism.

It is interesting that listener RZ exhibited a ρ estimate considerably above 1 particularly for a baseline ITD/IID of 0 μs/0 dB ($\rho=\sim 3$). The exceptionally high ρ value shown by listener RZ was treated as an outlier in some of the analyses in this study when it was regarded as trivial (i.e., the result was simply caused by, for example, measurement error, or an unknown factor specific to the listener). However, it is arguable whether listener RZ is "truly" exceptional, because there is no obvious evidence, other than the ρ values, indicating outlier. For example, the detection thresholds for listener RZ (Fig. 2) did not appear to be noticeably different from those for the other listeners. Since $\rho>1$ was not expected by the present working model (Fig. 1), the results for listener RZ may indicate that the way the cues interacted in the binaural pathway could not be approximated as the simple linear additivity of ITD and IID information. This issue is discussed in a later section.

Regardless of the validity of the linear model, the present study clearly demonstrated that the value of ρ, as an index of the detectability of combined cue changes with reference to the detectability of individual changes, was significantly greater for a high frequency stimulus than for a low frequency stimulus. This indicates the possibility that mechanisms responsible for integrating the two cues are different for low and high frequency signals.

The notions regarding the frequency dependence of the binaural mechanism and of a common ITD/IID channel for a high frequency signal (given that the linear additivity assumption is valid) are not consistent with conclusions drawn from a previous study on perceptual learning. Zhan and Wright (2007) measured human performance for detecting changes in the ITD and IID of high-frequency transposed sounds (comparable to the 4-kHz transposed tone in the present study), and compared the effect of training between the cues to be detected. They found that multihour training significantly improved the IID change detection performance, whereas there was negligible improvement in the ITD task. The learning patterns were similar for low and high frequency signals. Zhan and Wright interpreted this result as indicating separate mechanisms for ITD and IID. That is, no point along the entire neural pathway for ITD processing undergoes modifications that could be observed by their behavioral task, whereas there is at least one point along the IID processing pathway that is malleable. They also inferred that the binaural mechanisms are essentially the same at low and high frequencies. Zhan and Wright drew the same conclusion through evaluations of internal noise in the ITD and IID processing pathways by comparing within- and across-listener variations in performance. It is difficult to explain the apparent discrepancy between the study by Zhan and Wright (2007) and the present study, not only because of the large differences in the experimental paradigm, but also because it remains unclear what specific mechanisms in the IID processing are malleable in the learning process. It is possible that in the IID task described by Zhan and Wright (2007), listeners might learn to use cues not specific to laterality-related sensation. A candidate is the monaural loudness cue, which was made ineffective by level roving in the present study, although the monaural loudness cue appears to make little contribution to the performance of trained listeners (Bernstein, 2004).

C. Dependence on baseline ITD and IID

Generally, the baseline ITD and IID had little effect on the degree of ITD and IID interaction, with one exception (500-Hz tone; 600 μs/−12 dB). This indicates that the mechanism that combines ITD and IID is not tuned to a particular baseline ITD and IID combination. This is remarkable when considering that the parameters tested in the present study covered wide ranges of baseline ITD and IID relative to the ranges that we experience in natural settings, and included unnatural combinations of ITD and IID (Gaik, 1993). It should be pointed out that the listeners' sensitivities to ITD or IID changes *do* depend strongly on baselines (Fig. 2).

Exceptionally, when the stimulus was a 500-Hz tone and the baseline ITD/IID was 600 μs/−12 dB, the estimated ρ value was below zero. This combination of stimulus type and baseline values is ecologically and perceptually very unnatural (Gaik, 1993). The observation of a small ρ value for this condition could be related to the "cue-reversal point" in the plot of intracranial lateral position versus ΔITD (Domnitz and Colburn, 1977). For a sinusoidal stimulus, a small change in ITD toward one direction can move the intracranial image toward the opposite direction depending on whether the baseline ITD is positive or negative relative to the cue-reversal point. For a baseline IID of 0 dB, the cue-reversal point is usually around the ITD corresponding to the phase of $\pi/2$ (i.e., 500 μs for a 500-Hz tone). Domnitz and Colburn (1977) showed that the reversal point was closer to 0 μs, namely ~300 μs for a 500-Hz tone, when the baseline IID was around −12 dB. A baseline ITD of 600 μs was sufficiently beyond the cue reversal point when the baseline IID was −12 dB. Thus, at these baseline values, simultaneously changing ITD and IID in the same direction could have a subtractive (rather than an additive) effect. In fact, the d' value for combined ITD and IID changes was often below that for individual changes, which was particularly evident for listeners RZ and SF. This notion can be understood graphically by switching the vector for ITD to the opposite direction in Fig. 1(b) (see the vector with the dashed line labeled ITD′). Given a "true" angle between the two dimen-

sions of θ (corresponding to ρ), the angle between the vectors representing measured d'_T and d'_I would be $180°-\theta$, which corresponds to $-\rho$. Replotting the ρ values for listeners RZ and SF at the baseline ITD/IID of 600 μs/−12 dB in Fig. 6 so that they have the opposite sign, we can see that the replotted ρ values would no longer differ greatly from the values at the other baseline ITD/IID. With this line of explanation, the exceptionally small (often negative) value of ρ would be attributed to the property of the ITD-related "signal" [denoted as S(ITD) in Fig. 1], rather than to the channel interaction property. However, it should be noted that this observation could not be entirely accounted for by the present linear model. A complete explanation requires a mechanism that can account for the dependence of the cue-reversal point on baseline IID (e.g., Stern and Colburn, 1978).

D. Underlying physiological mechanisms

It has been argued thus far that, under the linear additivity model, the two binaural cues are processed by "partially independent/common mechanisms" $(0<\rho<1)$ at low frequencies and by a "common mechanism" $(\rho\cong 1)$ at high frequencies. This section speculates on what specific physiological mechanisms might correspond to such independent and common processes.

For low frequencies, the MSO and the LSO are candidates for the more or less independent channels. As mentioned in Sec. I, low frequencies are overrepresented in the MSO, where mainly ITD is encoded presumably by the excitatory–excitatory type of binaural neurons acting as a coincidence detector. The LSO, which is considered to encode primarily IID by the excitatory–inhibitory type of neurons, is also known to have low-frequency neurons (Yin, 2002; Tollin, 2003; Tollin and Yin, 2005). Interactions of ITD and IID, which correspond to the "common" component of the two channels, can occur within those nuclei. The MSO also has inhibitory inputs that may play roles in encoding ITD and/or IID (Grothe, 2003). The LSO is known to have a population of low frequency neurons that are sensitive to both ITD and IID (Yin, 2002; Tollin, 2003; Tollin and Yin, 2005). There, neural responses are probably shaped by the relative amplitude of the excitatory and inhibitory inputs from the two ears, reflecting mainly the intensity of the stimulus, and also by the relative arrival time of those inputs, reflecting not only the stimulus timing difference, but also the relative stimulus intensity ("the latency hypothesis," Park *et al.*, 1996; Irvine *et al.*, 2001). Cue interaction can also occur in more peripheral processes, e.g., as neuronal noise in auditory nerve example, the converging inputs to the inferior colliculus (Palmer *et al.*, 2007); Stanford *et al.*, 1992; McAlpine *et al.*, 1998; Shackleton *et al.*, 2000). Possibly, the net degree of ITD and IID interaction (quantified as ρ) is determined by the relative contributions of these separate and converging pathways to the behavioral performance.

On the other hand, for high frequencies, it may be that essentially a single pathway was involved in the present task. The contributions of independent processes, if they exist, were smaller than for low frequencies. It can be speculated that the LSO pathway was dominant, where both ITD and IID information is processed by a common excitatory–inhibitory mechanism, namely the previously described latency mechanism, which possibly involves a nonlinear processes (Park *et al.*, 1996; Irvine *et al.*, 2001). The inferior colliculus also is known to have high frequency neurons that are simultaneously sensitive to ITD and IID (Caird and Klinke, 1987; Irvine *et al.*, 1995). Theoretically, however, there are other possibilities. One is that the binaural mechanism as a whole has little noise, and the performance for binaural tasks is possibly limited mainly by the peripheral noise (Colburn, 1973). Another possibility is that ITD and IID are initially processed in separate nuclei (with little internal noise), but that the outputs converge completely in a higher auditory pathway. In this case, the psychophysical performance is governed by the internal noise in the higher pathway. The low level, separate ITD and IID processes may be not accessible by the decision device used in the present paradigm, but only by the decision device for certain other paradigms (e.g., Zhang and Wright, 2007).

E. Nonlinear processes?

Thus far, discussions have been made under the framework of a working model, which assumes the linear additivity of binaural cues. However, the present study revealed some results that were unexpected within this framework. Specifically, the listener RZ data for the 4-kHz transposed tone showed that the ρ value was markedly greater than one. Thus, it is possible that a nonlinear mechanism makes a non negligible contribution to the binaural processing. In fact, major successful models of binaural processing incorporate the idea of nonlinear interaction (e.g., Stern and Colburn, 1978; Lindemann, 1986; Breebaart *et al.*, 2001). We speculate that the ITD- and IID-related signals interact in some nonlinear fashion (e.g., effectively multiplication, expansion, etc.), to which a common noise, as assumed in the previous models (Stern and Colburn, 1978; Breebaart *et al.*, 2001), is added. This scheme could be intuitively understood by replacing the portion for the linear addition of input signals in Fig. 1(c) (left diagram) with some kind of nonlinear operation. The model analysis performed in Sec. IV A is in accord with this scheme. Here, $\rho>1$ would be observed if the resulting "signal" after the nonlinear interaction is greater than the simple sum of the signal when ITD or IID was presented alone. Unfortunately, this possibility could not be tested with currently available data and models. This is because the parameters of the existing elaborate models were adapted for low frequency signals, and thus are not readily applicable to the present data for the 4-kHz transposed tones.

V. CONCLUDING REMARKS

The present study demonstrated the difference between low and high frequency signals with respect to the effect of combining cues on the detectability of ITD/IID changes. Since the binaural system inputs were made comparable by using the transposed stimuli for high frequency (van de Par and Kohlrausch, 1997), the results imply that different binaural mechanisms are involved at low and high frequencies.

Future studies, incorporating various kinds of approaches (e.g., psychophysical learning, computer modeling, and human-physiological techniques) are expected to reveal the details of underlying mechanisms more specifically, regarding the place and manner of the cue interaction, and the relative contribution of internal noise at various stages of auditory processing (i.e., from the stochastic nature of the auditory nerve to high level cognitive processing).

ACKNOWLEDGMENTS

The author thanks Dr. Richard Freyman, the Associate Editor, and two anonymous reviewers for their thoughtful comments, which significantly improved the quality of the paper.

[1] $d'_C=\sqrt{d'^2_T+d'^2_I+2\rho d'_T d'_I}=\sqrt{d'^2_T+d'^2_I+2d'_T d'_I}=\sqrt{(d'_T+d'_I)^2}=d'_T+d'_I$.

[2] The formula used in the study, $d'=\sqrt{2}$ z-score(P_c) (Macmillan and Creelman, 2005), can be derived as follows: The internal variables for the stimulus at baseline value (x_N) and the stimulus with ITD and/or IID change (x_S) are assumed to be samples from Gaussian distributions with means of 0 and μ, respectively, and with equal variance (σ^2), hence $d'=(\mu-0)/\sigma$. In a 2AFC task, the listener is expected to choose the stimulus interval for which the internal variable is greater, and thus the listener's judgment is correct when $x_S-x_N>0$. The random variable x_S-x_N has a Gaussian distribution with a mean of μ and a variance of $(2\sigma^2)$, which could be normalized so that it has a unity variance and a mean of $\mu/(\sqrt{2}\sigma)$, which is $d'/\sqrt{2}$. Thus, the z score of the proportion of correct responses is $d'/\sqrt{2}$. With the present stimulus configuration, however, the listener might look at the difference of the second burst from the baseline burst within an interval, and compare the size of the difference between the two intervals. In this case, the proportion of correct answers should be the probability of $y_S-y_N>0$, where y_s is the difference between the random samples of x_S and x_N, and y_N is the difference between two random samples of x_N. y_S-y_N follows a Gaussian distribution with a mean of μ and a variance of $(4\sigma^2)$. This would lead to the equation $d'=2$ z-score(P_c). If this formula represents the "true" psychophysical processes, one can expect that the d' computed in the present study underestimated the true magnitude of d' by factor of $\sqrt{2}$.

[3] The recalculation was performed for each listener represented in Hafter *et al.* (1990, Fig. 3) in the following way: First, a straight line passing the origin was fitted to the plot of d' vs ΔITD with no ΔIID (IDI=0 dB by their convention). Then, for the data when a nonzero ΔIID was imposed, d'_I (for each ΔIID) and the ρ values were varied individually, and the d'_C value was predicted for each data point using Eq. (1) on the basis of d'_T (estimated by the linear fit), and each instance of the d'_I and ρ values. The combination of d'_I and ρ that would minimize the sum of the squared errors of the d'_C prediction was found. The ρ value found in this way was adopted as the final ρ estimate. The estimates for listeners KK, EW, TD, and RF were 1.3, 1.1, 0.9, and 3.5, respectively. It should be noted that the ρ estimates obtained here were slightly different from those given in Hafter *et al.* (1990) even when $\rho<1$, although the best-fitting curves of d'_C appeared similar in both studies. This small difference was probably due to detailed differences in the fitting method which are not described in Hafter *et al.* (1990).

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997a**). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. I. Heterogeneity of responses," J. Neurophysiol. **78**, 1222–1236.

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997b**). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. II. Coincidence detection," J. Neurophysiol. **78**, 1237–1247.

Bernstein, L. R. (**2001**). "Auditory processing of interaural timing information: New insights," J. Neurosci. Res. **66**, 1035–1046.

Bernstein, L. R. (**2004**). "Sensitivity to interaural intensitive disparities: Listeners' use of potential cues," J. Acoust. Soc. Am. **115**, 3156–3160.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli"," J. Acoust. Soc. Am. **112**, 1026–1036.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Bremer, C. D., Pittenger, J. B., Warren, R., and Jenkins, J. J. (**1977**). "An illusion of auditory saltation similar to the cutaneous 'rabbit'," Am. J. Psychol. **90**, 645–654.

Caird, D., and Klinke, R. (**1987**). "Processing of interaural time and intensity differences in the cat inferior colliculus," Exp. Brain Res. **68**, 379–392.

Cohen, D. J. (**1997**). "Visual detection and perceptual independence: Assessing color and form," Percept. Psychophys. **59**, 623–635.

Colburn, H. S. (**1973**). "Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination," J. Acoust. Soc. Am. **54**, 1458–1470.

Colburn, H. S., and Esquissaud, P. (**1976**). "An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli," J. Acoust. Soc. Am. **59**, S23–S23.

Domnitz, R. (**1973**). "The interaural time jnd as a simultaneous function of interaural time and interaural amplitude," J. Acoust. Soc. Am. **53**, 1549–1552.

Domnitz, R. H., and Colburn, H. S. (**1977**). "Lateral position and interaural discrimination," J. Acoust. Soc. Am. **61**, 1586–1598.

Gaik, W. (**1993**). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," J. Acoust. Soc. Am. **94**, 98–110.

Gilliom, J. D., and Sorkin, R. D. (**1972**). "Discrimination of interaural time and intensity," J. Acoust. Soc. Am. **52**, 1635–1647.

Grantham, D. W. (**1984**). "Interaural intensity discrimination—Insensitivity at 1000 Hz," J. Acoust. Soc. Am. **75**, 1191–1194.

Green, D. M. (**1958**). "Detection of multiple component signals in noise," J. Acoust. Soc. Am. **30**, 904–911.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Griffiths, T. D., Elliott, C., Coulthard, A., Cartlidge, N. E., and Green, G. G. (**1998**). "A distinct low-level mechanism for interaural timing analysis in human hearing," NeuroReport **9**, 3383–3386.

Grothe, B. (**2000**). "The evolution of temporal processing in the medial superior olive, an auditory brainstem structure," Progr Med. (Napoli) **61**, 581–610.

Grothe, B. (**2003**). "Sensory systems: New roles for synaptic inhibition in sound localization," Nat. Rev. Neurosci. **4**, 540–550.

Hafter, E. R., and Carrier, S. C. (**1970**). "Masking-level differences obtained with a pulsed tonal masker," J. Acoust. Soc. Am. **47**, 1041–1047.

Hafter, E. R., and Carrier, S. C. (**1972**). "Binaural interaction in low-frequency stimuli: The inability to trade time and intensity completely," J. Acoust. Soc. Am. **51**, 1852–1862.

Hafter, E. R., and Dye, R. H., Jr. (**1983**). "Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number," J. Acoust. Soc. Am. **73**, 644–651.

Hafter, E. R., Dye, R. H., Jr., Wenzel, E. M., and Knecht, K. (**1990**). "The combination of interaural time and intensity in the lateralization of high-frequency complex signals," J. Acoust. Soc. Am. **87**, 1702–1708.

Harris, G. G. (**1960**). "Binaural interactions of impulsive stimuli and pure tones," J. Acoust. Soc. Am. **32**, 685–692.

Hartmann, W. M., and Constan, Z. A. (**2002**). "Interaural level differences and the level-meter model," J. Acoust. Soc. Am. **112**, 1037–1045.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Irvine, D. R., Park, V. N., and Mattingley, J. B. (**1995**). "Responses of neurons in the inferior colliculus of the rat to interaural time and intensity differences in transient stimuli: Implications for the latency hypotheses," Hear. Res. **85**, 127–141.

Irvine, D. R. F. (**1992**). "Physiology of the auditory brainstem," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 153–231.

Irvine, D. R. F., Park, V. N., and McCormick, L. (**2001**). "Mechanisms underlying the sensitivity of neurons in the lateral superior olive to interaural intensity differences," J. Neurophysiol. **86**, 2647–2666.

Keen, K. (**1972**). "Preservation of constant loudness with interaural amplitude asymmetry," J. Acoust. Soc. Am. **52**, 1193–1196.

Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Erlbaum, Mahwah, NJ).
Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.
McAlpine, D., Jiang, D., Shackleton, T. M., and Palmer, A. R. (**1998**). "Convergent input from brainstem coincidence detectors onto delay-sensitive neurons in the inferior colliculus," J. Neurosci. **18**, 6026–6039.
Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.
Palmer, A. R., Liu, L. F., and Shackleton, T. M. (**2007**). "Changes in interaural time sensitivity with interaural level differences in the inferior colliculus," Hear. Res. **223**, 105–113.
Park, T. J., Grothe, B., Pollak, G. D., Schuller, G., and Koch, U. (**1996**). "Neural delays shape selectivity to interaural intensity differences in the lateral superior olive," J. Neurosci. **16**, 6554–6566.
Pratt, H., Polyakov, A., and Kontorovich, L. (**1997**). "Evidence for separate processing in the human brainstem of interaural intensity and temporal disparities for sound lateralization," Hear. Res. **108**, 1–8.
Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.
Ruotolo, B. R., Stern, R. M., Jr., and Colburn, H. S. (**1979**). "Discrimination of symmetric time-intensity traded binaural stimuli," J. Acoust. Soc. Am. **66**, 1733–1737.
Schröger, E. (**1996**). "Interaural time and level differences: Integrated or separated processing?," Hear. Res. **96**, 191–198.
Shackleton, T. M., McAlpine, D., and Palmer, A. R. (**2000**). "Modelling convergent input onto interaural-delay-sensitive inferior colliculus neurones," Hear. Res. **149**, 199–215.
Spitzer, M. W., and Semple, M. N. (**1995**). "Neurons sensitive to interaural phase disparity in gerbil superior olive: Diverse monaural and temporal response properties," J. Neurophysiol. **73**, 1668–1690.
Stanford, T. R., Kuwada, S., and Batra, R. (**1992**). "A comparison of the interaural time sensitivity of neurons in the inferior colliculus and thalamus of the unanesthetized rabbit," J. Neurosci. **12**, 3200–3216.
Stern, R. M. Jr., and Colburn, H. S. (**1985**). "Lateral-position-based models of interaural discrimination," J. Acoust. Soc. Am. **77**, 753–755, .
Stern, R. M., Jr., and Colburn, H. S. (**1978**). "Theory of binaural interaction based in auditory-nerve data. IV. A model for subjective lateral position," J. Acoust. Soc. Am. **64**, 127–140.
Tanner, W. P., Jr. (**1956**). "Theory of recognition," J. Acoust. Soc. Am. **28**, 882–888.
Tardif, E., Murray, M. M., Meylan, R., Spierer, L., and Clarke, S. (**2006**). "The spatio-temporal brain dynamics of processing and integrating sound localization cues in humans," Brain Res. **1092**, 161–176.
Tollin, D. J. (**2003**). "The lateral superior olive: A functional role in sound source localization," Neuroscientist **9**, 127–143.
Tollin, D. J., and Yin, T. C. T. (**2005**). "Interaural phase and level difference sensitivity in low-frequency neurons in the lateral superior olive," J. Neurosci. **25**, 10648–10657.
van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.
Whitworth, R. H., and Jeffress, L. A. (**1961**). "Time versus intensity in the localization of tones," J. Acoust. Soc. Am. **33**, 925–929.
Wright, B. A., and Fitzgerald, M. B. (**2001**). "Different patterns of human discrimination learning for two interaural cues to sound-source location," Proc. Natl. Acad. Sci. U.S.A. **98**, 12307–12312.
Yin, T. C. (**2002**). "Neural mechanisms of encoding binaural localization cues in the auditory brainstem," in *Integrative Functions in the Mammalian Auditory Pathway*, edited by D. Oertel, R. R. Fay and A. N. Popper (Springer, New York), pp. 99–159.
Yost, W. A. (**1972**). "Tone-on-tone masking for three binaural listening conditions," J. Acoust. Soc. Am. **52**, 1234–1237.
Yost, W. A., and Dye, R. H., Jr. (**1988**). "Discrimination of interaural differences of level as a function of frequency," J. Acoust. Soc. Am. **83**, 1846–1851.
Yost, W. A., and Hafter, E. R. (**1987**). "Lateralization," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 3–25.
Zhang, Y., and Wright, B. A. (**2007**). "Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies," J. Acoust. Soc. Am. **121**, 2207–2216.

Adaptive feedback cancellation in hearing aids with clipping in the feedback path[a)]

Daniel J. Freed[b)]
Department of Human Communication Sciences and Devices, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 25 October 2006; revised 27 December 2007; accepted 27 December 2007)

Adaptive linear filtering algorithms are commonly used to cancel feedback in hearing aids. The use of these algorithms is based on the assumption that the feedback path is linear, so nonlinearities in the feedback path may affect performance. This study investigated the effect on feedback canceller performance of clipping of the feedback signal arriving at the microphone, as well as the benefit of applying identical clipping to the cancellation signal so that the cancellation path modeled the nonlinearity of the feedback path. Feedback signal clipping limited the amount of added stable gain that the feedback canceller could provide, and caused misadjustment in response to high-level inputs, by biasing adaptive filter coefficients toward lower magnitudes. Cancellation signal clipping mitigated these negative effects, permitting higher amounts of added stable gain and less misadjustment in response to high-level inputs, but the benefit was reduced in the presence of the highest-level inputs. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2836767]

PACS number(s): 43.66.Ts, 43.60.Mn [DOS] Pages: 1618–1626

I. INTRODUCTION

Cancellation of feedback in hearing aids may be accomplished by means of adaptive linear filtering algorithms, such as the least mean square or normalized least mean square (NLMS) algorithms. These algorithms calculate a time-varying linear filter that estimates the transfer function of the hearing aid feedback path. The filter is applied to the hearing aid output signal in order to form an estimate of the feedback signal, which is then subtracted from the hearing aid input signal to cancel the feedback. This approach has been shown to be effective, both in laboratory studies and in commercial hearing aids (Bustamante *et al.*, 1989; Dyrlund and Bisgaard, 1991; Kates, 1991; Engebretson *et al.*, 1993; Joson *et al.*, 1993; Dyrlund *et al.*, 1994; Maxwell and Zurek, 1995; Kaelin *et al.*, 1998; Kates, 1999; Siqueira and Alwan, 2000; Greenberg *et al.*, 2000; Hellgren and Forssell, 2001; Hellgren, 2002; Chi *et al.*, 2003; Spriet *et al.*, 2005; Boukis *et al.*, 2006; Freed and Soli, 2006).

The use of an adaptive linear filter to model hearing aid feedback is based on the assumption that the feedback path is linear. In the context of digital feedback cancellation, the "feedback path" refers to the entire chain of elements from the output of the digital processing system back to its input, consisting of the digital-to-analog (D/A) converter, the hearing aid receiver (and any associated amplifier), the acoustical and/or mechanical feedback path from the receiver to the microphone, the microphone (and any associated preamplifier), and the analog-to-digital (A/D) converter. In fact, any of these elements may introduce nonlinearity into the feedback path. As nonlinearity violates the underlying assumption of the feedback cancellation algorithm, it may be expected to affect the performance of the algorithm.

The purpose of this study is to undertake an initial investigation of how feedback path nonlinearity affects feedback canceller performance, and how a feedback canceller may be modified to accommodate feedback path nonlinearity. Because of the difficulty of analyzing adaptive nonlinear systems mathematically, an empirical approach is taken. The study addresses one form of nonlinearity: clipping of the feedback signal arriving at the microphone.

Feedback signal clipping occurs whenever the feedback signal arriving at the microphone exceeds the input range of the A/D converter or the microphone preamplifier; this study focuses on the case where the A/D converter range is exceeded, which produces a simple form of hard clipping that is easy to model. Because feedback clipping causes a distorted signal to be presented to the user, a hearing aid designer generally should try to guarantee that this condition will never occur, but design constraints may make it impossible to provide such a guarantee. Ensuring against feedback signal clipping would require ensuring that the maximum acoustic output of the system, minus the attenuation provided by the acoustic feedback path, is less than the maximum acoustic input to the system. However, the hearing aid designer cannot predict the feedback path with certainty. Furthermore, the designer is not free to reduce the output limit, or to increase the input limit, arbitrarily. Reducing the maximum output level may compromise the benefit of the hearing aid, and increasing the maximum input level may raise quantization noise to an unacceptable level.

The likelihood of feedback signal clipping may be increased in some newer types of hearing aids. One example is an open-ear system, in which the conventional earmold of a behind-the-ear hearing aid is replaced by a sound delivery device that leaves the ear canal unoccluded (Chung, 2004).

[a)]Portions of this work were presented at the International Hearing Aid Research Conference (IHCON), Tahoe City, CA, August 2006.
[b)]Electronic mail: dfreed@hei.org

 0001-4966/2008/123(3)/1618/9/$23.00

The lack of occlusion reduces the attenuation in the acoustic feedback path. Another example is a middle ear implant system, in which a vibratory transducer is attached to the ossicular chain (e.g., Kasic and Fredrickson, 2001). Combining such a system with an implanted microphone (e.g., Miller, 2004) may give rise to a mechanical feedback path via bone conduction that demonstrates relatively little attenuation.

The problem of nonlinearity in a system being modeled by an adaptive linear filter has been addressed previously in the context of acoustic echo cancellation in hands-free telephones (speaker phones). In this application, the primary source of nonlinearity is the distortion that results from overdriving the speaker. Investigators have sought to improve the performance of acoustic echo cancellers by incorporating an adaptive nonlinearity into the cancellation path, with the intent that the nonlinearity will adapt to model the speaker distortion, yielding a better match between the cancellation path and the echo path. Several classes of adaptive nonlinearities have been tried, including neural networks (Birkett and Goubran, 1996; Ngia and Sjöberg, 1998), Volterra filters (Stenger *et al.*, 1999; Guérin *et al.*, 2003), spline functions (Uncini *et al.*, 2002), polynomial functions (Stenger and Kellermann, 2000), and "hard" and "soft" saturating nonlinearities (Nollett and Jones, 1997; Stenger and Kellermann, 2000), each of which has demonstrated benefit.

As Maxwell and Zurek (1995) and others have pointed out, feedback cancellation differs from typical applications of adaptive linear filtering (including acoustic echo cancellation) in that the adaptive filter is embedded in a closed-loop system. The loop is created by using the adaptation error signal to feed the hearing aid forward path, the output of which becomes the input to the adaptive filter. One consequence of the loop is to invalidate the assumption, typical in other adaptive filtering applications, that the canceller input signal and the "perturbing signal" are statistically independent. In acoustic echo cancellation, for example, the perturbing signal is the near-end talker signal and the canceller input is the far-end talker signal, so it is reasonable to assume that they are uncorrelated. By contrast, in feedback cancellation, the perturbing signal is the direct input to the hearing aid and the canceller input is the hearing aid output. For nonwhite inputs, these two signals are correlated. Consequently, it cannot be assumed that the nonlinear modeling techniques that have been shown to be effective in acoustic echo cancellation will also be effective in feedback cancellation.

In this study, a simplified version of the hard saturating nonlinearity (clipping) used by Nollett and Jones (1997) and Stenger and Kellermann (2000) is used to model feedback signal clipping in a feedback canceller. The feedback signal clipping behavior is often known *a priori*, in which case it is not necessary to use an adaptive nonlinearity in the cancellation path. Instead, fixed clipping, set to match the clipping in the feedback path, may be used in the cancellation path. The use of a fixed nonlinearity considerably simplifies the implementation.

II. METHODS

All experiments in this study were conducted as simulations, using MATLAB/SIMULINK software. Figure 1 shows the

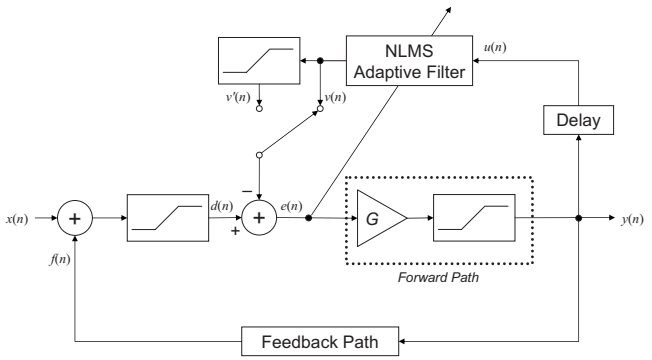


FIG. 1. Model used in simulation experiments.

model that was used in the simulations. The direct input signal $x(n)$ is added to the feedback signal $f(n)$, and the sum is clipped to form $d(n)$. Depending on the experimental condition, either the unclipped cancellation signal $v(n)$ or the clipped cancellation signal $v'(n)$ is then subtracted from $d(n)$ to form the error signal $e(n)$, which both controls the filter adaptation and serves as the input to the forward path processing of the hearing aid. For simplicity, the forward path is restricted to a gain G, followed by clipping to keep the output signal in the appropriate range. The resulting output signal $y(n)$ is used to generate the feedback signal, and a delayed version $u(n)$ feeds the adaptive filter.

The model represents the feedback path as a finite impulse response filter. As described in Sec. I, the feedback path includes the receiver and microphone responses, not just the acoustical/mechanical feedback, and thus the filter gain can exceed 0 dB. Three different impulse responses were used during simulations, representing three levels of feedback that are referred to herein as "low gain," "medium gain," and "high gain." The impulse responses were derived by applying three different scaling factors to a feedback impulse response measured on an experimental open-ear hearing aid system.

Figure 2 shows the impulse response of the low-gain feedback filter. The sampling rate is 16 kHz. The first 35 samples of the impulse response are zero, corresponding to a

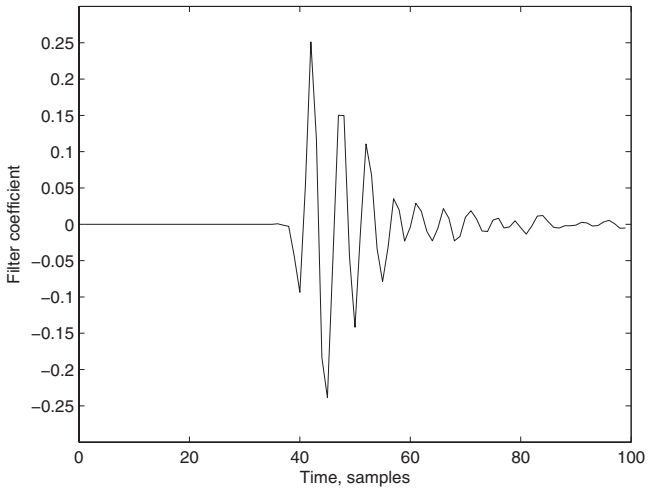


FIG. 2. Impulse response of "low-gain" feedback path filter.

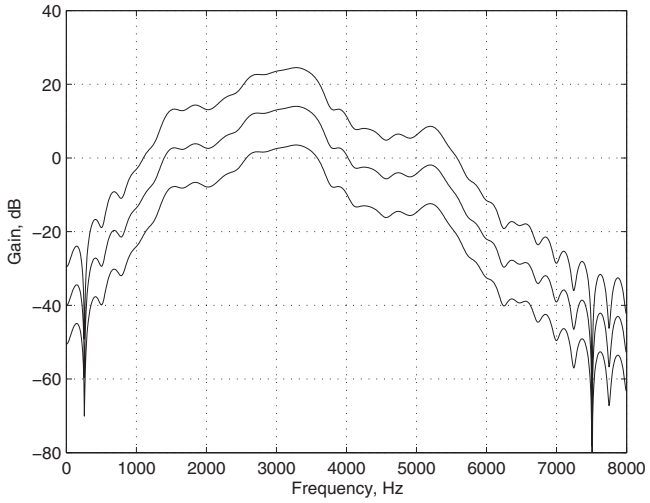


FIG. 3. Frequency response of "low-gain," "medium-gain," and "high-gain" feedback path filters.

pure delay of 35 samples. This delay results from a combination of the propagation time of the acoustic feedback and the processing delay of the A/D and D/A converters in the system that was used to measure the feedback impulse response. To compensate, the model applies a 35-sample delay to the adaptive filter input $u(n)$. Figure 3 shows the frequency response of all three filters.

All simulations were performed using floating-point arithmetic, which allows signal levels to grow virtually without bound. In order to model the range limitations of the A/D and D/A converters, however, signal levels must be limited at the points in the model corresponding to the converters. A magnitude of 1.0 is taken to represent the range limit of the converters; this value is referred to herein as 0 dBFS (0 dB re: full-scale digital signal level). The clipping level for all three of the clipping blocks in the model is 0 dBFS. Clipping of $d(n)$ represents the A/D converter range limit and imposes the feedback signal clipping that renders the feedback path nonlinear. Clipping of $y(n)$ represents the D/A converter range limit. Finally, clipping of $v'(n)$ is the nonlinearity introduced into the cancellation path to match the nonlinearity in the feedback path. At all points in the model where clipping is not explicitly indicated, signals are allowed to grow arbitrarily large.

The adaptive filter is calculated using a standard NLMS algorithm with two minor modifications. Commercial hearing aids typically use more complex algorithms that include band-limiting, coefficient constraints, or other substantial modifications to standard NLMS. In order to avoid limiting the applicability of the results, such modifications were not used in this study.

At sample n, the calculations are

$$v(n) = \sum_{k=0}^{N-1} h_k(n)u(n-k), \tag{1}$$

$$v'(n) = \begin{cases} -1 & \text{if } v(n) < -1 \\ +1 & \text{if } v(n) > +1 \\ v(n) & \text{otherwise,} \end{cases} \tag{2}$$

$$e(n) = \begin{cases} d(n) - v(n) & \text{if canc. clipping disabled} \\ d(n) - v'(n) & \text{if canc. clipping enabled,} \end{cases} \tag{3}$$

$$p(n) = (1-\rho)y^2(n) + \rho p(n-1), \tag{4}$$

$$h_k(n+1) = h_k(n) + \frac{\mu e(n)u(n-k)}{Np(n)+\varepsilon}, \tag{5}$$

where $h_k(n)$ is the kth-order filter coefficient at sample n; N is the number of filter coefficients; μ is the step-size parameter that controls the rate of adaptation; $p(n)$ is the estimated power of $y(n)$, computed by applying a 1-pole low-pass filter to $y^2(n)$; ρ is the pole of the 1-pole low-pass filter; ε is a small constant (approximately 2.2×10^{-16}) to prevent divide-by-zero; and all other symbols are as defined in Fig. 1.

For all simulations, N was set to 32, μ was set to 0.05, and ρ was set to $1-2^{-6}=0.984375$. Given that the nonzero portion of the feedback impulse response is longer than 32 samples, setting N to 32 causes the adaptive filter to undermodel the feedback path. Undermodeling was intentionally incorporated into the simulations because it can occur in feedback cancellers in real hearing aids, where computational constraints limit the number of filter coefficients (Kates, 2001).

The two modifications to standard NLMS both relate to the computation of the normalization factor $p(n)$ in Eq. (4). First, $p(n)$ estimates the power of the hearing aid output signal $y(n)$ rather than the adaptive filter input signal $u(n)$, which is a delayed version of $y(n)$. This modification addresses a problem noted by Chi *et al.* (2003). A sudden increase in the level of the hearing aid input $x(n)$ produces an immediate rise in the power of $e(n)$, but a delayed rise in the power of $u(n)$. If the normalization factor is based on the power of $u(n)$, the adaptation step size is not immediately reduced to compensate for the rise in $e(n)$, and inappropriately large coefficient updates are performed, resulting in a burst of excessive coefficient misadjustment noise. Basing the normalization factor on $y(n)$ eliminates this problem by ensuring that the normalization factor responds to the $e(n)$ power increase without delay.

Second, $p(n)$ is computed by means of a 1-pole low-pass filter rather than the more standard N-point moving average. This change is necessary to deal with a consequence of clipping. Ordinarily, when the adaptive filter is well-adapted, the power of $e(n)$ roughly matches the power of $x(n)$. However, clipping can disrupt this relationship. Consider the case where $x(n)$ is at a high level (well above the input clipping limit) and dominated by low frequencies. In this case, input clipping produces a signal $d(n)$ that contains runs of consecutive samples at the clipping limit of −1 or +1. If cancellation signal clipping is enabled, the cancellation signal contains similar runs of consecutive clipped samples. This in turn leads to runs of zero-valued samples in $e(n)$. During the zero-valued runs of $e(n)$, the power of $e(n)$ drops to zero—radically different from the power of $x(n)$—resulting in extreme short-term variability in the power of $e(n)$, $y(n)$, and $u(n)$. If the power estimator tracks this variability too rapidly, bursts of excessive coefficient growth may occur at the transitions from zero-power $e(n)$ to high-power $e(n)$. A 1-pole

low-pass estimator smooths the power variability better than an N-point moving average estimator, yielding better performance in the special case of high-level low-frequency input.[1]

Two experiments were performed. Experiment 1 measured the ability of the feedback canceller to adapt successfully in the presence of feedback signal clipping, both with and without cancellation signal clipping, as a function of forward path gain and feedback path gain. Experiment 2 investigated the additional effect of high-level direct input signals on a system that was already adapted.

Before performing the experiments, the maximum value of G for which the system was stable with the feedback canceller disabled was determined for each of the three feedback paths. This value is referred to as the maximum stable gain, or $\mathrm{MSG}_{\mathrm{off}}$, where "off" indicates that the canceller was disabled. $\mathrm{MSG}_{\mathrm{off}}$ was found to be −24, −13.5, and −3 dB for the high-, medium-, and low-gain feedback paths, respectively. (The negative values of $\mathrm{MSG}_{\mathrm{off}}$ indicate that attenuation in the forward path is required for stability. This compensates for the gain in the feedback path, which is due to the inclusion of the receiver and microphone responses, as noted earlier.)

For Experiment 1, the initial values of the adaptive filter coefficients $h_k(0)$ were set to zero for all k. All three feedback paths were used (low gain, medium gain, and high gain). For each choice of feedback path, the forward path gain G was varied from 15 to 27 dB above $\mathrm{MSG}_{\mathrm{off}}$ in 0.5 dB steps. These gains were high enough to ensure that the feedback signal exceeded the clipping limit during some portion of the simulation run. Two choices of direct input signal $x(n)$ were used: (1) white noise with uniform probability distribution and rms amplitude −34.7 dBFS (peak −30.0 dBFS), and (2) speech with rms amplitude −39.0 dBFS (peak −21.3 dBFS). To reduce variability in the results, the same stored noise signal was reused for each simulation run, rather than generating a new noise signal for each run. The speech was taken from the Hearing In Noise Test (HINT; Nilsson *et al.*, 1994) and consisted of a single repeated sentence ("The boy got into trouble") spoken by a male talker. The purpose of using both white noise and speech was to investigate the interaction between the effect of the nonlinearities and the effect of the correlation between the canceller input and perturbing signals that arises with nonwhite input signals (see Sec. I).

The ability of the feedback canceller to cancel feedback was measured by comparing the level of the error signal $e(n)$ to the level of the direct input signal $x(n)$. In the ideal case of perfect cancellation, $e(n)=x(n)$. Any excess energy in $e(n)$ compared to $x(n)$ represents feedback that has not been canceled. To quantify this, each simulation was run for 5 s, the rms levels of $e(n)$ and $x(n)$ were measured during the final 1 s, and the excess error level was defined as the ratio of $e(n)$ level to $x(n)$ level, in decibels. The excess error level results were then used to determine $\mathrm{MSG}_{\mathrm{on}}$, the maximum stable gain with the feedback canceller enabled, for each feedback path and clipping condition. $\mathrm{MSG}_{\mathrm{on}}$ was defined as the maximum value of G for which excess error level was $\leqslant$3 dB. (The 3 dB criterion was selected because it appeared to correspond to a kneepoint in the curves relating G and excess error level; see Fig. 4.) Finally, the added stable gain (ASG) was defined as $\mathrm{MSG}_{\mathrm{on}}$-$\mathrm{MSG}_{\mathrm{off}}$. ASG is the primary measure of benefit of a feedback canceller.

For Experiment 2, the initial values of the adaptive filter coefficients $h_k(0)$ were set to match the first 32 nonzero feedback filter coefficients, so that the coefficients started near their steady-state values. All three feedback paths were used. For each choice of feedback path, the forward path gain G was set to be 18 dB above $\mathrm{MSG}_{\mathrm{off}}$, which was determined in Experiment 1 to be $-\mathrm{MSG}_{\mathrm{on}}$ for all feedback paths and clipping conditions. The corresponding values of G were 15 dB for the low-gain feedback path, 4.5 dB for the medium-gain feedback path, and −6 dB for the high-gain feedback path.

As in Experiment 1, two choices of direct input signal $x(n)$ were used. In both cases, $x(n)$ consisted of three segments. The initial segment, lasting 2 s, was at a moderate level and served to allow the coefficients to adapt to their steady-state values. The middle segment, lasting 5 s, was at a high level to test the effect of a high-level input. The final segment, lasting 5 s, returned to the level of the initial segment to determine how the system recovered after the presentation of a high-level input. For the first input signal, all three segments were white noise with uniform probability distribution. The initial and final segment amplitude was −34.7 dBFS rms (−30.0 dBFS peak) and the middle segment amplitude was varied from −24.7 to +15.3 dBFS rms (−20 to +20 dBFS peak). To reduce variability in the results, the same stored noise signal was reused for each simulation run (with different scaling factors applied to the middle segment), rather than generating a new noise signal for each run. For the second input signal, speech with amplitude −39.0 dBFS rms (−21.3 dBFS peak) was presented during all three segments; during the middle segment, speech-shaped noise was added with amplitude varying from −24.7 to +15.3 dBFS rms (−12.9 to +27.1 dBFS peak). The speech was the same as that used in Experiment 1, and the speech-shaped noise was taken from the HINT (Nilsson *et al.*, 1994). A noise burst was added to the speech in the middle segment, rather than raising the speech amplitude in the middle segment, in order to produce a somewhat more realistic test signal.

The misadjustment of the feedback canceller caused by high-level inputs was quantified by comparing the adaptive filter coefficients to the impulse response of the feedback path. The normalized coefficient error power at sample n was defined as

$$E(n) = \frac{\sum_{k=0}^{31} [h_k(n) - h_k^0]^2 + \sum_{k=32}^{L-1} [h_k^0]^2}{\sum_{k=0}^{L-1} [h_k^0]^2}, \tag{6}$$

where $\{h_0^0, \ldots, h_{L-1}^0\}$ is the L-point impulse response of the feedback path (excluding the initial 35 zero-valued samples). The first sum in the numerator represents the mismatch due to nonoptimal adaptation, which is time varying, and the second sum represents the mismatch due to undermodeling, which is fixed. The total mismatch power is normalized by

dividing it by the total power of the feedback path impulse response.

$E(n)$ was used to derive two outcome measures. First, the mean value of $E(n)$ during the noise burst (middle segment) was measured as an indicator of the degree of misadjustment. Second, the time required to recover from misadjustment after the noise burst was measured. To do this, the mean value of $E(n)$ during the 1 s interval immediately before the noise burst was measured to serve as a baseline error power. Recovery time was then defined as the time required for $E(n)$ to drop back to baseline after the end of the noise burst.

III. RESULTS

Figures 4 and 5 show the excess error level results for Experiment 1 with white noise and speech inputs, respectively. Results without cancellation signal clipping are shown by open symbols, and results with cancellation signal clipping are shown by closed symbols, with symbol shape indicating the feedback path gain. Figure 6 shows the ASG results for both white noise and speech inputs. The white bars show ASG without cancellation signal clipping, and the black bars show ASG with cancellation signal clipping.

As shown in Figs. 4 and 5, the excess error level at lower gains was near 0 dB for all conditions, indicating successful feedback cancellation. With speech input, excess error level was slightly below 0 dB, suggesting that the speech signal itself was partially canceled. Some degree of signal cancellation is expected for nonwhite input signals (Joson *et al.*, 1993; Kates, 1999). At higher gains, excess error level rose, either gradually or rapidly, for all conditions. Gradual growth of excess error level indicates an increasing amount of suboscillatory uncanceled feedback, whereas rapid growth of excess error level indicates that the system had gone into full oscillation and was no longer stable. Results for white noise and speech inputs were similar, except that excess error level rose rapidly in all conditions with speech and rose gradually in some conditions with white noise.

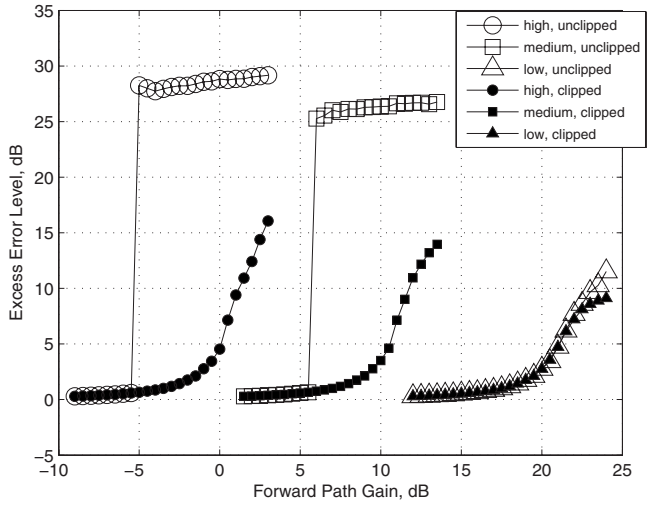


FIG. 4. Excess error level results for Experiment 1 with white noise input. Open symbols are results without cancellation signal clipping; closed symbols are results with cancellation signal clipping. Circles, squares, and triangles are results with high-gain, medium-gain, and low-gain feedback paths, respectively.

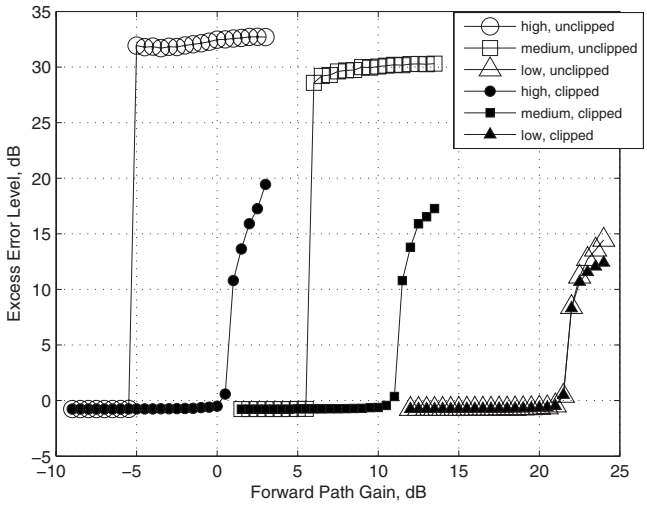


FIG. 5. Same as Fig. 4, but for speech input.

For the high- and medium-gain feedback paths, cancellation signal clipping provided a clear benefit. The gain at which excess error level began rising substantially was higher with cancellation signal clipping than without. For the low-gain feedback path, the benefit of cancellation signal clipping was minimal. It produced a modest reduction in excess error level at the very highest gains, but this is of little practical consequence because these gains are well above MSG_{on}. The same pattern is seen in Fig. 6: Cancellation signal clipping improved ASG for the high- and medium-gain feedback paths by 4–6 dB but provided no improvement for the low-gain feedback path.

Additional insight about the benefit of cancellation signal clipping may be gained by examining the power of the adaptive filter coefficients. Figure 7 shows a typical example.

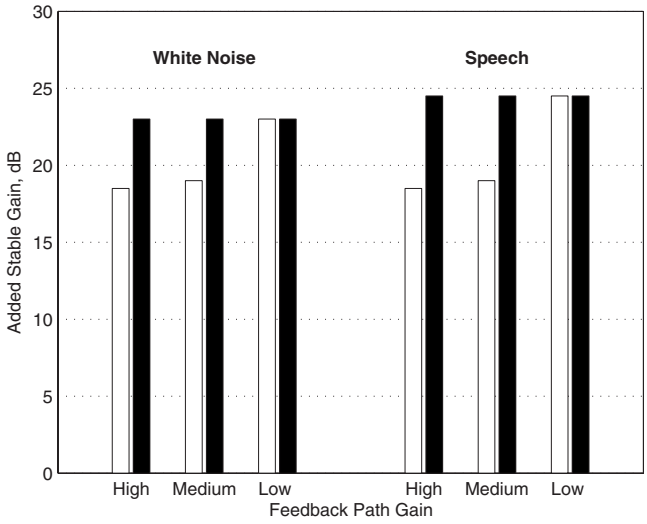


FIG. 6. Added stable gain results for Experiment 1. Results for white noise input are on the left and results for speech input are on the right. White bars are without cancellation signal clipping, black bars are with cancellation signal clipping.

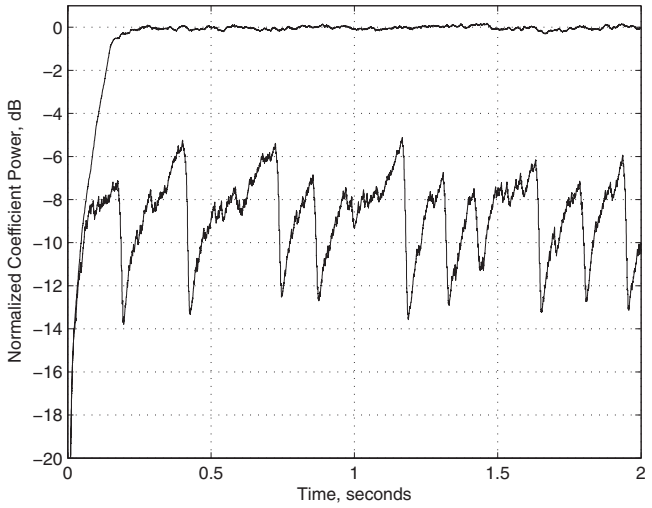


FIG. 7. Normalized coefficient power over time (first 2 s only) for two selected simulation runs of Experiment 1. High-gain feedback path was used, with forward path gain of −5 dB and white noise input. Lower curve is without cancellation signal clipping, upper curve is with cancellation signal clipping.

Figure 7 plots normalized coefficient power (in decibels) as a function of time, where normalized coefficient power at sample n is defined as

$$\frac{\sum_{k=0}^{31}[h_k(n)]^2}{\sum_{k=0}^{L-1}[h_k^0]^2}. \tag{7}$$

Figure 7 shows the results of the first 2 s of two simulation runs. Both runs used white noise input, the high-gain feedback path, and a forward path gain of −5 dB. The lower curve shows the result without cancellation signal clipping: The coefficients persistently underestimate the target (0 dB) by at least 5 dB and display large, undirected variation over time. The upper curve shows the result with cancellation signal clipping: The coefficients adapt smoothly and consistently toward the target, reaching it within approximately 0.25 s.

Figures 8 and 9 show the normalized coefficient error power results for Experiment 2 with white noise and speech/speech-shaped noise inputs, respectively. Results without cancellation signal clipping are shown by open symbols, and results with cancellation signal clipping are shown by closed symbols, with symbol shape indicating the feedback path gain.

The pattern of results was similar for the two types of inputs. The lowest-amplitude noise bursts caused little misadjustment. Once the noise burst amplitude rose above −20 dBFS rms for white noise or −15 dBFS rms for speech-shaped noise, the degree of misadjustment depended on feedback path and clipping condition. For the low-gain feedback path, misadjustment rose gradually with increasing noise burst amplitude, and results were nearly identical with and without cancellation signal clipping. For the high- and medium-gain feedback paths, by contrast, misadjustment rose rapidly without cancellation signal clipping and only

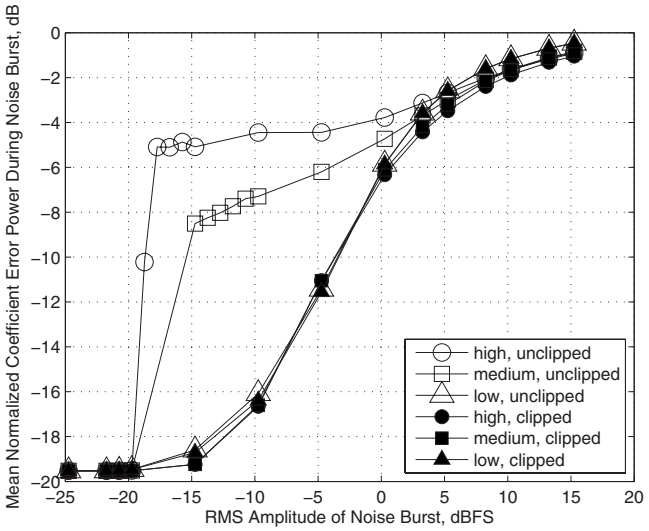


FIG. 8. Normalized coefficient error power results for Experiment 2 with white noise input. Open symbols are results without cancellation signal clipping; closed symbols are results with cancellation signal clipping. Circles, squares, and triangles are results with high-gain, medium-gain, and low-gain feedback paths, respectively.

gradually with it. Consequently, at moderate noise burst amplitudes, cancellation signal clipping had the benefit of substantially reducing misadjustment. This benefit decreased as noise burst amplitude rose and disappeared entirely at the highest amplitudes.

Figures 10 and 11 show the recovery time results for Experiment 2 with white noise and speech/speech-shaped noise inputs, respectively. The symbols are the same as in Figs. 8 and 9. For visual clarity, recovery times below 1 ms are not shown, so some points are missing from Fig. 11. Again, similar patterns are seen for the two types of inputs. Recovery time tended to rise with increasing noise burst amplitude and was generally longer with higher-gain feedback paths. For the low-gain feedback path, results with and without cancellation signal clipping were only slightly different. For the high- and medium-gain feedback paths, cancellation signal clipping provided benefit by reducing recovery time in

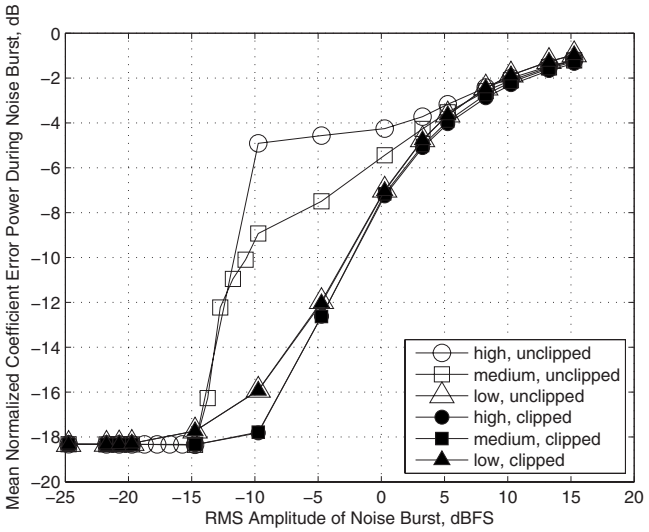


FIG. 9. Same as Fig. 8, but for speech/speech-shaped noise input.

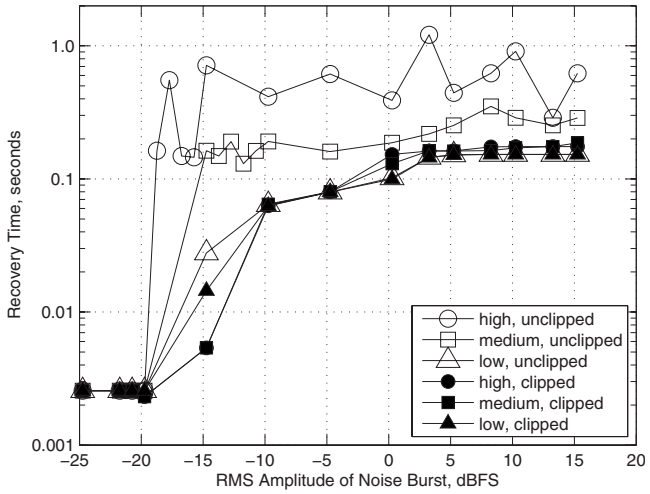


FIG. 10. Recovery time results for Experiment 2 with white noise input. Symbols are the same as in Figs. 8 and 9.

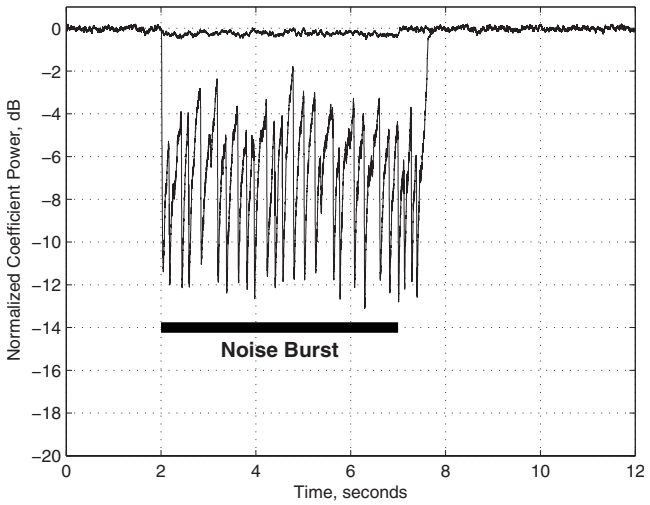


FIG. 12. Normalized coefficient power over time for two selected simulation runs of Experiment 2. High-gain feedback path was used, with noise burst rms amplitude of −14.7 dBFS and white noise input. Lower curve is without cancellation signal clipping, upper curve is with cancellation signal clipping. Heavy horizontal bar indicates time interval of noise burst.

almost every case, except at the lowest noise burst amplitudes where recovery time was already very brief.

Figure 12 shows an example of the benefit of cancellation signal clipping in Experiment 2. The figure shows normalized coefficient power over time during two simulation runs. Both runs used white noise input, the high-gain feedback path, and a noise burst rms amplitude of −14.7 dBFS. The lower curve shows the result without cancellation signal clipping: large drop and wide variation in coefficient power during the noise burst (2–7 s) with recovery approximately 0.5 s after the noise burst ends. The upper curve shows the result with cancellation signal clipping: small drop and minimal power variation during the noise burst with nearly immediate recovery after the noise burst ends.

IV. DISCUSSION

Figures 7 and 12 provide the key to explaining the results of both experiments by showing that, in the absence of cancellation signal clipping, feedback signal clipping biases the adaptive filter coefficients toward lower magnitudes. This may be explained by noting that clipping, being a form of gain reduction, reduces the effective gain of the feedback path; the adaptive filter compensates by reducing coefficient magnitudes. During initial adaptation (Experiment 1), as the adaptive filter coefficients rise in magnitude from zero toward target, a point is reached where the magnitude of the unclipped cancellation signal exceeds the magnitude of the clipped feedback signal. This mismatch yields an increase in the magnitude of the error signal $e(n)$, creating a resistance to further coefficient increases and a slowing of adaptation. If the feedback signal clipping is severe enough, the coefficients never reach target. Similarly, when a high-level input introduces feedback signal clipping into a previously adapted system (Experiment 2), the resulting mismatch between feedback and cancellation signal magnitudes causes an increase in the error signal magnitude. The adaptive filter responds by lowering coefficient magnitudes to reduce the mismatch.

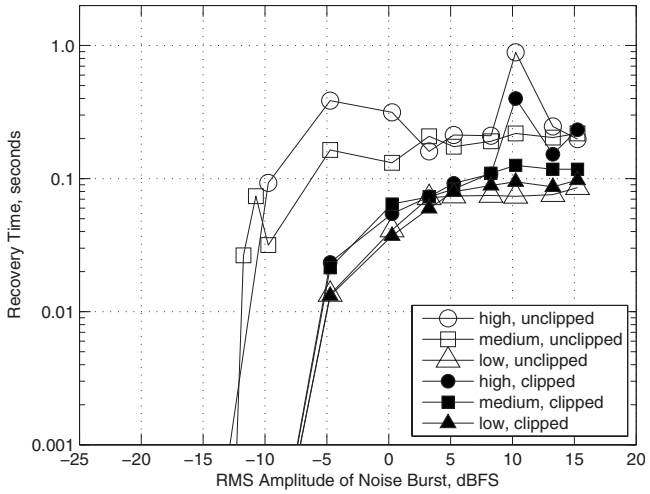


FIG. 11. Same as Fig. 10, but for speech/speech-shaped noise input.

Cancellation signal clipping mitigates the coefficient bias by imposing the same gain limitation in the cancellation path that is present in the feedback path, thus preventing the magnitude mismatch between the two paths. However, as the results of Experiment 2 show, the benefit of cancellation signal clipping is diminished as the level of the direct input signal increases. This phenomenon may be explained as follows. The purpose of clipping the cancellation signal is to introduce a nonlinearity in the cancellation path that matches the nonlinearity in the feedback path. However, the signal that is subjected to clipping at the input is not just the feedback signal, but rather the sum of the direct input signal and the feedback signal (see Fig. 1). As the direct input signal level approaches and exceeds the clipping limit, the clipping behavior will be increasingly affected by the direct input signal. By contrast, clipping of the cancellation signal is affected only by the estimated feedback signal; the direct input signal is not available to the canceller and cannot be included in the model. As a result, for higher-level direct input signals, clipping of the cancellation signal less accurately models the

nonlinearity of the feedback path, and the performance benefit of cancellation signal clipping is reduced.

In both experiments, cancellation signal clipping provided virtually no benefit for the low-gain feedback path. The lack of benefit is likely related to the clipping applied to the output signal $y(n)$ in the forward path (see Fig. 1). Because the maximum level of $y(n)$ is limited, the maximum level of the feedback signal $f(n)$ is only slightly above the input clipping level when the low-gain feedback path is used. In other words, most of the clipping is applied to the output signal rather than the feedback signal. Since the degree of feedback signal clipping is minimal, the benefit of cancellation signal clipping is also minimal. By contrast, with higher-gain feedback paths, most of the clipping is applied to the feedback signal rather than the output signal, so cancellation signal clipping is beneficial.

V. CONCLUSIONS

(1) In a NLMS-based feedback canceller, feedback signal clipping limits the amount of added stable gain that the feedback canceller can provide, and causes misadjustment in response to high-level inputs, by biasing the adaptive filter coefficients toward lower magnitudes.
(2) The negative effects of feedback signal clipping can be mitigated by clipping the cancellation signal, so that the cancellation path models the nonlinearity in the feedback path. However, the benefits of cancellation signal clipping are reduced in the presence of the highest-level inputs.

[1]Clipping of the cancellation signal raises additional questions about the correctness of the coefficient update algorithm. The NLMS algorithm adjusts each coefficient $h_k(n)$ by an amount proportional to $-\partial e^2(n)/\partial h_k(n)$. When the cancellation signal is in the unclipped range (i.e., $|v(n)| \leq 1$), this partial derivative is equal to $2e(n)u(n-k)$, leading to the standard coefficient update equation. But when $|v(n)| > 1$, the partial derivative becomes 0. This would seem to imply that the coefficients should not be updated whenever $|v(n)| > 1$. Indeed, as long as coefficient updates fail to move $v(n)$ out of the clipping range, the clipped cancellation signal $v'(n)$ will remain constant and the updates will have no beneficial effect on error magnitude. Nevertheless, the sign of $e(n)u(n-k)$ still points to the correct direction for coefficient updates; repeated adjustment of all coefficients in this direction may be expected eventually to move $v'(n)$ in the correct direction and reduce error magnitude. (Furthermore, avoiding coefficient updates altogether when $|v(n)| > 1$ runs the risk of causing the feedback canceller to get stuck in a permanently oscillating state.) While the sign of $e(n)u(n-k)$ provides the correct direction for coefficient updates, the appropriate update magnitude is less clear. It is possible that some form of step size adjustment when $|v(n)| > 1$ would improve algorithm performance. This question is not addressed in the present study.

Birkett, A. N., and Goubran, R. A. (**1996**). "Nonlinear loudspeaker compensation for handsfree acoustic echo cancellation," Electron. Lett. **32**, 1063–1064.

Boukis, C., Mandic, D. P., and Constantinides, A. G. (**2006**). "Bias reduction in acoustic feedback cancellation systems with varying all-pass filters," Electron. Lett. **42**, 556–557.

Bustamante, D. K., Worrall, T. L., and Williamson, M. J. (**1989**). "Measurement and adaptive suppression of acoustic feedback in hearing aids," Proceedings of the 1989 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 23–26, **1989**, Glasgow (IEEE, New York), pp. 2017–2020.

Chi, H.-F., Gao, S. X., Soli, S. D., and Alwan, A. (**2003**). "Band-limited feedback cancellation with a modified filtered-X LMS algorithm for hearing aids," Speech Commun. **39**, 147–161.

Chung, K. (**2004**). "Challenges and recent developments in hearing aids. II. Feedback and occlusion effect reduction strategies, laser shell manufacturing processes, and other signal processing technologies," Trends Amplif. **8**, 125–164.

Dyrlund, O., and Bisgaard, N. (**1991**). "Acoustic feedback margin improvements in hearing instruments using a prototype DFS (digital feedback suppression) system," Scand. Audiol. **20**, 49–53.

Dyrlund, O., Henningsen, L. B., Bisgaard, N., and Jensen, J. H. (**1994**). "Digital feedback suppression (DFS): Characterization of feedback-margin improvements in a DFS hearing instrument," Scand. Audiol. **23**, 135–138.

Engebretson, A. M., French-St. George, M., and O'Connell, M. P. (**1993**). "Adaptive feedback stabilization of hearing aids," Scand. Audiol. Suppl. **38**, 56–64.

Freed, D. J., and Soli, S. D. (**2006**). "An objective procedure for evaluation of adaptive antifeedback algorithms in hearing aids," Ear Hear. **27**, 382–398.

Greenberg, J. E., Zurek, P. M., and Brantley, M. (**2000**). "Evaluation of feedback-reduction algorithms for hearing aids," J. Acoust. Soc. Am. **108**, 2366–2376.

Guérin, A., Faucon, G., and Le Bouquin-Jeannès, R. (**2003**). "Nonlinear acoustic echo cancellation based on Volterra filter," IEEE Trans. Speech Audio Process. **11**, 672–683.

Hellgren, J. (**2002**). "Analysis of feedback cancellation in hearing aids with filtered-X LMS and the direct method of closed loop identification," IEEE Trans. Speech Audio Process. **10**, 119–131.

Hellgren, J., and Forssell, U. (**2001**). "Bias of feedback cancellation algorithms in hearing aids based on direct closed loop identification," IEEE Trans. Speech Audio Process. **9**, 906–913.

Joson, H. A. L., Asano, F., Suzuki, Y., and Sone, T. (**1993**). "Adaptive feedback cancellation with frequency compression for hearing aids," J. Acoust. Soc. Am. **94**, 3248–3254.

Kaelin, A., Lindgren, A., and Wyrsch, S. (**1998**). "A digital frequency-domain implementation of a very high gain hearing aid with compensation for recruitment of loudness and acoustic echo cancellation," Signal Process. **64**, 71–85.

Kasic, J. F., and Fredrickson, J. M. (**2001**). "The Otologics MET ossicular stimulator," Otolaryngol. Clin. North Am. **34**, 501–513.

Kates, J. M. (**1991**). "Feedback cancellation in hearing aids: Results from a computer simulation," IEEE Trans. Signal Process. **39**, 553–562.

Kates, J. M. (**1999**). "Constrained adaptation for feedback cancellation in hearing aids," J. Acoust. Soc. Am. **106**, 1010–1019.

Kates, J. M. (**2001**). "Room reverberation effects in hearing aid feedback cancellation," J. Acoust. Soc. Am. **109**, 367–378.

Maxwell, J. A., and Zurek, P. M. (**1995**). "Reducing acoustic feedback in hearing aids," IEEE Trans. Speech Audio Process. **3**, 304–313.

Miller, D. A. (**2004**). "Implantable hearing aid microphone," U.S. Patent No. 6,707,920, 16 March 2004.

Ngia, L. S. H., and Sjöberg, J. (**1998**). "Nonlinear acoustic echo cancellation using a Hammerstein model," Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 12–15, **1998**, Seattle, WA (IEEE, New York), pp. 1229–1232.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Nollett, B. S., and Jones, D. L. (**1997**). "Nonlinear echo cancellation for hands-free speakerphones," presented at 1997 IEEE Workshop on Nonlinear Signal and Image Processing (NSIP), Mackinac Island, MI, available at http://www.ecn.purdue.edu/nsip, last viewed 27 December 2007.

Siqueira, M. G., and Alwan, A. (**2000**). "Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids," IEEE Trans. Speech Audio Process. **8**, 443–453.

Spriet, A., Proudler, I., Moonen, M., and Wouters, J. (**2005**). "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," IEEE Trans. Signal Process. **53**, 3749–3763.

Stenger, A., and Kellermann, W. (**2000**). "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," Signal Process. **80**, 1747–1760.

Stenger, A., Trautmann, L., and Rabenstein, R. (**1999**). "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), March 15–19, **1999**, Phoenix, AZ (IEEE, New York), pp. 877–880.

Uncini, A., Nalin, A., and Parisi, R. (**2002**). "Acoustic echo cancellation in the presence of distorting loudspeakers," presented at the 2002 European Signal Processing Conference (EUSIPCO), Toulouse, France, Paper No. 473, available at http://www.eurasip.org/Proceedings/Eusipco/2002/articles/paper473.pdf, last viewed 27 December 2007.

A biphasic theory for the viscoelastic behaviors of vocal fold lamina propria in stress relaxation

Yu Zhang, Lukasz Czerwonka, Chao Tao, and Jack J. Jiang[a)]
Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, Wisconsin 53792-7375

(Received 31 May 2007; revised 6 December 2007; accepted 7 December 2007)

In this study, a biphasic theory is applied to investigate the viscoelastic behaviors of vocal fold lamina propria during stress relaxation. The vocal fold lamina propria tissue is described as a biphasic material composed of a solid phase and an interstitial fluid phase. The biphasic theory reveals the interaction between the solid and the fluid. For the one-dimensional case, the analytical solutions of solid displacement, fluid velocity, and stress are derived. The biphasic theory predicts the stress relaxation of the vocal fold lamina propria. The quasilinear viscoelastic model as well as its higher-order elastic parameters can be derived from this biphasic theory. Furthermore, the fluid is found to support the majority of the stress at the early stage of stress relaxation; however, when the time becomes sufficiently large, the solid eventually bears all the stress. The early fluid stress support is much higher than the eventual solid support and may be important for understanding the effects of dehydration on tissue damage. By considering the solid-fluid structure of the vocal fold lamina propria, the biphasic theory allows for a more physical theory of tissue viscoelasticity than a single phase solid description and may provide a valuable physical mechanism for the observed vocal fold rheologic behaviors. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2831739]

PACS number(s): 43.70.Bk, 43.35.Mr [BHS] Pages: 1627–1636

I. INTRODUCTION

Many soft, highly hydrated interstitial tissues in the body, including articular cartilage, intervertebral disks, menisci, and corneas, can essentially be divided into a porous permeable solid structure and the fluid that occupies the pores. Such a "biphasic" (solid-fluid) description has gained considerable interest in the last two decades for its successful application in investigating the viscoelastic behavior of these tissues, particularly creep and stress relaxation.[1–3] The biomechanical characterization of soft hydrated tissues using the biphasic theory is currently an active area of interdisciplinary research.[4–8]

Vocal fold tissue also shows the stress relaxation behavior typical of viscoelastic materials. Alipour and Titze[9] found a time-related dependence of stress relaxation of canine vocalis muscle, where stress decreased exponentially with time and decayed very slowly. Chan and Titze[10] measured the viscoelastic behaviors of vocal fold lamina propria under dynamic loading using a rotational rheometer and studied the effects of age and gender on the viscoelastic parameters. To theoretically describe the observed viscoelastic behavior, these studies applied a quasi-linear model combining traditional elements such as springs to represent the elastic component and dashpots to represent the viscous component. The viscoelastic characteristics of vocal fold tissue have recently received considerable interest with computer modeling and viscoelastic parameter measurements.[11–15]

The vocal folds share many structural similarities with other porous permeable tissues. The intermediate vocal fold layer, known as the lamina propria, is a highly hydrated structure containing collagen fibers, elastin, and proteoglycans, similar to articular cartilage.[16–23] The multiphasic structure of vocal fold lamina propria may contribute to its viscoelastic properties. A single phase description has been traditionally used in vocal fold models;[24–27] however, this may oversimplify the influence of the tissue's macromolecular structure on its viscoelastic behavior. Indeed, the single phase elastic theory does not readily explain the stress relaxation phenomenon observed in rheologic experiments on vocal fold lamina propria. Defining the interaction between the fluid and solid components using the biphasic theory may provide the mechanism for the viscoelastic behavior and lead to a more appropriate vocal fold model.

The purpose of this study is to apply the biphasic theory to study the viscoelastic behavior of vocal fold lamina propria in stress relaxation. Vocal fold lamina propria will be described as a biphasic material composed of a solid phase and an interstitial fluid phase, and a biphasic model of the tissue will be given. The analytical solutions of the one-dimensional case for the dependent variables (solid displacement, fluid velocity, fluid pressure, and stresses) will be derived. The relationship between the biphasic and quasi-linear models will be investigated. The dynamics of the solid and the fluid as well as their interaction will be discussed.

II. BIPHASIC THEORY OF VOCAL FOLD LAMINA PROPRIA

A. General biphasic description of vocal fold lamina propria

The biphasic (solid-fluid) description has become well recognized in investigating the viscoelastic behaviors of ar-

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

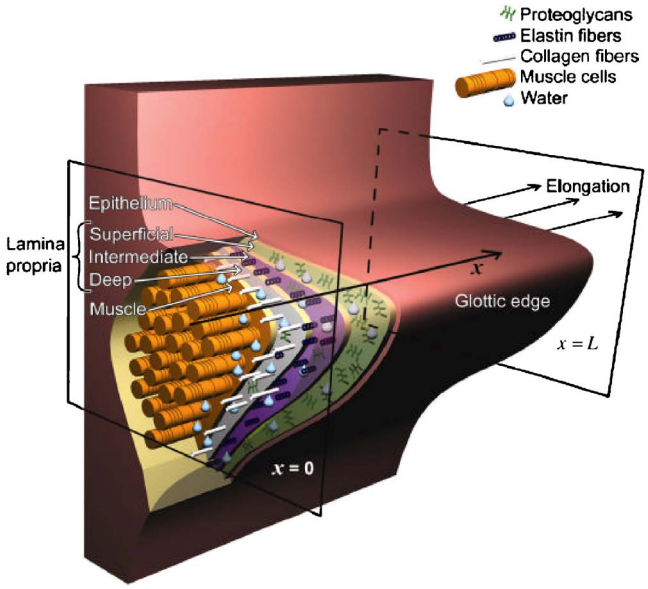


FIG. 1. (Color online) Schematic representation of the principal components of vocal fold lamina propria.

ticular cartilage.[1–6] Articular cartilage tissue is made up of about 80% water and 20% solid, consisting mostly of collagen fibers and proteoglycans. The movement of water through the tissue is governed by viscosity and permeability of the tissue. The biphasic theory[1,2] has been applied to explain the creep and stress relaxation phenomena in articular cartilage tissue.

Vocal fold lamina propria can be considered a porous solid matrix that is swollen by fluid.[19–23] It is comprised mostly of water, collagen and proteoglycans like articular cartilage, though it also contains elastin fibers.[16–19,21] The percentage of water in vocal fold tissue relative to solid is not precisely known; however, it is probably in the range of other cartilage tissue.[2] The organization of the solid constituents in the lamina propria is not random as in articular cartilage and is commonly split into three layers (superficial, intermediate, and deep), as shown in the schematic representation of the vocal folds in Fig. 1. Collagen is sparse in the superficial layer, more prevalent in the intermediate layer, and most numerous in the deep layer. The collagen in the intermediate and deep layers is clearly arranged longitudinally, parallel to the vocal fold edge. Elastin is the other major fibrous structural protein of the lamina propria. It exists in three structural forms, only one of which forms fibers. These fibers are extremely resilient and can stretch to twice their length in the vocal folds. Although there is elastin in every layer of the lamina propria, most of the elastin fibers are organized longitudinally within the intermediate layer, with a smaller amount in the deep layer and a sparse amount in the superficial layer. Proteoglycans are very large branching interstitial proteins that occupy most of the superficial layer and fill the remaining spaces between the fibrous proteins. They fill a volume of space much larger than would be predicted based on their molecular weight due to their long branches. The fibrous units form a cohesive porous composite solid phase in vocal fold lamina propria.

Vocal fold lamina propria contains a high concentration of liquids and is permeable to fluid flow.[19–23] The distribution and flow of fluid within the tissue is affected by the solid components. Hyaluronic acid attracts water to form large and space-filling molecules. The pores maintained by the fibers and proteoglycans can change in volume as well as degree of twisting depending upon compression and elongation.[16–23] Collagen and elastin fibers may provide a substantial level of resistance to directional tensile strain; however, the high water content maintained by the proteoglycans provides the support and bulk in the tissue. The interaction between solid deformation and interstitial fluid motion determines the viscoelastic behavior of vocal fold lamina propria. Due to the multiphase nature and the interdependence of the solid-fluid interactions, vocal fold lamina propria can show complex rheological behaviors, such as stress relaxation.

Based on the aforementioned biomechanical properties, we can describe the vocal fold lamina propria as a biphasic mixture consisting of a fluid phase (interstitial fluid) and a solid phase (collagen, proteoglycan, and other structural components). A biphasic theory may not capture all of the complexity of the structure of vocal fold lamina propria outlined above; however, it should provide a description adequate enough to deduce the general influence of the solid-fluid interaction on the viscoelastic properties of this tissue.

In order to derive the biphasic governing equation, we assume that the solid phase of vocal fold lamina propria is intrinsically incompressible, porous, and elastic, and the fluid phase is intrinsically incompressible. The volume fraction ϕ^α ($\alpha=s,f$ denote the solid and fluid phases) for each phase can be described as[2,4]

$$\phi^\alpha = \lim_{V\to 0} \frac{V^\alpha}{V}, \tag{1}$$

where ϕ^s and ϕ^f are the volume fractions of the solid and fluid, respectively. V^α is the volume occupied by the phase α, and $V=V^s+V^f$ is the total volume. Considering that each phase is incompressible, we can then describe their continuity equations as[2,4]

$$\frac{\partial \phi^\alpha}{\partial t} + \nabla\cdot(\phi^\alpha \mathbf{v}^\alpha) = 0, \tag{2}$$

where $\mathbf{v}^\alpha$ is the velocity of the phase α. From the definitions of ϕ^s and ϕ^f in Eq. (1), we have $\phi^s+\phi^f=1$, and can then deduce the continuity equation as

$$\nabla\cdot(\phi^f \mathbf{v}^f + \phi^s \mathbf{v}^s) = 0. \tag{3}$$

In typical creep and stress relaxation experiments, the whole tissue does not receive body force,[1–6] and we describe the momentum equations for each phase as

$$\rho^\alpha \frac{d\mathbf{v}^\alpha}{dt} = \nabla\cdot\boldsymbol{\sigma}^\alpha + \boldsymbol{\pi}^\alpha, \quad \alpha = s,f, \tag{4}$$

where ρ^α is the mass density for the α phase in the mixture and $\boldsymbol{\sigma}^\alpha$ is the Cauchy stress; $\boldsymbol{\pi}^\alpha$ is the diffusive momentum exchange for the α phase caused by the relative motion between the solid and fluid. Without external body forces, the momentum balance for mixtures leads to $\boldsymbol{\pi}^f=-\boldsymbol{\pi}^s$; ϕ_s, as the

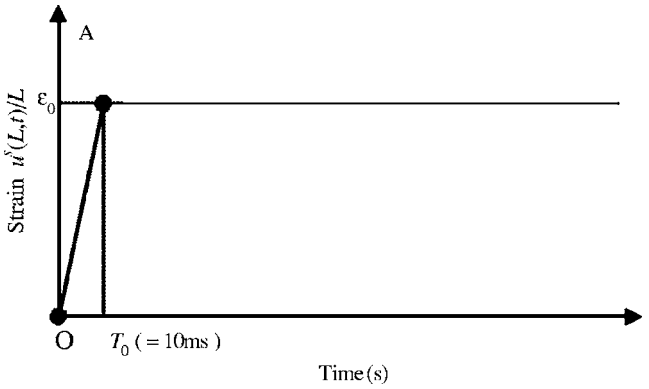


FIG. 2. Stepwise strain load process of vocal fold lamina propria at $x=L$, where O and A correspond to the phase $t<T_0=0.01$ s, and after time A, the strain ε_0/L is held constant.

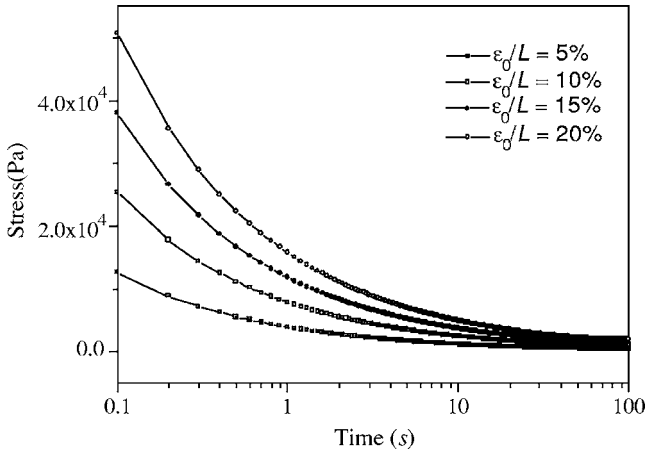


FIG. 3. Overall stress $\sigma(L,t)$ where the curves from the top to bottom correspond to the strain $\varepsilon_0/L=20\%$, 15%, 10%, and 5%.

volume fraction of the solid, is a constant, and its gradient $\nabla\phi_s$ is zero. In addition, according to Darcy's law, the diffusive interaction between the fluid and the solid is proportional to the difference in their velocity; thus the momentum exchange has been previously obtained as[1–6,28,29]

$$\boldsymbol{\pi}^s = -\boldsymbol{\pi}^f = \frac{(\phi^f)^2}{k}(\mathbf{v}^f - \mathbf{v}^s), \tag{5}$$

where $\mathbf{k}$ is the hydraulic permeability of the porous-permeable solid phase. Low permeability k means that large drag forces can be exerted by the interstitial fluid as it flows through the porous-permeable solid material; $k\to\infty$ for a single phase incompressible fluid and $k=0$ for a single phase solid. Although nonlinear factors are important for large strain in vocal fold tissue,[8,27] in this study, we consider the linear biphasic theory under the assumption of small strains and constant permeability in order to emphasize the importance of the solid-fluid interaction in determining the viscoelastic behaviors of the tissue.

For the linear elastic solid phase and the Newtonian fluid phase, we have the following constitutive relationship:[1–6]

$$\boldsymbol{\sigma}^s = -\phi^s p\mathbf{I} + \lambda_s(tr\mathbf{E})\mathbf{I} + 2\mu_s\mathbf{E}, \tag{6}$$

$$\boldsymbol{\sigma}^f = -\phi^f p\mathbf{I} + \lambda_f(tr\mathbf{D}^f)\mathbf{I} + 2\mu_f\mathbf{D}^f, \tag{7}$$

where $\boldsymbol{\sigma}^\alpha$ is the apparent stress for phase α, p is the interstitial fluid pressure. $\mathbf{I}$ is the unit diagonal matrix, λ_s and μ_s are the elastic constants of the solid phase, λ_f and μ_f are the viscosity coefficients of the fluid phase, and $\mathbf{u}^s$ is the solid displacement. $\mathbf{E}=(\nabla\mathbf{u}^s+\nabla^T\mathbf{u}^s)/2$ is the infinitesimal strain tensor of the solid phase, and $\mathbf{D}^f=(\nabla\mathbf{v}^f+\nabla^T\mathbf{v}^f)/2$ is the rate of deformation tensor of the fluid phase. From the above momentum equation, continuity equation, and constitutive relationship of Eqs. (3)–(7), we can obtain the general linear biphasic model of the vocal folds.

B. One-dimensional biphasic model

The one-dimensional model has been used in previous experiments to measure the biomechanical properties of vocal fold tissues.[9–15] Figure 1 shows the one-dimensional vocal fold lamina propria in stress relaxation, where the x axle corresponds to the vocal fold length. The side of the vocal fold tissue at $x=0$ is static, and the other side at $x=L$ receives a stepwise linear strain load. For the one-dimensional case, the biphasic model Eqs. (3)–(7) can be simplified as

$$\phi^f v^f + \phi^s v^s = 0, \tag{8}$$

$$\frac{\partial\sigma^s}{\partial x} + \pi^s = 0, \tag{9}$$

$$\frac{\partial\sigma^f}{\partial x} + \pi^f = 0, \tag{10}$$

$$\pi^s = -\pi^f = \frac{(\phi^f)^2}{k}(v^f - v^s), \tag{11}$$

$$\sigma^s = -\phi^s p + (\lambda_s + 2\mu_s)\frac{\partial u^s}{\partial x}, \tag{12}$$

$$\sigma^f = -\phi^f p + (\lambda_f + 2\mu_f)\frac{\partial v^f}{\partial x}. \tag{13}$$

Based on Eqs. (8)–(13), a partial differential equation describing vocal fold lamina propria movement can be derived as (See Appendix A),

$$\frac{\partial u^s}{\partial t} - \left[kH_A\frac{\partial^2 u^s}{\partial x^2} + kH_B\frac{\partial^3 u^s}{\partial x^2\partial t}\right] = 0, \tag{14}$$

where $H_A=\lambda_s+2\mu_s$ is the aggregate compressive modulus,[1,2] and $H_B=(\lambda_f+2\mu_f)(\phi^s/\phi^f)^2$ is the modulus due to the fluid viscosity.

In order to simulate previous stress relaxation experiments,[1,9] we set the initial condition of the solid displacement $u^s(x,t)$ as

$$u^s(x,0)=0, \quad 0 \leqslant x \leqslant L. \tag{15}$$

In this case, the side of the vocal fold at $x=0$ is impermeable to fluid and the displacement $u^s(0,t)$ is zero. The side at $x=L$ of the tissue receives a stepwise linear strain load within a very short period $T_0(=0.01s) \ll 1$ s from the time O to A, as shown in Fig. 2.[9] After the time A, the strain load is held constant as ε_0/L for a long time (as shown in Fig. 2). Thus, we get the following boundary conditions:

$$u^s(0,t)=0, \quad u^s(L,t) = \begin{cases} \varepsilon_0 t/T_0, & 0 \leqslant t < T_0 \\ \varepsilon_0 & t \geqslant T_0 \end{cases}, \quad \text{and } p(L,t)=0. \tag{16}$$

Here, we apply elongation or the positive strain ε_0/L, as is commonly done in vocal fold viscoelastic studies;[9–15] however, this mathematical theory also applies to compression or negative strain $(-\varepsilon_0/L)$, which has been well studied in articular cartilage.[1–6] For the governing equation Eq. (14) with the initial condition (15) and the boundary conditions (16), the following analytic solutions for the solid displacement and the fluid velocity can be obtained (See Appendix B):

$$u^s(x,t) = \begin{cases} \dfrac{\varepsilon_0 t x}{T_0 L} + \displaystyle\sum_{n=1}^{\infty} \frac{2\varepsilon_0 L^2(-1)^n}{n^3\pi^3 T_0 k H_A}\left[1-e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\right]\sin(n\pi x/L), & 0 \leqslant t < T_0 \\ \dfrac{\varepsilon_0 x}{L} + \displaystyle\sum_{n=1}^{\infty} \frac{2\varepsilon_0 L^2(-1)^n}{n^3\pi^3 T_0 k H_A}\left[e^{n^2\pi^2 k H_A T_0/L^2+n^2\pi^2 k H_B}-1\right]e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\sin(n\pi x/L), & t \geqslant T_0 \end{cases} \tag{17}$$

and

$$v^f(x,t) = -\frac{\phi^s}{\phi^f}v^s(x,t) = \begin{cases} -\dfrac{\phi^s}{\phi^f}\left\{\dfrac{\varepsilon_0 x}{T_0 L} + \displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L^2(-1)^n}{n\pi T_0(L^2+n^2\pi^2 k H_B)}e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\sin(n\pi x/L)\right\}, & 0 \leqslant t < T_0 \\ -\dfrac{\phi^s}{\phi^f}\left\{\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L^2(-1)^{n+1}}{n\pi T_0(L^2+n^2\pi^2 k H_B)}\left[e^{n^2\pi^2 k H_A T_0/L^2+n^2\pi^2 k H_B}-1\right]e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\sin(n\pi x/L)\right\}, & t \geqslant T_0 \end{cases} . \tag{18}$$

Under a certain strain ε_0/L, the exponential function of the stress at the top of the tissue $(x=h)$ can be given as

$$\sigma(L,t)=\sigma^s(L,t)+\sigma^f(L,t)=H_A\left.\frac{\partial u^s}{\partial x}\right|_{x=L}+\left(\frac{\phi^f}{\phi^s}\right)^2 H_B\left.\frac{\partial v^f}{\partial x}\right|_{x=L}$$

$$= \begin{cases} H_A\left\{\dfrac{\varepsilon_0 t}{T_0 L}+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L}{n^2\pi^2 T_0 k H_A}\left[1-e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\right]\right\} - \dfrac{\phi^f H_B}{\phi^s}\left\{\dfrac{\varepsilon_0}{T_0 L}+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L}{T_0(L^2+n^2\pi^2 k H_B)}e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\right\}, \\ 0 \leqslant t < T_0 \\ H_A\left\{\dfrac{\varepsilon_0}{L}+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L}{n^2\pi^2 T_0 k H_A}\left[e^{n^2\pi^2 k H_A T_0/L^2+n^2\pi^2 k H_B}-1\right]e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\right\} - \dfrac{\phi^f H_B}{\phi^s}\left\{\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L}{T_0(L^2+n^2\pi^2 k H_B)}\right. \\ \left.\left[1-e^{n^2\pi^2 k H_A T_0/L^2+n^2\pi^2 k H_B}\right]e^{-n^2\pi^2 k H_A t/L^2+n^2\pi^2 k H_B}\right\}, \\ t \geqslant T_0. \end{cases} \tag{19}$$

From the momentum Eqs. (9) and (10), it is noted that the total stress is uniform throughout x, i.e., $\sigma(x,t)=\sigma^s(x,t)+\sigma^f(x,t)=\sigma(L,t)$. Substituting $\sigma(L,t)$ into Eqs. (12) and (13), we can describe the solid and fluid stresses as (See Appendix A),

$$\sigma^s(x,t)=\phi^f H_A\frac{\partial u^s(x,t)}{\partial x}-H_B\frac{(\phi^f)^2}{\phi^s}\frac{\partial v^f(x,t)}{\partial x}+\phi^s\sigma(L,t), \tag{20}$$

$$\sigma^f(x,t)=-\phi^f H_A\frac{\partial u^s(x,t)}{\partial x}+H_B\frac{(\phi^f)^2}{\phi^s}\frac{\partial v^f(x,t)}{\partial x}+\phi^f\sigma(L,t). \tag{21}$$

Furthermore, from Eqs. (17)–(21), the analytic solutions for the solid and fluid stress can be obtained as

$$\sigma^s(x,t)=\begin{cases}
H_A\dfrac{\varepsilon_0 t}{T_0 L}+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f(-1)^n}{n^2\pi^2 T_0 k}[1-e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}]\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f H_B(-1)^n}{T_0(L^2+n^2\pi^2 kH_B)}e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^s}{n^2\pi^2 T_0 k}[1-e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}]\\
-\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f H_B}{T_0(L^2+n^2\pi^2 kH_B)}e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}, \quad 0\leqslant t<T_0\\
H_A\dfrac{\varepsilon_0}{L}+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f(-1)^n}{n^2\pi^2 T_0 k}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f H_B(-1)^{n+1}}{T_0(L^2+n^2\pi^2 kH_B)}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^s}{n^2\pi^2 T_0 k}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f H_B}{T_0(L^2+n^2\pi^2 kH_B)}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}, \quad t\geqslant T_0
\end{cases}\tag{22}$$

and

$$\sigma^f(x,t)=\begin{cases}
-\dfrac{\varepsilon_0 H_B\phi^f}{T_0 L\phi^s}+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f(-1)^n}{n^2\pi^2 T_0 k}[e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}-1]\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f H_B(-1)^{n+1}}{T_0(L^2+n^2\pi^2 kH_B)}e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f}{n^2\pi^2 T_0 k}[1-e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}]\\
-\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L(\phi^f)^2 H_B}{\phi^s T_0(L^2+n^2\pi^2 kH_B)}e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}, \quad 0\leqslant t<T_0\\
\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f(-1)^n}{n^2\pi^2 T_0 k}[1-e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f H_B(-1)^n}{T_0(L^2+n^2\pi^2 kH_B)}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\cos(n\pi x/L)\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L\phi^f}{n^2\pi^2 T_0 k}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\\
+\displaystyle\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L(\phi^f)^2 H_B}{\phi^s T_0(L^2+n^2\pi^2 kH_B)}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}, \quad t\geqslant T_0.
\end{cases}\tag{23}$$

Based on Eqs. (17)–(23), the analytic solutions for all dependent variables (displacement, velocity, and stress) can be obtained. It is clear that the biphasic model is applicable for both time conditions ($t<T_0$ and $t\geqslant T_0$). The ramp time mathematics and modeling are included in this model. In previous studies,[9–15] stress relaxation has been measured within the second time condition ($t>>T_0$); we thus focus on this time condition in this theoretical study. The stress relaxation behaviors of vocal fold lamina propria under the first time condition are also important and require further experimental study. When $t\to\infty$, the exponential terms in Eqs. (17)–(23) converge to zero, i.e., $e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\to 0$, then the following asymptotical solutions for all position $x\in[0,L]$ can be deduced:

$$u^s(x,t) \to \frac{\varepsilon_0 x}{L}, \quad v^f(x,t) \to 0, \quad \sigma(L,t) \to H_A \frac{\varepsilon_0}{L}, \quad \sigma^s(x,t) \to H_A \frac{\varepsilon_0}{L}, \quad \text{and } \sigma^f(x,t) \to 0. \tag{24}$$

III. NUMERICAL CALCULATIONS AND DISCUSSIONS

Stress relaxation has been experimentally observed in canine vocalis muscle,[9] however, empirical data of the stress relaxation behavior of the lamina propria may not be currently available in the literature.[10–14] In order to numerically simulate the viscoelastic behavior of the vocal fold lamina propria using the biphasic model, we set $H_A = 10^4$ Pa, $H_B = 2$ Pa s, and $T_0 = 10$ ms. The permeability k and fluid volume fraction ϕ^f have been theoretically studied and experimentally measured in articular cartilage;[1,2] however, to the best of our knowledge, these two parameters have never been examined in laryngeal research. In this theoretical study, we set the k value as 10^{-10} and the volume fractions of solid and fluid phases were used as $\phi^s = 0.2$ and $\phi^f = 0.8$ based on averages from articular cartilage.[1,2] The actual k and ϕ^f values of vocal folds require future experimental measurement although, theoretically, we have tested different values of k and ϕ^f and found qualitatively similar results. Substituting these biphasic system parameters into Eqs. (17) and (22), we can investigate the overall as well as the solid and fluid dynamics of vocal fold lamina propria.

A. Overall stress

Figure 3 shows the time evolution of overall stress $\sigma(L,t)$ at $x=L$ after applying the strain ε_0/L, where the curves from top to bottom correspond to $\varepsilon_0/L = 20\%$, 15%, 10%, and 5%. With the increase of strain ε_0/L, stress $\sigma(x,t)$ continues to increase. For sufficiently large t ($t \to \infty$), according to Eq. (24), the overall stress will approach a constant $H_A \varepsilon_0/L$. In Fig. 3, when $t = 100$ s, the overall stress approaches its equilibrium value which can be approximately described by using the traditional elastic theory.[24–27] Under this condition, the relative motion between the solid and fluid can be neglected; therefore, the vocal fold lamina propria behaves like an elastic system.

However, when t is not sufficiently high ($t < 10$ s), the total stress decreases with time at this early stage of stress relaxation, as shown in Fig. 3. According to the present biphasic model, the vocal fold lamina propria shows the viscoelastic behavior of stress relaxation qualitatively similar to that obtained in previous experimental measurements, except the strain process ($t < T_0$) between O and A, which was not measured in previous experimental studies.[9–15] To explain this phenomenon of stress relaxation, a quasi-linear viscoelastic model[9,11] has been applied by using springs and dashpots where the relaxation modulus was defined as $R(t) = \sigma(t)/\varepsilon_0$. According to the quasi-linear viscoelastic model, the relaxation modulus can be expanded as[3,9–15]

$$R(t) = E_0 + \sum_{i=1}^{N} E_i e^{-t/\lambda_i}, \tag{25}$$

where E_i is the ith elastic constant and λ_i is the ith time constant of the Maxwell elements. Using Eq. (19) derived from the biphasic theory, the exponential relationship Eq. (25) can be easily deduced when $t \geqslant T_0$, and then the elastic and time constants can be described as

$$\begin{aligned} E_0 &= H_A/L, \\ E_i &= \left(\frac{2L}{i^2\pi^2 T_0 k} + \frac{\phi^f H_B}{\phi^s} \frac{2L}{T_0(L^2 + i^2\pi^2 k H_B)} \right) \\ &\quad \times \left(e^{i^2\pi^2 k H_A T_0/L^2 + i^2\pi^2 k H_B} - 1 \right) \end{aligned} \tag{26}$$

and

$$\lambda_i = \frac{L^2 + i^2\pi^2 k H_B}{i^2\pi^2 k H_A} \quad (i = 1,2,\cdots). \tag{27}$$

Thus, the quasi-linear viscoelastic model can be derived from the biphasic theory. More importantly, the quasi-linear viscoelastic model cannot describe the fluid motion, the stress fields in the tissue, and the solid-fluid interaction. Physical or physiological relevance of high-order elastic and time constants in the quasi-linear viscoelastic model are not clear. The high-order elastic and time constants are not be straightforward to obtain using the quasi-linear model and may require using other experimental methods.[3,9] Alipour and Titze[9] applied a continuous spectrum assumption of the relaxation function; however, their model required predefining a certain probability distribution for elastic constants which is undetermined for practical systems. Recently, the statistical network model[11] has been applied to study the viscoelastic behaviors of vocal fold mucosa. As a macromolecular dynamic description, the statistical network model contributes the system parameters to molecular structure. The tissue is described as a network consisting of macromolecular chain segments through intermolecular interactions, and the constitutive equation is used to obtain the elastic and viscous shear moduli. However, the statistical network theory does not apply the continuity and momentum equations to describe the conservations of mass and momentum, and thus cannot describe the momentum exchange between the fluid and solid. In order to describe macro movements of fluid and solid phases in vocal fold lamina propria, we apply the biphasic theory in this study. The fluid and solid dynamics can be obtained using the continuity, momentum, and constitutive Eqs. (3), (4), (6), and (7). The biphasic theory introduced in this study presents the continuum mechanism describing the fluid-solid interaction and gives the analytic solutions for the elastic and time constants of the quasi-linear viscoelastic model. Note that from Eq. (26), the E_1 term in

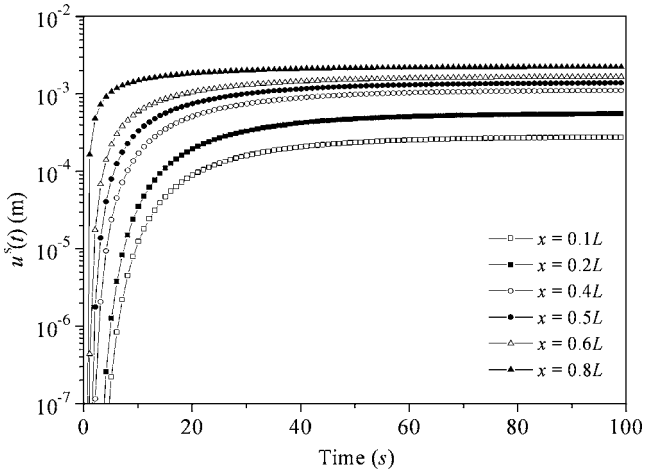


FIG. 4. Solid displacement u^s at six different depths, where $\varepsilon_0/L=20\%$, $\phi^s=0.2$, $\phi^f=0.8$. The curves from the top to bottom correspond to the positions $x=0.1L$, $0.2L$, $0.4L$, $0.5L$, $0.6L$, and $0.8L$, respectively.

Eq. (25) will be much higher than the E_0 term in the short time period due to the small value of k. Therefore, the solid-fluid interaction, which determines E_i, is very important in the early stage of stress relaxation, though this has not been experimentally studied. Physically, the rheological behavior of vocal fold lamina propria is due to the intrinsic interaction between the deformation of the solid phase and its relative motion with respect to the interstitial fluid. The single solid phase model[24–27] may not describe the multiphase nature of the vocal fold lamina propria and cannot describe the solid-fluid interaction. In comparison with the single phase solid model, the biphasic model includes both solid and fluid phases, and thus provides the solid-fluid interaction mechanism of stress relaxation, representing a more physical theory to describe the viscoelastic behaviors of vocal fold lamina propria.

B. Solid dynamics

Figures 4 and 5 show the time evolutions of the solid displacement $u^s(t)$ and the solid stress $\sigma^s(t)$ at six different positions after applying the strain $\varepsilon_0/L=20\%$, where $\phi^s=0.2$, $\phi^f=0.8$, and the curves from top to bottom correspond to the position $x=0.1L$, $0.2L$, $0.4L$, $0.5L$, $0.6L$, and $0.8L$, respectively, and the dotted line in Fig. 5 denotes the relaxation curve of the total stress $\sigma(t)$. When t is not sufficiently large, the solid displacement $u^s(t)$ and stress $\sigma^s(t)$ of vocal fold lamina propria show obvious time dependencies that cannot be seen in pure elastic materials. In comparison with total stress $\sigma(t)$, the solid stress is much lower. The curve shapes of $u^s(t)$ and $\sigma^s(t)$ at different positions x are also different when $t<10$ s; $u^s(t)$ near the position of the load application ($x=L$) reaches the equilibrium state faster than that at the other side ($x=0$). These results are associated with the solid-fluid interaction of vocal fold lamina propria.

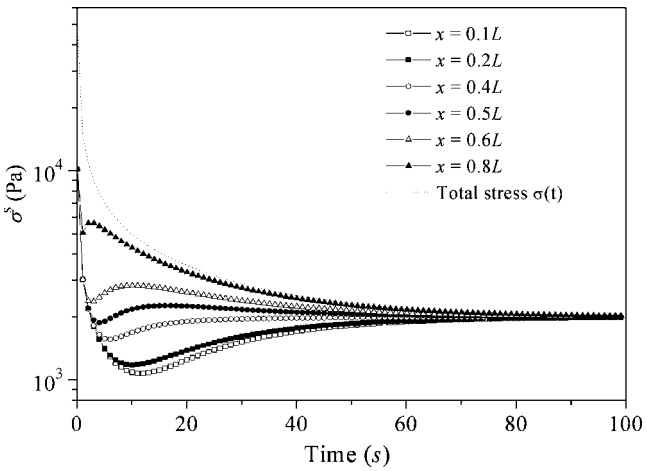


FIG. 5. The solid stress σ^s, where $\varepsilon_0/L=20\%$, where $\phi^s=0.2$, $\phi^f=0.8$. The curves from bottom to top correspond to the positions $x=0.1L$, $0.2L$, $0.4L$, $0.5L$, $0.6L$, and $0.8L$, respectively, and the dotted line denotes the relaxation curve of total stress $\sigma(t)$.

At the early stage of stress relaxation, in which the fluid and solid have the strong solid-fluid interaction and the total stress decreases with time, the relative motion and drag between the interstitial fluid and solid phases leads to an exchange of momentum; $u^s(t)$ increases with time due to the momentum exchange from fluid to solid. The porous structure of the solid material retards the fluid flow away from the side of $x=L$ to the other side of $x=0$, and thus the solid amplitude far from the strain load position shows a time delay in reaching equilibrium. The elongated solid matrix at $x=L$ relieves itself by propagating the displacement into the lower distances (or smaller x) of the tissue.

Furthermore, for sufficiently large t, based on Eq. (24), the asymptotical solutions of the solid displacement $u^s(t)$ and stress $\sigma^s(t)$ will approach $\varepsilon_0 x/L$ and $H_A\varepsilon_0/L$, respectively, which can also be obtained using the first-order linear elastic model of Eq. (26). Figures 4 and 5 show the time convergence tendency of $u^s(t)$ and $\sigma^s(t)$; $u^s(t)$ and $\sigma^s(t)$ approach the equilibrium values obtained using the linear elastic response of the solid components. Although the solid stress is much less than the total stress at the early stage of stress relaxation, its role in supporting the tissue stress becomes more and more important as time increases until eventually, the solid supports all of the stress. However, the final stress at equilibrium is much smaller than the total stress at the early stage of relaxation.

C. Fluid dynamics

Figure 6 shows the time history of the fluid velocity, where $\phi^s=0.2$, $\phi^f=0.8$, and the curves from top to bottom correspond to the positions $x=0.2L$, $0.4L$, $0.5L$, $0.6L$, and $0.8L$, respectively. For sufficiently large t, the asymptotical solutions of the fluid velocity $v^f(x,t)$ and stress $\sigma^f(t)$ will approach zero, according to Eq. (24). The fluid stops flowing and the momentum exchange between fluid and solid becomes negligible. However, when t is not large, stress relaxation proceeds via a dynamic process of fluid redistribution. Strain ε_0 leads to an expansion of fluid at one side ($x=L$) that subsequently redistributes toward the other side ($x=0$) of the vocal fold lamina propria. Because of the momentum exchange between the interstitial fluid and the solid, the absolute value of the fluid velocity $v^f(x,t)$ decreases with time and depth x.

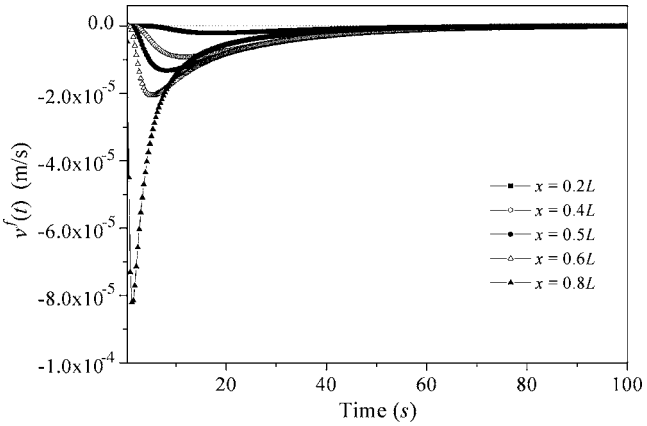


FIG. 6. The fluid velocity v^f, where $\phi^s=0.2$, $\phi^f=0.8$. The curves from the top to bottom correspond to the positions $x=0.2L$, $0.4L$, $0.5L$, $0.6L$, and $0.8L$, respectively.

Previous studies have shown that the tissue (or solid) stress is an important factor for vocal fold tissue damage;[26] however, using the biphasic theory, we see that the interstitial fluid is also important for stress load support, particularly in the early stage of external loading.[5,6] Figure 7 shows the time evolution of the relative fluid stress $100\sigma^f/(\sigma^s+\sigma^f)$ during stress relaxation, where $\phi^s=0.2$, $\phi^f=0.8$, and the curves from top to bottom correspond to the positions $x=0.1L$, $0.2L$, $0.4L$, $0.5L$, $0.6L$, and $0.8L$, respectively. In Fig. 7, when t is sufficiently small, $\sigma^f/(\sigma^s+\sigma^f)$ approaches 0.8, meaning that the stress supported by the fluid is four times larger than the stress supported by the solid in the first few seconds of strain loading. The fluid content ϕ^f of tissue is usually much higher than the solid content ϕ^s, and thus the interstitial fluid supports the majority of the stress in the initial phases of strain loading. Higher fluid content ϕ^f and permeability k leads to higher strength and duration of fluid support.

In addition, for a lower fluid content $\phi^s=0.6$ and $\phi^f=0.4$, we found that for sufficiently small t, the fluid's support of stress $\sigma^f/(\sigma^s+\sigma^f)$ decreases to about 40%, so the solid structure must support more stress. The ratio of fluid to solid stress is dependent on the ratio of fluid to solid volume, so that when there is less fluid, the solid must bear a greater fraction of the total stress. This suggests that interstitial fluid can shield the collagen-elastin-proteoglycan solid structure from excessive stresses which may otherwise lead to excessive tissue damage. It is well accepted that dehydration can have adverse effects on voice quality and health. Dehydration will decrease ϕ^f and thus reduce the fluid support to the stress. In addition, previous studies have demonstrated that dehydration raises the minimum subglottal pressure required to induce and sustain phonation, known as the phonation threshold pressure or PTP.[23,30] It has been shown that this may be related to increased overall stiffness with dehydration. This result may be understood with the biphasic theory if we consider the effects of dehydration on permeability. Since proteoglycans are osmotically active, they induce a swelling of the vocal fold tissue with water.[31–33] Removing this water leads to a collapse of the solid structure and a decrease in pore size. This will reduce the permeability k and thus increase the stiffness or the relaxation modulus of the tissue according to Eq. (25). The increased stiffness will increase the PTP of the vocal folds. More studies need to be performed looking at the effects of dehydration on permeability as well as damage to the solid-fluid structures of the vocal fold lamina propria.

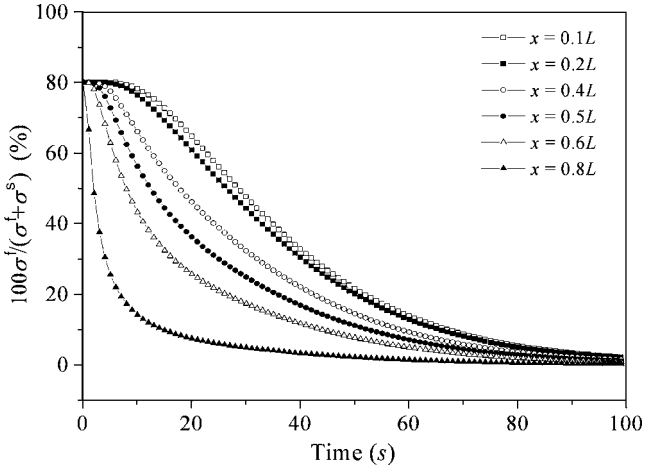


FIG. 7. The relative stress $100\sigma^f/(\sigma^s+\sigma^f)$ during stress relaxation for $\phi^s=0.2$ and $\phi^f=0.8$.

IV. CONCLUSION

In this study, we proposed a biphasic theory for the study of the viscoelastic responses of stress relaxation of the vocal fold lamina propria. Specifically, we looked at the significance of the fluid and solid interaction during stress relaxation. The vocal fold lamina propria was described as a mixture of a solid and a fluid. Under the action of strain load, the deformation of the solid matrix and the flow of the interstitial fluid gave rise to their relative motion and momentum exchange. The biphasic theory predicted the stress relaxation behavior in the vocal fold lamina propria that might not be reported in previous experimental study. Further experimental study is important to examine the stress relaxation behavior in the vocal fold lamina propria. The previous quasilinear viscoelastic model and the high-order elastic parameters can be derived using this theory. It was found that as time increased, the solid phase eventually supported the full stress. Furthermore, the interstitial fluid was found to be an important factor for stress load support, particularly at the early stage of stress loading. The early stress supported by the fluid was much higher than the eventual stress supported by the solid. The importance of this stress shielding by the fluid may be valuable in explaining the effect of dehydration on vocal fold tissue damage.

Through the recognition of the multiphase nature of vocal fold lamina propria, the biphasic theory presents a kind of valuable viscoelastic theory and is able to provide a continuum model describing fluid and solid dynamics. The single-solid description previously applied in voice study has limitations in explaining solid-fluid interaction in vocal fold lamina propria. The biphasic model presented in this study should serve as a framework for future experimental measurements and more complete theoretical developments, such as permeability parameter consideration, specific structure of

vocal fold tissue, and dynamic loading. Further study and application to laryngeal pathology is needed in this potentially important topic.

ACKNOWLEDGMENTS

This study was supported by NIH Grant Nos. 1-RO1DC006019 and No. 1-RO1DC05522 from the National Institute of Deafness and other Communication Disorders.

APPENDIX A: DERIVATION OF ONE-DIMENSIONAL BIPHASIC MODEL

Substituting Eqs. (8), (11), and (12) into Eq. (9), we have the constitutive equation of the solid phase as

$$-\phi^s\frac{\partial p}{\partial x}+(\lambda_s+2\mu_s)\frac{\partial^2 u^s}{\partial x^2}-\frac{\phi^f}{k}v^s=0. \tag{A1}$$

Similarly, substituting Eqs. (8), (11), and (13) into Eq. (10), we have the constitutive equation of the fluid phase as

$$-\phi^f\frac{\partial p}{\partial x}+(\lambda_f+2\mu_f)\frac{\partial^2 v^f}{\partial x^2}+\frac{\phi^f}{k}v^s=0. \tag{A2}$$

Combining Eqs. (A1) and (A2) and eliminating the fluid velocity variable v^f, we have the following partial differential equation:

$$\frac{\partial u^s}{\partial t}-\left[k(\lambda_s+2\mu_s)\frac{\partial^2 u^s}{\partial x^2}+k(\lambda_f+2\mu_f)\left(\frac{\phi^s}{\phi^f}\right)^2\frac{\partial^3 u^s}{\partial x^2\partial t}\right]=0. \tag{A3}$$

By defining $H_A=\lambda_s+2\mu_s$ and $H_B(\lambda_f+2\mu_f)(\phi^s/\phi^f)^2$, we can obtain the one-dimensional biphasic model Eq. (14) of vocal fold lamina propria.

Furthermore, from the constitutive relationship of Eqs. (11) and (12), we have the total stress as

$$\begin{aligned}\sigma(x,t)&=\sigma^s(x,t)+\sigma^f(x,t)\\&=-p(x,t)+H_A\frac{\partial u^s}{\partial x}+H_B\left(\frac{\phi^f}{\phi^s}\right)^2\frac{\partial v^f}{\partial x}.\end{aligned} \tag{A4}$$

According to the momentum Eqs. (9) and (10), it is noted that the total stress is uniform throughout x, i.e., $\sigma(x,t)=\sigma^s(x,t)+\sigma^f(x,t)=\sigma(L,t)$. Thus we can obtain

$$p(x,t)=H_A\frac{\partial u^s(x,t)}{\partial x}+H_B\left(\frac{\phi^f}{\phi^s}\right)^2\frac{\partial v^f(x,t)}{\partial x}-\sigma(L,t), \tag{A5}$$

$$\sigma^s(x,t)=-\phi^s p(x,t)+H_A\frac{\partial u^s(x,t)}{\partial x}, \tag{A6}$$

$$\sigma^f(x,t)=-\phi^f p(x,t)+H_B\left(\frac{\phi^f}{\phi^s}\right)^2\frac{\partial v^f(x,t)}{\partial x}. \tag{A7}$$

Substituting Eq. (A5) into Eqs. (A6) and (A7), we can obtain Eqs. (20) and (21) of the fluid and solid stress.

APPENDIX B: DERIVATION OF THE ANALYTICAL SOLUTIONS OF ONE-DIMENSIONAL BIPHASIC MODEL

In this study, one side ($x=L$) of the tissue receives a strain load of the ramping-displacement function described by $u^s(L,t)=\begin{cases}\varepsilon_0 t/T_0, & 0\leqslant t<T_0\\ \varepsilon_0, & t\geqslant T_0\end{cases}$. The different boundary conditions before and after T_0 lead to different analytical solutions of Eq. (14):

(1) When $0\leqslant t<T_0$, the initial and boundary conditions of Eq. (14) can be written as

$$u^s(x,0)=0,\quad u^s(0,t)=0,\ \text{and}\quad u^s(L,t)=\varepsilon_0 t/T. \tag{B1}$$

For the nonhomogeneous boundary condition of $u^s(x,t)$, the boundary condition can be made homogeneous by defining a new variable $w(x,t)$ as $u^s(x,t)=w(x,t)+\varepsilon_0 tx/T_0L$, and then we have the following equation for $w(x,t)$:

$$\frac{\partial w}{\partial t}-\left[kH_A\frac{\partial^2 w}{\partial x^2}+kH_B\frac{\partial^3 w}{\partial x^2\partial t}\right]=-\frac{\varepsilon_0 x}{T_0 L} \tag{B2}$$

with the initial condition of $w(x,0)=0(0\leqslant x\leqslant L)$ and the homogeneous boundary conditions of $w(0,t)=0$, $w(L,t)=0$. According to the standard method of separating variables for partial differential equations, the solution of the nonhomogeneous Eq. (B2) with the initial and boundary conditions can be expanded as the Fourier series

$$w(x,t)=\sum_{n=0}^{\infty}T_n(t)\sin(n\pi x/L). \tag{B3}$$

Substituting Eq. (B3) into Eq. (B2), we have

$$\sum_{n=0}^{\infty}\left\{T_n'(t)+kH_A\frac{n^2\pi^2}{L^2}T_n(t)+kH_B\frac{n^2\pi^2}{L^2}T_n'(t)\right\}\times\sin(n\pi x/L)=-\frac{\varepsilon_0 x}{T_0 L}. \tag{B4}$$

Expanding the term $-\varepsilon_0 x/T_0L$ by finding its Fourier series of $\sin(n\pi x/L)$, the nth coefficient is $2/L\int_0^L(-\varepsilon_0 x/T_0L)\sin(n\pi x/L)dx=(2\varepsilon_0/n\pi T_0)(-1)^n$, the solution of Eq. (B4) can be obtained by solving the linear first order ordinary differential equation as

$$T_n(t)=\frac{2\varepsilon_0(-1)^n L^2}{n^3\pi^3 T_0 kH_A}(1-e^{-kH_A n^2\pi^2 t/L^2+kH_B n^2\pi^2}). \tag{B5}$$

From Eqs. (B2) to (B5), the analytical solution of $u(x,t)$ can be expressed as

$$u(x,t)=\frac{\varepsilon_0 tx}{T_0 L}+\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L^2(-1)^n}{n^3\pi^3 T_0 kH_A}[1-e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}]\times\sin(n\pi x/L),\quad 0\leqslant t<T_0. \tag{B6}$$

(2) When $t\geqslant T_0$, the initial (at $t=T_0$) and boundary conditions of Eq. (14) can be described as

$$u^s(x,T_0)=\frac{\varepsilon_0 x}{L}+\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L^2(-1)^n}{n^3\pi^3 T_0 kH_A}\times[1-e^{-n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}]\sin(n\pi x/L),$$

$$u^s(0,t)=0, \quad \text{and} \quad u^s(L,t)=\varepsilon_0. \tag{B7}$$

Based on these conditions, we can expand the solution of the homogeneous Eq. (14) as

$$u^s(x,t)=C_0+\sum_{n=1}^{\infty} C_n e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B} \sin(n\pi x/L). \tag{B8}$$

From Eqs. (B7) and (B8), we have the following coefficients:

$$C_0=\frac{\varepsilon_0 x}{L} \quad \text{and} \quad C_n=\frac{2\varepsilon_0 L^2(-1)^n}{n^3\pi^3 T_0 kH_A}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1]. \tag{B9}$$

Therefore, the solution of Eq. (14) for $t \geqslant T_0$ can be described as

$$u^s(x,t)=\frac{\varepsilon_0 x}{L}+\sum_{n=1}^{\infty}\frac{2\varepsilon_0 L^2(-1)^n}{n^3\pi^3 T_0 kH_A}[e^{n^2\pi^2 kH_A T_0/L^2+n^2\pi^2 kH_B}-1] \times e^{-n^2\pi^2 kH_A t/L^2+n^2\pi^2 kH_B}\sin(n\pi x/L), \quad t \geqslant T_0. \tag{B10}$$

[1] V. C. Mow, M. H. Holms, and W. M. Lai, "Fluid transport and mechanical properties of articular cartilage: A review," J. Biomech. **17**, 377–394 (1984).
[2] V. C. Mow, S. C. Kuei, W. M. Lai, and C. G. Armstrong, "Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments," J. Biomech. Eng. **102**, 73–84 (1980).
[3] Y. C. Fung, *Biomechanics. Mechanical Properties of Living Tissues*, 2nd ed. (Springer-Verlag, New York, 1993), pp. 23–65, 242–320.
[4] M. H. Holmes, "Finite deformation of soft tissue: analysis of a mixture model in uni-axial compression," J. Biomech. Eng. **108**, 372–381 (1986).
[5] M. A. Soltz and G. A. Ateshian, "Experimental verification and theoretical prediction of cartilage interstitial fluid pressurization at an impermeable contact interface in confined compression," J. Biomech. **31**, 927–934 (1998).
[6] M. A. Soltz, I. M. Basalo, and G. A. Ateshian, "Hydrostatic pressurization and depletion of trapped lubricant pool during creep contact of a rippled indenter against a biphasic articular layer," J. Biomech. Eng. **125**, 585–593 (2003).
[7] D. D. Sun, X. E. Guo, M. Likhitpanichkul, W. M. Lai, and V. C. Mow, "The influence of the fixed negative charges on mechanical and electrical behaviors of articular cartilage under unconfined compression," J. Biomech. Eng. **126**, 6–16 (2004).
[8] C. W. J. Oomens, D. H. Van Campen, and H. J. Grootenboer, "A mixture approach to the mechanics of skin," J. Biomech. **20**, 877–885 (1987).
[9] F. Alipour-Haghighi and I. R. Titze, "Viscoelastic modeling of canine vocalis muscle in relaxation," J. Acoust. Soc. Am. **78**, 1939–1943 (1985).
[10] R. W. Chan and I. R. Titze, "Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021. (1999).
[11] R. W. Chan and I. R. Titze, "Viscoelastic shear properties of human vocal fold mucosa: theoretical characterization based on constitutive modeling," J. Acoust. Soc. Am. **107**, 565–580 (2000).
[12] R. W. Chan, "Estimation of viscoelastic shear properties of vocal fold tissues based on time-temperature superposition," J. Acoust. Soc. Am. **110**, 1548–1561 (2001).
[13] R. W. Chan and N. Tayama, "Biomechanical effects of hydration in vocal fold tissues," Otolaryngol.-Head Neck Surg. **126**, 528–537 (2002).
[14] R. W. Chan, "Measurements of vocal fold tissue viscoelasticity: Approaching the male phonatory frequency range," J. Acoust. Soc. Am. **115**, 3161–3170 (2004).
[15] K. Zhang, T. Siegmund, and R. W. Chan, "A constitutive model of the human vocal fold cover for fundamental frequency regulation," J. Acoust. Soc. Am. **119**, 1050–1062 (2006).
[16] S. D. Gray, M. Hirano, and K. Sato, "Molecular and cellular structure of vocal fold tissue," in *Vocal Fold Physiology: Frontiers of Basic Science*, edited by I. R. Titze (Singular, San Diego, 1993) pp. 1–34.
[17] S. D. Gray, I. R. Titze, R. Chan, and T. H. Hammond, "Vocal fold proteoglycans and their influence on biomechanics," Laryngoscope **109**, 845–854 (1999).
[18] S. D. Gray, I. R. Titze, F. Alipour, and T. H. Hammond, "Biomechanical and histologic observations of vocal fold fibrous proteins," Ann. Otol. Rhinol. Laryngol. **109**, 77–85 (2000).
[19] J. C. Stemple, L. E. Glaze, and B. Klaben, *Clinical Voice Pathology* (Singular, San Diego, 2000).
[20] S. D. Gray, M. Hirano, and K. Sato, "Molecular and cellular structure of vocal fold tissue," in *Vocal Fold Physiology: Frontiers of Basic Science*, edited by I. R. Titze (Singular, San Diego, 1993) pp. 1–34.
[21] J. C. Kahane, "A survey of age-related changes in the connective tissues of the human adult larynx," in *Vocal Fold Physiology. Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill Press, San Diego, 1983).
[22] J. Jiang, K. Verdolini, B. Acquino, J. Ng, and D. G. Hanson, "Effects of dehydration on phonation in excised canine larynges," Ann. Otol. Rhinol. Laryngol. **109**, 568–575 (2000).
[23] K. Verdolini, Y. Min, I. R. Titze, J. Lemke, K. Brown, M. Mersbergen, J. Jiang, and K. Fisher, "Biological mechanisms underlying voice changes due to dehydration." J. Speech Lang. Hear. Res. **45**, 268–281 (2002).
[24] I. R. Titze and E. J. Hunter, "Normal vibration frequencies of the vocal ligament," J. Acoust. Soc. Am. **115**, 2264–2269 (2004).
[25] I. R. Titze, "On the mechanics of vocal-fold vibration," J. Acoust. Soc. Am. **60**, 1366–1380 (1976).
[26] J. J. Jiang, C. E. Diaz, and D. G. Hanson, "Finite element modeling of vocal fold vibration in normal phonation and hyperfunctional dysphonia: Implications for the pathogenesis of vocal nodules," Ann. Otol. Rhinol. Laryngol. **107**, 603–609 (1998).
[27] J. J. Jiang, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," J. Acoust. Soc. Am. **110**, 2120–2128 (2001).
[28] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[29] B. R. Simon, "Multiphase poroelastic finite element models for soft tissue structures," Appl. Mech. Rev. **45**, 191–218 (1992).
[30] K. Verdolini, I. R. Titze, and A. Fennell, "Dependence of phonatory effort on hydration level," J. Speech Hear. Res. **37**, 1001–1007 (1994).
[31] J. R. Levick, "Flow through interstitium and other fibrous matrices," Q. J. Exp. Physiol. **72**, 409–437 (1987).
[32] J. R. Levick, "Relation between hydraulic resistance and composition of the interstitium," Adv. Microcirc. **13**, 124–133 (1987).
[33] K. V. Fisher, J. Ligon, J. L. Sobecks, and D. M. Roxe, "Phonatory effects of body fluid removal," J. Speech Lang. Hear. Res. **44**, 354–367 (2001).

The phonation critical condition in rectangular glottis with wide prephonatory gaps

Chao Tao and Jack J. Jiang[a)]
Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, Wisconsin 53792-7375

(Received 2 March 2007; revised 30 August 2007; accepted 13 December 2007)

The effect of the pressure recovery at glottal exit is introduced to modify the one-mass model. Using the modified one-mass model, the phonation critical condition, including phonation threshold pressure and phonation threshold flow, is analyzed by using the small-amplitude oscillation theory. It is found that the phonation threshold pressure is not sensitive to the change of the prephonatory glottal width at a wide glottal gap. This result agrees with previous experimental observations and suggests that the low slope of dependence of phonation threshold pressure on prephonatory gap found by Chan and Titze [J. Acoust. Soc. Am. **119**, 2351–2362 (2006)] could be a consequence of the pressure recovery effect at the glottal exit. In addition, it is predicted that the phonation threshold flow is always significantly increased with the prephonatory gap even at a wide prephonatory glottal gap. Therefore, the phonation threshold flow has an advantage in assessing the phonatory system at a wide prephonatory gap in comparison with the phonation threshold pressure. The phonation threshold flow can be a useful aerodynamic parameter for pathological conditions in which the incomplete glottal gap is often seen.
 [DOI: 10.1121/1.2832328]

PACS number(s): 43.70.Bk, 43.75.Rs [AL] Pages: 1637–1641

I. INTRODUCTION

The phonation critical conditions, including the phonation threshold pressure (Titze, 1988) and the phonation threshold flow (Jiang and Tao, 2007), are defined as the minimum subglottal pressure or glottal airflow required to initiate vocal fold oscillation. A lot of attention has been paid to these phonation critical conditions because they could provide considerable information about speech system dysfunctions. Applying the small-amplitude oscillation theory on a one-mass vocal fold model, Titze (1988, 1989, 1992) predicted that phonation threshold pressure (PTP) is directly related to tissue properties and glottal configuration. Lucero (1996) analytically suggested the existence of an optimal glottal width for ease of phonation by including the air pressure losses due to glottal viscous resistance. Recently, Jiang and Tao (2007) proposed the phonation threshold flow (PTF) as a new aerodynamic parameter of the speech production system. It is analytically predicted that PTF is a sensitive aerodynamic parameter dependent on tissue properties, glottal configuration, and vocal tract loading. Because the glottal airflow can be easily noninvasively measured, it was suggested that PTF may be more practical for clinical vocal disease assessment. Experimental measurements on excised larynx setups (Verdolini *et al.*, 2002; Hottinger *et al.*, 2007) and physical models (Titze *et al.*, 1995; Chan and Titze, 2006) further confirmed the potential value of these phonation critical conditions on vocal disease assessment.

Recently, Chan and Titze (2006) found that the slope of the dependence of PTP on the prephonatory gap measured on a physical model is lower than the previous theoretical prediction and the measured PTP value is lower than the theoretical prediction at a large glottal gap. The recent experiments on excised larynges also showed that PTP is not sensitive to changes of the prephonatory glottal gap (Hottinger *et al.*, 2007). Simultaneously, it was found in the same study that PTF is highly sensitive to the glottal gaps. Therefore, it was suggested that PTF may be better in indicating vocal fold abduction than PTP (Hottinger *et al.*, 2007). However, according to the previous theory, PTP should linearly increase with the prephonatory glottal gap, i.e., PTP should also be sensitive to the glottal gap. Therefore, the findings in excised larynges and physical models have not been theoretically well explained.

An appropriate theory to predict the PTP and PTF at a large glottal gap is helpful in comprehending these aerodynamic parameters and their effective application in clinical assessment. Therefore, in order to actually predict the phonation critical condition measured in experiments, the effect of the pressure recovery at glottal exit will be introduced to modify the one-mass model in this study. Applying the small-amplitude oscillation theory on the modified one-mass model, PTP and PTF are analytically studied in Sec. II. Then, in Sec. III, the PTP and PTF predicted by the modified one-mass model are compared with experimental data reported by Chan and Titze (2006) and Hottinger *et al.* (2007). At last, we draw our conclusion.

[a)]Additional address: Department of Otolaryngology, EENT Hospital, Fudan University, Shanghai, P. R. China, 200031; Electronic mail: jjjiang@wisc.edu

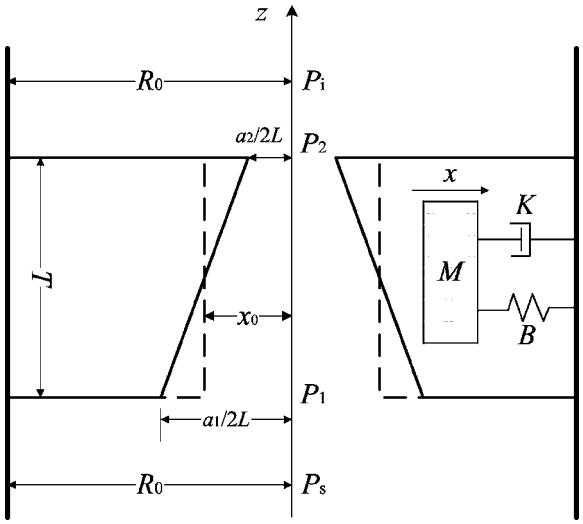


FIG. 1. The sketch map of one-mass model, where M, B, and K are the effective mass, damping, and stiffness per unit area of the vocal fold, P_s is the subglottal pressure, P_1 is the glottal entry pressure, P_2 is the glottal exit pressure, P_i is the input pressure to the vocal tract, x_0 is the prephonatory glottal half-width, a_2 is the glottal area at exit, a_1 is the glottal area at entry, $R_0=A/2L$, A represents the subglottal area and trachea area, L is the vocal fold length, and T is the vocal fold thickness. In addition, in this study, the glottal entry is defined at $z=0$, therefore, the glottal exit is at $z=T$.

II. THE MODIFIED ONE-MASS VOCAL FOLD MODEL

The vocal fold model is shown in Fig. 1. The vocal fold vibration can be approximately described by a simple mechanical oscillator with a spring, a mass block, and a damper:

$$M\ddot{x} + B\dot{x} + Kx = P_g(x,\dot{x},t), \tag{1}$$

where M, B, and K are the effective mass, damping, and stiffness per unit area of the vocal fold. The right term, $P_g = (1/T)\int_0^T P(z)dz$, is the mean intraglottal pressure acting on the vocal fold surface due to the interaction between the vocal fold and the airflow through glottis. $P(z)$ is the intraglottal pressure.

At glottal entry and throughout the glottis, the flow is basically nonturbulent and roughly described by the Bernoulli equation (van den Berg *et al.*, 1957; Scherer, 1981; Gauffin *et al.*, 1983):

$$P_s + \frac{1}{2}\rho\frac{U^2}{A_1^2} = P(z) + \frac{1}{2}\rho\frac{U^2}{a^2(z)} = P_2 + \frac{1}{2}\rho\frac{U^2}{a_2^2}, \tag{2}$$

where P_s is subglottal pressure, P_2 is the glottal exit pressure, A_1 is the subglottal area (trachea area), a_2 is the glottal area at exit, $a(z)$ is the glottal area at position z, U is the airflow volume velocity, and ρ is the air density. In addition, van den Berg *et al.* (1957) observed a good agreement between the measured viscous pressure loss (P_v) and the Poiseuille formula, $\Delta P_v = 12\mu L^2 TU/a_m^3$, where μ is the air viscosity and a_m is the glottal area as the area at the midpoint of the glottis. Then, it has (Lucero, 1996)

$$P_s = P_2 + \frac{1}{2}\rho\frac{U^2}{a_2^2}\left(1 - \frac{a_2^2}{A_1^2}\right) + \frac{12\mu L^2 UT}{a_m^3} \tag{3}$$

and

$$P(z) = P_2 + \frac{1}{2}\rho\frac{U^2}{a_2^2}\left(1 - \frac{a_2^2}{a^2(z)}\right) + \frac{12\mu L^2 U(T-z)}{a_m^3}. \tag{4}$$

In Eqs. (3) and (4), the glottal exit pressure P_2 can be related with P_i, the input pressure to the vocal tract, by $P_2 = P_i + P_r$, where P_r represents the recovered pressure at glottal exit ($z=T$). The recovered pressure P_r is mainly induced by the expansion of airflow from area a_2 (the glottal area at exit) to area A_2 (the vocal tract area) (Ishizaka and Flanagan, 1972). When the airflow velocity is high, the turbulence at glottal exit could occur (Triep *et al.*, 2005; Neubauer *et al.*, 2007) and affect the value of P_r (van den Berg *et al.*, 1957; Scherer, 1981; Gauffin *et al.*, 1983). The recovered pressure is related to the ratio between a_2 and A_2 (Ishizaka and Flanagan, 1972; Titze, 1988):

$$P_r = -\frac{1}{2}\rho\frac{U^2}{a_2^2}\left(\frac{2a_2}{A_2} - \frac{2a_2^2}{A_2^2}\right). \tag{5}$$

It is seen that the recovered pressure P_r will approach zero when $a_2 \ll A_2$. Therefore, in the previous studies (Titze, 1988; Jiang and Tao, 2007), it was assumed that P_2 is approximately equal to P_i. However, when $a_2 \ll A_2$ cannot be satisfied, i.e., the prephonatory gap is wide, the recovered pressure P_r at glottal exit cannot be ignored. Therefore, substituting $P_2 = P_i + P_r$ and Eq. (5) into Eq. (4), gives

$$P(z) = P_i + \frac{1}{2}\rho\frac{U^2}{a_2^2}\left(1 - \frac{2a_2}{A_2} + \frac{2a_2^2}{A_2^2} - \frac{a_2^2}{a^2(z)}\right) + \frac{12\mu L^2 U(T-z)}{a_m^3}. \tag{6}$$

For the sake of simplicity, the following assumption is used to simplify Eq. (6). First, the vocal tract input pressure is roughly equal to the atmospheric pressure ($P_i=0$) by ignoring the vocal tract. Second, the vocal tract and the trachea have the similar geometric size, that is, $A \approx A_1 \approx A_2$. Third, the vocal fold displacement has been linearized along glottis, that is, $a(z) = a_1 + (a_2 - a_1)z/T$, where T is the thickness of vocal fold. Fourth, because this study mainly focuses on the phonation threshold condition in glottis with wide gaps, the viscous term in Eq. (6) is small and can be ignored in this situation (Lucero, 1996). With the above-mentioned approximation, the relationship between the subglottal pressure and the glottal airflow can be obtained by substituting Eq. (5) into Eq. (3),

$$P_s = \frac{1}{2}\rho\frac{U^2}{a_2^2}\left(1 - \frac{a_2}{A}\right)^2. \tag{7}$$

And the glottal pressure P_g acting on vocal folds can be integrated as the mean glottal pressure along the glottis $P_g = (1/T)\int_0^T P(z)dz$:

$$P_g = P_s\left[1 - \frac{(a_2/a_1)}{(1 - a_2/A)^2} + \frac{(a_2/A)^2}{(1 - a_2/A)^2}\right], \tag{8}$$

where the glottal areas at entry and exit can be defined as (Titze, 1988)

$$a_1 = 2L(x_0 + x + \tau\dot{x}), \tag{9}$$

$$a_2 = 2L(x_0 + x - \tau\dot{x}). \tag{10}$$

Here x_0 is the prephonatory half-glottal width, $\tau = T/2c$ is the time delay of the mucosal wave from the lower edge of the vocal folds to the midpoint of the glottis, and c is the mucosal wave velocity. Equations (1) and (8) describe the modified one-mass vocal fold model, which is nonlinear and autonomous. In this model, no vocal tract is attached downstream of the vocal folds. Usually, a single mass model in the absence of a vocal fold tract load will not oscillate. However, a mucosal wave velocity is imposed instead of the rotational component (Titze and Story, 2002) to drive the vocal fold oscillation in this model. Therefore, this one-mass model can oscillate even in the absence of a vocal fold tract.

III. THE PHONATION CRITICAL CONDITIONS

A. Phonation threshold pressure

In order to get simple relationships for the phonation critical condition, we decompose x into a mean value $\bar{x}$ and a small oscillatory component $\tilde{x}$, i.e., $x = \bar{x} + \tilde{x}$. Moreover, it is assumed that the oscillatory component satisfies the small-amplitude oscillation condition, i.e., $(x_0 + \bar{x})^2 \gg (\tilde{x} + \tau\dot{\tilde{x}})^2$. By letting $\ddot{x} = \dot{x} = 0$ and $\tilde{x} = 0$ in Eqs. (1) and (8), the mean displacement $\bar{x}$ of the vocal fold can be determined from $K\bar{x} = P_g(\bar{x}, 0)$, that is,

$$K\bar{x} = P_s\left[1 - (R_0 + x_0 + \bar{x})/(R_0 - x_0 - \bar{x})\right]. \tag{11}$$

Then expanding Eq. (8) around $x = \bar{x}$ and $\dot{x} = 0$ with a Taylor series, it has

$$P_g(x, \dot{x}) = P_g(\bar{x}, 0) + (x - \bar{x})\frac{\partial P_g(\bar{x}, 0)}{\partial x} + \dot{x}\frac{\partial P_g(\bar{x}, 0)}{\partial \dot{x}} + \sum_{i=2}^{\infty}\sum_{j=2}^{\infty}\frac{\partial^{i+j}}{\partial x^i\, \partial \dot{x}^j}\frac{P_g(\bar{x}, 0)}{i!\cdot j!}(x - \bar{x})^i \dot{x}^j. \tag{12}$$

Retaining the linear terms [the first, second, and third terms on the right-hand side of Eq. (12)], Eq. (1) with Eq. (8) can be linearized as

$$M\ddot{x} + \left(B - \frac{\partial P_g}{\partial \dot{x}}(\bar{x}, 0)\right)\dot{x} + \left(K - \frac{\partial P_g}{\partial x}(\bar{x}, 0)\right)(x - \bar{x}) = 0. \tag{13}$$

It is clear from Eq. (13) that oscillation will grow or be sustained when the small-amplitude damping is negative or zero (Titze, 1988), i.e.,

$$B - \frac{\partial P_g}{\partial \dot{x}}(\bar{x}, 0) \leqslant 0. \tag{14}$$

Substituting Eq. (8) into Eq. (14), the phonation threshold pressure P_{th} can be obtained:

$$P_g \geqslant P_{th} = \frac{Bc(x_0 + \bar{x})}{T} \times \frac{(R_0 - x_0 - \bar{x})^2}{R_0(R_0 + x_0 + \bar{x})} \tag{15}$$

with $R_0 = A/2L$.

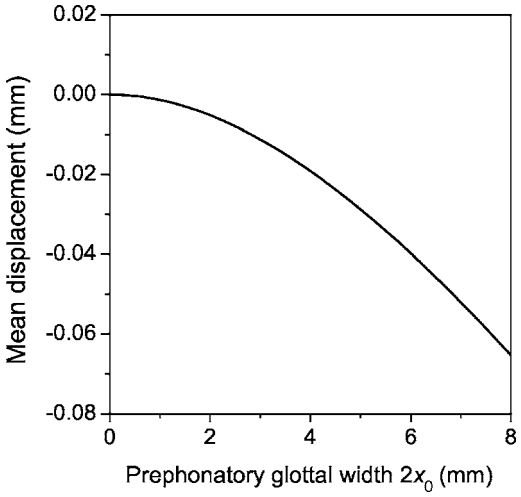


FIG. 2. Mean displacement at the phonation critical condition as a function of prephonatory glottal width.

B. Phonation threshold flow

The phonation threshold flow is defined as the minimum value of the mean airflow to initiate the phonation (Jiang and Tao, 2007). PTF can be easily derived from the relationship between the subglottal pressure and the glottal airflow. Decomposing the airflow U into a mean value $\bar{U}$ and a small oscillatory component $\tilde{U}$, i.e., $U = \bar{U} + \tilde{U}$, and substituting into Eq. (7), we have the relationship between the mean airflow $\bar{U}$ and the subglottal pressure P_s,

$$P_s = \frac{1}{2}\rho\left(\frac{\bar{U}}{2L(x_0 + \bar{x})}\right)^2\left(1 - \frac{x_0 + \bar{x}}{R_0}\right)^2. \tag{16}$$

Then, PTF can be obtained:

$$\bar{U}_{th} = \sqrt{\frac{8L^2}{\rho}\frac{Bc(x_0 + \bar{x})^3}{T} \times \frac{R_0}{(R_0 + x_0 + \bar{x})}}. \tag{17}$$

IV. NUMERICAL SIMULATION

Figure 2 presents the mean displacement at the phonation critical condition as a function of prephonatory glottal width. This result is obtained by numerically solving Eqs. (1) and (8) with the parameters $M = 0.476$ g/cm^2, $K = 200$ kdyn/cm^3, $\rho = 0.00114$ g/cm^3 (Titze, 1988; Ishizaka and Flanagan, 1972), $L = 2.2$ cm, $T = 1.1$ cm, $R_0 = 2.7$ cm (Chan and Titze, 2006), $B = 93.8$ dyn s/cm^3, and $c = 177$ cm/s, where the values of B and c used in this simulation correspond to 0.1% hyaluronic acid (HA) used in Chan and Titze's experiments (2006) and the value of R_0 comes from the dimension of experimental setup used in Chan and Titze's experiments (2006). Usually, the mean displacement for a rectangular glottis always stays at zero in the one-mass model ignoring the pressure recovery at glottal exit (Titze, 1988; Jiang and Tao, 2007). However, in the current model, due to the pressure recovery at glottal exit, the pressure P_2 is negative, which cause the mean displacement to be less than zero. The mean displacement decreases from 0 mm to about −0.07 mm as $2x_0$ increases from 0 to 8 mm. Moreover, it can be seen that the mean displacement is much

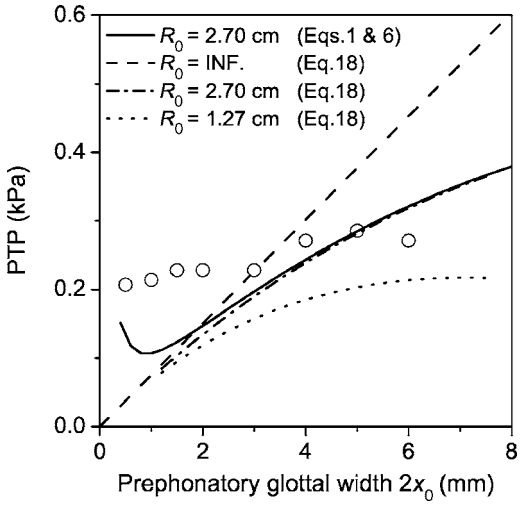


FIG. 3. Phonation threshold pressure as a function of prephonatory glottal width, where the dotted dashed line and dotted line are predicted by Eq. (18) with R_0=2.7 cm and R_0=1.27 cm, respectively. The dash line is predicted by the equation $P_{th}=Bcx_0/T$, which is equal to Eq. (18) with $R_0=\infty$. The solid line is predicted by numerically solving Eqs. (1) and (6) with R_0 =2.7 cm. The values of R_0=2.7 cm and R_0=1.27 cm come from the dimensions of Chan and Titze's (2006) experimental setups with and without uniform-tube vocal tract. The circles represent the onset PTP values measured from the physical model with 0.1% HA (Chan and Titze, 2006).

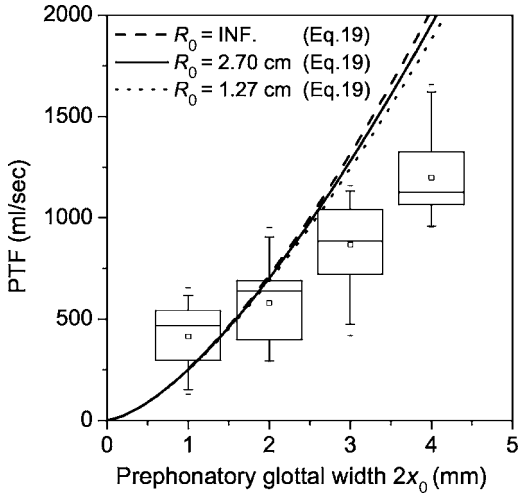


FIG. 4. Phonation threshold flow as a function of prephonatory glottal width, where the dashed line, the solid line, and the dotted line are predicted by Eq. (19) with $R_0=\infty$, R_0=2.70 cm, and R_0=1.27 cm, respectively. The box plots are the experimental results of excised larynges reported by Hottinger *et al.* (2007), where the boundaries of the boxes closest to zero indicate the 25th percentiles, a line within each box marks the median, and the boundaries farthest from zero indicate the 75th percentile, whiskers above and below indicate the 90th and 10th percentiles.

smaller than the prephonatory glottal width x_0. Therefore, in comparison with x_0, $\bar{x}$ in Eqs. (15) and (17) can be ignored and the expression of PTP and PTF can be simplified as

$$P_{th} \approx \frac{Bcx_0}{T}\frac{(R_0-x_0)^2}{R_0(R_0+x_0)} \tag{18}$$

and

$$\bar{U}_{th} \approx \sqrt{\frac{8L^2}{\rho}\frac{Bcx_0^3}{T}\frac{R_0}{(R_0+x_0)}}. \tag{19}$$

It is seen that letting $R_0/x_0 \to \infty$, Eqs. (18) and (19) will be reduced to the PTP equation and the PTF equation reported by Titze (1988) and Jiang and Tao (2007), respectively. Therefore, the previous PTP and PTF theory can be considered as the special case of Eqs. (18) and (19) with $R_0/x_0 \to \infty$.

Figure 3 shows PTP as a function of the prephonatory glottal width. The dash line is predicted by the equation $P_{th}=Bcx_0/T$ (Titze, 1988), which is equal to Eq. (18) with $R_0=\infty$. The dotted dashed line and dotted line are predicted by Eq. (18) with R_0=2.7 cm and R_0=1.27 cm, respectively. The solid line is predicted by numerically evaluating Eqs. (1) and (6) with R_0=2.7 cm and $\mu=1.86\times10^{-4}$ g/(s cm). The other parameters used to predict the curves in Fig. 3 are the same as those used in Fig. 2. It can be seen that when the prephonatory glottal width is small, the PTP value predicted by Eq. (18) is smaller than the numerical solution of Eqs. (1) and (6). This difference exists because the effect of the viscous pressure loss is ignored in Eq. (18). However, when the prephonatory glottal width is large, the dashed dotted line is very close to the solid line, which suggested that Eq. (18) is a good approximation of the phonation critical condition of the vocal fold with large glottal width. For the sake of comparison, the onset PTP results measured from the physical model with 0.1% HA (Fig. 4 in Chan and Titze, 2006) are also presented in Fig. 3 with circles.

As shown in Fig. 3, the PTP value is overpredicted at wide glottal width by the previous theory $P_{th}=Bcx_0/T$. Moreover, the predicted slope of the curve PTP vs x_0 is constant and much larger than the experimental measurement. The PTP value predicted by Eq. (18) is less than that predicted by the previous theory and the discrepancy is increased with the glottal gap. In addition, the slope of the dependence of PTP on x_0 is decreased with the glottal gap due to the effect of the pressure recovery at glottal exit. Because the effect of the pressure recovery is increased with the ratio a_2/A when $a_2<A/2$ [see Eq. (5)], the slope of the dependence of PTF on x_0 is quite small for a large a_2 or a small A (the solid line and the dotted line as shown in Fig. 3). In other words, PTP is not sensitive to a change of vocal abduction at a wide glottal gap. At a wide prephonatory gap, the result predicted by Eq. (18) is close to the measurements in the physical model (Chan and Titze, 2006), and also agrees with the observations in excised larynges (Hottinger *et al.*, 2007), where it was found that the PTP has a low sensitivity to vocal fold abduction.

In addition, one can notice that Eq. (18) underpredicts the PTP value for a narrow glottis as well as the previous PTP theory. This is because the viscous pressure losses associated with small widths have not been accounted for in the present model of vocal folds with wide glottal widths. Lucero (1996) has shown that the viscous pressure losses cause an increase of the threshold pressure at small values of the prephonatory glottal width.

Figure 4 presents PTF as a function of the prephonatory glottal width, where the dashed line is predicted by $\bar{U}_{th} \approx (8L^2Bcx_0^3/\rho T)^{1/2}$, the special case of Eq. (19) with $R_0=\infty$, the solid line and the dotted line are predicted by Eq. (19)

with $R_0=2.70$ cm and $R_0=1.27$ cm, respectively. The other parameters used to predict the curves in Fig. 4 are the same as those used in Figs. 2 and 3. The box plots are the experimental results of excised larynges reported by Hottinger *et al.* (2007), where the boundaries of the boxes closest to zero indicate the 25th percentiles, a line within each box marks the median, the boundaries farthest from zero indicate the 75th percentile, and whiskers above and below indicate the 90th and 10th percentiles.

It can be seen from Fig. 4 that PTF values predicted by Eq. (19) are always increased with the prephonatory glottal width $2x_0$, even when the prephonatory glottal width is wide. The theoretical predictions are in agreement with the measurements from excised larynx experiments. Therefore, the current theory supports the conclusion that PTF could be more effective in indicating vocal fold abduction than PTP. The differences between the theory and the experimental measurements in Fig. 4 could be explained by the simplification of mathematical model, the measurement error existing in experiments, and the discrepancy between the parameters used in Eq. (19) and the actual biomechanical parameters of the excised larynges used in the experiments.

V. CONCLUSION

In this study, the effect of the pressure recovery at glottal exit is introduced to modify the one-mass model. Applying the small-amplitude oscillation theory to the modified one-mass model, we analytically study the phonation critical conditions, including phonation threshold pressure and phonation threshold flow.

It is found that the effect of the pressure recovery at the glottis decreases the PTP value; moreover, the predicted slope of the dependence of PTP on x_0 is also decreased by the effect of the pressure recovery. At large glottal widths, the slope of the dependence of PTP on x_0 is quite small, which indicates that PTP is not sensitive to the change of vocal abduction. These results are close to the experimental measurements of PTP in a physical vocal fold model (Chan and Titze, 2006) and in excised larynges (Hottinger *et al.*, 2007), which suggests that the low sensitivity of PTP to the glottal width may be a consequence of the pressure recovery at glottal exit. The match between the theory prediction and the experimental measurements also verifies the abilities of the pressure recovery hypothesis and the modified one-mass model in predicting the phonation critical conditions at wide gap conditions.

The phonation threshold flow is also analytically studied. It is found that, in contrast to PTP, PTF always significantly increases with the prephonatory glottal width even when the prephonatory glottal width is large. The theoretical predictions agree with excised larynx experiments and suggest that PTF could be a sensitive indicator for vocal abduction. Therefore, PTF is advantageous in assessing the phonatory system at a wide prephonatory gap in comparison with the phonation threshold pressure. PTF can be a useful aerodynamic parameter for clinical vocal disease assessment, especially in pathological conditions such as vocal fold paralysis, vocal nodules and polyps, in which the glottal gap is relatively large.

It is noted that, for the sake of simplicity, the current model lacks consideration of the vocal tract, flow inertia, flow separation effects, and convergent and divergent glottal shape; therefore, the future studies should aim to incorporate these variables into the theory of phonation threshold conditions.

ACKNOWLEDGMENTS

This study was supported by NIH Grant No. 1-RO1DC006019 and No. 1-RO1DC05522 from the National Institute of Deafness and other Communication Disorders and NSFC Grant No. 30328029. The authors also want to thank Lukasz Czerwonka for helpful discussion.

Chan, R. W., and Titze, I. R. (**2006**). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362.

Gauffin, J., Binh, N., Anathapadmanabha, T., and Fant, G. (**1983**). "Glottal geometry and volume velocity waveforms," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. Bless and J. Abbs (College Hill, San Diego, CA).

Hottinger, D. G., Tao, C., and Jiang, J. J. (**2007**). "Comparing phonation threshold flow, pressure by abducting excised larynges," Laryngoscope **117**, 1695–1699.

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.

Jiang, J. J., and Tao, C. (**2007**). "The minimum glottal airflow to initiate vocal fold oscillation," J. Acoust. Soc. Am. **121**, 2873–2881.

Lucero, J. C. (**1996**). "Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis," J. Acoust. Soc. Am. **100**, 2551–2554.

Neubauer, J., Zhang, Z., Miraghaie, R., and Berry, D. (**2007**). "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118.

Scherer, R. (**1981**). "Laryngeal fluid mechanics: Steady flow considerations using static models," Ph.D. dissertation, University of Iowa, Iowa City, IA.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1555.

Titze, I. R. (**1989**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906.

Titze, I. R. (**1992**). "Phonation threshold pressure: A missing link in glottal aerodynamics," J. Acoust. Soc. Am. **91**, 2926–2935.

Titze, I. R., Schmidt, S. S., and Titze, M. R. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084.

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.

Triep, M., Brücker, C., and Scröder, W. (**2005**). "High-speed PIV measurements of the flow downstream of a dynamic mechanical model of the human vocal folds," Exp. Fluids **39**, 232–245.

van den Berg, J., Zantema, J., and Doornenbal, P. (**1957**). "On the air resistance and Bernoulli effect of the human larynx," J. Acoust. Soc. Am. **29**, 626–631.

Verdolini, K., Min, Y., Titze, I. R., Lemke, J., Brown, K., Mersbergen, M. V., Jiang, J. J., and Fisher, K. (**2002**). "Biological mechanisms underlying voice changes due to dehydration," J. Speech Lang. Hear. Res. **45**, 268–281.

Investigation of a glottal related harmonics-to-noise ratio and spectral tilt as indicators of glottal noise in synthesized and human voice signals

Peter J. Murphy[a)]
Department of Electronic and Computer Engineering, University of Limerick, Limerick, Ireland

Kevin G. McGuigan
Department of Physiology and Medical Physics, Royal College of Surgeons in Ireland, 123 St. Stephen's Green, Dublin 2, Ireland

Michael Walsh
Royal College of Surgeons in Ireland, Educational Research Centre, Beaumont Hospital, Dublin 9, Ireland

Michael Colreavy
Children's University Hospital, Temple St., Dublin, Ireland

(Received 24 January 2007; revised 18 December 2007; accepted 18 December 2007)

The harmonics-to-noise ratio (HNR) of the voiced speech signal has implicitly been used to infer information regarding the turbulent noise level at the glottis. However, two problems exist for inferring glottal noise attributes from the HNR of the speech wave form: (i) the measure is fundamental frequency ($f0$) dependent for equal levels of glottal noise, and (ii) any deviation from signal periodicity affects the ratio, not just turbulent noise. An alternative harmonics-to-noise ratio formulation [glottal related HNR (GHNR')] is proposed to overcome the former problem. In GHNR' a mean over the spectral range of interest of the HNRs at specific harmonic/between-harmonic frequencies (expressed in linear scale) is calculated. For the latter issue [(ii)] two spectral tilt measures are shown, using synthesis data, to be sensitive to glottal noise while at the same time being comparatively insensitive to other glottal aperiodicities. The theoretical development predicts that the spectral tilt measures reduce as noise levels increase. A conventional HNR estimator, GHNR' and two spectral tilt measures are applied to a data set of 13 pathological and 12 normal voice samples. One of the tilt measures and GHNR' are shown to provide statistically significant differentiating power over a conventional HNR estimator.
 [DOI: 10.1121/1.2832651]

PACS number(s): 43.70.Dn, 43.70.Gr [BHS] Pages: 1642–1652

I. INTRODUCTION

A. Harmonics-to-noise ratio for voice signals

Acoustic analysis of voice offers the possibility of noninvasive, objective analysis of voice disorders. Acoustic indices have shown promise as noninvasive descriptors for discriminating between patient and normal voice data (cf. Kasuya *et al.*, 1986). A commonly used index is the harmonics-to-noise ratio (HNR), which by definition provides an indication of the ratio of the periodic to aperiodic component in voiced speech signals. The aperiodicity of the voiced speech signal may be due to several causes as neuromuscular, biomechanical, aerodynamic and aeroacoustic forces interact at the level of the larynx (cf. Titze, 1994 for an overview). Typical acoustic manifestations of vocal fold instability/irregularity include jitter, shimmer, turbulent noise and nonlinear effects such as beat frequencies or subharmonics (cf. Schoentgen, 2006).

Two key challenges in the acoustic analysis of voice are:

(i) inferring source variability from a vocal tract filtered signal (i.e. the voiced speech signal) and
(ii) estimating the individual contributory components to source variability.

A relatively straightforward example of challenge (i) is inferring the HNR for a glottal flow wave form containing turbulent noise based on analysis of the microphone recorded voiced speech signal. A consequence of source-filter theory is that the HNR of the speech signal (SHNR) differs from the HNR of the glottal signal (GHNR). For example, consider two time-scaled glottal flow wave forms ($f0=100$ Hz and $f0=200$ Hz) containing 1% glottal noise. GHNR in each case is 40 dB. However, SHNR differs for each signal as the vocal tract (e.g., set to produce the vowel /a/) is excited differently in each case as dictated by the amplitude and spacing of the underlying glottal harmonics. Hence, when SHNR is used in the context of differentiating between normal and pathological voice it is potentially problematic because there is a fundamental frequency ($f0$) dependence on the HNR of the speech wave form i.e., for equal GHNR

[a)]Electronic mail: peter.murphy@ul.ie

values across a range of $f0$ values the SHNR estimates differ as $f0$ changes. SHNR is the measure that is used in practice because it is the voiced speech signal that is available for analysis; GHNR has not previously been estimated from human speech.

In the previous situation only one form of signal aperiodicity was considered—glottal turbulent noise. In the more general case a number of types of signal aperiodicity may be present in the voice signal, e.g., turbulent noise, jitter, shimmer, etc., SHNR is a general measure of speech wave form periodicity: it does not discriminate between the various types of aperiodicity in the voice signal. This has been noted in a number of investigations (Hillenbrand, 1987; Milenkovic, 1988; Qi, 1992; Michaelis *et al.*, 1998; Murphy, 1999). Murphy (1999) observed that in the acoustic analysis of the voiced speech wave form, glottal source variation is only indirectly estimated—under the cover of the vocal tract filter. Hence, if a specific measure of the glottal signal-to-turbulent noise ratio is desired in the general case where all forms of perturbation are present in the voice signal then alternative processing strategies are required. An approach to provide a perturbation-free GHNR measurement was suggested in Murphy (1999). However, the practical implementation of this method requires accurate inverse filtering that retains both the glottal flow and glottal noise characteristics. The implementation of precise inverse filtering strategies is the subject of continued research efforts (c.f. Akande and Murphy, 2005; Fu and Murphy, 2006). In the meantime convenient methods for extracting glottal source information/characteristics from the microphone recorded speech wave form provide a pragmatic alternative. A measure termed glottal to noise excitation ratio (GNE) was proposed by Michaelis *et al.* (1997) and the index was reported in Michaelis *et al.* (1998) to have perturbation-free signal-to-noise measurement properties. However, GNE employs the linear prediction residual signal rather than a glottal flow signal and hence the resulting measure is indirectly related to GHNR. Two candidate measures for estimating glottal noise characteristics are investigated in the present study: GHNR'—glottal related harmonics-to-noise ratio and spectral tilt. Before detailing the theoretical basis for the methods a brief review of spectral tilt measurement in the context of noise estimation is provided.

B. Spectral tilt measures for voice quality assessment and noise estimation

Spectral tilt estimates have found application in the assessment and characterization of various aspects of voice quality. For example, the spectral measurements H_1-A_1 and H_1-A_3 (level difference between first harmonic and first and third formant levels, respectively) have been shown to correlate with measures of first formant bandwidth and glottal adduction, respectively (Hanson, 1997; Hansen and Chuang, 2000; Stevens, 2000). In addition loudness variation has been shown to correlate with measures of spectral tilt (Sundberg and Nordenberg, 2006). In other studies the correlation of spectral tilt with listeners' ratings of breathiness has produced differing results. Hillenbrand *et al.* (1994) provide a review of spectral tilt measures in the context of their correlation with perceived degree of breathiness ratings. In summary, their review of the literature reveals that there are somewhat conflicting reports on the relationship between spectral tilt and breathiness, with certain investigations (e.g. Frokjaer-Jensen and Prytz, 1976; Fukazawa *et al.*, 1988; Klich, 1982) implying that breathiness correlates with an increase in high frequency energy, whereas other studies (Hillenbrand, 1987; Klatt and Klatt, 1990) suggest that spectral tilt has less effect on the perception of breathy voice. Further studies of voice types (e.g. pressed, normal and breathy) have shown that voice quality can be differentiated on the basis of spectral tilt with breathy voice being characterized by a greater than 12 dB/octave falloff in energy of the harmonic structure (Childers and Lee, 1990; Alku *et al.*, 1997).

Spectral tilt measures have also been applied to connected speech for the specific purpose of identifying dysphonic speakers (Wendler *et al.*, 1980; Prytz, 1977; Hammarberg *et al.*, 1980; Löfqvist, 1986; Kitzing, 1986; Kitzing and Akerlund, 1993). The long term average spectrum (LTAS), defined as the average of successive spectra of the connected speech signal, is employed in these studies. However opinions on the potential use of spectral tilt measures extracted from the LTAS as objective indicators of dysphonia have been divided (Wendler *et al.*, 1980; Prytz, 1977). Hammarberg *et al.* (1980) successfully used the LTAS to differentiate breathiness conditions of both hypofunction and hyperfunctional voice. Wendler *et al.* (1980) also found it to be a very promising measure but later the first author tailored his initial optimistic opinions on its use (Wendler *et al.*, 1986). In a study performed by Löfqvist (1986) an attempt was made to differentiate a set of 37 clinical voices from a set of 36 normal voices. The two measures taken from the LTAS were the ratio of the energy below 1 kHz to the energy from 1 to 5 kHz and the energy level between 5 and 8 kHz. The result of the analysis did not separate the data sets. Löfqvist and Mandersson (1987) investigated the effect of the time length on the LTAS and found that a minimum of 10 s is required to provide predictable estimates. In a study performed by Kitzing (1986) an extensive set of LTAS derived measured were investigated. Their measures: (i) the ratio of energy below and above 1 kHz, (ii) a measure of the spectral slope inclination in the first formant range, and (iii) the ratio of the peak level of the fundamental and the first formant region) proved useful for separating strained and sonorous vocal qualities. In a later study (Kitzing and Akerlund, 1993) LTAS derived measures were shown to correlate moderately well with perceived improvements in patients undergoing voice therapy.

In summary, as per Hillenbrand *et al.* (1994) relating spectral tilt to listeners' impression of breathiness, the present review with respect to pathological voice reveals somewhat mixed findings reported regarding spectral tilt as an indicator of disordered voice quality. Looking at these studies in general, some reasons for the difference in findings may be attributable to the different types of data employed; synthesis versus real speech, sustained vowel versus connected speech or, to the differences in the technique used to estimate spectral tilt. Further reasons are suggested in the light of the current work relating spectral tilt to glottal noise.

The bases for the glottal related HNR measure, GHNR' and the spectral tilt measures R_{14} and R_{24}—the ratio between the speech spectral energies from 0 to 1 kHz to the energy from 1 to 4 kHz and the ratio between the speech spectral energies from 0 to 2 kHz to the energy from 2 to 4 kHz, respectively, are first described theoretically. The measures are then tested systematically with synthesis data and against a set of pathological and normal real voice data. GNHR' is shown to be largely $f0$ independent but sensitive to all forms of waveform aperiodicity. For the synthesis data spectral tilt is shown to increase as the noise variance increases, whereas it is relatively insensitive to the perturbations of jitter and shimmer. This suggests that spectral tilt can potentially be used as a perturbation-free measurement of noise levels in voice signals. However, the measure is shown to be $f0$ dependent. The results for real speech indicate that GHNR' and R_{14} show statistically significant improvement over existing measures in separating patient/normal data.

II. THEORY

A. Glottal noise in sustained vowel phonation

As a useful first approximation, a sampled glottal flow waveform containing turbulent noise ($g_n[i]$) can be written as

$$g_n[i] = g[i] + n[i], \tag{1}$$

where g is the noise free, periodic glottal wave form, n represents Gaussian white noise, and i is the sample index running from $i=0$ to $I-1$ (where I is the total sample length).

Taking the discrete Fourier transform (DFT) and normalizing by dividing by the window length, I gives

$$G_n[k] = G[k] + N[k], \tag{2}$$

where capital letters indicate the normalized Fourier transforms of the sampled time domain equivalents and k is the frequency sample index of the DFT running from $k=0$ to $I/2-1$.

Taking energy values gives,

$$|G_n[k]|^2 = |G[k] + N[k]|^2 \tag{3}$$

$$= |G[k]|^2 + |N[k]|^2 + 2|G[k]||N[k]|\cos(\theta_{GN}[k]), \tag{4}$$

where $\theta_{GN}[k]$ is the phase angle between $G[k]$ and $N[k]$.

From Eq. (4) it can be inferred that at harmonic frequencies ($G[k] \neq 0$) all three terms contribute to the signal amplitude (noise is broadband), whereas between harmonic locations ($G[k]=0$) the first and last terms disappear. Therefore in the energy spectrum representation of a single segment of voiced speech, even if the noise is constant at all frequencies, the effect of the noise is different at harmonic and between harmonic locations. The last term indicates that for low levels of noise the effect of the noise on the harmonics is greatest at low frequencies where the signal energy is largest. For higher noise levels the noise-only term dominates. In each case the harmonic level may be increased or decreased due to the presence of noise.

For the speech signal (Fant, 1970) this becomes

$$s_n[i] = (g[i] + n[i]) * v[i] * r[i], \tag{5}$$

where s_n is the voiced speech wave form produced by exciting the vocal tract (v) and radiation load (r) with a periodic glottal source plus Gaussian noise. The asterisk indicates convolution.

Taking the discrete Fourier transform gives

$$S_n[k] = (G[k] + N[k]) \times V[k] \times R[k], \tag{6}$$

where again capital letters indicate the normalized Fourier transforms of the sampled time domain equivalent measures.

Taking magnitude-squared values gives

$$|S_n[k]|^2 = |G[k]V[k]R[k] + N[k]V[k]R[k]|^2, \tag{7}$$

$$|S_n[k]|^2 = \{|G[k]|^2 + |N[k]|^2 + 2|G[k]||N[k]|\cos(\theta_{GN}[k])\}|V[k]|^2|R[k]|^2. \tag{8}$$

Assuming the vocal tract acts as a linear, time-invariant system during the analysis interval, on averaging successive short term spectra (where $|S_n[k]|^2_{hi}$ is the ith spectrum) the third term within the braces in Eq. (8) disappears (zero mean Gaussian noise) and the noise $N[k]$ at a given frequency approaches the true noise variance, $\sigma_n^2[k]$ at that frequency, to give

$$|S_n[k]|^2_{h\ \mathrm{AV}} = |G[k]|^2_{\mathrm{AV}}|V[k]|^2_{\mathrm{AV}}|R[k]|^2_{\mathrm{AV}} + \sigma_n^2[k]|V[k]|^2_{\mathrm{AV}}|R[k]|^2_{\mathrm{AV}} \tag{9}$$

and

$$|S_n[k]|^2_{\mathrm{bh\ AV}} = \sigma_n^2[k]|V[k]|^2_{\mathrm{AV}}|R[k]|^2_{\mathrm{AV}}, \tag{10}$$

where $|S_n[k]|^2_{h\mathrm{AV}}$ and $|S_n[k]|^2_{\mathrm{bhAV}}$ indicate the average of the power spectra of voiced speech containing glottal noise, at harmonic and at between-harmonic frequencies, respectively, and $\sigma_n[k]$ represents the standard deviation of the noise at frequency index, k.

B. Glottal related harmonics-to-noise ratio

As stated in Sec. I the HNR taken from the speech wave form or spectrum is a commonly used measure in assessing voice disorders. As one goal of the acoustic analysis of voice signals is to supply noninvasive information regarding the physical characteristics at the level of the glottis it is pertinent to ask, in what sense is the harmonics-to-noise ratio of the speech wave form indicative of the harmonics-to-noise ratio of the glottal source signal. Rearranging Eq. (9) gives

$$\mathrm{HNR}[k] = \frac{(|G[k]| \times |V[k]||R[k]|)^2}{(\sigma_n[k] \times |V[k]||R[k]|)^2} = \frac{|G[k]|^2}{|\sigma_n[k]|^2} \tag{11}$$

which is the HNR at frequency bin k. Within the limits of the source filter model, the HNR at a given harmonic location is the same for the source and filtered wave form, i.e.,

$$\mathrm{HNR}[k](\text{speech wave form}) = \mathrm{HNR}[k](\text{glottal source}). \tag{12}$$

However, the HNR is usually determined over the complete frequency range or over a bandlimited region and not at a single discrete frequency. Therefore the numerator and de-

moninator in Eq. (11) are independently summed over the frequency range of interest in order to obtain the overall harmonics-to-noise ratio for the signal segment. The ratio is generally expressed in decibels:

$$\text{HNR(speech wave form)} = 10\log_{10}\left[\frac{\frac{2}{I}\sum_{k=0}^{I/2-1}(|G[k]||V[k]||R[k]|)^2}{\sum_{k=0}^{I/2-1}(\sigma_n[k]|V[k]||R[k]|)^2}\right]. \quad (13)$$

As $V[k]$ and $R[k]$ no longer cancel in the calculation, this ratio varies as the fundamental frequency of the glottal source varies, exciting the vocal tract at different frequencies. Hence

$$\text{HNR(speech wave form)} \neq \text{HNR(glottal source)}, \quad (14)$$

where

$$\text{HNR(glottal source)} = 10\log_{10}\left[\frac{\frac{2}{I}\sum_{k=0}^{I/2-1}(|G[k]|)^2}{\sum_{k=0}^{I/2-1}(|\sigma_n[k]|)^2}\right] \quad (15)$$

Equation (13) is a generic formulation of the harmonics-to-noise ratio as calculated by various investigators. Variations include giving the signal-to-noise ratio by including the noise and harmonics as signal or inverting this ratio, giving the noise-to-signal ratio. Band limiting the range over which the ratio has been calculated has also been investigated. Despite the fact that several variations of Eq. (13) have been implemented, no study has attempted to relate the HNR of the speech wave form to the HNR of the glottal source (GHNR). Taking an alternative summation allows a ratio closely related to GHNR to be recovered:

$$\text{GHNR}' = 10\log_{10}\left(\frac{2}{I}\sum_{k=0}^{I/2-1}\frac{|G[k]|^2|V[k]|^2|R[k]|^2}{\sigma_n^2[k]|V[k]|^2|R[k]|^2}\right) \quad (16)$$

In practice, for voiced speech, $|G[k]|^2|V[k]|^2$ is estimated at harmonic locations and $\sigma_n^2[k]\ |V[k]|^2$ can be estimated at between-harmonic locations. To a first order approximation (in reality noise is broadband and the signal energy is located at harmonic locations—the present calculation is limited to harmonic locations) the following relation holds on taking the summation:

$$\text{GHNR}' \cong 10\log_{10}\left(\frac{2}{I}\sum_{k=0}^{I/2-1}\frac{|G[k]|^2}{|\sigma_n[k]|^2}\right) \quad (17)$$

Through comparison of Eq. (15) and Eq. (17) it can be inferred that GHNR′ reflects the HNR of the glottal source (GHNR) in a true sense only when either the noise is equal for all frequencies or the HNR is constant for all frequencies. The former is true for zero mean, Gaussian white noise, whereas the latter is satisfied when shimmer (only) is present in the source signal (Murphy, 2000). Cox *et al.* (1989) similarly investigated this ratio, also noting the bias given to formant frequencies in traditional HNR calculations. However the authors did not relate the ratio to the glottal source harmonics-to-noise ratio as presently detailed. Another somewhat interesting aspect of GHNR′ is that it effectively gives greater weight to the higher frequency components. Each harmonic/between harmonic estimate is treated equally in the summation of Eq. (16) and hence the noise at high frequencies (low glottal harmonic energy) has a greater influence than it does in the calculation of HNR. This perhaps also supplies GHNR′ with more auditory perceptual relevance than either HNR or GHNR.

C. Spectral tilt

Two spectral tilt measures [the ratio of the energy below 1 kHz to that from 1 to 4 kHz—R_{14}—Eq. (18), and the ratio of the energy below 2 kHz to that from 2 to 4 kHz—R_{24}—Eq. (19)] are estimated in the present study:

$$R_{14} = \frac{\sum_{k=1}^{N/5} S_n[k]^2_{\text{AV}}}{\sum_{k=N/5}^{4N/5} S_n[k]^2_{\text{AV}}}, \quad (18)$$

$$R_{24} = \frac{\sum_{i=1}^{N2/5} S_n[k]^2_{\text{AV}}}{\sum_{k=N\times 2/5}^{4\times N/5} S_n[k]^2_{\text{AV}}}, \quad (19)$$

where $S_n[k]^2_{\text{AV}}$ is the spectral energy at the kth frequency location taken from the averaged spectrum. $N=I/2$ =number of spectral estimates covering up to 5 kHz.

From Sec. II A section it can be inferred [Eqs. (10) and (11)] that the R_{14} tilt varies with noise as per the following ratio:

$$\frac{|S[k]|^2_{01} + (|N[k]||V[k]||R[k]|)^2_{01}}{|S[k]|^2_{14} + (|N[k]||V[k]||R[k]|)^2_{14}}, \quad (20)$$

where $|S[k]|^2_{01}$ and $|S[k]|^2_{14}$ represent the noise free energy from 0 to 1 kHz and from 1 to 4 kHz, respectively, and $(|N[k]||V[k]||R[k]|)^2_{01}$ and $+(|N[k]||V[k]||R[k]|)^2_{14}$ represent the noise energy from 0 to 1 kHz and from 1 to 4 kHz, respectively. Therefore the noise at the glottis is distributed as a 1/4:3/4 ratio (0–1 kHz/1–4 kHz) which then receives different vocal tract and radiation responses before being added to the noise free estimates. The radiation at the lips preferentially enhances the high frequency noise with its +6 dB/octave characteristic. Therefore the theory predicts that the overall result is for spectral tilt to decrease as glottal noise increases.

III. METHOD

A. Synthesis

The vowel /a/ was synthesized using a discrete time model for speech production (Rabiner and Schafer, 1978) with a Rosenberg glottal flow pulse (type C, Rosenberg, 1971) used as the source function. Radiation at the lips was modeled by the first order difference equation $R(z)=1-z^{-1}$. Wave form aperiodicity was introduced by altering the source function. Random shimmer was introduced by adding

TABLE I. Patient listing and details.

Patient No.	Age	Sex	Pathology
1	39	F	Vocal cord nodules (bilateral)
2	70	M	No organic pathology (hoarseness reported)
3	43	F	Vocal cord oedema /nodules
4	33	M	Vocal nodule
5	22	F	Vocal cord nodules (bilateral)
6	22	F	No organic pathology (hoarseness reported—on/off)
7	43	M	Verrucous carcinoma of both folds
8	65	M	Hyperkeratosis and parakeratosis
9	57	F	Mild swollen vocal cords
10	74	M	Carcinoma post radiation right vocal cord immobile
11	23	F	Left vocal cord palsy Immobile—well compensated right cord
12	54	M	Laryngeal papillomaptosis
13	57	F	Abductor palsy

a random variable gain factor (six doubling levels from 1% to 32% s.d.) to the amplitude of the fundamental period impulse train prior to convolution with the glottal pulse. The fundamental period was multiplied by a random number generator of given variance in order to provide the required amount of random jitter (1% to 6% s.d.). Gaussian noise was introduced to the glottal pulse via a random noise generator arranged to give noise of a user specified variance (six doubling levels from 1% to 32% s.d.). Further signals were created for three levels of Gaussian noise for fundamental frequencies beginning at 80 Hz and increasing in six, approximately equi-spaced steps of 60 Hz up to 350 Hz. A sampling frequency of 10 kHz and 16-bit quantization was used throughout.

B. Human voice signals

Recordings were made of the participants phonating the sustained vowel /a/ at their comfortable pitch and loudness level. All recordings were made using an electret microphone and a Tandberg audio recorder (AT 771, Audio Tutor Educational, Japan) prior to the participants (13 in all) undergoing laryngovideostroboscopic (LVS, Endo-Stroboskop, Atmos) evaluation at the outpatient's ENT (Ear, Nose and Throat) clinic in Beaumont Hospital, Dublin, Ireland.

Along with the results of the stroboscopic examination, full medical details regarding the vocal pathology were taken for each patient as well as any further diagnostic comments at the time of assessment. Patient details are outlined in Table I. The audio (and video) data from the stroboscopic evaluation were recorded using SONY SVHS (E-180) cassettes. Twelve normal speakers were subsequently recorded under the same conditions. All recordings were low-passed filtered at 4 kHz and digitized at a sampling rate of 10 kHz into a data acquisition expansion card (Integrated Measurement Systems, PCL-814, Southampton, UK) with 14-bit resolution and stored for subsequent analysis. In the acoustic analysis only the spectral measures (tilt and HNR) were estimated from the digitized recordings. Reliable perturbation estimation necessitates an alternative recording protocol (c.f. Doherty and Shipp, 1988; Titze and Winholtz, 1993).

The aim of the acoustic evaluation of human speech signals in the present study is to separate the patients and normals based solely on acoustic analysis of the audio recording (Valencia-Naranjo *et al.*, 1995; Fex *et al.*, 1995). The data obtained in the stroboscopic and perceptual evaluations were used to demarcate the separation of patient and normal data sets. A simple perceptual rating scale, based on a system proposed by Hammarberg and Gauffin (1995) was used for both the normal and patient data. In the case of the normal data, two of the normals showed slightly dysphonic voice qualities, the equivalent of a rating of one on a five-point scale, where zero represents normal and four indicates severe dysphonia. All patient data showed dysphonic qualities (as rated by two speech and language pathologists) but unfortunately the degree is not given and therefore the perceptual ratings are used simply as an accompaniment to the acoustic findings, rather than as the basis for correlation. Furthermore, there was only 66% agreement between each speech pathologist's ratings of the individual categories of the patient data and hence the data did not provide a good platform for correlation purposes. Therefore, for investigation of the potential usefulness of spectral tilt and GHNR' as an indicators of dysphonic voice it is simply stated that the patient data was perceptually assessed as being dysphonic (though varying in severity), while the normal data was rated normal except for two of the subjects who exhibited very mild dysphonia.

C. Analysis parameters

A sustained phonation of the vowel /a/ was analyzed. A Hamming window of length 2048, padded up to 4096 and hopped by 1024 points providing 8 independent spectral estimates (Welch, 1970) for about 1.2 s of speech, was used for the synthesis data and for the GHNR' measures on real speech. Most frequency domain HNR estimators can be adapted to provide a measure of GHNR'. A method similar to Kasuya *et al.* (1986) was used in the present study. In calculating SHNR the noise estimates are calculated by directly estimating the energy at between harmonic main lobes and interpolating these adjacent values to provide an estimate at frequencies within harmonic main lobes. Harmonic energy is estimated by subtracting the interpolated noise estimates from the harmonic estimates within the main lobe width ($8 \times 4096/2/2048$). In determining the noise estimate at a given harmonic in GHNR' an average noise estimate at each adjacent between harmonic main lobe is estimated. The peak value within a harmonic main lobe minus the harmonic noise estimate is taken as the harmonic energy estimate. The following input parameters were used for the spectral tilt measures of real speech: Hamming window size—256, hop size—100 and total length 0.75 s. A conventional HNR [Eq.

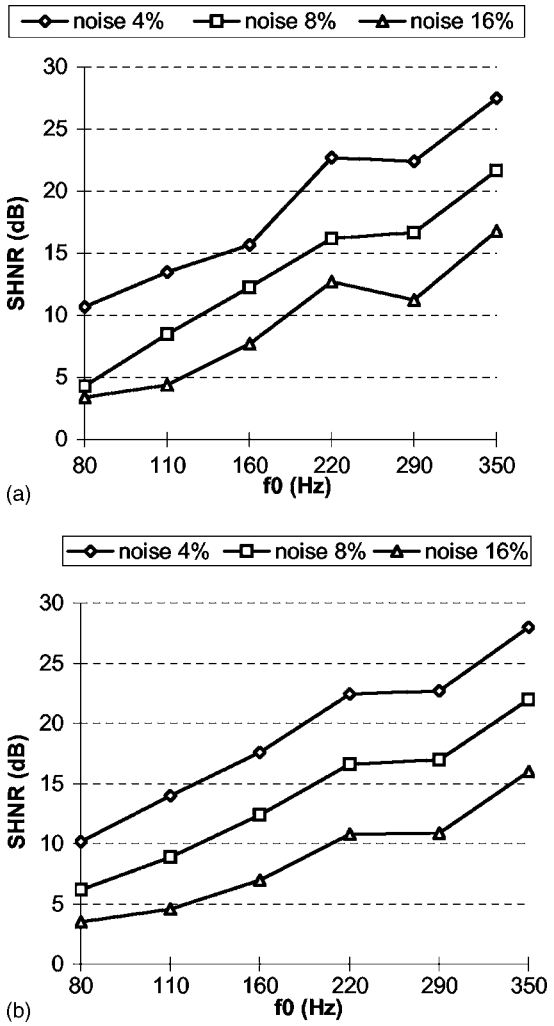


FIG. 1. Speech harmonics-to-noise ratio (SHNR) for the synthesized vowel /a/ vs $f0$ for three levels of glottal noise taken from (a) a single spectrum and (b) from averaged spectra

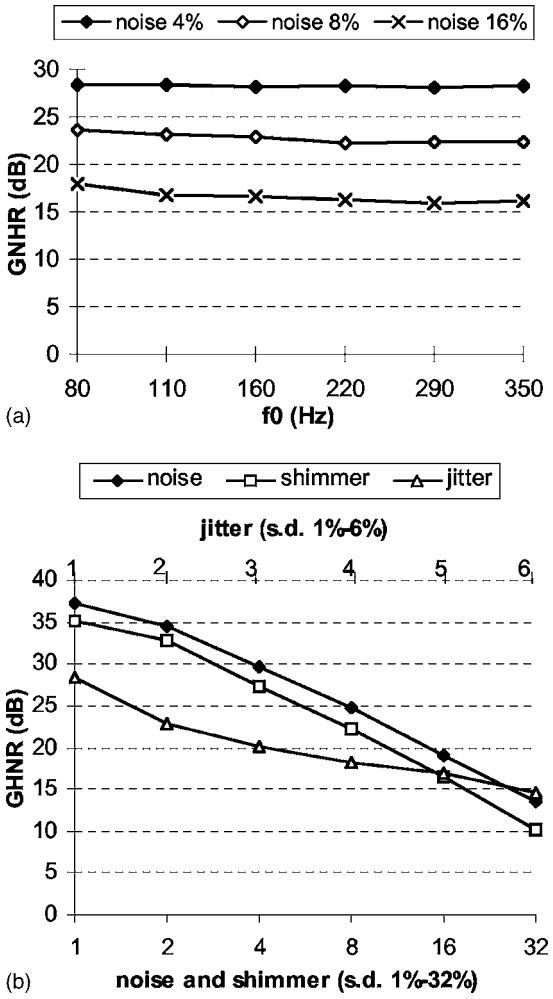


FIG. 2. Glottal harmonics-to-noise ratio (GHNR) vs (a) $f0$ for three noise levels and (b) noise, shimmer $f0$ and random and cyclic jitter of the glottal source

(13)], the source related HNR [GHNR'—Eq. (16)] and the two spectral tilt measures [R_{14}—Eq. (18) and R_{24}—Eq. (19)] were estimated for all data.

IV. RESULTS

A. Synthesis

Figure 1 shows the variation of the HNR of the speech signal (SHNR) versus $f0$ for three levels of glottal noise using (a) a single spectrum and (b) an averaged spectrum. It can be seen that spectral averaging provides more reliable HNR estimates. However, SHNR is $f0$ dependent because the vocal tract is excited at different harmonic frequencies as dictated by $f0$. The variation of the HNR of the glottal signal (GHNR) versus $f0$ for three levels of glottal noise is shown in Fig. 2(a), whereas Fig. 2(b) illustrates the variation of GHNR with noise, shimmer and jitter for a signal with a fundamental frequency of 110 Hz. Figure 2 affords a validation of the HNR estimation procedure. The noise contaminated glottal signal has noise levels of 1%, 2%, 4%, 8%, 16%, and 32%. These noise levels correspond to GHNRs of 40, 34, 28, 22, 16, and 10 dB, respectively. Similarly, the shimmered glottal signals (doubling in magnitude levels from 1% to 32%) have corresponding GHNRs of 40 dB decreasing in 6 dB steps to 10 dB. The variation of GHNR' with respect to $f0$ for three levels of glottal noise is shown in Fig. 3(a), whereas Fig. 3(b) shows that GHNR' is sensitive to glottal noise and perturbation. Comparing Fig. 2 with Fig. 3 allows for a validation of the GHNR' measure. As stated previously GHNR' matches GHNR when noise is constant at all frequencies or when shimmer (only) is present in the signal. For noise or shimmer levels of 1%, 2%, 4%, 8%, 16%, and 32% the corresponding theoretical GHNRs are 40, 34, 28, 22, 16, and 10 dB, respectively. It is observed that GHNR' approximates these theoretical values across $f0$ for three noise levels [Fig. 3(a)] and for six noise and shimmer levels at 110 Hz [Fig. 3(b)].

Figure 4 shows (a) R_{14} and (b) R_{24} plotted against glottal noise and perturbation. The tilt indices decrease for increasing noise levels while remaining relatively insensitive to the perturbations of jitter and shimmer. Figure 5 shows that (a) R_{14} varies from 0 to 6 and that (b) R_{24} varies from 0 to 70, for a range of $f0$ values ranging from 80 to 350 Hz.

B. Human voice signals

Histograms of the patient/normal data for the HNR and spectral tilt measures are given in Figs. 6 and 7, respectively.

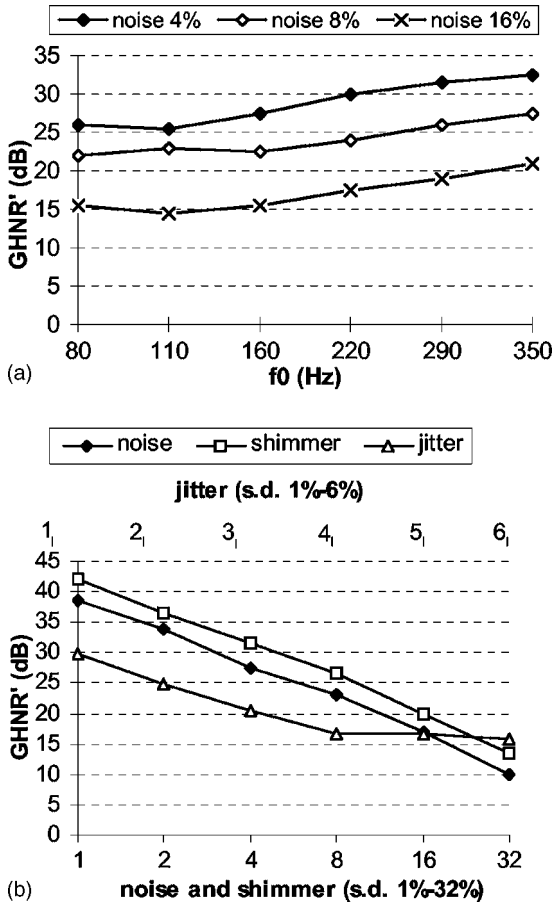


FIG. 3. Glottal related harmonics-to-noise ratio (GHNR') vs (a) $f0$ for three noise levels and (b) noise, shimmer $f0$ and random and cyclic jitter of the glottal source

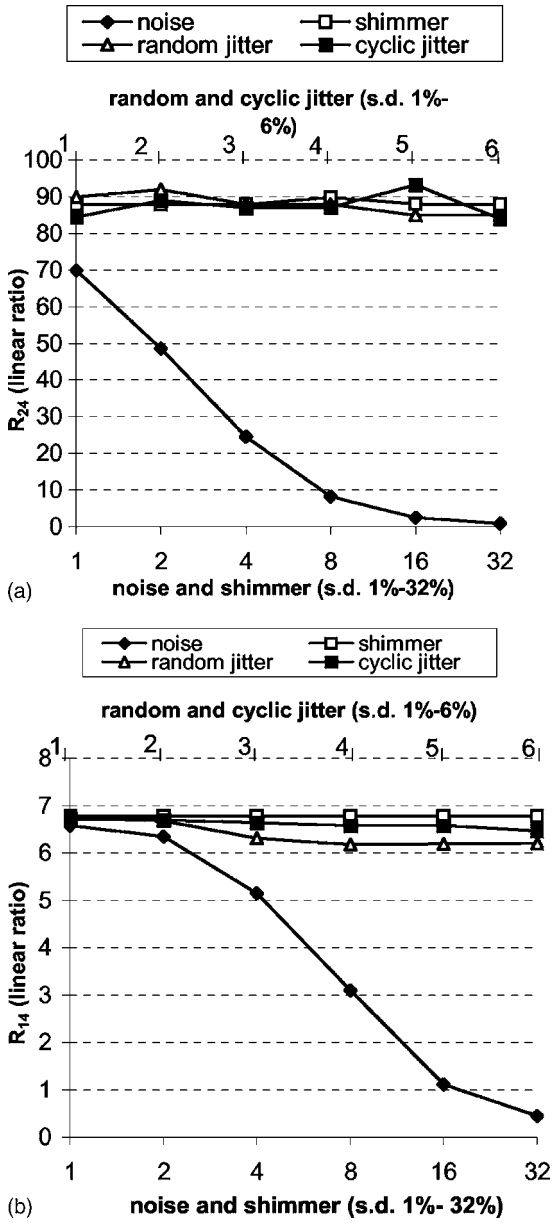


FIG. 4. (a) Ratio of energy below 2 kHz to energy from 2 to 4 kHz for the speech signal (R_{24}) vs noise, shimmer and random and cyclic jitter of the glottal source. (b) Ratio of energy below 1 kHz to energy from 1 to 4 kHz for the speech signal (R_{14}) vs noise, shimmer and random and cyclic jitter of the glottal source

Figure 6 shows (a) SHNR and (b) GHNR' for the patient/normal data. The latter shows statistically significant separability of the data sets (i.e., significant at the 5% level in a one-tailed, equal variance, student's t-test), whereas the former does not. Referring to Fig. 7, it can be seen that the R_{24} ratio shows poor separability, whereas the R_{14} ratio has separated all but one of the patients from the normal data. Interestingly, however, the results are in complete opposition to the ratios obtained with the Gaussian noise (Fig. 4)—for synthesis as noise increased R_{14} reduced, whereas for the human voice signals, the R_{14} values were in general higher for the patient data. In respect of the R_{14} ratio, the "normal" data are in agreement with values obtained from the low glottal noise or perturbation results whereas the patient data shows a marked increase in this ratio.

V. DISCUSSION

In comparing Figs. 1(a) and 2(a) it can be seen that the $f0$ variation is greatly reduced in the latter case as desired. The variation of GHNR' with $f0$ is not completely flat however. If the number of harmonics is specified, as opposed to taking the complete frequency range, the remaining $f0$ influence should be removed, e.g., compare a 100 Hz glottal signal and a 200 Hz glottal signal over a set frequency range—for a flat noise level more of the 100 Hz harmonics are subsumed by the noise level than for the 200 Hz signal, giving rise to a reduced GHNR'. In investigating the variation of the GHNR' (and spectral tilt) measures with $f0$ a constant filtering action is assumed. However, the constancy of the filtering action over the entire $f0$ range is an oversimplification; e.g., formant frequencies of male speakers are on average 25% lower than the formant frequencies of female speakers. For real speech considerable improvement in the separation of the patient/normal data is observed when using GHNR' in place of HNR. This is as might be expected as GHNR' more closely represents the HNR of the glottal source.

The spectral tilt ratios R_{14}, R_{24} have been shown to decrease with increasing noise levels. At the same time, the indices are shown to be largely independent of jitter and shimmer. This is encouraging as the extraction of

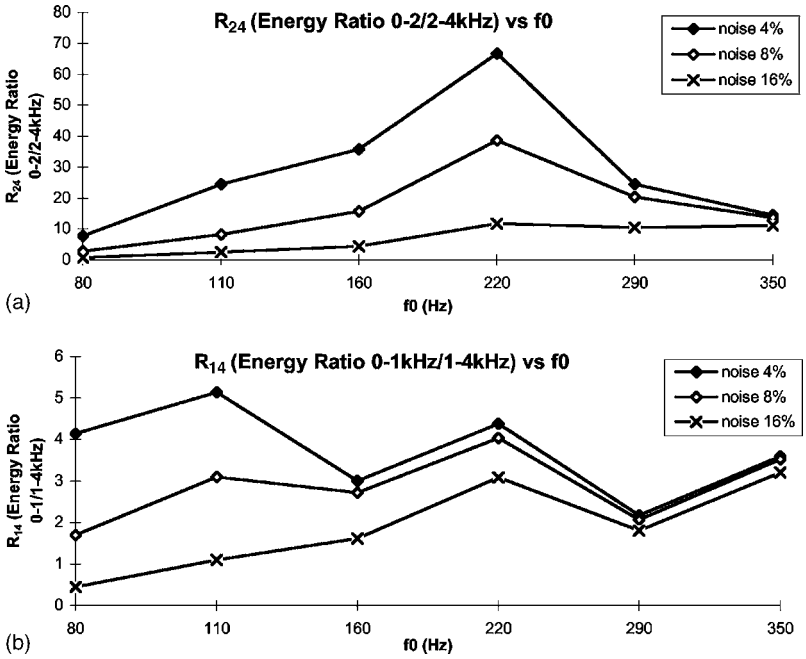


FIG. 5. Variation of (a) R_{14} (speech signal) and (b) R_{24} (speech signal) with fundamental frequency for three levels of glottal noise of std. dev. 4%, 8%, and 16%

perturbation-free glottal turbulent noise estimates are an ongoing concern. In general this is a difficult problem given the complicated effects of perturbations on the spectra of voiced speech (Fig. 8). From Sec. II A it can be inferred [Eqs. (10) and (11)] that the R_{14} tilt varies with noise as per the following ratio:

$$\frac{|S[k]|_{01}^2 + (|N[k]||V[k]||R[k]|)_{01}^2}{|S[k]|_{14}^2 + (|N[k]||V[k]||R[k]|)_{14}^2}, \tag{21}$$

where $|S[k]|_{01}^2$ and $|S[k]|_{14}^2$ represent the noise free energy from 0 to 1 kHz and from 1 to 4 kHz, respectively and $(|N[k]||V[k]||R[k]|)_{01}^2$ and $+(|N[k]||V[k]||R[k]|)_{14}^2$ represent the noise energy from 0 to 1 kHz and from 1 to 4 kHz, respectively. Therefore the noise at the glottis is distributed as a 1/4:3/4 ratio (0–1 kHz/1–4 kHz) which then receives different vocal tract and radiation responses before being added to the noise free estimates. The radiation at the lips preferentially enhances the high frequency noise with its +6 dB/octave characteristic. This alters the 1/4:3/4 noise ratio for 0–1 kHz/1–4 kHz. But whatever this ratio is, it remains constant as noise increases. The tilt reduces because of the additive factor.

The histogram of the patient/normal data shows statistically significant separation of the data for R_{14} but not for R_{24}. For R_{14} the patient/normal values vary from 0 to 36, whereas the synthesis values vary from 0 to 7. Of particular note is that the patient R_{14} tilt values are higher than the normal values whereas the synthesis shows that R_{14} decreases with increasing noise levels. This suggests that for real speech the introduction of greater amounts of glottal turbulent noise takes place concurrently with other glottal changes or/and that the noise is not simply Gaussian white noise. Regarding the latter hypothesis a preliminary investigation by O'Leidhin and Murphy (2005) shows that the response of spectral tilt to segment noise (noise on for a certain percent-

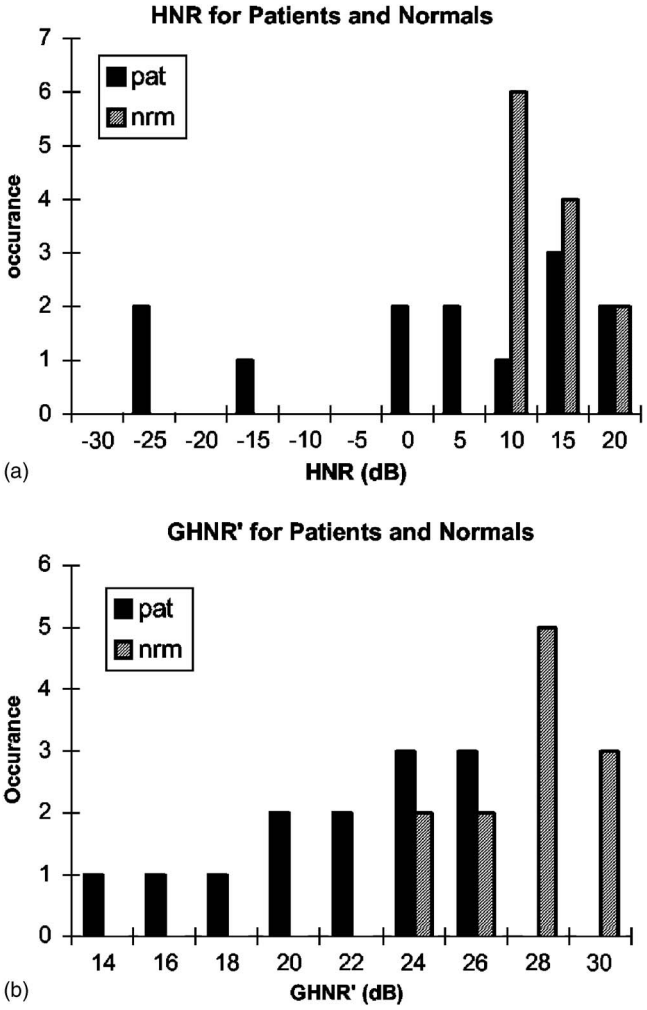
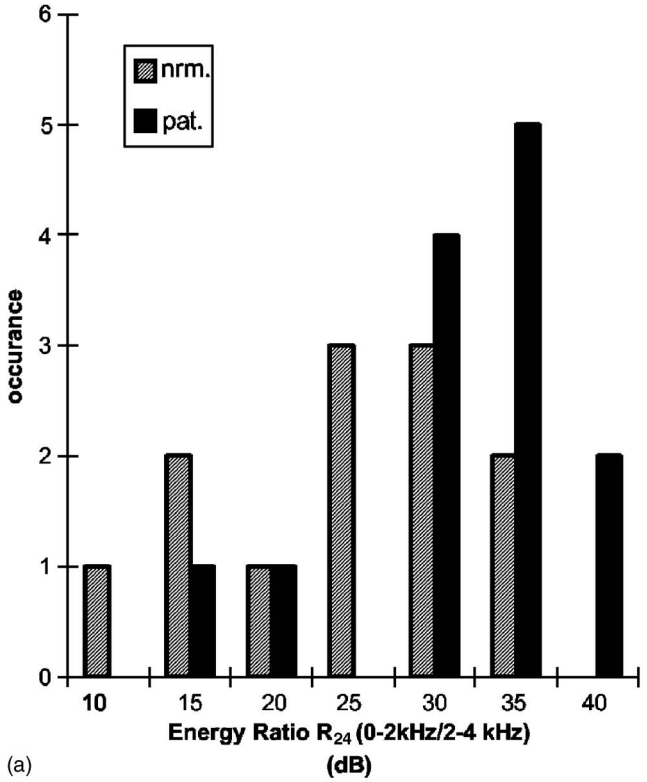


FIG. 6. (Color online) (a) Histogram of HNR (harmonics-to-noise ratio of the speech signal) for the patient/normal data set—not statistically significant at the 5% level (one tailed, two sample, equal variance, student's t-test). (b) Histogram of GHNR' (glottal related harmonics-to-noise ratio) for the patient/normal data set—highly statistically significant at the 5% level (one tailed, two sample, equal variance, student's t-test).



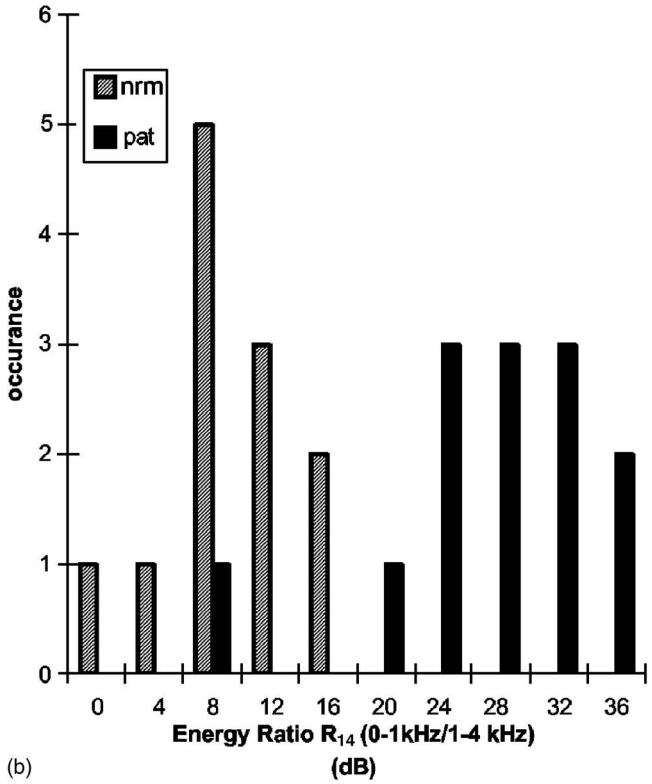


FIG. 7. (Color online) (a) Histogram of R_{24} (speech signal) for the patient/normal data set—highly statistically significant at the 5% level (one tailed, two sample, equal variance, student's t-test). (b) Histogram of R_{14} (speech signal) for the patient/normal data set—not statistically significant at the 5% level (one tailed, two sample, equal variance, student's t-test).

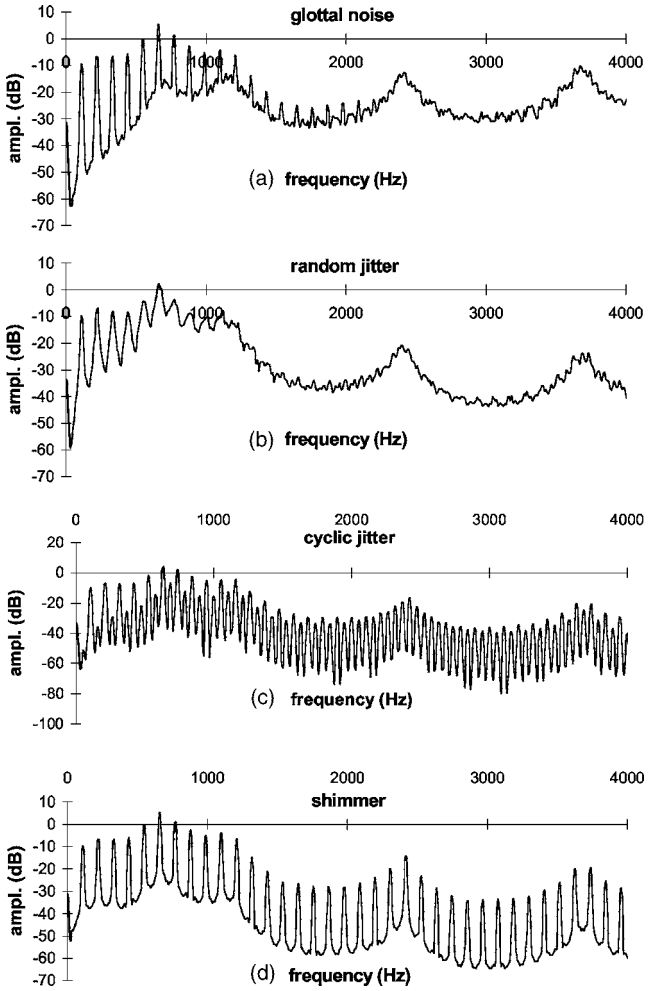


FIG. 8. Averaged spectra for (a) 8% s.d. glottal noise, (b) 4% s.d. random jitter, (c) 4% cyclic jitter and (d) s.d. 8% random shimmer (window length=1024, hop size=512, sampling frequency=10 kHz) illustrating the reduction in spectral tilt wrt noise and the invariance of spectral tilt wrt temporal and amplitude perturbations.

age of the glottal cycle), multiplicative noise (noise increasing with glottal waveform amplitude) and additive noise was found to be similar.

The objective of the current synthesis is to allow for systematic testing of variables and not to specifically attempt to replicate a given voice quality. Hence in the synthesis the glottal waveform maintains its shape with the open quotient remaining fixed as Gaussian white noise is introduced. However, in reality it is expected that the closed phase may be less pronounced in many cases involving voice pathologies. This is due to, for example, the effect of mass lesions such as nodules, polyps, etc. that result in appreciable airflow leakage even when the vocal processes come into contact during the attempted close phase of the cycle (Hanson *et al.*, 2001). Incomplete or reduced closed phase also occurs in hypofunctional and breathy voice or as a result of vocal cord paralysis or paresis. The acoustic effect of this is a reduction in harmonic structure in the higher frequency end of the spectrum, leading to an overall reduction in energy of the upper partials despite having an increase in noise energy in this region (Fig. 9). Also, the lower frequency region may have an increase in the lower frequency components due to the more sinusoidal nature of the glottal flow wave form. The longer open time gives rise to increased damping of the formants. Hanson (1997) has shown that the tilt measure, H1–A1 increases with increased bandwidth of the first formant, B1. The increased open time may also introduce tracheal resonances

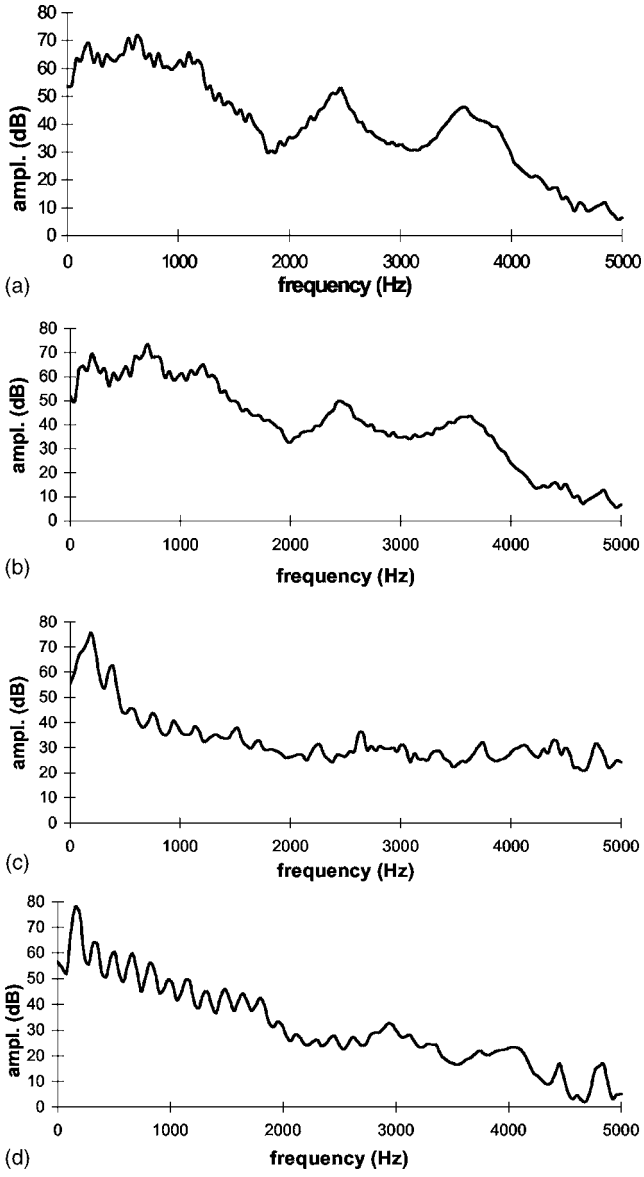


FIG. 9. Averaged spectra for two "normals," (a) and (b) and two patients, (c) and (d), of the present study, illustrating the greater spectral tilt for the patient data (window length 25.6 ms, hop size 10 ms and total analysis length 0.75 s).

and antiresonances (Klatt and Klatt, 1990). The overall result of this is an increase in the R_{14} ratio. The present development and results may help to explain the apparently conflicting results regarding spectral tilt correlations with breathiness and pathological voice as reviewed in Sec. I. Noise may simply appear more prominent at higher frequencies because the harmonic energy is less in this region or conversely a measurement of tilt may be greater when more noise is present but the increase in tilt is due, in fact, to other glottal factors as mentioned earlier. From the theoretical development it can also be inferred that for tilt measures taken from single point estimates (e.g., H1–A3) the values are not expected to change based solely on an increase in noise levels as the noise adds equally throughout the averaged spectrum. Hence these measures can be used to indicate glottal abductory changes even in the presence of noise (assuming the noise is approximately flat throughout the spectrum). However, care must be taken when interpreting a measure such as R_{14} in terms of degree of adduction as the index is sensitive to noise. The effect of the noise can be taken into consideration or alternatively a noise-free R_{14} tilt measure could be implemented. The latter can, in theory at least, be achieved using the method presented in Murphy (1999).

A practical limitation of the spectral tilt measurements is their variation with $f0$ (Fig. 5). The harmonics receive different resonant contributions as the fundamental frequency varies giving rise to a nonlinear variation of tilt with fundamental frequency. In addition, the amplitudes of the harmonics for the low fundamental frequency glottal signals are greatly attenuated by approximately 2 kHz, whereas the harmonic amplitudes for the higher fundamental frequency signals are still prominent at 4 kHz. The number of harmonics within the 0–1 kHz range to the 1–4 kHz range changes with fundamental frequency. Table II illustrates these ratios for the synthesis data files. An alternative tilt ratio could be obtained through use of harmonic numbers (e.g., ratio of the first five harmonic amplitudes to the next ten/twenty harmonic amplitudes). An obvious objection to this approach is that the harmonics receive different resonant contributions if the fundamental frequency is different. A possible solution would be to calculate the ratio between the glottal related harmonics-to-noise ratio from 0 to 1 kHz to the glottal-related harmonics-to-noise ratio from 1 to 4 kHz.

VI. CONCLUSION

A glottal related harmonics-to-noise ratio (GHNR') is shown to be approximately free of the influence of the vocal tract for the vowel /a/ and hence more closely related to the HNR of the glottal source. The synthesis shows that the measure is largely $f0$ insensitive and the measure is also shown to be a superior indicator of pathological voice in comparison with SHNR. Spectral tilt is a potentially useful candidate as a perturbation-free noise estimator in voice signals. Two measures, the ratio of the energy below 1 kHz to the energy above 1 kHz (R_{14}) and the ratio of the energy below 2 kHz

TABLE II. Ratio of number of harmonics from 0–1 kHz to the number of harmonics to 1–4 kHz.

Fundamental frequency / number of harmonics	Vowel /a/ with first three resonances at ~660, ~1100, and ~2400 Hz					
	80	110	160	220	290	350
0–1 kHz	12	9	6	4	3	2
1–4 kHz	38	27	19	14	10	9
Ratio	0.316	0.333	0.316	0.286	0.3	0.22

to the energy above 2 kHz (R_{24}), were considered for analysis. Both the R_{14} and R_{24} ratios decreased with increasing levels of glottal turbulent noise. Both methods were also relatively insensitive to jitter and shimmer. Further, R_{14} has been shown to be a useful indicator of voice pathology. Future work will look at refining the acoustic measures described herein and possibly combining them with complimentary measures with a view to differential categorization of both pathological and normal voice types.

ACKNOWLEDGMENTS

Partial funding for the preset work was supplied through a Heath Research Board grant (No. 01-95) and an Enterprise Ireland Research Innovation Fund Grant No. 2002/037. The authors would like to thank Professor Story and two anonymous reviewers for many helpful suggestions and comments.

Akande, O., and Murphy, P. J. (**2005**). "Estimation of the vocal tract transfer function with application to glottal wave analysis," Speech Commun. **46**, 15–36.

Alku, P., Strik, H., and Vilkman, E. (**1997**). "Parabolic spectral parameter—A new method for quantification of the glottal flow," Speech Commun. **22**, 67–79.

Childers, D. G., and Lee, C. K. (**1990**). "Voice quality factors: Analysis, synthesis, and perception," J. Acoust. Soc. Am. **90**, 2394–2410.

Cox, N. B., Ito, M. R., and Morrison, M. D. (**1989**). "Technical considerations in computation of spectral harmonics-to-noise ratios for sustained vowels," J. Speech Hear. Res. **32**, 203–218.

Doherty, E. T., and Shipp, T. (**1988**). "Tape recorder effects on jitter and shimmer extraction." J. Speech Hear. Res. **31**, 485–490.

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

Fex, B., Fex, S., Shiromoto, O., and Hirano, M. (**1995**). "Acoustic analysis of functional dysphonia: Before and after voice therapy (accent method)," J. Voice **8**, 163–187.

Frokjaer-Jensen, B., and Prytz, S. (**1976**). "Registration of voice quality," Bruel and Kjaer Technical review **3**, 3–17.

Fu, Q., and Murphy, P. J. (**2006**). "Robust glottal source estimation based on joint source-filter model optimization," IEEE Transactions on Audio, Speech and Language Processing **14**, 492–501.

Fukazawa, T., El-Assuooty, A., and Honjo, I. (**1988**). "A new index for the evaluation of the turbulent noise in pathological voice," J. Acoust. Soc. Am. **83**, 1189–1192.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., and Wedin, L. (**1980**). "Perceptual and acoustic correlates of abnormal voice qualities," Acta Oto-Laryngol. **90**, 441–451.

Hammarberg, B., and Gauffin, J. (**1995**). "Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects," *Vocal Fold Physiology: Voice Quality Control*, edited by O. Fujimura and M. Hirano (Singular, San Diego), pp. 283–303.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers; Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Hanson, H. M., and Chuang, E. S. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**, 1064–1077.

Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., and Slifka, J. (**2001**). "Towards models of phonation," J. Phonetics **29**, 451–480.

Hillenbrand, J. (**1987**). "A methodological study of perturbation and additive noise in synthetically generated voice signals," J. Speech Hear. Res. **30** 448–461.

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech Hear. Res. **37**, 769–778.

Kasuya, H., Ogawa, S., Mashima, K., and Ebihara, S. (**1986**). "Normalized noise energy as an acoustic measure to evaluate pathologic voice," J. Acoust. Soc. Am. **80**, 1329–1334.

Kitzing, P. (**1986**). "LTAS criteria pertinent to the measurement of voice quality," J. Phonetics **14**, 477–482.

Kitzing, P., and Akerlund, 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.

Klich, R. J. (**1982**). "Relationships of vowel characteristics to listener ratings of breathiness," J. Speech Hear. Res. **25**, 574–580.

Löfqvist, A. (**1986**). "The long time average spectrum as a tool in voice research," J. Phonetics **14**, 471–475.

Löfqvist, A., and Mandersson, B. (**1987**). "Long-time average spectrum of speech and voice analysis," Folia Phoniatr. **39**, 221–229.

Michaelis, D., Fröhlich, M., and Strube, H. W. (**1998**). "Selection and combination of acoustic features for the description of pathologic voices," J. Acoust. Soc. Am. **103**, 1628–1639.

Michaelis, D., Gramss, T., and Strube, H. W. (**1997**). "Glottal-to-noise excitation ratio—A new measure for describing pathological voices," Acust. Acta Acust. **83**, 700–706.

Milenkovic, P. (**1987**). "Least mean square measures of voice perturbation," J. Speech Hear. Res.," **30**(4), 529–538.

Murphy, P. J. (**1999**). "Perturbation free measurement of the harmonics-to-noise ratio in voice signals using pitch-synchronous harmonic analysis," J. Acoust. Soc. Am. **105**, 2866–2881.

Murphy, P. J. (**2000**). "Spectral characterization of jitter, shimmer and additive noise in synthetically generated voice signals," J. Acoust. Soc. Am. **107**, 978–988.

O'Leidhin, E., and Murphy, P. J. (**2004**). "Acoustic analysis of voiced speech with varying noise and perturbation levels," Proceedings Irish Signals and Systems Conference, Belfast, 47–51.

Prytz, S. (**1977**). "Long time average spectra (LTAS) of normal and pathological voices," Proceedings of the 17th International Congress on on Logopedics Phoniat. Vol. **1**, pp. 458–475.

Qi, Y. (**1992**). "Time normalization in voice analysis," J. Acoust. Soc. Am. **92**, 2569–2576.

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

Rosenberg, A. E. (**1971**). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **84**, 583–588.

Schoentgen, J. (**2006**). "Vocal cues of disordered voices: an overview," Acust. Acta Acust. **92**, 667–680.

Stevens, K. (**2000**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Sundberg, J., and Nordenberg, M. (**2006**). "Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech," J. Acoust. Soc. Am. **120**, 453–457.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).

Titze, I. R., and Winholtz, W. S. (**1993**). "Effect of microphone type and placement on voice perturbation measurements." J. Speech Hear. Res. **36**, 1177–1190.

Valencia-Naranjo, N., Mendoza, E., and Carballo, G. (**1995**). "Diagnostic voice disorders," Current Opinion in Head and Neck Surgery **3**, 164–168.

Welch, P. D. (**1970**). "The use of the fast Fourier transform for the estimation of power spectra," IEEE Trans. Audio Electroacoust. **AU-15**, 70–73.

Wendler, J., Doherty, E., and Hollien, H. (**1980**). "Voice classification by means of long term spectra," Folia Phoniatr. **32**, 51–60.

Wendler, J., Rauhut, A., and Kruger, H. (**1986**). "Classification of voice qualities," J. Phonetics **14**, 477–482.

Development and evaluation of methods for assessing tone production skills in Mandarin-speaking children with cochlear implants

Ning Zhou and Li Xu[a)]
School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701

(Received 1 February 2007; revised 6 November 2007; accepted 18 December 2007)

The aim of the study was (1) to develop methods for evaluating tone production of children with cochlear implants (CIs) who speak Mandarin Chinese and (2) to evaluate the efficacy of using these methods to assess tone production. The subjects included two groups of native-Mandarin-Chinese-speaking children: 14 prelingually deafened children who had received CIs and 61 normal-hearing (NH) children as controls. The acoustic analysis focused on quantification of the degree of differentiation among lexical tones based on tonal ellipses and the overall similarity of tone contours produced by the children with CIs to normative contours derived from the 61 NH children. An artificial neural network was used to recognize tones produced by the children with CIs after trained with tone tokens produced by the NH children. Finally, perceptual judgments on the tone production of both groups were obtained from eight native-Mandarin-speaking NH adults to evaluate the efficacy of the methods. The results showed that all measures using the acoustic, neural-network, and perceptual analyses were highly correlated with each other and could be used to effectively evaluate tone production of children with CIs.
 [DOI: 10.1121/1.2832623]

PACS number(s): 43.70.Ep, 43.64.Me, 43.70.Dn [AL] Pages: 1653–1664

I. INTRODUCTION

The main purpose of this study was twofold. One was to develop methods to assess Mandarin-Chinese tone production in children with cochlear implants (CIs). The second purpose was to validate these methods and evaluate their relative utility in the assessment of tone production in children with CIs.

Mandarin Chinese is a tonal language that uses pitch differences to contrast word meanings. The four tones in Mandarin Chinese are commonly referred to as tones 1, 2, 3, and 4. Tone 1 is a high level tone that has a flat fundamental frequency (F0) contour. Tone 2 has a rising F0 contour. Tone 3 is a falling and rising tone with a dip in the middle of the vowel. Tone 4 is a high falling tone (Duanmu, 2000; Whalen and Xu, 1992). Chao (1948) introduced a nomenclature that delineates the F0 patterns for the four Mandarin-Chinese tones. The four tones are named tone 55, tone 35, tone 214, and tone 51. The numbers reflect the relative levels of the starting and ending points of the F0 contours, or those of the starting, middle, and ending points of the F0 contour in the case of tone 3. Although F0 is the primary acoustic correlate for Mandarin tone perception, other cues such as duration and amplitude contour may also assist tone identification (Whalen and Xu, 1992; Xu *et al.*, 2002).

It is believed that Mandarin-Chinese tone is acquired by typically developing children at a very young age (e.g., Li and Thompson, 1997), probably even earlier than the mastery of segmentals (i.e., vowels and consonants). Research has shown that children with CIs who speak tonal languages have deficits in tone perception (e.g., Wei *et al.*, 2000; Ciocca *et al.*, 2002; Lee *et al.*, 2002; Wong and Wong, 2004) due to the lack of F0 information in the electrical stimulation in current CI technology [see Moore (2003) for a review]. The degree of their deficits in tone perception varied tremendously from individual to individual. Tone perception performance of these children ranged from chance level to as high as 80% correct. Some studies observed postoperative improvement of tone perception to about 65% correct (e.g., Wei *et al.*, 2000), whereas other reports demonstrated that children with CIs could only perceive tones with an accuracy slightly above chance (Lee *et al.*, 2002; Ciocca *et al.*, 2002). With the limited tone exposures that the current CIs provide, prelingually deafened children also experience difficulties in tone production due to the lack of acoustic feedback. Recent studies (Xu *et al.*, 2004; Han *et al.*, 2007; Peng *et al.* 2004) all showed that a majority of the prelingually deafened children with CIs did not master Mandarin-Chinese tones. Moreover, age at implantation and duration of device use appeared to correlate with tone development of children with CIs (Wu and Yang; 2003 Wu *et al.*, 2006; Peng *et al.*, 2004; Han *et al.*, 2007).

The current methods used for evaluating Mandarin tone production involve mainly acoustic analysis and perceptual judgments from adult native speakers. Methods based on acoustic analysis for evaluating Mandarin tone production of either normal hearing (NH) or CI users are still limited. Xu *et al.* (2004), in their preliminary reports of Mandarin tone production in children with CIs, described the CI children's Mandarin-Chinese tone production by presenting the F0 contour of each tone produced by individual children. Wang

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

et al. (2003) studied nonnative speakers' tone production by comparing their F0 contours to an averaged native form. They normalized the duration and height of F0 contours of the speakers to fit their F0 range in levels arranging from 1 to 5 corresponding to the five-point pitch scale for Mandarin tone proposed by Chao (1948). Five evenly spaced F0 values were chosen to compare with the native forms. Analysis of variance was performed to evaluate whether the mean deviation between the native and nonnative production differed across positions on an F0 contour. Although the pitch contours were examined, neither study quantified the acoustic properties of the tone production. In contrast, a more sophisticated method of acoustic analysis was developed by Barry and Blamey (2004) for Cantonese tones. They studied the tonal ellipses generated over the scatter plots of the onset versus offset values of the F0 contours. The spread and degree of overlap among tonal ellipses were quantified to indicate the degree of differentiation among Cantonese tones.

A more commonly used method to the assessment of Mandarin tone, however, has been perceptual judgments. It is widely used to evaluate speech outcome for NH subjects, and especially for CI users. We used a four-alternative forced-choice perceptual test (Xu *et al.*, 2004; Han *et al.*, 2007) in which native adult speakers of Mandarin were presented with audio recordings of the subjects' tone production and were required to identify the tones they heard. Methods using auditory judgment involved measuring the correctness of tone production on equal interval scales (e.g., Peng *et al.*, 2004). The problem with the perceptual studies is the limited information they offer about the property of tone production besides percentage scores.

In summary, the existing methods for evaluating tone production in Mandarin-speaking implant users are limited and are largely restricted to the use of perceptual intelligibility judgments. We aimed at further developing methods for evaluating tone production of Mandarin-speaking CI users and to assess the efficacy of the methods.

We followed the approach of Barry and Blamey (2004) that was used for assessing Cantonese tone production, and adjusted and expanded that approach to accommodate its use for Mandarin Chinese. In the present study, the two indices of differentiation first developed by Barry and Blamey were modified based on the differences of the two languages, the differences being the number of tones as well as their acoustic properties [see Ciocca *et al.* (2002) for a review]. In addition, a new index that allows the examination of the degree of differentiation of an individual tone from the others was proposed.

Although examination of the tonal ellipses can reflect the degree of differentiation among tones, it does not provide a complete description of the Mandarin-Chinese tone contours. The approach used by Wang *et al.* (2003), which attempted to examine the tone contours of nonnative speakers by calculating their deviation from the native forms, involved also only five positions on the contours. We proposed a cross-correlation method that uses a common technique for random signal processing to study the correlation of two contours. It makes possible a direct comparison of F0 contours produced by each child with a CI with the normative contours and yields a correlation coefficient to indicate the degree of similarity between the contours.

In addition to the acoustic analysis, we also developed an artificial neural network to classify the tones produced by CI users. Neural network classifiers are able to emulate perception of human sensory functions. In such neural networks, simple nodes, or neurons are interconnected to form a network, which operates on a principle similar to biological neurons. Training of the neural network enables it to learn by adjusting its parameters and to generalize the acquired knowledge to new situations like humans do. A neural network bases its working mechanism on a defined mathematical algorithm and classifies tones into categories much like how perception works. Several studies have used artificial neural networks to classify Mandarin or Cantonese tones (Chang *et al.*, 1990; Wang and Chen, 1994; Lee *et al.*, 1995; Lan *et al.*, 2004; Xu *et al.*, 2006, 2007; Zhou *et al.*, 2007). In our recent study, a feedforward multilayer perceptron was used to recognize the tones produced by a group of Mandarin-speaking children with NH (Xu *et al.*, 2007). In the study, it was found that, in terms of recognition accuracy, performance of the multilayer perceptron was comparable to that of human listeners and was even superior to human ears when the classification task involved a large number of speakers and tokens (Xu *et al.*, 2007). The rationale for using a neural network in addition to the acoustic analysis in the present study was that the neural network could not only produce human perception-like performance, but also operates automatically and economically.

To examine the relative utility of the proposed acoustic analysis and neural-network approaches, a traditional tone perception task was carried out using native-Mandarin-speaking listeners. Further, a correlational analysis was conducted for each pair of the measures including the perceptual results. Information that these measures provide and their strength and limitations were further discussed.

Considering the relative independence of each approach, we describe and discuss the tonal ellipse analysis, contour correlation analysis, neural network, and perception tests separately in Secs. II–V. General discussions are provided in Sec. VI.

II. ACOUSTIC ANALYSIS BASED ON TONAL ELLIPSES

A. Methodology

1. Subjects and recording procedures

Two groups of native Mandarin-Chinese-speaking children were recruited. One group consisted of 14 prelingually deafened children who had received CIs from the Cochlear Implant Center of Beijing Tongren Hospital, China. There were no specific criteria for recruitment in terms of age, device experience and type, or other contributing factors of tone production. Our sampling provided cross-section data on tone development. Their chronological ages ranged from 2.9 to 8.3 years old (5.2 ± 1.8, mean $\pm$ SD). Five of the fourteen children used hearing aids before the implantation. The implants that they had received included Clarion CII,

TABLE I. Demographic information of the CI subjects. The order of the subject number is consistent with that in Fig. 1. The order follows the perception scores by the adult listeners on tone production of individual children with CIs, from high to low.

Subject number	Gender	Age at implantation (years)	Chronological age (years)	Duration of implant use (years)	CI device
1	F	3.10	5.51	2.41	N24M
2	M	1.16	2.93	1.77	Clarion CII
3	M	1.50	3.60	2.10	N24M
4	F	2.05	4.27	2.22	N24M
5	M	1.32	3.38	2.06	N24M
6	M	5.73	8.33	2.60	N24M
7	M	3.90	5.50	1.60	Clarion CII
8	M	2.60	2.91	0.30	N24R
9	F	5.19	7.53	2.34	Clarion CII
10	F	2.59	4.41	1.82	N24M
11	F	5.56	6.53	0.97	N24R
12	F	4.55	5.35	0.79	N24R
13	M	1.70	4.27	2.57	N24M
14	M	7.09	7.67	0.58	N24R

N24M, and N24R. The duration of CI use ranged from 0.3 to 2.6 years (1.7 ± 0.8, mean $\pm$ SD). All of them had received rehabilitation at professional rehabilitation centers in Beijing after implantation with the amount of rehabilitation varied. Their demographic information is summarized in Table I. The control group consisted of 61 NH children from kindergartens and elementary schools in Beijing. The ages of the NH children ranged from 3.1 to 9.0 years old (6.2 ± 1.7, mean $\pm$ SD). All the 61 NH children had a pure tone average (at 500, 1000, and 2000 Hz) $\leq$20 dB HL.

To collect their tone production samples, both groups of children were instructed to produce tones 1–4 for the following monosyllables: *ai*, *bao*, *bi*, *can*, *chi*, *du*, *duo*, *fa*, *fu*, *ge*, *hu*, *ji*, *jie*, *ke*, *la*, *ma*, *na*, *pao*, *pi*, *qi*, *qie*, *shi*, *tu*, *tuo*, *wan*, *wen*, *wu*, *xian*, *xu*, *ya*, *yan*, *yang*, *yao*, *yi*, *ying*, *you*, *yu*, *yuan*, *zan*, *zhi*. These syllables were chosen because when associated with the four tones, they all produce real words in Mandarin Chinese. The children heard each of the syllables in its high flat tone, and were asked to produce the syllables in four tones. The children's familiarity with the words was not a concern because language education in kindergartens and elementary schools in Beijing greatly emphasizes tone drills. The tone drills require producing four tones for any syllables even though the children might not know the words. The elicited production of the monosyllabic words was digitally recorded at a sampling rate of 44.1 kHz with a 16-bit resolution in quiet rooms with ambient noise typically round 40 dB SPL. In total, 9760 tone tokens were obtained (40 syllables $\times$ 4 tones $\times$ 61 speakers) from the NH control group, and 2240 tone tokens were obtained (40 syllables $\times$ 4 tones $\times$ 14 speakers) from the CI group. The mean and standard deviation of the durations of all the 9760 and 2240 tokens from the NH and CI groups were 511.31 ± 30.18 and 610.50 ± 35.26 ms, which were significantly different [$t(73) = 10.75$, $p < 0.001$]. It was evident that the typically developing NH children sometimes produced inaccurate tones. One legitimate approach would be to exclude those inaccurate tone tokens from the normative data. However, we regarded the errors as part of the development process in the typically developing children. Therefore, all the tone tokens from the NH children were included as normative data in the analysis.

2. F0 extraction and F0 onset–offset plotting

The F0 contours of the vowel part of the monosyllabic words were extracted using an autocorrelation algorithm. Although there are other cues that contribute to Mandarin tone perception, the focus of this research was F0, as the major problem with tone development in CI users has been the unresolved F0. The update rate of the F0 extraction was 8 ms with a frame size of 24 ms. The lower and upper boundaries for the extraction were 50 and 500 Hz. The extracted F0 contours were plotted onto and compared with the narrowband spectrograms (window size = 13.61 ms) generated from MATLAB (MathWorks, Natick, MA) to verify the accuracy of F0 extraction. The extraction might sometimes include a small part of the preceding voiced consonant, but the voicing of the consonant usually does not produce reliable F0 (Duanmu, 2000). The unreliable F0 data points from the voiced consonant could be easily identified and manually deleted based on the narrowband spectrograms. Occasional errors of halving or doubling were also manually corrected with reference to the narrowband spectrograms.

The offset values of the F0 contours were plotted against the onset values for each tone token. To summarize the distribution of the onset–offset plots for each child, tonal ellipses around each tone type were plotted in a manner similar to that introduced by Barry and Blamey (2004). First, the inclination of an ellipse (i.e., the direction of the semi-major axis of the ellipse) was determined by the positive angle of the linear fit of the data points of a particular tone. The ellipse center was located using the mean of the F0 onset values along the fitting line. A perpendicular line to the fitting line defined the direction of the semiminor axis. The lengths of the semimajor and the semiminor axes were determined by two standard deviations of all data points away from the center along the linear fitting line and the perpendicular line, respectively. Thus, each tonal ellipse encompassed 95% of the onset–offset data points of a tone category. The plot of the F0 onset and offset values capture the extreme values of the F0 contours of a Cantonese speaker. Barry and Blamey (2004) defined such a plot as the tonal space. In Mandarin Chinese, the plot of the F0 onset and offset values should also capture most of the extreme values of the F0 contours except sometimes for tone 3. Thus, we regarded such a plot as approximating the tonal space in Mandarin Chinese.

3. Definition of the three indices

The four tonal ellipses that outlined the F0 onset–offset data points were used for studying the differentiability of the tones. Following Barry and Blamey (2004), Index 1 was defined by the ratio of two areas, i.e., *At* over $\overline{Ae}$. *At* was the area of the triangle formed by joining the centers of the three most differentiated tonal ellipses in Cantonese, which was

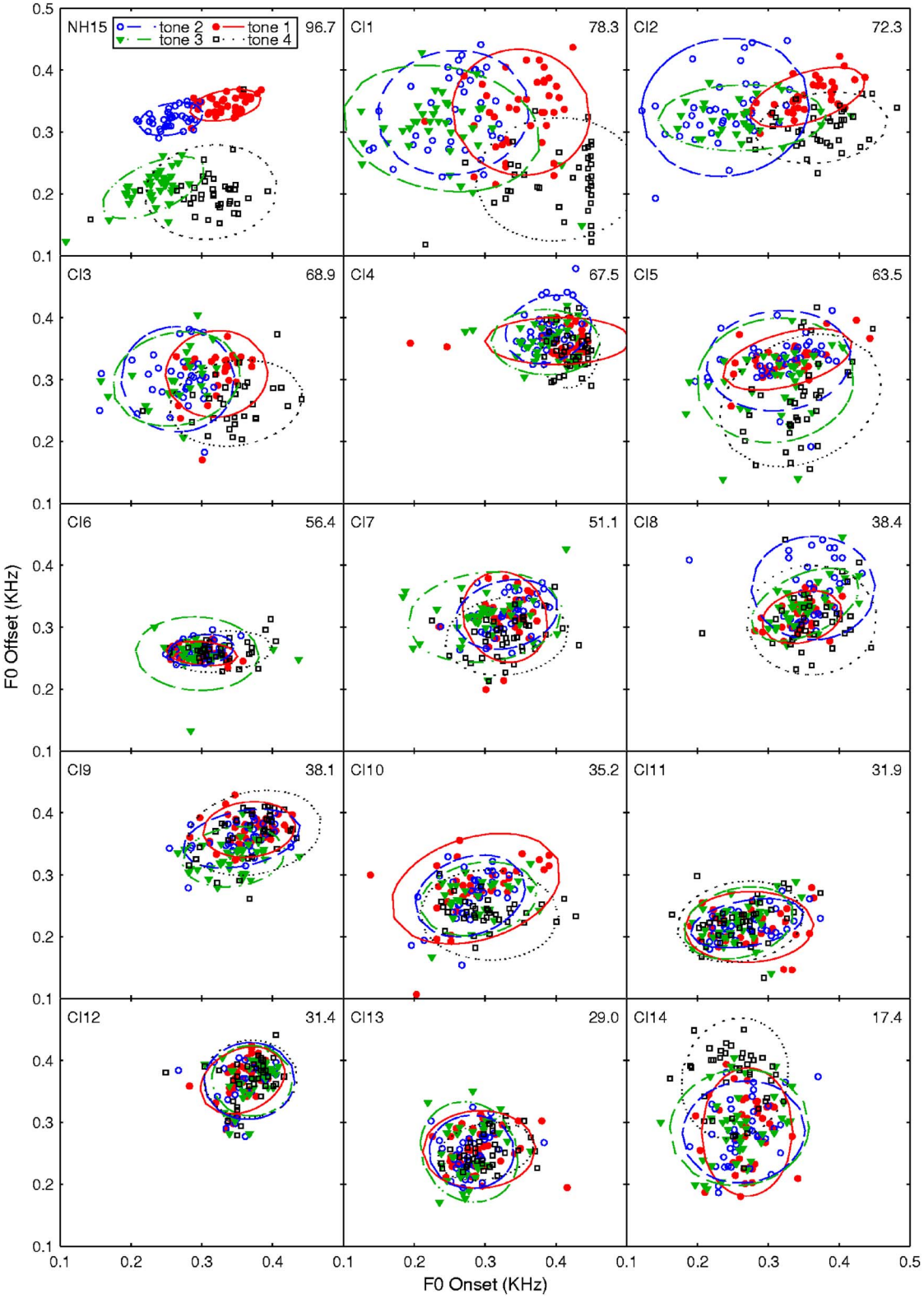


FIG. 1. (Color online) Tonal ellipses of one NH child (subject NH15) and the 14 children with CIs. The data from the representative NH child is shown in the top left-hand panel. The remaining panels are organized in the order of tone production accuracy scores provided by human listeners (see Sec. V) for each CI child. The tone production accuracy score (percent correct) is shown in the upper right-hand corner of each panel. Each data point represents a pair of F0 onset–offset value of a monosyllabic word. Different symbols represent different tones as indicated by the legend in the top left-hand panel.

referred to as the tonal area. $\overline{Ae}$ was the averaged area of these three ellipses. Based on the lengths of the semiminor (a) and semimajor (b) axes of the ellipse, an ellipse area was calculated using $a \times b \times \pi$.

The reason that only three tonal ellipses were chosen for analysis was that Cantonese has six tones and their tonal ellipses are relatively crowded in the tonal space. In contrast, the tonal ellipses for Mandarin Chinese tones are more sepa-

rable in that each of the four tonal ellipses is located in one of the four quadrants of the tonal space. Specifically, the flat tone 1 with an F0 onset–offset pair that is high and similar in level, is located in the upper right-hand quadrant; the rising tone 2 with a low onset and high offset is in the upper left-hand quadrant; the low dipping tone 3 with both median onset and offset is in the lower left-hand quadrant; and the high falling tone 4 with the high onset and low offset is in the lower right-hand quadrant (see Fig. 1, top left-hand panel). Thus, connecting the centers of the four tonal ellipses, which forms a quadrangle, measures the span of the F0 range for Mandarin Chinese tones. Index 1 of Barry and Blamey (2004) was then modified for Mandarin Chinese using:

$$\text{Index } 1 = \frac{Aq}{\overline{Ae}}, \tag{1}$$

where Aq is the area of the quadrangle and $\overline{Ae}$ is the averaged area of the four ellipses.

Index 2, also proposed by Barry and Blamey (2004), was the ratio of the averaged lengths of the two axes for six Cantonese tonal ellipses and the averaged distance of the centers of the six tonal ellipses from each other. In the present study, this index was adopted for the four Mandarin Chinese tones:

$$\text{Index } 2 = \frac{\text{Ave Dist}}{\text{Ave Ax}_{1+2}}. \tag{2}$$

Because Index 1 and Index 2 used the parameters of the four tonal ellipses to measure the overall differentiability of the four tones, these indices did not measure the separability of a particular tone from the other three. A new index was developed to examine the false detections of any individual tone. This index enabled us to infer the degree of differentiation of that tone from the others. As a result of the overlap of tonal ellipses, a given tonal ellipse might enclose not only data points from that tone category but also data points from the other tone categories. Data points from the target tone were considered hits, whereas the data points from all other three tones were categorized as false detections. To determine whether a data point was enclosed by a given tonal ellipse, the position of the data point relative to the ellipse was evaluated using the ellipse function (i.e., $x^2/a^2+y^2/b^2-1=0$, where a and b are the semimajor and semiminor of the ellipse, and x and y are the coordinates of a data point). The coordinates of the data point were first adjusted using the ellipse center as the origin of the axes. If $x^2/a^2+y^2/b^2-1>0$, then the data point was outside the ellipse. If $x^2/a^2+y^2/b^2-1<0$, then the data point was counted to be within the ellipse. For all the data points that were in a particular ellipse, they were further categorized as hits if they were identified to be from the target tone, or false detections if they were from other tones. The false detection probability (P_{fd}) of the target tone was defined as the probability of other tones being falsely recognized as the target tone. It was calculated as the number of false detections over the total number of data points from the other three tones. Index 3 was thus defined as

$$\text{Index } 3 = 1 - (\text{Ave } P_{fd}). \tag{3}$$

B. Results and discussion

Tonal ellipses were generated for each tone for individual children. The tonal ellipses of one typical NH child (NH15, 6.2 years old) and the 14 children with CIs are shown in Fig. 1. The tonal ellipses of the NH child demonstrated an easily separable pattern, with four tonal ellipses being located in the four quadrants of the child's tonal space. The ellipses of tones 1 and 2 were completely separable from those of tones 3 and 4. Although the ellipses of certain tone pairs overlapped, they were still spatially separable. The data points for tones 1 and 2 were relatively densely clustered and restricted to a location in the tonal space. As a result, the sizes of these tonal ellipses were small. The small size of the ellipses suggests small variances in these data points, which in turn indicates that the child had consistent use of F0 range for tone 1 and tone 2. Such an easily separable pattern of ellipses was compromised to various degrees in the plots for the children with CIs. For some children with CIs, their data points scattered from the center of the ellipses, leading to larger sizes of the tonal ellipses. This suggests that they lacked a consistent use of F0 range to produce a certain tone. For those children with CIs whose tonal ellipses were comparatively small, the centers of their ellipses, however, shifted toward each other from the locations observed in the NH children. Consequently, the tonal ellipses largely overlapped with each other. The loss of contrast among tones in their F0 onset and offset characteristics could lead to the overlapping ellipses, thus a decreased degree of differentiation.

Two components of Index 1 quantify the size and separation of the tonal ellipses. Recall that $\overline{Ae}$ is the averaged area of the four tonal ellipses. $\overline{Ae}$ of the NH children did not differ from that of the children with CIs (Wilcoxon rank sum test, $z=1.54$, $p=0.1228$), which suggests that the two groups had comparable variability in F0 use for individual tones. Interestingly, $\overline{Ae}$ of the NH children showed a nonlinear decrease with their ages (Fig. 2), a trend of developing toward a smaller variation in F0 use for individual tones. The development appeared to saturate at around 6 years of age. Data analysis revealed that it was mainly the ellipses of tones 3 and 4 that decreased in size with age. It is worth noting that Barry and Blamey (2004) found the tonal ellipse size of their NH children was even larger than that of the CI group. This discrepancy might be due to a younger age of their NH children group (all $\leq$6 years old). The majority of their group might still be in the process of developing tone normalization skills (Barry and Blamey, 2004). There was no correlation found between ellipse size and chronological age or duration of implant use for the CI group ($p>0.05$). It is possible though that the children with CIs may need more device experience to develop tone production skills. Although the F0 use of individual tones used by the NH children did not show a more confined pattern than the children with CIs, Aq (i.e., the tonal area that connects the centers of the four ellipses) for the NH children was in average three times larger than that of the CI group and this difference was statistically

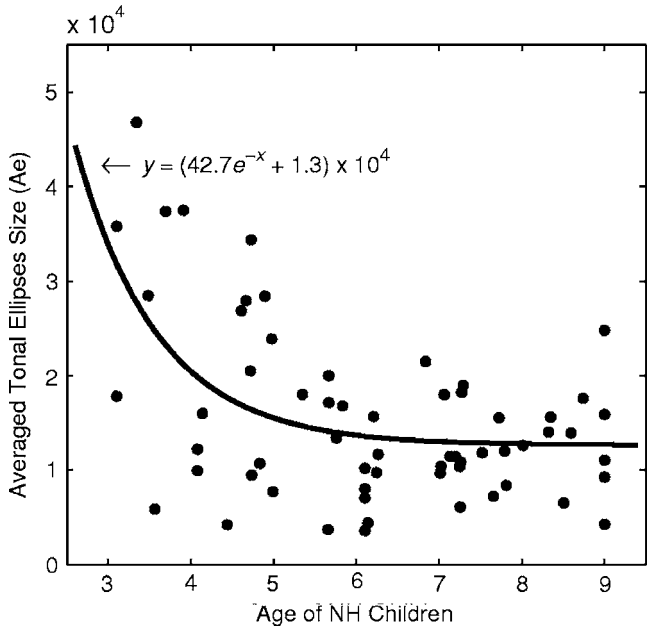


FIG. 2. Averaged tonal ellipses size ($\overline{Ae}$) as a function of age of the NH children. An exponential fitting is plotted with the solid line. Each symbol represents one NH child.

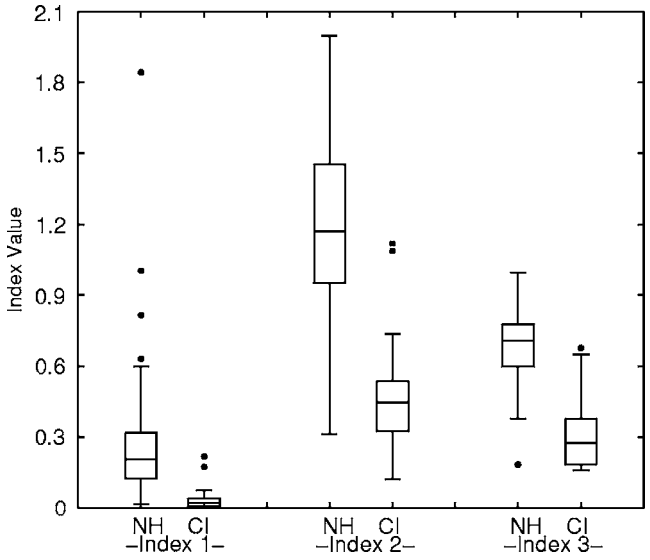


FIG. 3. Box plots of the values of three indices (from left to right) for the NH and CI groups. Each box depicts the lower quartile (Q1), median, and upper quartile (Q3). The whiskers show the range of the rest of the data. The outliers plotted with filled circles are data points that fall more than 1.5 box-lengths (Q3–Q1) away from the lower or upper quartile.

significant (Wilcoxon rank sum test, $z=4.11$, $p<0.0001$). The broader tonal area of the NH children compensated for their diffused tonal ellipses such that the tonal ellipses were still differentiated from each other. In contrast, the equally diffuse tonal ellipses coupled with the small tonal area of the children with CIs resulted in their poorly differentiable tones. The shrunken tonal area of the children with CIs was also observed in the study by Barry and Blamey (2004) of native Cantonese speakers.

For all the three indices, the greater the value, the more differentiated the tones were. Values of the three indices obtained from the 14 children with CIs were compared with those of the 61 NH children (Fig. 3). The NH data were grouped together to be compared with those of the CI group, because although there was age differences in pitch range use (i.e., $\overline{Ae}$) by the NH children, their Aq and scores of Index 1 were not found to correlate with age ($p>0.05$). Besides, the number of children with CIs was too small to be further divided into age or hearing age subgroups. The results of the nonparametric two-sample tests showed that the two groups differed on all three index comparisons (Wilcoxon rank sum test, Index 1: $z=-4.78$, $p<0.0001$; Index 2: $z=-4.96$, $p<0.0001$; Index 3: $z=-4.98$, $p<0.0001$). The tones produced by the CI group were significantly less differentiable than those produced by the NH group as indicated by the index comparisons.

The mean value of the probabilities of false detection (i.e., P_{fd}) for each tone category was obtained for both groups of children. For the CI group, the probabilities of other tones being falsely recognized as tones 1–4 were 0.644, 0.711, 0.767, and 0.725, respectively. The corresponding averaged values of P_{fd} for tones 1–4 for the NH children were 0.228, 0.249, 0.406, and 0.389, respectively. Results of t test confirmed that P_{fd} for the NH children was significantly lower than that for the children with CIs on all tone comparisons [tone 1: $t(73)=-6.55$, tone 2: $t(73)=-7.89$, tone 3: $t(73)=-3.83$, tone 4: $t(73)=-3.82$, all $p<0.001$].

Indices 1 and 2 from Barry and Blamey (2004) were modified to measure the degree of differentiation for Mandarin Chinese tones. One limitation of Index 1 for Cantonese was that the overall differentiability of the six tones was evaluated based on the measures for only three of them. Our Index 1 evaluated all four tones in Mandarin Chinese to obtain their overall differentiation, thus the evaluation might be more accurate than what Index 1 for Cantonese provides. Another limitation of Index 1 for Cantonese as pointed out by Barry and Blamey was that the centers of the three tonal ellipses that defined the triangle could fall onto a line, which resulted in a zero tonal area. With four tonal ellipses, Index 1 for Mandarin Chinese used a quadrangle to represent the tonal area, thus greatly reducing the probability of having a zero area.

From the point of view of Signal Detection Theory (Green and Swets, 1966), Indices 1 and 2 are essentially analogous to the discriminability index d' (Barry and Blamey, 2004). The d' describes how differentiable a signal is from noise by evaluating the separation between the signal-plus-noise distribution and the noise-alone distribution (Green and Swets, 1966). The d' is defined by dividing the difference between means of the two distributions by their variances. For Index 1, the tonal area, (i.e., Aq) represents the separation among the distributions, whereas the averaged area of the ellipses (i.e., $\overline{Ae}$) reflects the pooled variance of the distributions. For Index 2, the averaged distances between any two ellipse centers (i.e., Ave Dist.) likewise represents the separation between the distributions and the averaged lengths of the semiminor and semimajor axes (i.e., Ave Ax_{1+2}) reflect the pooled variance of the distributions. Suppose only two distributions were examined instead of four (i.e., four tonal ellipses), or only one instead of two variables (i.e., onset and offset) were studied, the two indices

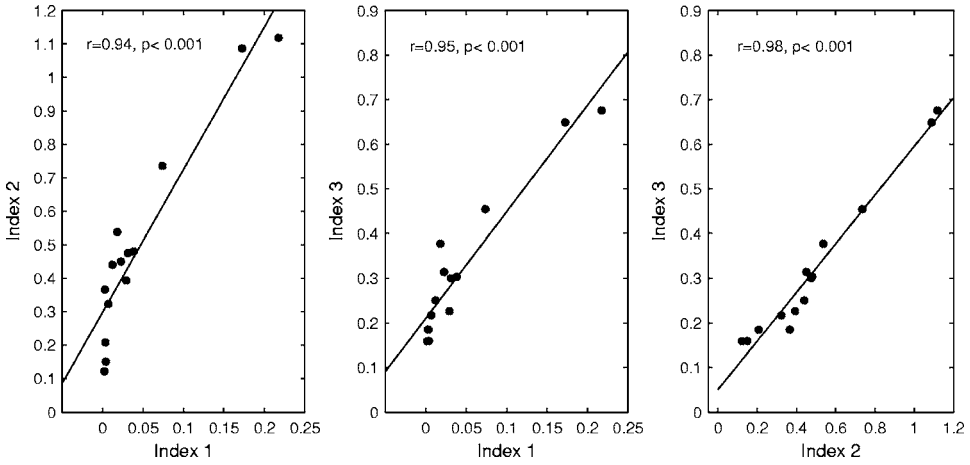


FIG. 4. Correlation between each pair of the three indices. Each data point represents one child with a CI.

would be equivalent to the d' described in the classic Signal Detection Theory (Green and Swets, 1966). In essence, Indices 1 and 2 measured the overall separability of the four tones based on the two variables.

Index 3 provided a direct description of the overlap of tonal ellipses in terms of an averaged false detection probability of four tones. The main strength of this measure is that such description (P_{fd}) can be made for each individual tone, as opposed to an overall description of the differentiation of tones as provided by Indices 1 and 2. Further, the ellipse size used for determining the counts of false detections could be optimized according to a desired level of detection and false alarm probabilities. Such optimization can be performed separately for individual tones, thus allowing greater flexibility in adjusting the threshold of detection.

In order to evaluate the relationship between the indices, the correlation coefficients between each pair of the indices were computed. Figure 4 shows the scatter plots of pairs of indices obtained from the 14 children with CIs. Each pair was highly correlated and the correlations were all significant after controlling for family-wise type I error using Bonferroni correction [$r_{12}=0.94$, $r_{13}=0.95$, $r_{23}=0.98$, $p<0.0167$]. The high correlation between each pair of the indices indicates that they are consistent with each other in evaluating tone differentiation of the CI children's production.

III. ACOUSTIC ANALYSIS OF F0 CONTOURS

A. Methodology

1. Standard F0 contour from NH children

The F0 contours of the Cantonese tones are either flat, rising, or falling. The use of the onset–offset pair for analyzing Cantonese tones is justified because the two endpoints can catch most of the F0 contour information of Cantonese tones such as direction and level. As shown in the previous section, the use of the onset–offset endpoints is also effective in differentiating Mandarin Chinese tones because the tonal ellipses have restricted locations in the tonal space. However, this method does not characterize the F0 contours for Mandarin Chinese tones that have complex contours such as tone 3. That is because these tones do not have simple rising or falling patterns, the direction and slope of which cannot be characterized by the two endpoints of the contours. This section explores a way to characterize a complex F0 contour by comparing it to a normative standard. The approach is based on cross correlation, a measure of similarity between two signals.

The normative contour for each of the four tones was obtained from the 61 NH children by averaging all the contours for each tone produced by these children. The 9760 F0 contours extracted from the 61 NH children were uneven in duration thus different in sampling numbers. To obtain an averaged value at a given time, normalization of F0 contour duration and interpolation of F0 samples were performed. The mean duration of the contours for each tone category was calculated and the longest among the four (i.e., 334 ms) was chosen as the target duration for normalization. Then, all the tone tokens were normalized to the target duration. After the normalization of duration, that is, stretching or squeezing the contours to the same length, the F0 samples on all the contours became no longer aligned on the time axis. In order to average them at any given point on the time axis, linear interpolation, or resampling of the F0 contours was performed. The interpolation rate for a tone category was determined by the number of samples greatest in that tone category divided by the normalization duration. After normalization and interpolation, the contours with time-wise matched F0 samples were then averaged for each tone to generate four normative contours (Fig. 5, upper panel). The normative contours of the four tones were then used as the standard for comparisons with the F0 contours produced by each CI child.

2. Comparing F0 contours using cross correlation

To compare the contours produced by the CI children with the normative ones, four averaged contours for each CI child were generated using the same methods of normalization and linear interpolation as described earlier. The normalization duration and interpolation rate were matched with those used for generating the normative contours. This was to ensure the same length and same number of samples for any pairs of contour comparison. Because the contours of children with CIs and the normative contours may differ in F0 levels, they were all normalized in level by forcing the mean F0 sample values of the contours to zero.

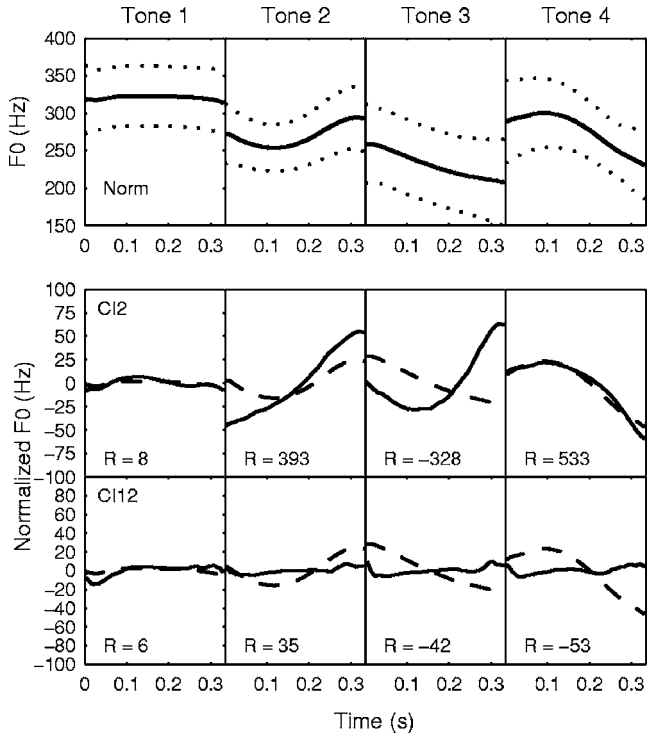


FIG. 5. (Top) Normative contours of the four tones plotted from left to right in solid lines. The contours were averaged from the 61 NH children. The dotted lines represent one standard deviation from the mean. (Middle and bottom): Mean F0 contours of two children with CIs (CI2 and CI12) plotted onto the normative contours for tones 1–4. The normative contours are in dashed lines and the mean F0 contours of the children with CIs are plotted in solid lines. The *R* values of the contour correlation are shown for each comparison. Both duration of the tones and the F0 level of all the contours were normalized.

The overall similarity in the patterns of the F0 contours produced by children with CIs to the normative ones was evaluated by cross correlation with a zero-time shift. Cross correlation is commonly used to find features in an unknown signal by comparing it to a known one. It involves shifting the unknown signal in time and multiplying it by the known signal, thus it is a function of the relative time between the two signals (Couch, 1996). In our case, however, the similarity between two F0 contours was measurable only when they were aligned in time. Therefore, we were only interested in the output of cross-correlation function at zero-time shift, which was thereafter referred to as contour correlation and was calculated using

$$R(0) = \frac{1}{N}\sum_{n=1}^{N} F0_T(n) \times F0_S(n), \tag{4}$$

where $F0_T$ stands for the normalized F0 contour of a child, $F0_S$ stands for the normative F0 contour, and N denotes the total number of samples.

Note that $F0_T$ and $F0_S$ are two zero-mean time sequences. If the two contours show opposite patterns, multiplication of the two [i.e., $F0_T \times F0_S$] will result in a negative *R* value. If a contour deviates randomly from the normative contour, the result of the multiplication [i.e., $F0_T \times F0_S$] will have similar properties to a zero-mean random sequence, which in turn causes the *R* value to approximate zero. On the other hand, if $F0_T$ and $F0_S$ show similar patterns, multiplication of the two will result in a positive *R* value. Note that for the previously described multiplication to work, the two sequences or contours must be normalized to have zero mean. The F0 height normalization procedure developed by Rose (1987, 1993) transforms F0 values to *Z* scores based on the overall mean F0 value of a particular speaker. Rose's method reduces between-subject variations and clusters their F0 contours to better summarize the acoustic characteristics of a given tonal language. This normalization method however does not produce zero-mean F0 contours thus does not suit our purpose here.

B. Results and discussion

The overall similarity of each child's contours to the normative contours was examined. Two examples of the comparisons were plotted in Fig. 5 (lower panels), where the contours from two children with CIs (CI2 and CI12) were plotted over the normative contours. The contour of tone 4 of subject CI2 followed the normative contour closely and the calculation of *R* yielded a large positive value. In contrast, the contour of tone 3 of the child and the normative contour had a quite different pattern, thus the *R* value became negative. The contours of all tones produced by subject CI12 were essentially flat. Consequently, the *R* values for all four tones approximated zero.

Contour correlation was very sensitive to comparing two F0 contours that had noticeable F0 change with time. These include contours of tones 2, 3 and 4. *R* value can clearly indicate the similarity in the patterns of these tones. An example can be seen from the comparison of tone 4 for subject CI2 (Fig. 5). On the other hand, this approach did not show comparable sensitivity to comparing tone 1. That is because F0 values of tone 1 fluctuate only slightly around a constant level. The similarity in patterns of tone 1 contours cannot be adequately evaluated by the contour correlation, because multiplication of two near-zero vectors produces *R* value approaching zero (Fig. 5).

Although the *R* values could not precisely estimate the similarities between the contours of tone 1, an averaged *R* value across tones could still successfully reflect the quality of tone production of a CI child. To compare whether the contours produced by the NH children had patterns more similar to the normative contours than that of children with CIs, the contour correlation was also calculated individually for the NH children. The averaged *R* values of children with CIs ranged from −240.5 to 388.9, with a median of 120.1. The *R* values of the NH children ranged from 92.3 to 397.5 with a median of 249.2 (Fig. 6). The two distributions were easily differentiable. The *R* values from the NH were all positive, whereas that of children with CIs were spread across zero. The results of the two-sample independent *t* test confirmed that such a difference between the two groups was statistically significant [$t(73)=5.49$, $p<0.001$]. The significant difference suggested that the use of contour correlation was effective in detecting the differences between the two groups in terms of the overall similarity of their F0 contours to a normative standard. The *R* values for the CI group av-

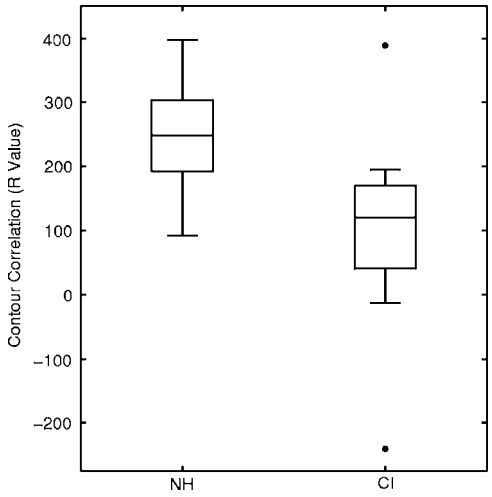


FIG. 6. Box plots of the contour correlation (R values) for the NH and CI groups. Each box depicts the lower quartile (Q1), median, and upper quartile (Q3). The whiskers show the range of the rest of the data. The outliers plotted with filled circles are data points that fall more than 1.5 box-lengths (Q3–Q1) away from the lower or upper quartile.

eraged for each tone were 5.2, 83.7, −70.9, and 394.2, respectively. Although the result of the contour correlation was not particularly indicative for the quality of production of tone 1, tone 4 was otherwise singled out as the best produced tone by the CI children.

IV. NEURAL NETWORK ANALYSIS OF F0 CONTOURS

A. Methodology

1. Neural network and its structure

A feedforward backpropagation multilayer perceptron (MLP) was implemented in MATLAB with the Neural Network Toolbox and used to recognize the tone production of the children with CIs. The neural network had three layers, i.e., an input layer, a hidden layer, and an output layer. Inputs to the MLP were F0 contours. Based on our previous studies on optimization of neural-network configurations for Mandarin Chinese tone recognition (Xu *et al.*, 2006, 2007; Zhou *et al.*, 2008), the number of inputs was set at 12. The F0 contour was evenly segmented to 12 parts, and the averaged F0 value for each part made up the inputs to the neural network. The number of neurons in the hidden layer was set at 16. The output layer of the neural network consisted of four neurons representing the four Mandarin Chinese tones. The Levenberg–Marquardt optimization was used as the training algorithm (Hagan and Menhaj, 1994). Our previous study (Xu *et al.*, 2007) indicated that this neural network was highly tolerant to between or within F0 variation in children speakers. Therefore, no normalization procedure was necessary for adjusting the height of input F0s.

2. Evaluation of tone production of children with CIs with the neural network

The neural network was trained with all the tone tokens from the 61 NH children. Training was stopped when the sum of squared errors became less than 0.01. Given the large amount of training data compared to the testing data, to avoid overfitting, the number of iterations for the training was set at 50. Overfitting refers to the situation in which the neural network starts to learn patterns in the noise signal due to too many training iterations or too many hidden neurons used, which results in poor performance in real test conditions.

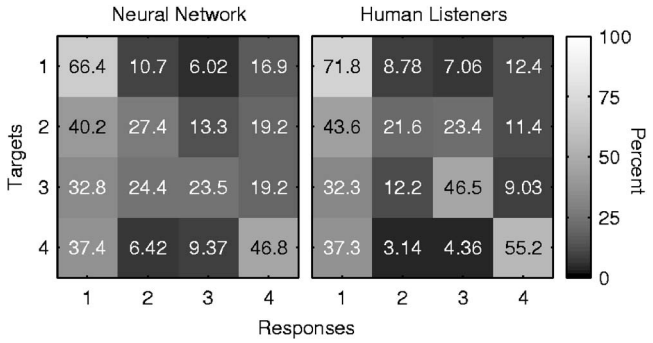


FIG. 7. The tone recognition error patterns by the neural network (left) and by human listeners (right). The value in the cell of row j and column k is the percent of times stimulus tone j was recognized as tone k ($j=1,2,3$, or 4; $k=1,2,3$, or 4). The gray scale in each cell reflects the value in it with reference to the color bar on the right.

The tone tokens of each CI child were tested with the neural network upon the completion of training with the tone tokens from the NH children. Half of the tone tokens from each CI child were tested 10 times with different randomization of inputs. A tone confusion matrix was used to describe the error patterns by the CI group.

B. Results and discussion

The mean recognition rate for each CI child was obtained. The recognition scores ranged from 13.5% to 69.6% correct (41.1% ± 15.8% correct). The averaged recognition scores for tones 1–4 by the children with CIs were 66.4%, 27.4%, 23.5%, and 46.8% correct, respectively. Note that the level tone 1 was more accurately recognized than any of the contour tones, and the falling tone 4 was better recognized than the rising tone 2. The confusion matrix revealed that 30%–40% of the time, the contour tones (i.e., tones 2, 3, and 4) produced by the children with CIs were recognized as tone 1 (Fig. 7). The same neural network was used for recognizing the tones of the 61 NH children in our recent work (Xu *et al.*, 2007). The corresponding recognition scores for the four tones by the NH children were 91.3%, 88.5%, 71.7%, and 83.9% correct, respectively. A group of t tests indicated that the recognition scores for the NH children were significantly higher than those of children with CIs for all tones [tone 1: $t(73)=5.38$, tone 2: $t(73)=13.87$, tone 3: $t(73)=7.95$, tone 4: $t(73)=6.28$, all $p<0.001$].

V. PERCEPTUAL STUDY

A. Methodology

Eight NH adult native-Mandarin-Chinese speakers were recruited from Ohio University for the tone perception tests. The group included seven female and one male adult listeners, with ages ranging from 26 to 43 years. A hearing test

was performed for each adult listener to confirm that his/her pure tone thresholds were ≤20 dB HL at octave frequencies from 250 to 8000 Hz.

A custom graphical user interface was developed in MATLAB to present the tone tokens and to collect the listeners' responses. The tone perception tests were done in an IAC double-walled sound booth. All the tone tokens (i.e., 12 320=9760+2560) from both NH and CI groups were randomized and half of them (i.e., 6160 tokens) were presented at the most comfortable level to the listeners via a circumaural headphone (Sennheiser, HD 265). The listeners were instructed to use a computer mouse to click on a button labeled 1, 2, 3, or 4 after each presentation of the stimulus to indicate the tone of the speech token that they had heard. Clicking on any of the buttons activated the presentation of the next tone token. It took approximately 5–6 h for each listener to complete the tone perception test.

B. Results and discussion

The tone perception scores were sorted out for each CI and NH child and the mean scores across all eight adult listeners were used to represent the tone production accuracy. The percent correct scores for the children with CIs ranged from 17.4% to 78.3% correct (48.5% ± 19.2% correct). Scores for the NH children ranged from 57.2% to 96.7% correct (79.9% ± 7.9% correct), which was significantly higher than that of the CI group [$t(73)=9.8$, $p<0.001$].

Percent correct scores varied for individual tones. They were 71.8%, 21.6%, 46.5%, and 55.2% correct for tones 1–4, respectively, for the children with CIs. For comparison, percent correct scores for individual tones produced by the NH children were 94.6%, 86.6%, 45.1%, and 93.2% correct for tones 1–4, respectively. Both groups showed patterns of better scores on the level tone 1 and falling tone 4. The rising tone 2 was perceived to be particularly poorly produced by the CI group. Consistent with the findings of the confusion matrix from the neural network, the confusion matrix from the perception tests also showed that the intended contour tones by the children with CIs were perceived as tone 1 about 30%–40% of the time (Fig. 7). It is worth noting that the preferred choice of tone 1 by the adult listeners reflects the monotonic features of the production of these children. It suggests that they either attempt to manipulate pitch in vain or that the limited tone information they receive through the implants hinders them from developing satisfactory tone production. Except for the 20 percentage points higher perception score on tone 3, the error patterns from the perception tests were in accordance with those from the neural network analysis.

VI. GENERAL DISCUSSION AND CONCLUSIONS

Several approaches were proposed and used for evaluating Mandarin Chinese tone production. The measures included three indices based on the acoustic analysis of the F0 onset–offset pair, contour correlation based on the examination of the patterns of F0 contours, and perception like classification based on the neural network analysis. These methods measure different aspects of tone production. The three indices evaluate the differentiation of tones. Indices 1 and 2 measure the overall differentiation of tones, whereas Index 3 breaks down to false detection probabilities of each tone category thus measures the differentiation of one particular tone from the others. The two components of Index 1 provide further information of pitch use and pitch range of a speaker. The tonal ellipse size measures variation of F0 use for individual tones, whereas the tonal area measures the F0 span used for all tones. The coefficients (i.e., R values) of the contour correlation evaluate the degree of similarity between two F0 contours. Although the previous acoustic measures describe certain acoustic properties of tone production, the neural network provides direct classification results, from which the recognition percent correct scores as well as tone confusion matrix can be generated.

These measures analyze different components of tone production. Pearson's correlation was performed for any two approaches, including results of the perceptual tests. The results of the correlations showed that any pairs of the approaches were significantly correlated after controlling for the family-wise type I error using Holm correction (i.e., Bonferroni step-down correction, family-wise $\alpha=0.05$). In the order of level of correlation, neural-network recognition and perception test correlated with the greatest level, followed by neural-network recognition with contour correlation, contour correlation with perception test, neural network with Indices 1, 2, and neural network, and perception test with Index 3 (Fig. 8).

The differences in the level of correlation are related to the nature of the analysis each measure entails. The three indices using the F0 onset–offset scatter plot mainly concern the level and slope of the F0 contour, whereas the contour correlation examines the patterns of the F0 contour. Contour correlation uses all the available samples from an F0 contour and it showed greater consistency with human perception and the neural network than the indices did. Regarding the strength of correlation to the perception test, neural network correlated better with the performance of human listeners than the contour correlation. Given the nature of the input to the neural network, it takes into account both dimensions (i.e., endpoints and pattern) of the F0 contours in the classification. As perception is also likely to involve the use of both level and contour cues, it is not surprising to see that the performance of the neural network has a greater correlation with the perceptual performance of human listeners than with the other approaches. Even though we consider the auditory judgment by adult listeners to be the fairest judge to tone production, we do not intend to imply that the usefulness of the other measures should be scaled on the strength of their correlation with the perception test. These measures emphasize different aspects of tone production whereas perception involves the use of weighted total of these aspects. We should also use caution when interpreting the correlations because the results of the perception tests may be favored with other cues such as duration that was not examined in other measures.

These measures were quite successful and consistent with each other in scoring individual children with CIs. As for the averaged scores for different tone categories, again,

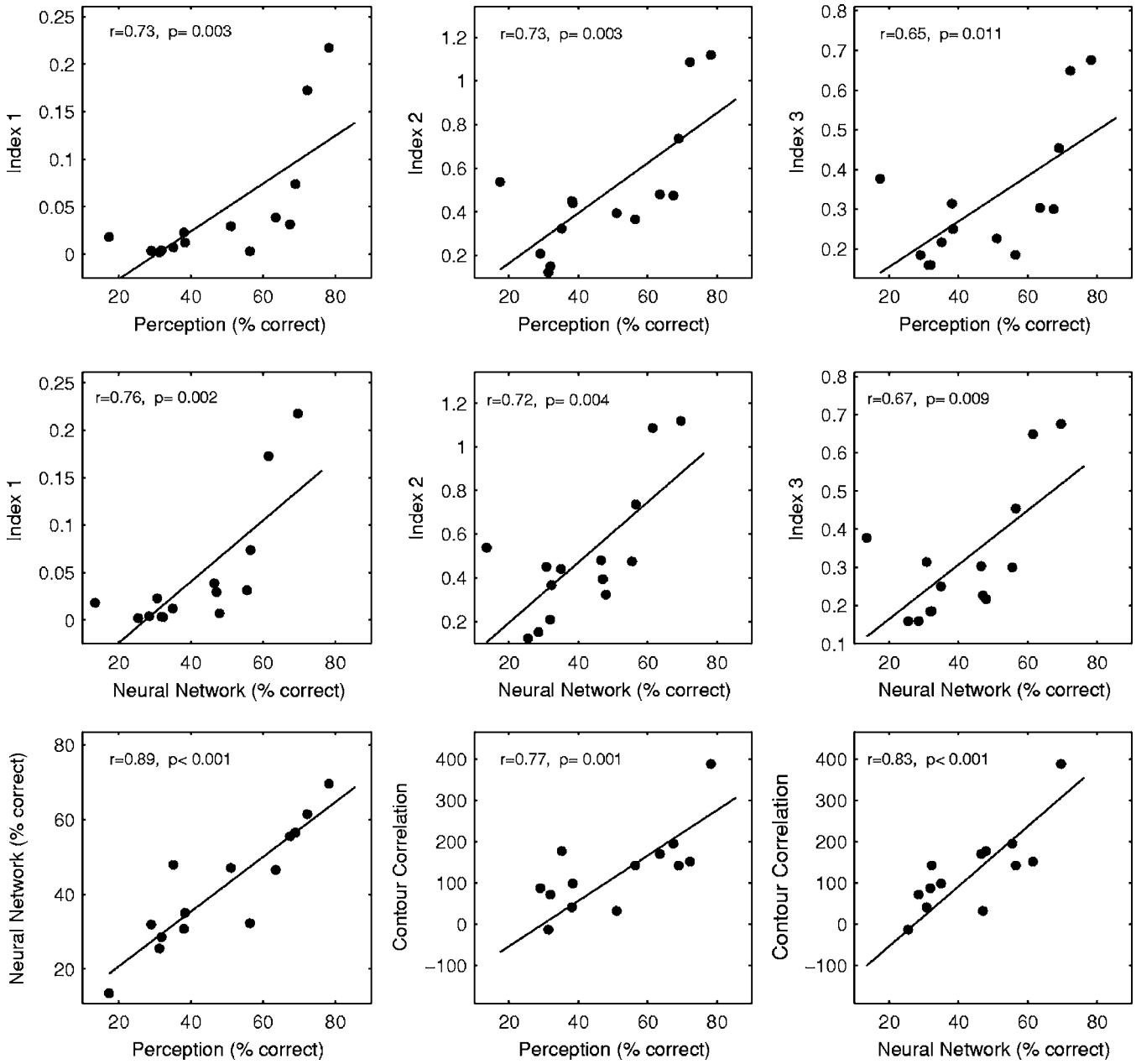


FIG. 8. Correlation among measures. (Top) Pearson's linear correlation between the perceptual scores and the three indices; (middle): Pearson's linear correlation between the scores of neural network and the three indices; and (bottom): Pearson's linear correlation between the perceptual scores and the scores of the neural network, contour correlation and the perceptual scores, and the scores of neural network and the contour correlation. Each symbol represents one child with a CI. The solid line represents the linear fit of the data in each panel.

the results of the neural network and the perception tests were the closest. It was shown that tones 1 and 4 were the relatively better produced tones. Confusion matrices of the two approaches also closely matched except for tone 3. Human ear could more correctly identify tone 3 than the neural network probably because tone 3 has a quite reliable duration cue that was available to human ears but was normalized in the neural network input. The contour correlation also indicated that tone 4 was the best produced tone albeit failed to evaluate tone 1. The production of tone 1, however, was better estimated with the false detection probabilities. It was shown that it was the least possible tone being falsely recognized.

Each measure has its strength and limitations. In general, the neural network demonstrated performance comparable to that of the human listeners with a high correlation of percent correct scores for individuals and very similar tone confusions. It is partly because the neural network works on a complex mathematical model with high tolerances of errors that emulate a biological perception system (Bishop, 1995; Arbib, 1995). It is thus ideal to use neural network to obtain perception like classification results. Nonetheless, the underlying process of the function approximation that the neural network relies on is hardly transparent. On the contrary, the acoustic analysis built on the examination of the level and pattern of the F0 contours provides more straightforward accounts for tone production performance. More importantly, the acoustic analysis offers more detailed and useful information of the acoustic properties of tone production than the perceptual judgments or neural networks do. The acoustic

analysis is particularly useful for studying the children's development in tone production. The tonal ellipses visualize the tonal space and pitch use for individual tones by a particular speaker. The indices can also quantify these parameters and evaluate the degree of differentiation among tones. Tonal ellipse analysis has its merit in particular. It greatly simplifies the analysis using just the onset and offset points of an F0 contour. Nonetheless, this simplification, as pointed out, prevents this method from giving a comprehensive representation of the contour tones in particular. The contour correlation is a useful complementary analysis in addition to the tonal ellipses. It quantifies the similarity of one tone contour to a standard form. Thus, the contour tones of Mandarin Chinese can be more adequately evaluated using this method.

In conclusion, the proposed methods can be used to evaluate the tone production of children with CIs effectively. The proposed methods can also be used to evaluate tone production of hearing impaired people, or to study tone acquisition in developmental research. They all focus on, however, different aspects of the F0's. It will be interesting in future studies to address aspects other than F0 (e.g., duration) in the assessment of tone production.

ACKNOWLEDGMENTS

The authors are grateful to Xiuwu Chen and Xiaoyang Zhao for their assistance in data collection in Beijing and to Erin Furman, Hui-Hui Li, Chessy Seebohm, Jessica Wolfanger, and Yunfang Zheng, for their technical support. They would like to thank the reviewers and the associate editor for their constructive comments in an earlier version of this paper. The study was supported in part by NIH NIDCD Grant No. R03-DC006161.

Arbib, M. A. (**1995**). *Handbook of Brain Theory and Neural Networks*, 2nd ed., (MIT Press, Cambridge, MA).

Barry, J. G., and Blamey, P. J. (**2004**). "The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese," J. Acoust. Soc. Am. **116**, 1739–1748.

Bishop, C. M. (**1995**). *Neural Networks for Pattern Recognition* (Oxford University Press, Oxford, UK).

Chang, P. C., Sun, S. W., and Chen, S. H. (**1990**). "Mandarin tone recognition by multi-layer perception," *Proceedings of the 1990 IEEE Conference on Acoustic Speech Signal Processes*, 3–6 April 1990, Albuquerque, NM, p. 517–520.

Chao, Y. R. (**1948**). *Language and Symbolic Systems* (Oxford University Press, Oxford, UK).

Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (**2002**). "The perception of Cantonese lexical tones by early-deafened cochlear implantees," J. Acoust. Soc. Am. **111**, 2250–2256.

Couch, L. W. (**1996**). *Digital and Analog Communication Systems* (Prentice Hall, Upper Saddle River, NJ).

Duanmu, S. (**2000**). *The Phonology of Standard Chinese* (Oxford University Press, Oxford, UK).

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hagan, M. T., and Menhaj, M. (**1994**). "Training feedforward networks with the Marquardt algorithm," IEEE Trans. Neural Netw. **5**, 989–993.

Han, D., Zhou, N., Li, Y., Chen, X., Zhao, X., and Xu, L. (**2007**). "Tone production of Mandarin-Chinese-speaking children with cochlear implants," Int. J. Pediatr. Otorhinolaryngol. **71**, 875–880.

Lan, N., Nie, K. B., Gao, S. K., and Zeng, F.-G. (**2004**). "A novel speech-processing strategy incorporating tonal information for cochlear implants," IEEE Trans. Biomed. Eng. **51**, 752–760.

Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N., and Cheung, D. M. C. (**2002**). "Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children," Int. J. Pediatr. Otorhinolaryngol. **63**, 137–147.

Lee, T., Ching, P. C., Chan, L. W., Cheng, Y. H., and Mak, B. (**1995**). "Tone recognition of isolated Cantonese syllables," IEEE Trans. Speech Audio Process. **3**, 204–209.

Li, C. N., and Thompson, S. A. (**1977**). "The acquisition of tone in Mandarin-speaking children," J. Child Lang, **4**, 185–199.

Moore, B. C. J. (**2003**). "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants," Otol. Neurotol. **24**, 243–254.

Peng, S., Tomblin, J. B., Cheung, C., Lin, Y. -S., and Wang, L. (**2004**). "Perception and production of Mandarin tones in prelingually deaf children with cochlear implants," Ear Hear. **25**, 251–264.

Rose, P. (**1987**). "Considerations in the normalization of the fundamental frequency of linguistic tone," Speech Commun. **6**, 343–352.

Rose, P. (**1993**). "A linguistic phonetic acoustic analysis of Shanghai tones," Austr. J. Ling. **13**, 185–219.

Wang, Y., Jongman, A., and Sereno, J. (**2003**). "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training," J. Acoust. Soc. Am. **113**, 1033–1043.

Wang, Y. R., and Chen, S. H. (**1994**). "Tone recognition of continuous Mandarin speech assisted with prosodic information," J. Acoust. Soc. Am. **96**, 2637–2645.

Wei, W. I., Wong, R., Hui, Y., Au, D. K. K., Wong, B. Y. K., Ho, W. K., Tsang, A., Kung, P., and Chung, E. (**2000**). "Chinese tonal language rehabilitation following cochlear implantation in children," Acta Oto-Laryngol. **120**, 218–221.

Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.

Wong, A. O. C., and Wong, L. L. N. (**2004**). "Tone perception of Cantonese-speaking prelingually hearing-impaired children with cochlear implants," Otolaryngol.-Head Neck Surg. **130**, 751–758.

Wu, J. -L., Lin, C. -Y., Yang, H. -M., and Lin, Y. -H. (**2006**). "Effect of age at cochlear implantation on open-set word recognition in Mandarin speaking deaf children," Int. J. Pediatr. Otorhinolaryngol. **70**, 207–211.

Wu, J. L., and Yang, H. M. (**2003**). "Speech perception of Mandarin Chinese speaking young children after cochlear implant use: Effect of age at implantation," Int. J. Pediatr. Otorhinolaryngol. **67**, 247–253.

Xu, L., Chen, X., Zhou, N., Li, Y., Zhao, X., Zhao, X., and Han, D. (**2007**). "Recognition of lexical tone production of children with an artificial neural network," Acta Oto-Laryngol. **127**, 365–369.

Xu, L., Li, Y., Hao, J., Chen, X., Xue, S. A., and Han, D. (**2004**). "Tone production in Mandarin-speaking children with cochlear implants: A preliminary study," Acta Oto-Laryngol. **124**, 363–367.

Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.

Xu, L., Zhang, W., Zhou, N., Lee, C.-Y., Li, Y., Chen, X., and Zhao, X. (**2006**). "Mandarin Chinese tone recognition with an artificial neural network," J. Otol. **1**, 30–34.

Zhou, N., Zhang, W., Lee, C.-Y., and Xu, L. (**2008**). "Lexical tone recognition by an artificial neural network," Ear Hear. in press.

Selectivity of modulation interference for consonant identification in normal-hearing listeners

Frédéric Apoux[a] and Sid P. Bacon
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 22 March 2007; revised 28 November 2007; accepted 4 December 2007)

The present study sought to establish whether speech recognition can be disrupted by the presence of amplitude modulation (AM) at a remote spectral region, and whether that disruption depends upon the rate of AM. The goal was to determine whether this paradigm could be used to examine which modulation frequencies in the speech envelope are most important for speech recognition. Consonant identification for a band of speech located in either the low- or high-frequency region was measured in the presence of a band of noise located in the opposite frequency region. The noise was either unmodulated or amplitude modulated by a sinusoid, a band of noise with a fixed absolute bandwidth, or a band of noise with a fixed relative bandwidth. The frequency of the modulator was 4, 16, 32, or 64 Hz. Small amounts of modulation interference were observed for all modulator types, irrespective of the location of the speech band. More important, the interference depended on modulation frequency, clearly supporting the existence of selectivity of modulation interference with speech stimuli. Overall, the results suggest a primary role of envelope fluctuations around 4 and 16 Hz without excluding the possibility of a contribution by faster rates.
 [DOI: 10.1121/1.2828067]

PACS number(s): 43.71.Es, 43.66.Dc, 43.66.Mk [KWG] Pages: 1665–1672

I. INTRODUCTION

During the past 10 years, envelope filtering has proven an easy and straightforward technique to assess the amount by which temporal modulations can be reduced without affecting speech intelligibility. In a pioneering study, Drullman *et al.* (1994) measured speech recognition in normal-hearing listeners when fast envelope modulations were systematically removed from speech. The authors used the Hilbert transform to decompose the original speech signal into an envelope and a carrier (fine structure). The carrier was then multiplied by the low-pass filtered envelope. Their results indicated that intelligibility increases progressively with low-pass cutoff frequency up to about 16 Hz, beyond which intelligibility does not improve significantly. This suggests that modulation frequencies above 16 Hz do not contribute to overall intelligibility, providing that lower rate fluctuations are preserved. The results from recent studies, however, strongly suggest that the envelope filtering techniques might be inappropriate to estimate the range of the pertinent fluctuation rates in speech (Ghitza, 2001; Atlas *et al.*, 2004). The main argument is that when a listener is presented with the speech fine structure only, part of the original envelope information may be recovered from the phase information. Similarly, when listeners are presented with envelope-filtered stimuli, envelope information well beyond the nominal cutoff frequency of the envelope filter may be recovered from the speech fine structure as well.

Because envelope filtering techniques have been considered dependable until very recently, few other approaches have been proposed to measure the limit of useful temporal information in speech. Moreover, most of these techniques also presented some important limitations. For example, Ghitza (2001) proposed an original technique based on dichotic speech synthesis using pure cosine carriers to generate smeared envelopes at the output of the listener's cochlea. However, these stimuli were of poor quality, limiting the extent to which they could be used in intelligibility tasks. More recently, Atlas and collaborators (Atlas *et al.*, 2004; Atlas and Janssen, 2005; Qin and Atlas, 2005) proposed a new approach to envelope filtering based on the concept of a complex modulator (in contrast to previous approaches assuming that the modulator is real and non-negative) to circumvent the limitations of common envelope extraction techniques. This approach is called coherent envelope detection. Schimmel and Atlas (2005a) tested the ability of this approach to remove envelope frequencies by quantifying the effective stop-band attenuation for several frequencies of the low-pass modulation filter and compared the results to levels of attenuation achieved with the conventional Hilbert approach in similar conditions. They reported that the coherent approach has a stop-band rejection of about 6–28 dB where the conventional Hilbert approach can only reject 3–14 dB. Although the effective stop-band attenuation is reportedly larger for the coherent approach than for the conventional Hilbert approach, it might remain inadequate to prevent listeners from using information beyond the envelope filter cutoff frequency. To our knowledge, other attempts to circumvent this limitation have not provided conclusive results yet both in terms of effective stop-band attenuation and signal quality (e.g., Schimmel and Atlas, 2005a, b; Schimmel *et al.*, 2006).

From this, it is apparent that the appropriate method to evaluate the role of envelope fluctuations in the identification

[a]Electronic mail: apoux@psycho.univ-paris5.fr

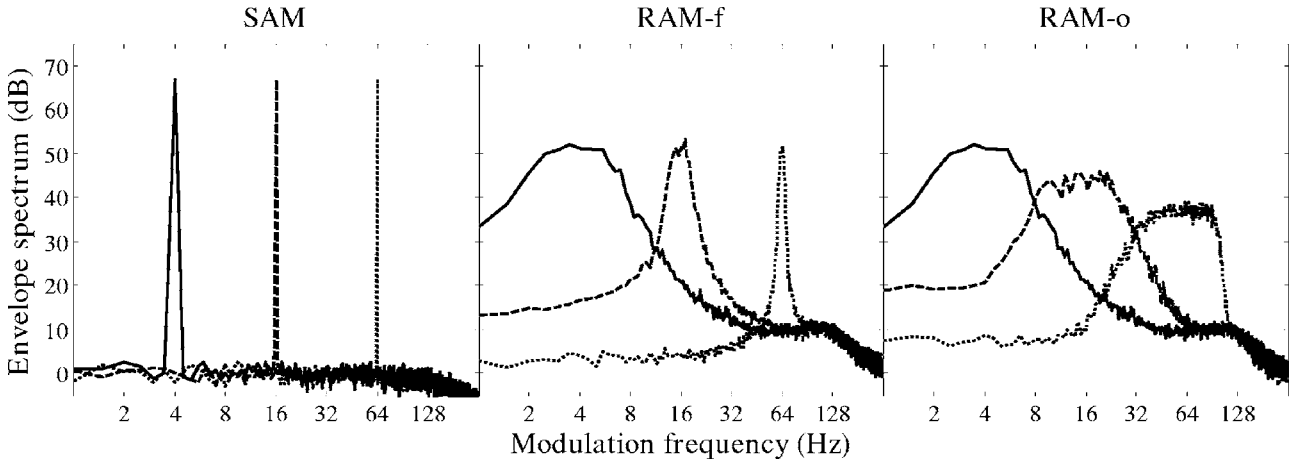


FIG. 1. The left, middle, and right panels show examples of the modulation spectra for the SAM, RAM-f, and RAM-o conditions. In each panel, the solid, short-dashed, and long-dashed lines correspond to the modulation spectrum for the 4-, 16-, and 64-Hz modulator frequency condition, respectively. Each given spectrum was computed from the average of 20 realizations.

of speech *when the speech fine structure is preserved* is yet to be found. Potentially, psychophysical studies on modulation detection interference (MDI) offer a way to assess the important modulation rates for speech intelligibility. Indeed, these studies indicate that the processing of amplitude modulation (AM) in one spectral region can be disrupted by simultaneous AM in another region (Yost and Sheft, 1989, 1994; Bacon and Moore, 1993). Accordingly, MDI is usually defined as the difference between threshold modulation depth in the presence of a modulated interferer and that obtained in the presence of an unmodulated interferer. More important, MDI experiments show that the amount of interference is maximal when the interferer and target carriers are modulated at similar rates and gradually declines as the target and interferer modulation frequencies diverge (e.g., Yost *et al.*, 1989). These two aspects of MDI, across-channel interaction and selectivity in the modulation domain, are of great interest for the purpose of studying the role of envelope fluctuations in the identification of speech. A paradigm such as MDI if applied to speech stimuli would allow us to circumvent the above-discussed limitations. To our knowledge, only one study has investigated modulation interference (MI) with speech stimuli (Kwon and Turner, 2001). However, the second experiment conducted in that study is the only MI experiment *per se*. In the other experiments, the original speech fine structure was replaced with noise or the speech and masker spectrum greatly overlapped. In Experiment 2, Kwon and Turner measured consonant identification in the presence of a random noise either sinusoidally modulated at 8 Hz or unmodulated. Speech and noise were divided into seven bands, each roughly 2.5-octave wide. Three speech conditions were tested: speech at band 3, at band 4, and at band 5. In each condition, a band of noise was presented simultaneously either in one of the six other bands or in the same band. Overall, performance was lowest when the speech and masker were in the same band and it increased as the masker band was moved away from the speech band. When the speech and masker were in two different bands performance in the unmodulated condition was usually higher than that in the modulated condition, indicating that MI occurred. Because the authors tested one modulation frequency only in this experiment, they could not show evidence for selectivity of MI. The present study seeks to establish whether speech recognition can be disrupted by the presence of simultaneous AM at a remote spectral region, and more important, whether selectivity of MI can be observed with speech stimuli using a paradigm modeled after the psychophysical MDI paradigm. If selectivity is observed, it may be possible to use this paradigm to determine which modulation frequencies are most important for speech intelligibility.

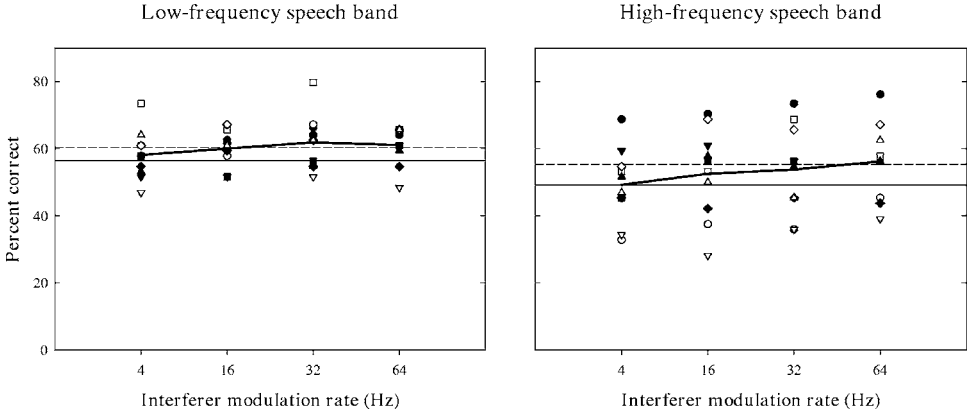


FIG. 2. Individual (symbols) and mean (bold line) percent correct scores for consonant identification as a function of the modulation frequency of the interferer as measured with the sinusoidal modulator (SAM). The left and right panels show the data for the LF and the HF conditions, respectively. In each panel, performance in the quiet and the unmodulated conditions is indicated by a solid and a dashed line, respectively.

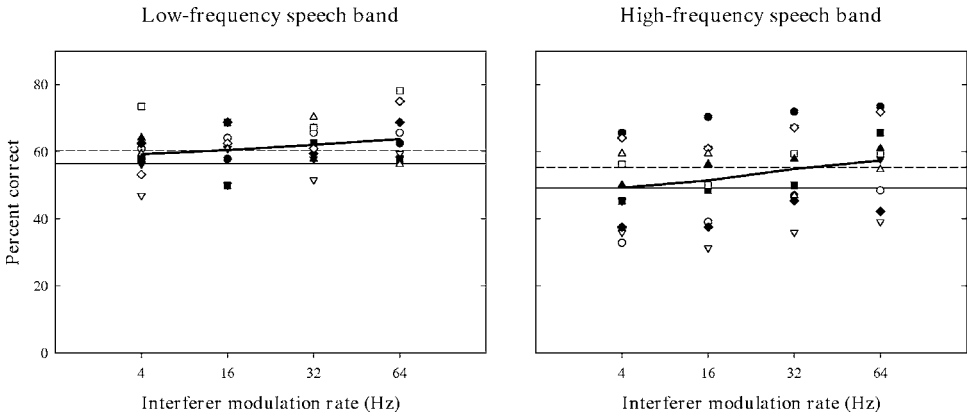


FIG. 3. The same as Fig. 2, but for the random modulator with the 4-Hz bandwidth (RAM-f).

II. METHODS

A. Stimuli

The stimuli consisted of 16 consonants in /a/-consonant-/a/ environment recorded by four speakers (two for each gender) for a total of 64 vowel-consonant-vowel (VCV) disyllables (Kwon and Turner, 2001). Different speakers were used to limit learning effects and to limit the influence of a given utterance. Each speech stimulus was split into two distinct broad frequency bands, one low-frequency speech band (LF) and one high-frequency speech band (HF). The cutoff frequencies of the two bands were chosen on the basis of the articulation index (AI) (French and Steinberg, 1947), so that each band had an importance value of approximately 0.25. The LF band was low-pass filtered at 1 kHz, whereas the HF band was bandpass filtered between 3.5 and 8 kHz. The speech band was either presented alone (Quiet) or with an interferer. The interferer was a band of noise with a width of 4 ERBs (equivalent rectangular bandwidths; Glasberg and Moore, 1990); a new sample of the random-noise carrier was generated on each trial. The speech and noise bands were separated by 5 ERBs. Therefore, the interferer carrier ranged from 1.9 to 3 kHz and from 1.1 to 1.9 kHz in the LF and the HF conditions, respectively. We used 5 ERBs because our pilot impressions indicated that this frequency separation was large enough to minimize peripheral interactions between the speech and the noise carriers. We reasoned that modulation of the masker might enhance speech recognition via a masking release (Miller and Licklider, 1950; Gustafsson and Arlinger, 1994) in case of excessive overlap of masker and speech spectrum (i.e., in case of within-channel interactions). No masking release was observed when the speech and noise bands were separated by 5 ERBs. The speech and interferer were always gated synchronously so that the duration of the interferer was always the same as for the speech stimulus. As a consequence, the duration of the interferer varied from trial to trial from 466 to 952 ms. The carrier was either left intact (unmodulated) or modulated for its entire duration. The study by Kwon and Turner (2001) used sinusoidal modulators. A MDI study by Mendoza *et al.* (1995) indicated that random-noise modulators produce more interference than sinusoidal modulators. It is unclear whether a similar effect of modulator type would be observed with speech signals. Accordingly, it was decided to test both types of modulator. Eventually, three types of modulator were tested. The first modulator was a sinusoid, resulting in sinusoidal amplitude modulation (SAM). The modulation depth was 1.0 and the rate was 4, 16, 32, or 64 Hz. The two other modulators were narrow bands of noise, resulting in random AM (RAM). For one, RAM-f, the band was centered at 4, 16, 32, or 64 Hz and the bandwidth was 4 Hz. In the other, RAM-o, the modulator was centered at 4, 16, or 64 Hz and the bandwidth was one octave. The 32-Hz condition was excluded because its modulation spectrum greatly overlapped with that of the other interferers. These two different bandwidths for the random-noise modulators allowed an examination of the trade-off between modulation depth and frequency span. Figure 1 shows the modulation spectrum computed for each modulator type. Each panel shows the mean modulation spectrum averaged across 20 realizations for modulator center frequen-

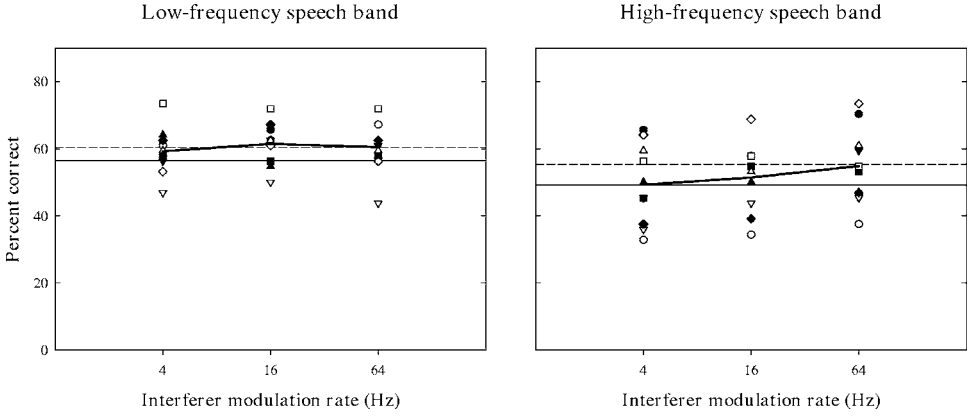


FIG. 4. The same as Fig. 2, but for the random modulator with the octave-wide bandwidth (RAM-o).

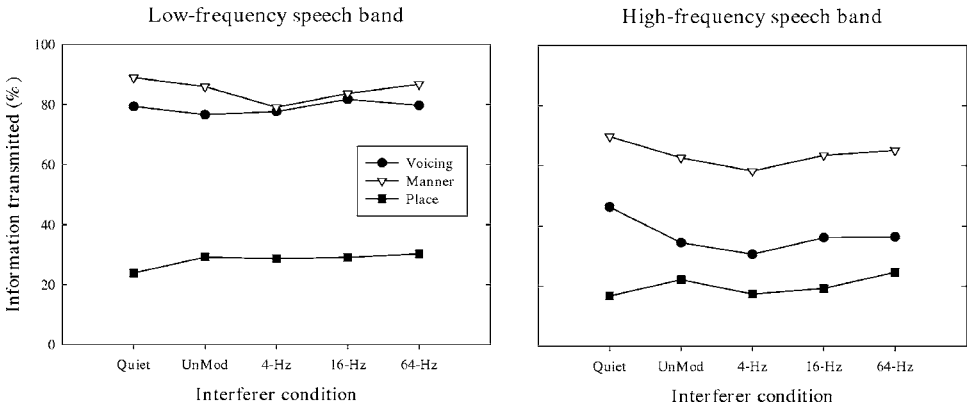


FIG. 5. Percent information transmitted for the features voicing, manner, and place of articulation as a function of the interferer condition for the low-frequency speech band (left panel) and the high-frequency speech band (right panel) condition.

cies of 4, 16, and 64 Hz. The left, middle, and right panels show the results for the SAM, RAM-f, and RAM-o conditions, respectively. Figure 1 provides a good illustration of how the three modulators cover the range of possible combinations between modulation depth and frequency span for the interferer. It can be seen that the modulation depth at a given center frequency decreases as the modulator bandwidth increases from SAM to RAM-f to RAM-o. Similarly, for the RAM-o condition, the modulation depth decreases as the center frequency (and hence the bandwidth) of the modulator increases.

B. Subjects

Ten subjects participated, six females and four males. One was author FA. Their age ranged from 22 to 35 years. All had absolute thresholds of 20 dB HL or better (ANSI, 1996) for octave frequencies from 500 to 8000 Hz. All participants, except author FA, were native American English speakers and were paid an hourly wage for their participation.

C. Procedure

Listeners initially completed a familiarization session in which the 64 unprocessed (i.e., broadband) VCV were presented in quiet. Then, the two speech band location conditions were administrated separately. Five listeners completed the LF condition first and five completed the HF condition first. Each run or block presented the entire set of VCV and corresponded to the combination between an interferer modulator type (SAM, RAM-f or RAM-o) and rate (4, 16, 32, or 64 Hz). Note that in the following the frequency or center frequency of the modulator will be referred to as the interferer modulation rate. Because the 4-Hz modulation rate condition was identical for the RAM-f and the RAM-o conditions, this particular condition was administered only once. In addition, listeners completed 2 blocks in the quiet condition and 2 blocks in the unmodulated condition, randomly distributed amongst the modulated interferer conditions. Therefore, in a given speech band location condition, each listener completed 14 blocks in which the 64 VCV were presented in random order. The presentation order for the different blocks was also chosen randomly.

Listeners were tested individually in a double-walled, sound-attenuated booth using a single-interval, 16-alternative procedure. The experiment was performed on a PC equipped with a sound card (Echo Gina3G). Stimuli were played to the listeners binaurally through Sennheiser HD 250 Linear II circumaural headphones. Listeners were instructed to report the perceived consonant. No feedback was given during the familiarization or the experimental sessions.

III. RESULTS

Figures 2–4 show the consonant identification scores for the SAM, RAM-f, and RAM-o conditions, respectively. In each figure, the left and right panels correspond to the data for the LF and the HF conditions, respectively, with each panel showing individual (symbols) and averaged (bold line) data as a function of the modulation rate of the interferer. Scores obtained in the quiet and the unmodulated conditions are indicated by a solid and a dashed horizontal line, respectively. Averaged performance in the quiet condition was 56% and 49% correct for the low- and high-frequency band of speech, respectively. The similarity in performance between location conditions was expected given that speech stimuli were partitioned based on equal intelligibility as predicted by the AI. Overall, the pattern of results was very similar in the two location conditions. The only noticeable difference was a slightly larger variability in the HF condition compared to the LF condition with an averaged standard deviation of 12 and 7 points, respectively. For both speech band locations, an *increase* in performance was observed when the (off-frequency) unmodulated noise was added. The improvement was about 5 percentage points. A paired t-test revealed that performance in the unmodulated condition was significantly different from that in the quiet condition ($p<0.05$). Similar beneficial effects when spectral (Warren *et al.*, 1997) or temporal (Powers and Wilcox, 1977) gaps in speech are replaced with noise have been reported previously. The quiet condition actually resulted in the lowest consonant identification score of all the conditions tested. The effect of modulation rate was small but consistent across all location and modulation rate conditions. Consonant identification generally increased with increasing rate. The improvement, as measured between the 4- and the 64-Hz condition, ranged from 1.25 to 9 percentage points. It seems, however, that the performance in the 32- and the 64-Hz conditions was similar to that in the unmodulated condition, suggesting a negligible amount of interference for these two rates.

Statistical analyses were first performed separately for each modulator type using repeated-measures analysis of variance (ANOVA), with factors of interferer modulation rate and speech band location. All three analyses indicated a significant effect of interferer modulation rate (all $p<0.05$) and no significant effect of speech band location. The interaction between the two main factors was not significant in all three analyses. Because no differences were found between the scores obtained with the low- and the high-frequency speech bands, data from these two conditions were pooled and a new ANOVA was performed with factors of interferer modulation rate and interferer modulation type. In this new analysis, a balanced set with missing data was assumed (no data at 32 Hz in the RAM-o condition). Again, the effect of modulation rate was found to be significant [$F(3,109)=10.95$, $p<0.001$]. Results of the analysis indicated no significant effect of the modulation type ($p=0.32$). Post hoc comparisons (Tukey) revealed significant pairwise differences between the following modulation rates: 4 and 32 Hz ($p=0.005$), 4 and 64 Hz ($p=0.001$), and 16 and 64 Hz ($p<0.05$).

The averaged consonant confusion matrices were analyzed in terms of relative information transmitted [(RIT); Miller and Nicely, 1955]. The specific reception of three speech features (voicing, manner, and place of articulation) was evaluated for each experimental condition. Results of the feature analysis, however, were averaged across modulator type conditions. Figure 5 presents the results this way for the low-frequency speech band (left panel) and the high-frequency speech band condition (right panel). Overall, the RIT for the three speech features in quiet was consistent with a previous report by Miller and Nicely (1955). In the low-frequency speech band condition, considerable voicing and manner information was available (79% and 89%, respectively). In the high-frequency speech band condition, the transmission of voicing and manner features was only 46% and 70% correct, respectively. In both conditions, the transmission of place feature was poor (around 20%). Although the amount of voicing and manner information was not equal in both speech band conditions, the pattern of results was very similar. The feature analysis indicated a small drop in RIT for the voicing and manner features when the unmodulated band of noise was added. In contrast, the transmission of place was higher in the presence of this interferer. The transmission of place information was poorer in the Quiet condition in general. This last result suggests that the relatively low performance in the Quiet condition is primarily related to poorer perception of place information. The analysis also revealed that RIT increased with increasing interferer modulation rate for all three features. In other words, the presence of low-frequency rates interfered with the perception of all three speech features.

IV. DISCUSSION

The primary goal of the present study was to establish whether MI and selectivity of MI could be observed with speech stimuli. Consistent with the existence of MI, our results showed poorer consonant recognition scores when a masker at a remote frequency region was amplitude modulated as compared to a condition in which the same masker was unmodulated. Consistent with the existence of selectivity of MI, our results showed that the presence of high-frequency AM in the masker's envelope is less detrimental for consonant recognition than the presence of low rates. Indeed, the fact that high-frequency AM in the masker's envelope produced less interference than low-frequency AM strongly supports the assumption of a form of selectivity in MI similar to what is typically observed in MDI studies. A secondary goal of the present study was to determine whether it might be possible to use the MDI paradigm to determine which modulation frequencies are most important for speech intelligibility. There are, however, potential limitations of the MDI paradigm when used to determine which modulation frequencies are most important for speech intelligibility. One limitation is the necessary restriction of the stimuli in the spectral domain in order to reduce within-band interactions between the target and the interferer. The spectral characteristics of the speech stimuli possibly limit the extent to which the results observed using the MDI paradigm can be applied to natural speech. Indeed, in natural-listening environments listeners typically have access to the broadband speech spectrum. Moreover, several studies clearly show that methods developed to predict speech intelligibility based on filtered speech such as the Articulation Index (French and Steinberg, 1947) do not reflect accurately how listeners actually combine and attend to speech information in natural-listening conditions (Turner *et al.*, 1998; Calandruccio and Doherty, 2007). However, it should be possible to make speech stimuli more natural by using narrow bands of speech interleaved with narrow bands of noise as in Kwon and Turner (2001).

The time range during which modulations conveying a given speech cue are available is highly variable and, as suggested by Kwon and Turner (2001), differences in duration may significantly affect the amount of interference. Modulations in the envelope of speech are characterized by their transiency. For example, the brief bursts of noise generated when the articulatory closure is released, the release bursts, have a very short time span (10–30 ms). This is extremely short compared to 500-ms stimuli used in most MDI studies. However, some cues conveyed by envelope fluctuations exist for very long periods of time and therefore should not be affected by the above-mentioned limitation. In particular, low-frequency envelope fluctuations (around 4 Hz) related to the average syllable rate, by definition, exist for the entire duration of the stimulus. Our results support this interpretation in that the largest amount of interference was observed with the 4-Hz modulator. Because to our knowledge no one has examined the effect of duration on MDI, it is difficult to determine the shortest time span necessary to observe interference. However, it is reasonable to assume that the large amounts of interference as measured in MDI studies cannot be observed with speech stimuli because of the very limited time during which the modulations conveying a given cue are available. Accordingly, another limitation of the MDI paradigm to examine which modulation frequencies

in the speech envelope play a role in speech recognition might be to reduce artificially the importance of envelope cues existing for a limited time range.

A last limitation is the relative uncertainty about the amount of interference that can be produced at various modulation rates in the speech envelope. MDI studies clearly show that the amount of interference decreases as the common modulation frequency of the target and interferer increases (Yost *et al.*, 1989; Bacon and Konrad, 1993), suggesting that high modulation rates might not be disrupted as effectively as lower rates. For example, Bacon and Konrad (1993) measured the minimum depth necessary to detect a modulation imposed on a 2-kHz tone in the presence of either a 1.8- or 4.0-kHz tone modulated at the same rate. The common modulation frequency was increased from 5 to 100 Hz. Their results showed that the amount of MDI decreased as modulation frequency was increased from 5 to 20 Hz, beyond which it remained constant. Although Bacon and Konrad observed some MDI with interferer modulations faster than 20 Hz, the amount of interference was relatively small (about 5 dB) at these rates in conditions for which within-channel interactions were limited. This suggests a reduced sensitivity of the auditory system to modulation interference at rates above 20 Hz. It is very likely that a comparable reduced sensitivity to interference at high modulation rates exists for speech stimuli.

In view of the above, can our data and more generally data gathered with the MDI paradigm be used to assess the critical modulation frequencies for consonant identification? As discussed previously, two limitations inherent to the use of speech stimuli and to our paradigm could reasonably explain (i) the lack of a significant difference between the 32- and the 64-Hz conditions and (ii) the similarity between the performance in these two conditions and those in the unmodulated condition, suggesting that our data do not necessarily reflect a marginal contribution of rates equal or faster than 32 Hz. Actually, we suggested in a previous study that temporal cues contained in the high-frequency region are relatively more important for consonant identification and that these cues might be conveyed by envelope fluctuations greater than 16 Hz (Apoux and Bacon, 2004). Accordingly, it is not possible to draw conclusions about the contribution of these fast envelope fluctuations to consonant identification from the present data. In contrast, it seems reasonable to conclude a role of envelope fluctuations around 4 and 16 Hz in consonant identification. Indeed, performance in the 4-Hz condition was significantly lower than that in the 32- and 64-Hz conditions and performance in the 16-Hz condition was significantly lower than that in the 64-Hz condition.

Unexpectedly, no differences in performance were found with the three types of modulators. Considering that the main physical difference between our three modulators can be described in terms of modulation depth and bandwidth, one interpretation is that, to some extent, MI in speech is not sensitive to these two factors. This interpretation is partially supported by the results from previous studies assessing the effect of the interferer modulation depth on the detection of SAM. Indeed, there is no clear consensus as to the exact influence of modulation depth on MDI. On the one hand, Bacon and Konrad (1993) showed that an increase in interferer depth from 0.0 to 1.0 (in 0.25 steps) results in a nearly proportional increase in signal depth. On the other hand, Yost and Sheft (1989) showed that the amount of interference increases as the modulation depth of the interferer increases from 0.0 to 0.5. Their data, however, also showed that additional increases in modulation depth did not lead to additional interference. Therefore, it may not be surprising that consonant identification was not significantly affected by the modulation depth of the interferer.

Possibly a better explanation for the absence of a differences in performance with the three types of modulators may be related to the concept of a modulation filterbank. Several researchers have suggested the existence of independent frequency channels in the envelope-frequency domain (Bacon and Grantham, 1989; Houtgast, 1989; Dau *et al.*, 1997a, b). One assumption is that these modulation filters are broadly tuned to modulation frequency and that either the modulation filters are broadly tuned to carrier frequency, or the outputs from modulation filters tuned to distinct carrier frequencies converge at a higher level of processing. According to the concept of a modulation filterbank, one can assume the amount of interference (i.e., masking in the modulation domain) to be determined by the overall interferer power within the modulation filter centered on the target modulation. Although there is no clear consensus about the bandwidth of these modulation filters, Q values of 1 are generally assumed (Dau *et al.*, 1997a, b; Ewert and Dau, 2000). Because (i) the bandwidth of the modulators was equal or narrower than that of a modulation filter with a Q value of 1 and (ii) the total power of intrinsic envelope fluctuations remained constant, the overall interferer power within the modulation filter centered on the target modulation was presumably held constant in the present experiment. As a consequence, all modulators produced a comparable amount of interference. In other words, the lesser modulation depth in some conditions was compensated by a larger modulation frequency span, resulting in similar amount of interference with the sinusoidal and the random modulations.

The implications for the intelligibility of speech in noise are complex. On the one hand, it is well established that adding an on-frequency masker typically degrades speech intelligibility (masking). On the other hand, we mentioned in Sec. III a study by Warren *et al.* (1997) showing that speech understanding may increase when a spectral gap which would normally contain speech is filled with noise (spectral restoration). In both cases, the magnitude of the effect is sensitive to the temporal properties of the masker. In the first (on-frequency) situation, imposing a modulation on the masker results in a diminution of masking (masking release). In the second (off-frequency) situation, imposing a modulation on the masker may completely eliminate the "phonemic restoration" effect (MI) as suggested by our findings. Finally, the magnitude of the effect (either masking release or MI) is optimum for very low-frequency AM in the masker in both cases. As indicated by Kwon and Turner (2001), these two opposite effects, masking release and MI, are presumably

involved whenever the masker is modulated and although masking release seems more dominant, more work needs to be done to separate their individual contributions to overall intelligibility.

V. SUMMARY AND CONCLUSION

The present study investigated how consonant identification in normal-hearing listeners is affected by the presence of simultaneous AM at a remote spectral region. Of particular interest was whether the amount of interference depends upon the frequency of the modulator imposed on an off-frequency masker, and whether a paradigm modeled after the psychophysical MDI paradigm could be used to assess the critical modulation frequencies for consonant identification.

Our results showed that consonant identification generally increases in the presence of an off-frequency masker. However, our results also showed that imposing a modulation on that off-frequency masker typically deteriorates performance. More important, the amount of interference was found to be larger for low-frequency AM and to decrease with increasing modulation rate, suggesting (i) the existence of selectivity of MI in speech and (ii) a reduced sensitivity to interference at high modulation rates. This relationship between the amount of interference and the frequency of the interferer's modulator is broadly similar to what is typically observed in MDI studies with nonspeech stimuli.

These findings were used to infer which modulation frequencies are most important for consonant identification. Based upon the amount of interference produced by each modulation rate tested, it is suggested that envelope fluctuations around 4 and 16 Hz are of particular significance for consonant identification. The absence of interference in the presence of AM equal or faster than 32 Hz suggests a limited contribution of these rates to consonant identification. However, the possibility exists that factors other than their actual importance for speech intelligibility may have obscured the contribution of envelope fluctuations equal or faster than 32 Hz.

In view of the limitations of envelope filtering techniques (i.e., the possibility to recover prescribed envelope information from the speech fine structure) and of the uncertainty in the amount of interference produced at high modulation frequencies in the speech envelope, it is apparent that neither the filtering techniques nor the MDI paradigm can be used to determine precisely which modulation frequencies are critical for speech, indicating that additional research is needed. By establishing the existence of MI for low-frequency AM in the speech envelope, the present study, however, demonstrates that the MDI paradigm may be used to effectively disrupt the processing of those low-frequency amplitude modulations in speech recognition tasks.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute of Deafness and Other Communication Disorders (NIDCD Grant No. DC01376). We thank Ken Grant and two anonymous reviewers for their helpful comments on a previous version of the manuscript.

ANSI (**1996**). "ANSI S3.6-1996, Specifications for audiometers," American National Standards Institute, New York.

Apoux, F., and Bacon, S. P. (**2004**). "Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise," J. Acoust. Soc. Am. **116**, 1671–1680.

Atlas, L. E., and Janssen, C. (**2005**). "Coherent modulation spectral filtering for single-channel music source separation," Proceedings of the 30th International Conference on Acoustics, Speech, and Signal Processing, Philadelphia.

Atlas, L. E., Li, Q., and Thompson, J. (**2004**). "Homomorphic modulation spectra," Proceedings of the 29th International Conference on Acoustics, Speech, and Signal Processing, Montreal, Canada.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Bacon, S. P., and Konrad, D. L. (**1993**). "Modulation detection interference under conditions favoring within- and across-channel processing," J. Acoust. Soc. Am. **93**, 1012–1022.

Bacon, S. P., and Moore, B. C. J. (**1993**). "Modulation detection interference: Some spectral effects," J. Acoust. Soc. Am. **93**, 3442–3453.

Calandruccio, L., and Doherty, K. A. (**2007**). "Spectral weighting strategies for sentences measured by a correlational method," J. Acoust. Soc. Am. **121**, 3827–3836.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Ewart, S.D. and Dau, T., Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196(**2000**).

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

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

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). "Masking of speech by amplitude modulated noise," J. Acoust. Soc. Am. **95**, 518–529.

Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.

Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.

Mendoza, L., Hall, J. W., and Grose, J. H. (**1995**). "Modulation detection interference using random and sinusoidal amplitude modulation," J. Acoust. Soc. Am. **97**, 2487–2497.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Powers, G. L., and Wilcox, J. C. (**1977**). "Intelligibility of temporally interrupted speech with and without intervening noise," J. Acoust. Soc. Am. **61**, 195–199.

Qin, L., and Atlas, L. E. (**2005**). "Properties for modulation spectral filtering," Proceedings of the 30th International Conference on Acoustics, Speech, and Signal Processing, Philadelphia.

Schimmel, S. M., and Atlas, L. E. (**2005a**). "Coherent envelope detection for modulation filtering of speech," Proceedings of the 30th International Conference on Acoustics, Speech, and Signal Processing, Philadelphia.

Schimmel, S. M., and Atlas, L. E. (**2005b**). "Advanced signal processing algorithms, architectures, and implementations XV," edited by, F. T. Luk, Proceedings of the Society of Photo-Optical Instrumentation Engineers, **5910**, 59100H-10.

Schimmel, S. M., Fitz, K. R., and Atlas, L. E. (**2006**). "Frequency reassignment for coherent modulation filtering," Proceedings of the 31st International Conference on Acoustics, Speech, and Signal Processing, Toulouse, France.

Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. A. (**1998**). "Frequency-weighting functions for broadband

speech as estimated by a correlational method," J. Acoust. Soc. Am. **104**, 1580–1585.
Warren, R. M., Hainsworth, K. R., Brubaker, B. S., Bashford, J. A., and Healy, E. W. (**1997**). "Spectral restoration of speech: Intelligibility is increased by inserting noise in spectral gaps," Percept. Psychophys. **59**, 275–283.
Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," J. Acoust. Soc. Am. **85**, 848–857.
Yost, W. A., and Sheft, S. (**1994**). "Modulation detection interference: Across-frequency processing and auditory grouping," Hear. Res. **79**, 48–58.
Yost, W. A., Sheft, S., and Opie, J. (**1989**). "Modulation interference in detection and discrimination of amplitude modulation," J. Acoust. Soc. Am. **86**, 2138–2147.

Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction

Ning Li and Philipos C. Loizou[a)]
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

(Received 26 June 2007; revised 14 December 2007; accepted 16 December 2007)

The application of the ideal binary mask to an auditory mixture has been shown to yield substantial improvements in intelligibility. This mask is commonly applied to the time–frequency ($T-F$) representation of a mixture signal and eliminates portions of a signal below a signal-to-noise-ratio (SNR) threshold while allowing others to pass through intact. The factors influencing intelligibility of ideal binary-masked speech are not well understood and are examined in the present study. Specifically, the effects of the local SNR threshold, input SNR level, masker type, and errors introduced in estimating the ideal mask are examined. Consistent with previous studies, intelligibility of binary-masked stimuli is quite high even at −10 dB SNR for all maskers tested. Performance was affected the most when the masker dominated $T-F$ units were wrongly labeled as target-dominated $T-F$ units. Performance plateaued near 100% correct for SNR thresholds ranging from −20 to 5 dB. The existence of the plateau region suggests that it is the pattern of the ideal binary mask that matters the most rather than the local SNR of each $T-F$ unit. This pattern directs the listener's attention to *where* the target is and enables them to segregate speech effectively in multitalker environments. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2832617]

PACS number(s): 43.71.Es, 43.71.Gv [MSS] Pages: 1673–1682

I. INTRODUCTION

Human listeners are able to understand speech even when it is masked by one or more competing voices. The speech segregation process is often approximated at the basic level by two distinct stages. In the first stage, the auditory periphery decomposes the auditory mixture to an array of individual time-and-frequency ($T-F$) units, with each unit representing the acoustic signal occurring at a particular instance in time and frequency, and with the size (e.g., bandwidth and duration) of each unit representing the smallest auditory event that can be resolved. The listener is able to reliably detect the acoustic energy of the target voice in $T-F$ regions, wherein the target is at least as strong as the masker. In the second stage, the listener examines all $T-F$ units in the mixture and uses *a priori* information about the target signal as well as a multitude of other cues to segregate the $T-F$ units of the target and integrate (or somehow group) all these units into a single auditory image of the target signal. Cues such as common periodicity across frequency, common offsets and onsets, amplitude and frequency modulations, are believed to be involved and used by human listeners in the previous auditory scene analysis (Bregman, 1990).

Several computational auditory scene analysis (CASA) techniques were proposed in the literature modeling the previous two-stage segregation process (Wang and Brown, 2006). The goal of CASA techniques was to segregate only the target signal, rather than all interfering sources, from the sound mixtures, and the means suggested for achieving this goal was the ideal $T-F$ binary mask (Wang, 2005). The ideal binary "mask" takes values of zero and one, and is constructed by comparing the local signal-to-noise ratio (SNR) in each $T-F$ unit against a threshold (e.g., 0 dB). The ideal mask is commonly applied to the $T-F$ representation of a mixture signal and eliminates portions of a signal (those assigned to a "zero" value) while allowing others (those assigned to a "one" value) to pass through intact.

A recent study (Brungart *et al.*, 2006) demonstrated the potential of using ideal binary masks for improving the intelligibility of speech masked by one or more interfering voices. A number of other studies (Roman *et al.*, 2003; Roman and Wang, 2006; Cooke, 2006; Brungart *et al.*, 2006; Anzalone *et al.*, 2006) have shown that speech synthesized from the ideal binary mask is highly intelligible even when extracted from multisource mixtures (Roman *et al.*, 2003) or in reverberant conditions (Roman and Wang, 2006). Ideal masks that are not binary but take real values have also been found to improve speech intelligibility (Li and Loizou, 2007). Roman *et al.* (2003) assessed the performance of an algorithm that used location cues and an ideal time–frequency binary mask to synthesize speech. Large improvements in intelligibility were obtained from partial spectrotemporal information extracted from the ideal time–frequency mask. Similar findings were also reported by Brungart *et al.* (2006), for a range of SNR thresholds (from −12 to 0 dB) used for constructing the ideal binary mask. A different method for constructing the ideal binary mask was used by Anzalone *et al.* (2006) based on comparisons of the speech energy detected in various bands against a preset threshold. The threshold value was chosen such that a fixed percentage (99%) of the total energy contained in the entire stimulus was above this threshold. Results with the ideal speech energy detector indicated significant reductions in speech reception thresholds for both normal-hearing and hearing-impaired listeners. Cooke (2006) used a computa-

a)Author to whom correspondence should be addressed. Electronic mail: loizou@utdallas.edu

tional model of glimpsing along with behavioral data collected from normal-hearing listeners on a consonant identification task. Close fits to listener's performance on a consonant task were obtained with local SNR thresholds in the range from −2 to 8 dB.

It is clear from the above-mentioned studies that the application of ideal binary masks is critically important for improving intelligibility of speech corrupted by interfering voices (or other types of maskers) by both normal-hearing and hearing-impaired listeners. In general, the study of ideal binary masks is important because it can be used for modeling the human auditory scene analysis process (Wang, 1996; Cooke, 2006), for evaluating the relative contributions of informational and energetic masking to overall perception of auditory mixtures (Brungart *et al.*, 2006), for improving the accuracy of automatic speech recognition systems (Cooke *et al.*, 1994, 2001; Srinivasan *et al.*, 2006) and for elucidating neurophysiological mechanisms underlying auditory scene analysis (McCabe and Denham, 1997; Alain, 2005).

In many of the previous studies, it is assumed that an ideal binary mask is available. In a practical system, the binary mask needs to be estimated from the noisy data, and that is a challenging task, particularly in adverse noisy conditions. As it is practically impossible to compute accurately the ideal binary mask for all T–F units, it is of interest to assess the effect of estimation errors on speech intelligibility. At issue is how accurate do we need to estimate the binary mask without compromising speech intelligibility. No studies have yet addressed that question. Other factors that may influence intelligibility of speech synthesized by the ideal (or estimated) binary mask include the choice of the local SNR threshold(s), the masker type (modulated versus steady state), speech materials, and input global SNR level. The effect of the local SNR threshold on speech intelligibility (and to some extent the masker type) was assessed in the study by Brungart *et al.* (2006) using the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000) as test material. The CRM test, however, is a closed-set test which is restricted to four phonetically distinct color alternatives and eight phonetically distinct number alternatives, both of which are easy to understand even in extremely noisy environments (Brungart *et al.*, 2001a, b). It is not known whether the intelligibility benefit seen in Brungart *et al.* (2006) carries through to other more challenging speech materials using the same range of SNR thresholds and input SNR levels. In the present study, we assess the intelligibility of ideal binary-masked speech using IEEE sentences as test material and speech-shaped noise or competing voices as maskers.

II. EXPERIMENT 1: EFFECTS OF SNR THRESHOLD AND INPUT SNR LEVEL

A. Methods

1. Subjects

Seven normal-hearing listeners participated in this experiment. All subjects were native speakers of American English, and were paid for their participation. The subjects' age ranged from 18 to 40 yrs, with the majority being undergraduate students from the University of Texas at Dallas.

2. Stimuli

The speech material consisted of sentences taken from the IEEE database (1969). All sentences were produced by a male speaker. The sentences were recorded in a sound-proof booth (Acoustic Systems) in our lab at a 25-kHz sampling rate. Details about the recording setup and copies of the recordings are available in Loizou (2007). The sentences were corrupted by a 20-talker babble (Auditec CD, St. Louis) at −5 and −10 dB SNR. These SNR levels were chosen to avoid floor effects. The babble interferer started at 100 ms before the beginning of each sentence and stopped at least 100 ms after the end of the sentence. The same babble segment was used for all sentences. The effect of using different types of maskers is investigated in Experiment 2.

3. Signal processing

The processing involved in the synthesis of ideal binary-masked (IdBM) stimuli made use of three signals: the target signal (prior to mixing), the masker (or interfering) signal and the resulting mixture. Each of these signals was first processed using a fast Fourier transform (FFT) applied to 20-ms segments of the signal (Hamming windowed) with 50% overlap between segments. A bank of 128 gammatone filters, with auditory-like frequency resolution, was used in Brungart *et al.* (2006) in place of the FFT. Following the earlier T–F decomposition, a comparison is made between the energy of the target and that of the masker. The resulting local SNR of each T–F unit is compared against a preset threshold value T to determine whether to retain the T–F unit (binary mask value is 1) or to eliminate it (binary mask value is 0). The computed pattern of binary mask values, consisting of 0's and 1's, is applied to the FFT magnitude spectrum of the mixture signal. The inverse FFT is finally applied to the modified magnitude spectrum to synthesize the IdBM stimuli. The phases of the mixture FFT spectra were used in the inverse FFT. Stimuli are synthesized in each 20-ms segment using the overlap-and-add method.

In the present experiment, we varied the local SNR threshold value T from −40 to +15 dB in steps of 5 and 10 dB, and assessed performance for each value of T. When the SNR threshold value is set to 0 dB, for instance, only target T–F units that have larger energy than the masker are kept, and the remaining units are zeroed out. Figure 1 illustrates the IdBM processing for two different values of SNR threshold (−10 and 0 dB) for a mixture embedded in multitalker babble at −5 dB SNR (measured from the rms energy of the utterance). The middle panels show the ideal binary masks, with white indicating a 1 and black indicating a 0. The bottom panels show the segregated mixtures. As can be seen from Fig. 1, the smaller the value of the SNR threshold is, the larger the number of T–F units retained, and in principle the original mixture corresponds to an SNR threshold value of $-\infty$. At the other extreme, increasing the SNR threshold reduces the total number of T–F units retained. The question addressed in this experiment, is what range of SNR thresholds is optimal in terms of obtaining the highest levels of speech intelligibility.

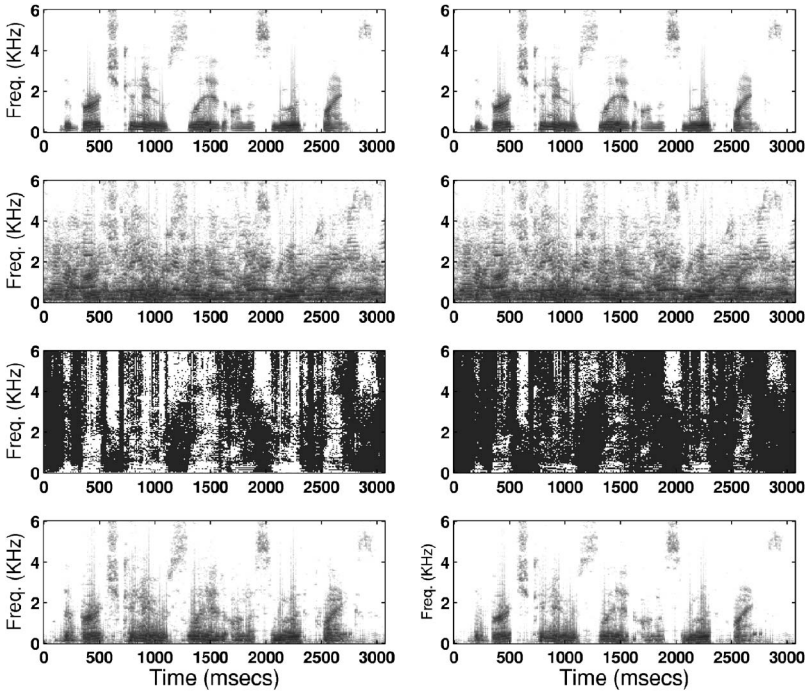


FIG. 1. Top row shows the spectrogram of a sentence in quiet from the IEEE corpus. The second row (from top) shows the spectrograms of the sentence embedded in multitalker babble at −5 dB SNR. The second row (from bottom) shows the ideal binary mask obtained using an SNR threshold of −10 dB (left) and 0 dB (right), with white pixels indicating a 1 (target stronger than the masker) and black pixels indicating a 0 (target weaker than the masker). Bottom row shows the segregated mixtures obtained with SNR thresholds of −10 dB (left) and 0 dB (right).

4. Procedure

The experiments were performed in a sound-proof room (Acoustic Systems, Inc) using a PC connected to a Tucker-Davis system 3. Stimuli were played to the listeners monaurally through Sennheiser HD 250 Linear II circumaural headphones at a comfortable listening level. Prior to the test, each subject listened to a set of noisy sentences to familiarize them with the testing procedure. During the test, the subjects were asked to write down the words they heard. Subjects participated in a total of 18 conditions (=2 SNR levels ×8 SNR thresholds +2 conditions involving the unprocessed mixtures). Two lists of sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. Sentences were presented to the listeners in blocks, with 20 sentences/block in each condition. The order of the test conditions was randomized across subjects.

B. Results and discussion

The mean scores for all conditions are shown in Fig. 2. Performance was measured in terms of percent of words identified correctly (all words were scored). Two-way analysis of variance (ANOVA) (with repeated measures) indicated a significant effect of local SNR threshold value ($F[7,42]=398.5$, $p<0.0005$), a significant effect of input SNR level ($F[1,6]=267.0$, $p<0.0005$), and a significant interaction ($F[7,42]=30.6$, $p<0.0005$).

Overall, the general pattern of performance is similar to that obtained by Brungart *et al.* (2006) with the CRM corpus. Significant gains in intelligibility are obtained with the IdBM processed speech for a range of SNR threshold values. Intelligibility of the −10 dB mixture improved from near 0% correct (unprocessed) to near 100% correct when processed using the ideal binary mask. Similarly, the intelligibility of the −5 dB mixture improved from 24% (unprocessed) to near 100%. The range of threshold values for which performance plateaued (near 100% correct) is slightly wider for mixtures at −5 dB SNR than for mixtures at −10 dB SNR. Protected post-hoc tests (Fisher's LSD) were run to determine the plateau region for each SNR level. Analysis indicated that the plateau region for mixtures in −5 dB SNR ranged from −20 to 5 dB, whereas the plateau region for mixtures in −10 dB SNR ranged from −20 to 0 dB. Brungart *et al.* (2006) reported a smaller region, −12 to 0 dB, however they used different test speech materials (CRM corpus) which lacked contextual cues.

Performance degraded markedly with IdBM stimuli for values of T smaller than −30 dB. We attribute this to the listeners being confused as to which T–F units belonged to the target and which units belonged to the masker. Interestingly enough, the ability of the listeners to segregate the target was not impaired at all when very strong masker units (by 20 dB) were included in the IdBM stimuli, i.e., with T

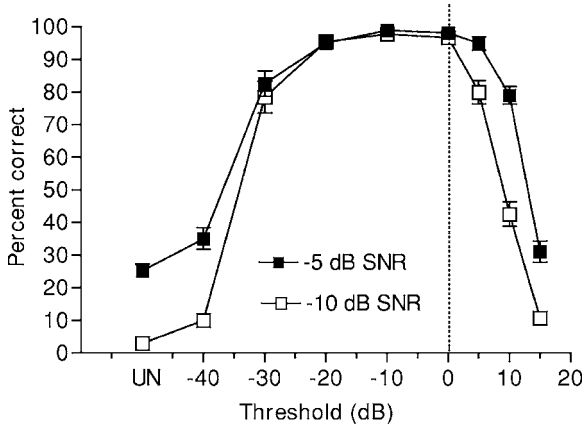


FIG. 2. Performance (percent of words identified correctly) as a function of SNR threshold (dB) for two input global SNR levels. The masker was 20-talker babble. Performance obtained with unprocessed mixtures is indicated as UN. Error bars indicate standard errors of the mean.

=−20 dB. Auditory masking effects, access to contextual cues and the overall pattern of the binary mask (in terms of location of masker units relative to the target units) could have contributed to that (see discussion in later section). Performance with $T=-40$ dB was not statistically different ($p>0.05$) from that attained with the unprocessed mixtures for both SNR levels.

Performance of the processed mixtures at −10 dB SNR dropped precipitously for positive values of T. This can be attributed to the fact that as the input SNR level of the mixture decreases, the number of $T-F$ units retained also decrease. In fact, for every 1 dB increase in T ($T>0$ dB), we observe a decrement in performance equivalent to that obtained when the input SNR level decreases by 1 dB and the value of T remains unchanged. This means that each 1 dB increase in T eliminates exactly the same $T-F$ units that would have been eliminated if T remained unchanged but the input SNR decreased by 1 dB. This can be seen in Fig. 2. Performance of IdBM stimuli obtained for mixtures at −5 dB SNR dropped from near 100% to 80% correct when T increased from 5 to 10 dB. Note that the same level of performance was obtained when the SNR of the mixtures decreased by 5 dB for the same value of T ($T=5$ dB). This outcome is consistent with that observed in Brungart *et al.* (2006) and has important implications in terms of assessing the energetic and informational components of speech-on-speech masking (see discussion in later section).

III. EXPERIMENT 2: EFFECT OF MASKER TYPE ON SPEECH INTELLIGIBILITY

In the previous experiment, we assessed performance using a single type of masker (20-talker babble). Acknowledging that performance might be affected by the use of different types of maskers, we examine in this experiment the performance of IdBM stimuli using modulated noise, steady-state noise and 2-talker maskers. This experiment will tell us whether the IdBM technique is more effective when the masking has both informational and energetic components, as that introduced by competing voices, or when the masking is purely energetic, as that introduced by steady-state noise.

A. Methods

1. Subjects and material

Seven new normal-hearing listeners participated in this experiment. All subjects were native speakers of American English, and were paid for their participation. Same speech material (IEEE, 1969) was used as in experiment 1.

2. Signal processing

Same signal processing technique was used as described in experiment 1. Three types of maskers were used. The first was continuous (steady-state) noise, henceforth referred to as SSN noise, which had the same long-term spectrum as the test sentences in the IEEE corpus. The second was two equal-level interfering talkers (female) based on two of the longest sentences in the corpus (this was done to ensure that all target sentences were shorter than the interferers). The third masker was a speech-shaped noise masker that was modulated to match the overall envelope of the IEEE sentences used for the 2-talker masker. We refer to the modulated speech-shaped noise as MSSN. The MSSN noise was constructed by modulating speech-shaped noise with the waveform envelope of the 2-talker masker. The envelope was extracted using a method similar to that described by Festen and Plomp (1990) by full-wave rectifying the speech masker and low-pass filtering (40-Hz cutoff) the rectified signal. The sentences were corrupted by the three types of maskers at −5 dB SNR.

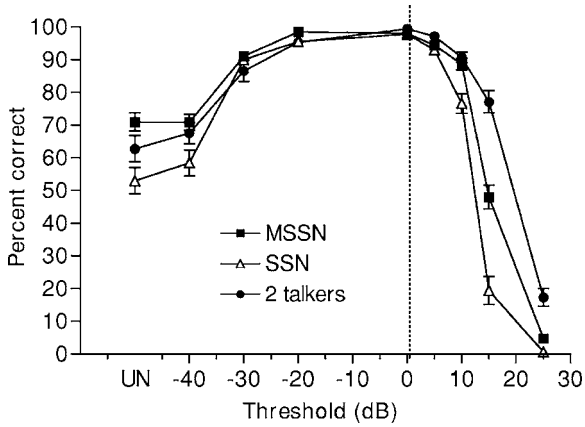


FIG. 3. Performance (percent of words identified correctly) as a function of SNR threshold (dB) for three types of maskers: modulated speech-shaped noise (MSSN), steady-state speech-shaped (SSN) noise and 2-talker masker. Performance obtained with unprocessed mixtures (at −5 dB SNR) is indicated as UN. Error bars indicate standard errors of the mean.

3. Procedure

The procedure was identical to that used in experiment 1. Subjects participated in a total of 27 conditions (=3 maskers ×8 SNR thresholds +3 conditions involving unprocessed mixtures). The SNR thresholds tested were the same as in Experiment 1, except for the addition of the $T=25$ dB condition and the elimination of the $T=-10$ dB condition. The latter condition was eliminated as it yielded (based on pilot data) the same performance as the $T=-20$ dB condition. Due to the large number of conditions involved, subjects performed the listening tests in two independent sessions on different days, with each session lasting approximately 2.5 to 3 h. Subjects were given 5 min breaks every 30 min of testing. Two lists of sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. The order of the test conditions was randomized across subjects.

B. Results and discussion

The mean scores for all conditions are shown in Fig. 3. Performance was measured in terms of percent of words identified correctly (all words were scored). Two-way ANOVA (with repeated measures) indicated a significant effect of local SNR threshold value ($F[7,42]=499.2$, $p<0.0005$), a significant effect of masker type ($F[2,12]=114.7$, $p<0.0005$), and a significant interaction ($F[14,84]=40.1$, $p<0.0005$).

The range of threshold values for which performance plateaued near 100% correct was similar for the three types of maskers. Protected post-hoc tests (Fisher's LSD) were run to determine the plateau region for each masker. For the SSN masker it ranged from −30 to +5 dB, for the MSSN masker it ranged from −20 to +5 dB and for the 2-talker masker it ranged from −20 to +10 dB.

Performance obtained with the three different types of maskers was very similar for negative values of T. Scores obtained with IdBM stimuli using $T=-40$ dB were not statistically different ($p>0.05$) from the scores obtained with the unprocessed mixtures for all three types of maskers. The difference in performance with the three types of maskers was more evident for positive values of T. Scores dropped significantly with the SSN masker when $T>5$ dB. The performance drop seen with the two modulated maskers (modulated noise and 2-talker masker) was also significant but more gradual. This outcome clearly illustrates the intelligibility benefit obtained when listening "in the gaps" of a fluctuating masker (Festen and Plomp, 1990). These gaps have presumably more favorable SNR which listeners exploit to hear out the target signal.

As demonstrated in the previous Experiment, the performance obtained with increasing T by say $+L$ dB is equivalent to that obtained by decreasing the input SNR by L dB at a fixed value of T (Brungart *et al.*, 2006). Making use of this approximation and after comparing the performance obtained with unprocessed mixtures with that obtained with positive values of T, we can conclude that the intelligibility improvement (in terms of SNR benefit) brought by the IdBM technique for SSN maskers is about 7 dB. Similarly, the improvement for MSSN maskers is 10 dB and for 2-talker maskers is near 15 dB. Thus it seems that the IdBM technique is more effective, in terms of improving intelligibility, when the target speech is masked by speech than when it is masked by noise. This outcome is consistent with that observed by Brungart *et al.*, (2006). The improvement brought by the IdBM technique for speech masked by noise is smaller (compared to the speech maskers), nevertheless it is quite significant, about 50 percentage points (Fig. 3).

IV. EXPERIMENT 3: EFFECT OF OVERALL BINARY MASK ERROR

In the previous experiments, we assumed that we had access to the ideal binary mask. In practice, however, the binary mask needs to be estimated from the mixtures. Algorithms (e.g., Hu and Wang, 2004) can be used in practice to estimate the SNR of each $T-F$ unit and subsequently the binary mask pattern. Such algorithms will possibly make errors in labeling each $T-F$ unit to 0 or 1, as we lack access to the masker signal. In the present experiment, we assess the effect of overall binary mask error on speech intelligibility. At issue is how accurate do algorithms need to be in estimating the binary mask without compromising the intelligibility brought by the IdBM technique.

A. Methods

1. Subjects and material

Seven new normal-hearing listeners participated in this experiment. All subjects were native speakers of American English, and were paid for their participation. The same speech material (IEEE, 1969) was used as in Experiment 1.

2. Signal processing

The ideal binary mask is first computed as in Experiment 1 with the use of the FFT operating on 20-ms segments of the signals (target, masker, and mixture). To create stimuli with varying degrees of binary mask error, we randomly selected a fixed percentage of $T-F$ units in each 20-ms frame and flipped the corresponding values of the ideal binary mask from 0 to 1 or from 1 to 0. We varied the percentage of binary mask error introduced in each 20-ms frame from 5% to 40%. More precisely, we tested the following amounts of error (in terms of percentage of $T-F$ units in each 20-ms frame): 5%, 10%, 20%, 30%, and 40%. The 5%-error condition, for instance, had 5% of the $T-F$ units marked wrongly (i.e., 0 was labeled as 1 and vice versa) and the remaining 95% of the $T-F$ units marked correctly (i.e., ideal binary mask was used) in each 20-ms frame. The new binary mask pattern containing fixed amounts of error in each frame was used to synthesize the stimuli using the same method described in Experiment 1.

To assess the effect of overall binary mask error on various maskers, we applied the previous technique to mixtures corrupted by three different types of masker: steady-state speech-shaped noise, 2-talker masker and 20-talker babble (same as in experiment 1). The former two maskers were the same as in Experiment 2.

3. Procedure

The procedure was identical to that used in experiment 1. Subjects participated in a total of 18 conditions (=3 maskers ×5 error values +3 conditions involving unprocessed mixtures). Two lists of sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. The order of the test conditions was randomized across subjects.

B. Results and discussion

The mean scores for all conditions are shown in Fig. 4. Performance was measured in terms of percent of words identified correctly (all words were scored). Performance obtained with the unprocessed mixtures is shown to the right (marked as "UN") for comparison. Two-way ANOVA (with repeated measures) indicated a significant effect of masker type ($F[2,12]=7.1$, $p=0.009$), a significant effect of binary mask error ($F[4,24]=432.3$, $p<0.0005$), and a nonsignificant interaction ($F[8,48]=0.5$, $p=0.791$).

The pattern of performance was similar for all three maskers. Scores remained high (near 100% correct) when the binary mask error was less or equal to 10%, and dropped relatively fast thereafter. In fact, for every 10% error introduced, performance dropped roughly by 20 percentage

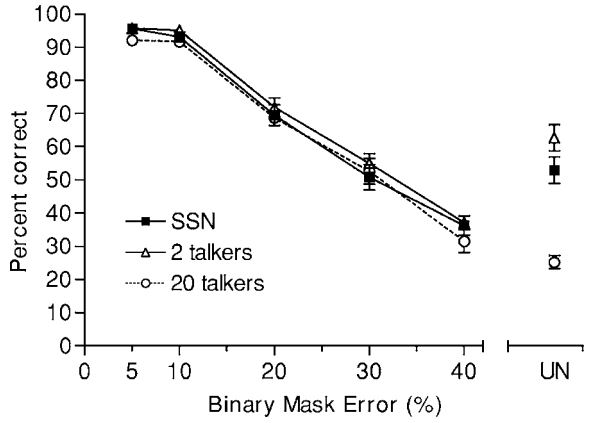


FIG. 4. Performance (percent of words identified correctly) as a function of the overall percentage of binary mask error introduced for three types of maskers: steady-state speech-shaped (SSN) noise, 2-talker masker, and 20-talker masker. Performance obtained with unprocessed mixtures (at −5 dB SNR) is indicated as UN. Error bars indicate standard errors of the mean.

points for both modulated and non-modulated maskers. Overall, there seems to be a strong (and negative) correlation between the amount of error introduced in the binary mask pattern and the intelligibility scores attained. In fact, the computed correlation coefficient between the binary mask error (percentage) and intelligibility score for the 15 conditions tested was quite high, $\rho = -0.98$ ($p < 0.005$). We should acknowledge, however, that this correlation was computed using the IdBM stimuli from which we can compute the overall error.

Regarding the question posed in this experiment as to how accurate do binary-mask estimation algorithms need to be, we observe from Fig. 4 that the answer depends on the type of masker and the set expectations. If the goal is to restore speech intelligibility (to the level attained in quiet), then the algorithm needs to produce at most 10% error when estimating the binary mask. This applies for all three types of maskers tested. If the goal is to improve speech intelligibility relative to that of the unprocessed mixtures, then different amounts of error can be tolerated depending on the type of masker. For the 20-talker masker, the overall error needs to be less than (or equal to) 30%, whereas for the 2-talker masker and steady-state noise, the error needs to be less than (or equal to) 20%.

V. EXPERIMENT 4: EFFECT OF TYPE OF BINARY MASK ERROR

In the previous experiment, we assessed the effect of the overall error in the binary mask pattern making no distinction between the two types of error that can occur. The first type of error occurs when a $T-F$ unit that was originally labeled as 0 (i.e., local SNR of $T-F$ unit is less than threshold T) is purposefully modified to 1. The second type of error occurs when a $T-F$ unit that was originally labeled as 1 (i.e., local SNR of $T-F$ unit is greater than threshold T) is purposefully modified to 0. From signal detection theory, we can say that the first type of error is similar to type I error (false alarm) and the second type of error is similar to type II error[1] (miss). Hence, for the purpose of discussion, we will refer to these two errors as type I and type II errors. The type I error will possibly introduce more noise distortion or more target-masker confusion, as $T-F$ units that would otherwise be zeroed-out (presumably belonging to the masker or dominated by the masker) would now be retained. The type II error will likely introduce target speech distortion, as it will zero out $T-F$ units that are dominated by the target signal and should therefore be retained. The perceptual effect of these two types of errors introduced in the binary masking pattern is likely different, and this is assessed in the present experiment.

A. Methods

1. Subjects and material

The same subjects used in experiment 3 participated in this experiment on a different day. Same speech material (IEEE, 1969) was used as in Experiment 1. None of the sentence lists was repeated.

2. Signal processing

The ideal binary mask is first computed as in experiment 1 with the use of the FFT operating on 20-ms segments of the signals (target, masker and mixture). To create stimuli with varying degrees of type I and type II binary mask errors, we followed a procedure similar to that in experiment 3. As we wanted to assess independently the effect of type I and II errors, we kept for the type-II stimuli all $T-F$ units originally labeled as 0 (according to the ideal binary mask) and introduced varying degrees of error only to units originally labeled as 1. Hence, we created type-II stimuli by introducing a fixed percentage of errors only to the $T-F$ units labeled as 1 (according to the ideal binary mask). No errors were introduced to the $T-F$ units originally labeled as 0. Similarly, we created type-I stimuli by introducing a fixed percentage of errors only to the $T-F$ units labeled as 0. No errors were introduced to the $T-F$ units originally labeled as 1. We varied the percentage of type I/II errors introduced to the $T-F$ units in each 20-ms frame, from 20% to 95%. More specifically we tested the following amounts of error (in percentage of $T-F$ units available in each frame): 20%, 40%, 60%, 70%, 80%, 85%, 90%, and 95%. In the 20% type-II error condition, for instance, 20% of the $T-F$ units in each 20-ms frame that were originally marked as 1 were flipped to 0, whereas the remaining units were kept intact. That is, no errors were introduced to the $T-F$ units originally labeled as 0. The new binary mask pattern containing fixed amounts of type I/II error was applied to the mixtures. The same method described in experiment 1 was used to synthesize the stimuli.

Given the limited number of lists available in the IEEE corpus, we applied the above technique only to mixtures embedded in 20-talker babble at −5 dB SNR. As the local SNR threshold of $T = 0$ dB was found to be quite effective in the previous experiments for all types of maskers and SNR levels, we assessed the effect of type I and II errors using $T = 0$ dB.

3. Procedure

The procedure was identical to that used in experiment 1. Subjects participated in a total of 17 conditions (=2 types

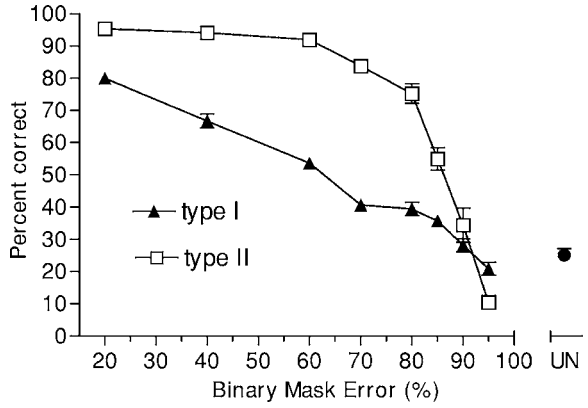


FIG. 5. Performance (percent of words identified correctly) as a function of the percentage of binary mask error introduced for two types of error, type I and type II (see the text for details). The masker was 20-talker babble. Performance obtained with unprocessed mixtures (at −5 dB SNR) is indicated as UN. Error bars indicate standard errors of the mean.

of error × 8 error values + 1 condition involving unprocessed mixtures). Two lists of sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. The order of the test conditions was randomized across subjects.

B. Results and discussion

The mean scores for all conditions are shown in Fig. 5. Performance was measured in terms of percent of words identified correctly (all words were scored). Performance obtained with the unprocessed mixtures is shown to the right (marked as UN) for comparison. Two-way ANOVA (with repeated measures) indicated a significant effect of the type of error ($F[1,6]=163.7$, $p<0.0005$), a significant effect of the amount of error ($F[7,42]=242.2$, $p<0.0005$), and a significant interaction ($F[7,42]=44.2$, $p<0.0005$).

It is clear from Fig. 5 that the type of error introduced in the binary mask pattern affected performance differently and to a different extent. There was a gradual degradation in performance when type II errors were introduced. Performance remained high (near 100% correct) even when 60% of the $T-F$ units originally labeled as 1 were purposefully switched to 0 (assuming no errors were introduced in the $T-F$ units originally labeled as 0). Post-hoc tests (Fisher's LSD) indicated that performance obtained with 60% type-II error was not statistically different ($p=0.377$) from performance obtained with 20% error. As one would expect, performance dropped substantially when the error exceeded 80%, as the overwhelming majority of the target-dominated $T-F$ units were eliminated.

In stark contrast, the type I error affected performance dramatically, even when small amounts (20%) of error were introduced. Performance dropped from nearly 100% correct obtained with the ideal binary mask (0% error, Fig. 2) to 80% correct with 20% error. Note that with this type of error, all $T-F$ units originally labeled as 1 remained intact, i.e., no errors were introduced in the target-dominated $T-F$ units. We attribute the dramatic decrease in performance with type I errors to the following two reasons. First, the number of $T-F$ units labeled by the ideal binary mask as 0 (i.e., with local SNR <0 dB) is substantially larger than the corresponding number of $T-F$ units labeled as 1 (i.e., with local SNR >0 dB). Consequently, a larger number of $T-F$ units are wrongly labeled when type I error is introduced in the binary masking pattern than when type II is introduced despite the fact that percentage of error is the same for the two types of error. To corroborate this observation, we plot in Fig. 6 the histogram of the percentage of $T-F$ units in each frame that have local SNR threshold greater or smaller than 0 dB (input mixture was corrupted by 20-talker babble at −5 dB SNR). Based on the histogram average (Fig. 6), 80% of the $T-F$ units in each 20-ms frame are labeled 0 (i.e., local SNR <0 dB) and the remaining 20% are labeled as 1 (i.e., local SNR >0 dB). Hence, the overwhelming majority of the $T-F$ units are labeled as 0. The second reason for the

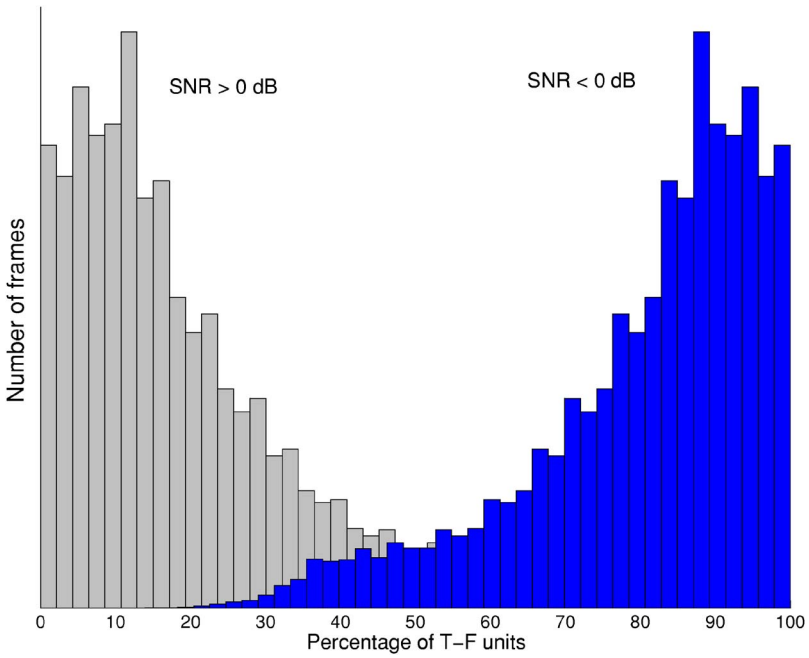


FIG. 6. (Color online) Histograms of the percentage of $T-F$ units in each 20-ms frame falling below the SNR threshold (0 dB) and above the SNR threshold. The mixture was corrupted at −5 dB SNR in 20-talker babble. The histogram was computed using all 720 sentences in the IEEE corpus and was based on a total of 190 386 frames.

dramatic decrease in performance with type I errors is that it increases the target-masker confusions. For the masker (competing voices) used in this experiment, type I errors increase the amount of informational masking. As more $T-F$ units are introduced with large and negative local SNRs, listeners have greater difficulty distinguishing the target voice from the interfering voices.

Figure 5 superimposes the performance obtained with the unprocessed mixtures. Analysis (paired samples t-tests) indicated that in order to obtain significant ($p<0.005$) improvement in speech intelligibility, the type I and II errors need to be lower than 85%. There is no statistically significant ($p>0.05$) difference between the performance obtained with unprocessed mixtures and that obtained with 90% type I or type II error. These results are interesting as they provide insights as to the amounts of type I or II "errors" typically present in unprocessed mixtures at −5 dB SNR. They also provide performance bounds on algorithms that can potentially be used to estimate the ideal binary mask. In the context of such algorithms, it is clear from Fig. 5 that it is more important to find techniques to keep the type I error low than keep the type II error low, although understandably a balance needs to be struck between the magnitude of the two errors. The consequences of type I error are counterintuitive as one would expect that it is more important to ensure that all target-dominated $T-F$ units are labeled correctly rather than ensuring that all masker-dominated $T-F$ units are labeled correctly. That is not the case, however, as can be seen from Fig. 5 by contrasting the performance obtained with type I and II errors for the same percentage of error. Taking the 60% error as an example, we see that performance dropped to 50% correct when type I error was introduced despite the fact that all target-dominated units were classified correctly. In contrast, when type II error was introduced with all masker-dominated units classified correctly, there was no noticeable decrease in performance (remained near 100% correct).

VI. DISCUSSION AND CONCLUSIONS

The outcomes in experiments 1 and 2 have important implications for understanding speech segregation in multitalker environments. Equally important are the implications of experiments 3 and 4 for designing algorithms capable of estimating the ideal binary mask with the intent of suppressing interfering noise or competing voices for improved speech intelligibility.

A. Speech segregation

The present study replicated and extended the findings of the study by Brungart *et al.* (2006). There are a number of similarities and a number of differences between the outcomes of the two studies. The overall pattern of performance was very similar in the two studies. Performance reached a plateau (near 100% correct) in the region near 0 dB SNR threshold and dropped for large and small values of the local SNR threshold. This pattern was observed for all types of maskers tested whether be modulated or steady-state. The application of the IdBM technique is more effective, in terms of intelligibility improvement, when speech is masked by speech (largely informational masking with some energetic masking) than when it is masked by steady-state noise (purely energetic masking). Experiment 1 confirmed the approximation that each 1 dB increase in the SNR threshold eliminates the same $T-F$ units that would be have been eliminated by each 1 dB decrease in the global SNR. With this approximation, we can say that the amount of informational masking produced by the speech maskers (2 talkers) in experiment 2 was roughly 20 dB, whereas that produced by the (energetic) noise masker is about 10–12 dB. Note that Experiment 2 used different-sex masker and target voices; hence, a greater amount of informational masking could potentially be introduced by using same-sex masker and target voices (Brungart, 2001a).

The main difference between the outcomes of this study and that of Brungart *et al.* (2006) was the plateau region. In the Brungart *et al.* study, the plateau region ranged from −12 to 0 dB, whereas in the present study, the plateau region ranged in most cases from −20 to 5 dB and in the steady-state noise case from −30 to 5 dB. We attribute this difference primarily to the use of different speech materials. In our study, listeners had access to contextual cues which we believe enabled them to segregate the target speech easily even after including strong (by 20 dB) masker $T-F$ units in the synthesized stimuli. In contrast, there is no contextual information present in the CRM phrases. Taking the outcomes of both studies into account, we can conclude that the width of the performance plateau region (near 100% correct) is greatly affected by the speech material used. Another potential factor contributing to the differences in the plateau region between the two studies might be the use of different $T-F$ decomposition techniques. The present study used the FFT, which implicitly uses a uniform frequency spacing to construct the IdBM stimuli. In contrast, Brungart *et al.* (2006) used a bank of 128 gammatone filters, with auditory-like frequency resolution (i.e., progressively wider filters in the high frequencies) in place of the FFT. Further experiments are warranted to examine the impact of the $T-F$ decomposition technique (FFT versus auditory-like filterbank) on the intelligibility of IdBM stimuli.

The existence of a wide plateau region has important implications for speech segregation in multitalker environments. In this region, the stimuli contain not only target $T-F$ units, but also units with stronger masker energy. In fact, in some cases the masker is stronger than the target by 20 dB. Yet, these masker units did not interfere with the recognition of the target speech (performance remained near 100% correct). We believe that this is not because there were not enough of them (see Fig. 6) but rather because they occurred in regions that already contained enough target energy. The existence of the plateau regions suggests that it is the pattern of the ideal binary mask (i.e., the pattern of target-dominated and masker-dominated $T-F$ units) that matters the most and not the local SNR of each $T-F$ unit, since conceivably different local SNR thresholds may yield the same binary mask pattern (Brungart *et al.*, 2006). It is this pattern that directs the listener's attention to *where* (in a $T-F$ auditory space) the target is, and this can be viewed as a perceptual mecha-

TABLE I. Errors (in percentage of $T-F$ units misclassified) produced by two noise reduction algorithms in estimating the ideal binary mask for different values of the SNR threshold (dB). The overall error is computed (as in Experiment 3) by counting the misclassifications across all $T-F$ units in each frame making no distinction between the two types of error.

		SNR threshold T			
Algorithm	Error	−10 dB	−5 dB	0 dB	5 dB
Wiener[a]	Type I (%)	31.73	26.27	18.30	10.29
(Scalart & Filho, 1996)	Type II (%)	21.69	18.73	14.65	10.45
	Overall (%)	53.42	45.00	32.95	20.74
Log MMSE[b]	Type I (%)	24.66	16.49	9.72	5.57
(Ephraim & Malah, 1985)	Type II (%)	25.13	21.54	16.40	11.27
	Overall (%)	49.79	38.04	26.12	16.85

[a]Scalart and Filho (1996).
[b]Ephraim and Malah (1985).

nism that does not require the listeners to extract specific target information within each $T-F$ unit and somehow group all pieces of the detected $T-F$ units to hear out the target. The ideal binary mask pattern provides the cue to *where* the target is, an essential cue that is probably used in the second (or later) stage of the speech segregation process. The importance of the *where* cue is also observed and mirrored (perhaps more clearly) in spatial hearing. Information about where the target is located in space can greatly enhance its identification (Kidd *et al.*, 2005; Best *et al.*, 2007).

B. Noise reduction

The present study, as well as others, have demonstrated the full potential of using the ideal binary mask to improve (and in some cases restore) intelligibility of speech in multitalker or other noisy environments. Algorithms capable of estimating the ideal binary mask accurately can therefore yield significant gains in intelligibility, which according to experiment 1 can range from 70 to 100 percentage points (experiment 1) improvement for input global SNRs of −5 and −10 dB, respectively. As demonstrated in experiment 3, these algorithms need to be quite accurate (>90% accurate) in estimating the ideal binary mask, at least for acoustic mixtures in −5 dB SNR. A 10% (or smaller) overall error in estimating the ideal binary mask is acceptable without compromising speech intelligibility (see Fig. 4), at least for mixtures in −5 dB multitalker babble. Experiment 4 showed that the type of error made in classifying target and masker dominated $T-F$ units is perhaps more important than the overall amount of error. Performance is affected the most when the masker dominated $T-F$ units are wrongly labeled as target-dominated $T-F$ units. In contrast, performance remains high (near 100% correct) if all masker-dominated $T-F$ units are correctly classified but the target-dominated $T-F$ units are misclassified with 60% (or less) error. The outcome in experiment 4 suggests that a right balance needs to be struck between the two types of errors, as both can affect performance, although to a different degree. This tradeoff between type I and type II errors is often summarized in detection theory via the receiver operating characteristic (ROC) curves (e.g., Kay, 1998). Each pair of type I and (1-type II) values provides a different point in the ROC curve, with best performance obtained when both types of error are small.

A number of techniques have been proposed in the CASA literature (see review in Wang and Brown, 2006) for estimating the ideal binary mask and include methods based on pitch continuity information (Hu and Wang, 2004; Roman and Wang, 2006) and sound-localization cues (Roman *et al.*, 2003). In the method by Hu and Wang (2004), the individual $T-F$ units are labeled according to the similarity between the periodicity pattern of the correlogram response and the dominant pitch of each frame. This method (and the method in Roman and Wang, 2006), however, is only applicable to voiced utterances and was evaluated using only objective measures and not with listening tests. Most of the CASA techniques proposed thus far are based on elaborate auditory models and make extensive use of grouping principles (e.g., pitch continuity, onset detection) to segregate the target from the mixture. Alternatively, the ideal binary mask can be estimated using simpler algorithms that compute the SNR in each $T-F$ unit and compare the estimated SNR values against a threshold. Several such algorithms do exist and are commonly used in speech enhancement applications to improve the quality of degraded speech (see review in Loizou, 2007). To assess how accurate are such algorithms, we processed the −5 dB SNR mixtures (20-talker babble) via two conventional noise reduction algorithms, which we found in a previous study to preserve intelligibility (Hu and Loizou, 2007a), and computed the percentage of errors made in classifying target-dominated and masker-dominated $T-F$ units (details are given in the Appendix). Overall, the amounts of errors made by either algorithm are quite high (see Table I), thus providing a plausible explanation as to why current noise reduction algorithms cannot improve speech intelligibility, although they improve speech quality (Hu and Loizou, 2007b). More research is thus warranted in developing algorithms capable of estimating more accurately the ideal binary mask.

ACKNOWLEDGMENTS

The authors would like to thank Yang Lu for his help in some experiments. The authors would also like to thank the associate editor Dr. Mitch Sommers for all his comments and

suggestions. This research was supported by Grant No. R01 DC007527 from the National Institute of Deafness and other Communication Disorders, NIH.

APPENDIX

To assess the accuracy of conventional noise reduction algorithms in estimating the binary mask, we processed the −5 dB SNR mixtures (20-talker babble) via two conventional noise reduction algorithms, which we found in a previous study to preserve intelligibility (Hu and Loizou, 2007a). The two algorithms were first used to estimate the instantaneous SNR in each $T-F$ unit as follows:

$$\mathrm{SNR}(t,k) = a\frac{X_k^2(t-1)}{\hat{D}_k^2(t-1)} + (1-a)\max\left[\frac{Y_k^2(t)}{\hat{D}_k^2(t)} - 1, 0\right] \tag{A1}$$

where $\mathrm{SNR}(t,k)$ is the estimated SNR at time frame t and frequency bin k, $a=0.98$, $X_k^2(t-1)$ denotes the power-spectrum of the enhanced target signal obtained in the previous frame, $\hat{D}_k^2(t)$ denotes the estimated power-spectrum of the masker [obtained using a noise-estimation algorithm (Rangachari and Loizou, 2006)], and $Y_k^2(t)$ denotes the mixture power spectrum. The two noise reduction algorithms differed in the way they estimated $X_k^2(t-1)$ in Eq. (A1), but both used Eq. (A1) to estimate the instantaneous SNR. The estimated SNR of each $T-F$ unit was compared against a threshold (varying from −10 to 5 dB), and $T-F$ units with positive SNR were classified as target-dominated $T-F$ units and units with negative SNR were classified as masker-dominated units. The binary mask pattern estimated using the two algorithms was compared against the (true) ideal binary mask pattern. Errors were computed in each frame by comparing the true decision made by the ideal binary mask with the decision made by the SNR-estimation algorithm [Eq. (A1)] for each $T-F$ unit. The percentage of type I/II errors were averaged across the 20 utterances tested. The results are given in Table I for different values of the local SNR threshold T. With $T=0$ dB (which is the value used in experiment 4) the type I and II errors are relatively large (>15%) and the overall error is even larger (~30%). From Fig. 5, we see that in order to obtain significant improvement in intelligibility, the type II error needs to be near 0% when the type I error is smaller than 20%. Changing the SNR threshold value affects markedly the amounts of type I and II errors introduced and does so in a manner that signifies the tradeoff between the probability of false alarm (type I error) and probability of detection (1-type II error), typically seen in ROC curves.

[1]Type I error (also called *false alarm*) is produced when deciding hypothesis H_1 (signal is present) when H_0 is true (signal is absent). Type II error (also called *miss*) is produced when deciding H_0 when H_1 is true (Kay, 1998).

Alain, C. (**2005**). "Speech separation: Further insights from recordings of event-related brain potentials in humans," *Speech Separation by Humans and Machines* edited by P. Divenyi (Kluwer Academic, Dordrecht), pp. 13–30.

Anzalone, M., Calandruccio, L., Doherty, K., and Carney, L. (**2006**). "Determination of the potential benefit of time-frequency gain manipulation," Ear Hear. **27**, 480–492.

Best, V., Ozmeral, E., Shinn-Cunningham, B. (**2007**). "Visually-guided attention enhances target identification in a complex auditory scene," J. Assoc. Res. Otolaryngol. **8**(2), 294–304

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

Bregman, A. (**1990**). "Auditory scene analysis," Cambridge, MA, MIT Press.

Brungart, D. (**2001a**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D., Chang, P., Simpson, B., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Brungart, D., Simpson, B., Ericson, M., and Scott, K. (**2001b**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

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

Cooke, M. P., Green, P. D., and Grawford, M. D. (**1994**). "Handling missing data in speech recognition," Proceedings of the Third International Conference of Spoken Language Processes, 1555–1558.

Cooke, M. P., Green, P. D., Josifovski, L., and Vizinho, A. (**2001**). "Robust automatic speech recognition with missing and uncertain acoustic data," Speech Commun. **34**, 267–285.

Ephraim, Y., and Malah, D. (**1985**). "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-23**, 443–445.

Festen, J., 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.

Hu, G., and Wang, D. (**2004**). "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw. **15**, 1135–1150.

Hu, Y., and Loizou, P. (**2007a**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am. **122**(3), pp. 1777–1786.

Hu, Y., and Loizou, P. (**2007b**). "Subjective comparison and evaluation of speech enhancement algorithms," Speech Commun. **49**, 588–601

IEEE. (**1969**). "IEEE Recommended Practice for Speech Quality Measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Kay, S. (**1998**). *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice Hall, Upper Saddle River, NJ).

Kidd, G., Arbogast, T., Mason, C., and Gallun, F. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**(6), 3804–3815

Li, N., and Loizou, P. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**(2), 1165–1172.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, Taylor Francis Group, Boca Raton, FL).

McCabe, S., and Denham, M. (**1997**). "A model of auditory streaming," J. Acoust. Soc. Am. **101**, 1611–1621.

Rangachari, S., and Loizou, P. (**2006**). "A noise estimation algorithm for highly non-stationary environments," Speech Commun. **28**, 220–231.

Roman, N., and Wang, D. (**2006**). "Pitch-based monaural segregation of reverberant speech," J. Acoust. Soc. Am. **120**, 458–469.

Roman, N., Wang, D., and Brown, G. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.

Scalart, P., and Filho, J. (**1996**). "Speech enhancement based on a priori signal to noise estimation," *Proceedings of the IEEE International Conference on Acoustic, Speech, Signal Processing*, 629–632.

Srinivasan, S., Shao, Y., Jin, Z., and Wang, D. (**2006**). "A computational auditory scene analysis system for robust speech recognition," *Proceeding of Interspeech*, pp. 73–76.

Wang, D. (**1996**). "Primitive auditory segregation based on oscillatory correlation," Cogn. Sci. **20**, 409–456.

Wang, D. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Dordrecht), pp. 181–187.

Wang, D., and Brown, G. (**2006**). *Computational Auditory Analysis* (Wiley, New York).

Acoustical measurements of expression devices in pipe organs[a)]

Jonas Braasch[b)]
School of Architecture, Rensselaer Polytechnic Institute, Troy, New York

(Received 4 January 2007; revised 23 July 2007; accepted 2 December 2007)

In this investigation, three different swell systems known in pipe organs, the swell box, the crescendo wheel, and the historic wind swell were measured and compared to each other. The dynamic range of the crescendo wheel was found to be most effective, and for frequencies near 2 kHz the increase in sound pressure level could be up to 50 dB between the softest and the loudest adjustment. The maximum dynamic range for the wind swell and the swell box were found to be 10–20 dB in the same frequency range. With its step-wise crescendo procedure, the crescendo wheel simulates the type of orchestra crescendo which is reached by successively adding further musical instruments. In contrast, the swell box and the wind swell produce a crescendo effect similar to the crescendo in which individual musical instruments perform a dynamic movement. This type of crescendo requires a continuous level increase but allows a smaller dynamic range. The disappearance of the wind swell is not surprising because it offers no advantage over the swell box, while being restricted to stops with free reeds. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2828062]

PACS number(s): 43.75.Np, 43.75.Lm [NHF] Pages: 1683–1693

I. INTRODUCTION

This study reports on measurements on the effectiveness of different expression devices in pipe organs. For this purpose, data were collected on different swell boxes, crescendo wheels, and wind swells to determine acoustical reasons for why these expression devices are used in different musical contexts. The second purpose of this investigation was to explain on the basis of these measurements why two of the three devices, the swell box and the crescendo wheel, are still in use today, while the third device, the wind swell, is not commonly used anymore. This study focuses on the German pipe organ tradition, but the results can be applied to other traditions as well. To the author's knowledge, only a small conference abstract and two brief patent reports have been previously published on the acoustics of the swell box (Meyer and Wayne, 1961; Peterson, 1982; Peterson and Mornar, 1983). The article will begin with a brief history on the development of expression devices, which will later become important to follow the conclusions derived from this work.

Soon after the introduction of the orchestra crescendo during the "Mannheim school" in the second half of the 18th century, the ability to play a pipe organ expressively became a major issue in pipe organ building because the pipe organ was then considered to be an unlively instrument. To overcome this limitation, one of the leading pipe organ virtuosos at this time, Georg Joseph Vogler, was dedicated to find a way to equip the pipe organ with an expression device (Grave and Grave, 1987). According to the literature, Vogler found a solution to his problem while visiting Russia, where he was familiarized with free-reed pipes which had just been invented by Christian Gottlieb Kratzenstein to serve in his new speaking machine (Kratzenstein, 1780; Ahrens and Braasch, 2002). In contrast to the beating reeds which are usually used in pipe organs, free reeds swing through their mounting frame and aside from having a different tone quality, their tuning changes only insignificantly when their driving wind pressure is varied. Vogler realized that when using free reeds, the so-called gazé or wind swell can be utilized in pipe organs. The wind swell is a valve with a variable opening that controls the air pressure with which the pipe is supplied. Until Vogler's discovery, attempts to equip instruments with a wind swell were unsuccessful, because both beating reeds and flue pipes change their fundamental frequency audibly when the wind pressure is varied.

Interestingly, Vogler came up with a second solution to operate the pipe organ dynamically: the swell box. In this device, the stops are built into a box, which is equipped with doors that can be opened and closed by a hand lever or pedal in order to change the sound level of the instrument. While Vogler was the first person to employ the swell box in the German pipe-organ tradition, it had already been invented in England toward the beginning of the 18th century by Renatus Harris or the Abraham Jordans, a father and son sharing the same name. While it is still not known whether Harris or the Jordans were the first to build the swell box (Jordan, 2002), other sources suggest that the cradle of the swell box might be found in Spain (Bicknell, 1999).

The third device that was invented to enable a pipe organ to be played dynamically was the crescendo wheel. Nearly at the same time as Vogler's discoveries, Justin Heinrich Knecht described how to achieve a crescendo effect by manually adding a number of stops successively (Knecht, 1795). It took several years after the appearance of Knecht's treatise until the first mechanical device in the form of a

[a)]Portions of this work were presented in "A comparison of different expression devices in pipe organs," at the 143rd Meeting of the Acoustical Society of America, San Diego, CA, 2004 and "Expression Devices in Pipe Organs," Proceedings of the Joint Congress CFA/DAGA, Strasbourg, France, March 2004.

[b)]Electronic mail: braasj@rpi.edu

knee-lever system was developed by Georg Christian Friedrich Schlimbach to relieve the hands from setting the stops manually (Schlimbach, 1806). Nowadays, all pipe-organ stops typically are included in this process, and the progression of the stops is controlled via a foot wheel, giving the device its name: *crescendo wheel*. In modern instruments, the order of stop progression can be programmed freely. Usually, the crescendo will start with softly intonated 8 and/or 16 ft flue pipes. At the end, loud beating-reed pipes, for example a trumpet and large stops like the Trombone 32 ft, are added.

Today, the swell box and the crescendo wheel are commonly employed in newly built pipe organs, while, with few exceptions, the wind swell has gone out of use. The description of the acoustic measurements will be divided into two sections. Section II covers the swell box and the crescendo wheel, and Sec. III addresses the wind swell as well as the harmonium. An overall comparison of the swell devices and a general discussion concludes the paper.

II. CURRENT EXPRESSION DEVICES: THE SWELL BOX AND THE CRESCENDO WHEEL

A. Methods

For the measurements of the pipe organs, a microphone (AKG, C-414) was placed outside the pipe organ case in the auditorium area at a sitting position close to the instrument. In most conditions, a c-major chord was played for both the lowest and the highest volume of the swell device. For each instrument, the sound of a stop or a combination of stops was captured at a sampling rate of 48 kHz and a resolution of 16 bit, using a DAT recorder (Tascam, DA-P1). The frequency spectra were calculated from 2 s long samples within the MATLAB 7.0 programming environment. For this task, the fast Fourier transform algorithm was applied with a frequency resolution of 0.366 Hz (131,072 coefficients). To determine the frequency-dependent dynamic range of each expression device, the following procedure was applied. First, the two sound samples for the lowest and highest settings were analyzed using a filter bank with third-octave band filters. Altogether, 109 Chebyshev Type-I IIR filters were used with 0.5 dB of peak-to-peak ripple. For the high frequencies seventh-order filters were used, but the order was gradually lowered with decreasing frequency until third order at 27.5 Hz to avoid filter instability. The center frequencies were spaced in semitones from 27.5 Hz to 14 kHz for a fine frequency resolution. Next, the sound pressure level was computed for each frequency band, and the smoothed spectrum for the lowest setting was subtracted from the smoothed spectrum of the highest setting to calculate the dynamic range of the expression device.

Altogether three pipe organs were measured. The largest instrument measured was the Klais pipe organ of the Auditorium maximum of the Ruhr-University Bochum, Germany. The pipe organ was built in 1998 and is, with 82 stops, one of the largest concert pipe organs in the area. The second instrument was built in 1874 by Franz Wilhelm Sonreck for St. Laurentius church in Essen-Steele, Germany. The pipe organ was extended and changed several times. In the last restoration of 1996, the instrument was brought back to the state of 1906, the year it had been extended by Klais. The instrument presently has 37 stops, and the original design included a crescendo wheel, which was never built for financial reasons. In this study, the dynamic range of the crescendo wheel was therefore simulated by manually setting the stop combinations that typically would be addressed with a crescendo wheel. The third instrument analyzed is a romantic pipe organ with 42 stops built by Eggert in 1905. The pipe organ is located in the catholic church of Wattenscheid-Höntrop near Bochum, Germany. The pipe organ was recently restored, and it is principally in the original condition of 1905.

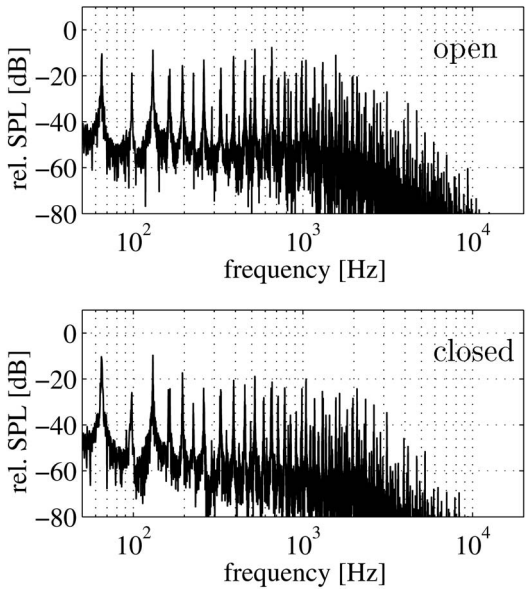


FIG. 1. Frequency spectrum of a c-major chord (C_2,C_3,C_4,E_4,G_4,C_5) with a typical German stop combination of the Klais pipe organ at the Ruhr-University Bochum. Both plots show the relative sound pressure levels (rel. SPL) in decibels. The top panel shows the spectrum for the open swell box, the bottom panel the spectrum for the closed swell box.

B. Results

In Fig. 1 (top graph), the frequency spectrum for a c-major chord is displayed for the Klais instrument at the Ruhr-University Bochum. The swell box was open during the recording, and the chord was played with a stop configuration that is typical for German pipe-organ music: *Bourdon* 16 ft, *Diapason* 8 ft, *Flûte traversiere* 8 ft, *Konzertflöte* 8 ft, *Octave* 4 ft, *Flûte octav*. 4 ft, *Violine* 4 ft, *Doublette* 2 ft, *Nasat* $2\frac{2}{3}$ ft, *Kleincornet* 3*f*, *Plein jeu harm.* 3-5*f*, *Basson* 16 ft, *Tromp. Harm.* 8 ft. The bottom graph of Fig. 1 shows the frequency spectrum for the same chord and instrumentation except that the swell box was closed during the measurement. In comparison to the open-swell-box case, the particles in the midfrequency range around 1 kHz are attenuated by up to 10 dB, while the low frequencies are hardly affected.

To demonstrate this characteristic influence of the swell box on the pipe organ sound more effectively, the signal transformation through the swell box was measured for three door positions: fully closed door, half-open door (45° door

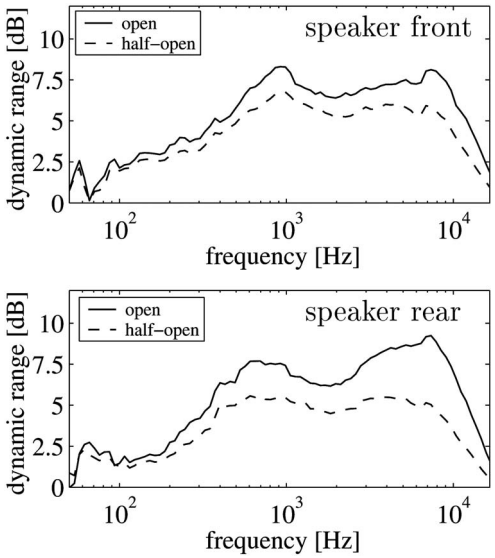


FIG. 2. Spectral difference between the open and the closed door setting of the swell box (solid line), and the spectral difference between the half-open and the closed door setting (dashed line). In the top graph, the loudspeaker was placed on the floor of the case near the swell door. In the bottom graph, it was placed 3 m behind the first measurement position toward the rear of the organ case. In both graphs, the transfer function between a loudspeaker placed in the swell box and the left ear of a dummy head is shown for the Klais pipe organ at the Ruhr-University Bochum.

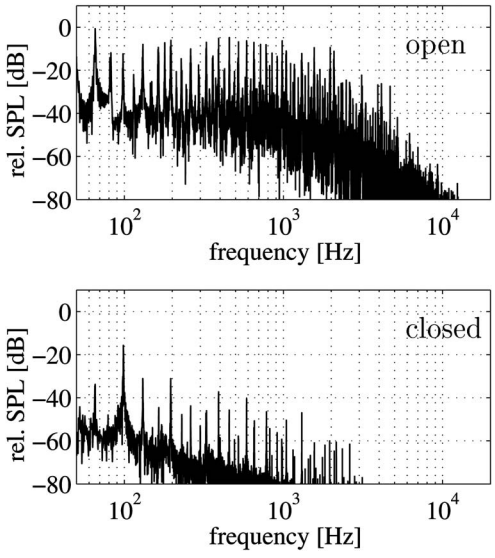


FIG. 3. Frequency spectrum of a c-major chord ($C_2, C_3, C_4, E_4, G_4, C_5$) with open and closed crescendo wheel (top and bottom panel, respectively). Data shown for the Klais instrument at the Ruhr-University Bochum.

angle), and fully open door. In each measurement, the frequency-dependent sound level of a full-range loudspeaker that was located inside the swell box was measured with a custom-built dummy head that was placed in the audience space using a sine-sweep excitation signal. For the analysis, the recorded signal for the left ear of the dummy head was used. Figure 2 (top graph) shows the spectral difference between the half-open and open door configurations versus the closed door configuration. For this purpose, the magnitude frequency spectrum for the closed door case was subtracted from the magnitude spectra for the open-door cases. The solid line shows the gain for the fully opened swell box, the dashed line depicts the gain when the doors of the swell box were opened to an angle of 45°. The largest level increase of approximately 7.5 dB is found at 1 kHz, below this frequency the level increase drops with decreasing frequencies. Above 8 kHz, the gain also declines rapidly. The frequency-dependent level enhancement leads to perceived coloration and the impression of the pipe-organ sound moving away when the doors of the swell box are being closed. The measurement verifies the observations that were made for the analysis of Fig. 1. The bottom graph of Fig. 2 shows the same measurement, but for a loudspeaker position that was located 3 m behind the first measurement position. The values are slightly different from the previous case. In particular, the second peak at 8 kHz is a few decibels higher.

Figure 3 depicts the crescendo wheel action of the Klais pipe organ at the Ruhr-University Bochum. The bottom graph shows the frequency spectrum of the lowest crescendo wheel setting. In this case, only two stops are activated: *Zartbaß* 16 ft and *Aeoline* 8 ft. When the crescendo wheel is set to the highest setting, the sound-pressure level increases by up to 50 dB at 2 kHz when all available stops are involved in this configuration (see Table I). This relatively high increase in level results from progressing from mellow flue stops at the lowest crescendo wheel setting (typically an 8 ft or 16 ft stop with low sound-pressure level and only little energy in the higher partials) to louder stops including mixed stops and reeds when the crescendo wheel is opened. If all stops were of the same tone quality—of course something that contradicts the organ builders' goal to create an instrument with a rich variety of different stops and tone colors—the level increase for doubling the number of stops would only be 3 dB, if all stops were of the same type (assuming the addition of two uncorrelated but otherwise identical sound sources). In this case the level increase from 1 to 64 stops would have resulted in a level increase of only: $\log_2(64)\cdot 3\ \text{dB} = 18\ \text{dB}$, much less than the actual measurement results of up to 50 dB.

The most dominant harmonic in the initial crescendo wheel setting is the twelfth of the *Zartbaß* 16 ft at 98 Hz which has a relative sound-pressure level of approximately −15 dB (Fig. 3, bottom graph). In the open crescendo wheel condition, the level of this frequency component is only a few decibels higher, while most other partial components below and above this frequency have increased by tens of decibels (Fig. 3, top graph). Figure 4 shows the dynamic range for the crescendo wheel (top curve, solid line). As stated before, the minimum in level difference is found for frequencies at 100 Hz, while the highest values of 50 dB occur near 2 kHz with a strong roll-off toward the higher frequencies. At the low frequencies, the sound pressure gain is 28 dB at 50 Hz. The three lower curves show the dynamic range of the swell box for three different stop configurations: the previously described German setting (dashed line), a stop configuration that was oriented toward French pipe organ music (dotted line), *Konzertflöte* 8 ft, *Octave* 4 ft, *Basson* 16 ft, *Tromp. harm.* 8 ft, *Hautbois* 8 ft, *Clairon harm.* 4 ft, and a configuration with the free-reed stop *Klarinette* 8 ft

TABLE I. Stop disposition for the Klais pipe organ at the Auditorium Maximum of the Ruhr-University Bochum, Germany.

I. HAUPTWERK·C_2–C_7		II. POSITIV·C_2–C_7		III. RÉCIT C_2–C_7	
Preastant	16 ft	Rohrgedackt	16 ft	Salicional	16 ft
Gedackt	16 ft	Principal	8 ft	Bourdon	16 ft
Principal	8 ft	Quintadena	8 ft	Diapason	8 ft
Flûte harmonique	8 ft	Portunalflöte	8 ft	Flûte traversière	8 ft
Doppelgedackt	8 ft	Unda maris [from c0]	8 ft	Konzertflöte	8 ft
Salicional	8 ft	Octave	4 ft	Aeoline	8 ft
Pifaro [from C_3]	8 ft	Koppelflöte	4 ft	Vox coelestic [from C_3]	8 ft
Octave	4 ft	Viola	4 ft	Octave	4 ft
Blockflöte	4 ft	Sesquialter 2f	$2\frac{2}{3}$ ft	Flûte octaviante	4 ft
Fugara	4 ft	Bauernflöte	2 ft	Violine	4 ft
Großquinte	$5\frac{1}{3}$ ft	Quinte	$1\frac{1}{3}$ ft	Nasat	$2\frac{2}{3}$ ft
Großterz	$3\frac{1}{3}$ ft	Mixtur 5f	2 ft	Doublette	2 ft
Qunite	$2\frac{2}{3}$ ft	Scharff 4f	$1\frac{1}{3}$ ft	Kleincornet 3f	
Octave	2 ft	Dulcian[a]	16 ft	Plein jeu harm. 3–5f	2 ft
Waldflöte	2 ft	Cromhorne[a]	8 ft	Harmonia aetheria 3f	$2\frac{2}{3}$ ft
Mixtur 5–7f	$2\frac{2}{3}$ ft	Klarinette[b]	8 ft	Basson[a]	16 ft
Cymbel 4f	1 ft	Horn[a]	8 ft	Trompette harm.[a]	8 ft
Cornet 5f [from G_3]	8 ft			Hautbois[a]	8 ft
Trompete[a]	16 ft			Voix humaine[a]	8 ft
Trompete[a]	8 ft			Clairon harmonique[a]	4 ft
Clairon[a]	4 ft				
IV. MANUAL·C_2–C_7		PEDAL·C_2–G_4			
Bombardwerk					
Tuba[a]	16 ft	Principalbaß	32 ft	Flöte	4 ft
Tuba[a]	8 ft	Untersatz	32 ft	Nachthorn	2 ft
Trompette en		Principal	16 ft	Großcornet 3f	$10^2/3$ ft
chamades[a]	8 ft	Violonbass	16 ft	Hintersatz 5f	4 ft
Trompette en		Subbass	16 ft	Rauschquinte 3f	$2^2/3$ ft
chamades[a]	4 ft	Zartbass	16 ft	Bombarde[a]	32 ft
		Octave	8 ft	Posaune[a]	16 ft
		Flöbaß	8 ft	Fagott[a]	16 ft
		Cello	8 ft	Trompete[a]	8 ft
		Choralbaß	4 ft	Clairon[a]	4 ft

[a]Striking reed.
[b]Free reed.

only (dash-dotted line). The variations between the three configurations can be explained by the different directivity patterns and locations among the different types of pipes (compare also Fig. 2).

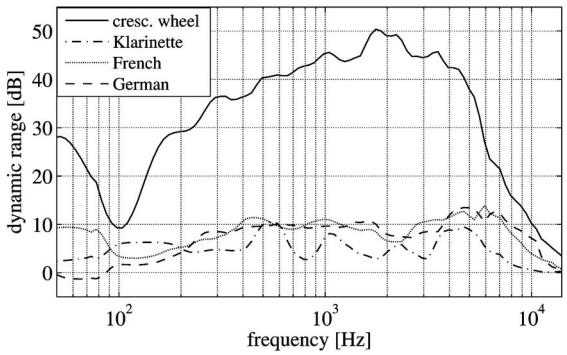


FIG. 4. Dynamic ranges for different expression devices of the Klais pipe organ of the Audimax at the Ruhr-University Bochum, Germany. The three lowest curves show the dynamic range of the swell box for various stop combinations. The top curve shows the difference between the lowest and highest crescendo wheel setting. All values are calculated from differences in the third-octave band spectra.

Figures 5 and 6 show the same measurements for the other two examined pipe organs. The data for the Sonreck/Klais pipe organ in Essen-Steele are in agreement with the measurements on the Klais pipe organ in Bochum, and again the dynamic range of the swell box (Fig. 5, lower four curves) is much smaller than the dynamic range of the simulated crescendo wheel (Fig. 5, upper two curves). However, the maximal dynamic range of the crescendo wheel is only 40 dB near 2 kHz, compared to 50 dB for the Klais pipe organ in Bochum. The smaller range can be explained easily by the fact that the Sonreck/Klais pipe organ has only 37 stops compared to the 82 stops of the Klais pipe organ in Bochum (see Table II for the stop disposition of the Sonreck/Klais organ). The graph also depicts the strong influence of the initial stop combination for the crescendo wheel setting. The dynamic range could be improved by several decibels at frequencies below 700 Hz, when the combination *Bordun* 16 ft/*Rohrflöte* 8 ft (wheel setting 1) was used instead of the combination *Bordun* 16 ft/*Aeoline* 8 ft (wheel setting 2). For the swell-box measurements, the dynamic range for the single free-reed stop *Klarinette* 8 ft (short dash-dotted line)

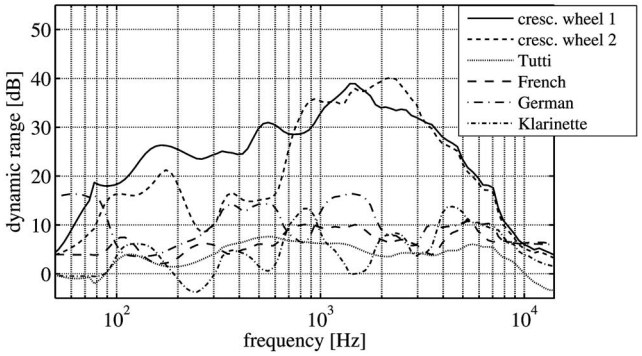


FIG. 5. Dynamic ranges of different expression devices of the Sonreck/Klais pipe organ in Essen-Steele, Germany. The four lowest curves show the dynamic range of the swell-box settings for various stop combinations. The two upper curves show the difference between the lowest and highest crescendo wheel setting. All values are calculated from differences in the third-octave band spectra.

is slightly lower for most frequencies than those for the typical French and German stop combinations (dashed line and long dash-dotted line), and the tutti setting (dotted line). The French combination consisted of: *Flauto amabile* 8 ft, *Fugara* 4 ft, *Oboe* 8 ft, *Klarinette* 8 ft, Sub II/II., while the German stop combination included the stops: *Lieblich Gedackt* 16 ft, *Principal* 8 ft, *Flauto amabile* 8 ft, *Rohrflöte* 8 ft, *Salicional* 8 ft, *Flauto traverso* 4 ft, *Fugara* 4 ft, *Flautino* 2 ft, *Harmonia aetheria* 3–4*f*. Again the radiation patterns of the pipes and the distance to the swell door can explain the measured variations.

The swell box of the Eggert pipe organ in Wattenscheid/Höntrop provides a larger dynamic range than the other two measured pipe organs with values above 10 dB for most frequencies (Fig. 6, four lower curves). The measurement reflects differences in constructional details. In case of the Eggert pipe organ, the swell section was placed in a solid brick enclosure, with only the swell door being a thin wood construction, while the swell boxes of the other two pipe organs are constructed in the tradition of a free-standing wood-case design. Within the frequency range from 300 to 4000 Hz, the dynamic range of the swell box was nearly independent from

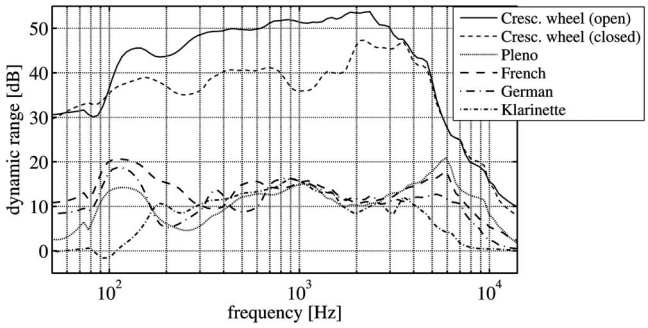


FIG. 6. Dynamic ranges for different expression devices of the Eggert pipe organ in Wattenscheid-Höntrop, Germany. The four lowest curves show the dynamic range of the swell box for various stop combinations. The two upper curves show the dynamic range of the crescendo wheel. The very top curve includes swell-box action in the crescendo process, while the swell box was deactivated in the curve second from top. All values are calculated from differences in the third-octave band spectra.

TABLE II. Stop disposition for the Sonreck/Klais pipe organ of St. Laurentius in Essen-Steele, Germany, (1874/1906/1968/1996).

I. MANUAL·C_2–G_6/G_7 *Hayotwerk (Grand-orgue)*		II. MANUAL·C_2–G_6 *(Schwellwerk) (Récit expressif)*		PEDAL·C_2–F_4 *(Pédale)*	
Praestant	16 ft	Lieblich Gedackt	16 ft	Contrabaß	16 ft
Bordun	16 ft	Principal	8 ft	Violon	16 ft
Principal	8 ft	Salicional	8 ft	Subbaß	16 ft
Flaut major	8 ft	Aeoline	8 ft	Octavbaß	8 ft
Gamba	8 ft	Vox coelestics	8 ft	Violoncello	8 ft
Gemshorn	8 ft	Flauto amabile	8 ft	Superoctave	4 ft
Gedackt	8 ft	Rohrflöte	8 ft	Posaune[a]	16 ft
Quintatön	8 ft	Fugara	4 ft	Tromba[b]	8 ft
Octave	4 ft	Flauto traverso	4 ft		
Hohlflöte	4 ft	Flautino	2 ft		
Quinte	$2\frac{2}{3}$ ft	Sesquialter 2f	$2\frac{2}{3}$ ft		
Octave	2 ft	Harmonia aetheria 4f			
Mixtur 4-5f	2 ft	Oboe[a]	8 ft		
Cornett 4-5f	8 ft	Klarinette[b]	8 ft		
Trompete[a]	8 ft				

[a]Striking reed.
[b]Free reed.

the choice of stop combination. Outside this range, all four curves show strong variations. In general, the single beating-reed stop *Clarinette* 8 ft showed the lowest dynamic range with a formant-like enhancement at 1 kHz, partly because the frequency range of the stop was exceeded by the frequency ranges of the other combinations. Interestingly, the pleno combination exceeded a value of 20 dB at 6000 Hz. The measured French stop combination included the following stops: *Flaut major* 8 ft, *Octave* 4 ft, *Trompete* 8 ft, *Oboe* 8 ft, *Clarinette* 8 ft, Super II/II, Sub II/II, while the German stop combination consisted of: *Gedackt* 16 ft, *Geigenprincipal* 8 ft, *Flaut major* 8 ft, *Liebl. Gedackt* 8 ft, *Gamba* 8 ft, *Octave* 4 ft, *Flaut travers* 4 ft, *Violine* 4 ft, *Quintflöte* $2\frac{2}{3}$ ft, *Waldflöte* 2 ft, *Mixtur* 2 ft 4*f*., pleno stop setting: *Gedackt* 16 ft, *Geigenprincipal* 8 ft, *Flaut major* 8 ft, *Liebl. Gedackt* 8 ft, *Gamba* 8 ft, *Aeoline* 8 ft, *Vox coelestis* 8 ft, *Octave* 4 ft, *Flaut travers* 4 ft, *Violine* 4 ft, *Quintflöte* $2\frac{2}{3}$ft, *Waldflöte* 2 ft, *Terzflöte* $1\frac{3}{5}$ ft, *Mixtur* 2 ft 4*f*, *Trompete* 8 ft, *Oboe* 8 ft, *Clarinette* 8 ft, Super II/II, Sub II/II.

The solid curve in Fig. 6 shows the dynamic range for the crescendo wheel in combination with the swell box. An additional gain of up to 15 dB can be reached when compared to the same measurement with constant swell-box setting (second curve from top in Fig. 6, short dashed line). Without activating the swell box, the maximum dynamic range is found to be in between the values of the other two pipe organs with 45 dB at 2.5 kHz. The stop disposition of the instrument is shown in Table III.

III. HISTORIC EXPRESSION DEVICES: THE WIND SWELL INCLUDING THE HARMONIUM

A. Methods

Four additional instruments were measured to understand how the wind swell dynamically affects the sound of free reeds. The first instrument, a pipe organ with 10 stops,

TABLE III. Stop disposition for the Eggert pipe organ in St. Maria Magdalena, Wattenscheid-Höntrop (1905), Germany.

I. MANUAL·C_2–G_6 *Hauptwerk*		II. MANUAL·C_2–G_6 *Schwellwerk*		PEDAL·C_2–F_4	
Bordun	16 ft	Gedackt	16 ft	Untersatz	32 ft
Principal	8 ft	Geigenprincipal	8 ft	Offenbaß	16 ft
Hohlflöte	8 ft	Flaut major	8 ft	Subbaß	16 ft
Rohrgedackt	8 ft	Liebl. Gedackt	8 ft	Echobaß	16 ft
Salicional	8 ft	Gamba	8 ft	Salicetbaß	16 ft
Unda Maris	8 ft	Aeoline	8 ft	Octavbaß	8 ft
Octave	4 ft	Vox coelestis	8 ft	Flötenbaß	8 ft
Rohrflöte	4 ft	Octave	4 ft	Cello	8 ft
Octave	2 ft	Flaut travers	4 ft	Choralbaß	4 ft
Cornett 3f	$2\frac{2}{3}$ ft	Violine	4 ft	Contraposaune[a]	32 ft
Mixtur 4f	$2\frac{2}{3}$ ft	Quintflöte	$2\frac{2}{3}$ ft	Posaune[a]	16 ft
Fagott[a]	16 ft	Waldflöte	2 ft	Trompete[a]	8 ft
Trompete[a]	8 ft	Terzflöte	$1\frac{3}{5}$ ft		
Clarine[a]	4 ft	Mixtur 4f	2 ft		
		Trompete[a]	8 ft		
		Oboe[a]	8 ft		
		Clarinette[a]	8 ft		

[a]Striking reed.

was established by Steinmeyer in 1883 for the catholic church in Gebsattel near Rothenburg o.d. Tauber. The Steinmeyer pipe organ was analyzed for its physharmonica, a free-reed stop without resonators which is equipped with a wind swell. Physharmonicas are a speciality of the South German pipe-organ tradition. The Steinmeyer physharmonica is based on the principle of a Schiedmayer physharmonica, for which the wind pressure is varied at the outlet. For this reason, the device is not a pure wind swell but rather a combination of a wind swell and a swell box, because the acoustical radiation of the reed sound is affected as well when the swell is being closed, and the direct connection to the outside is cut off. Figure 7 clarifies this constructional detail further. The cross section shows the reed (a), the tuning wire (b), and the keyboard valve (c) that is used to address individual reeds. The air flows from the compression chamber (d) through the reed into the sound chamber (e). The amount of airflow can be either controlled at the inlet of the compression chamber (f) or at the outlet of the sound chamber (g) as is the case for the Steinmeyer physharmonica. In the first case, the airflow into the compression chamber is often controlled through the foot pedals which are connected to the bellows, rather than through a special valve.

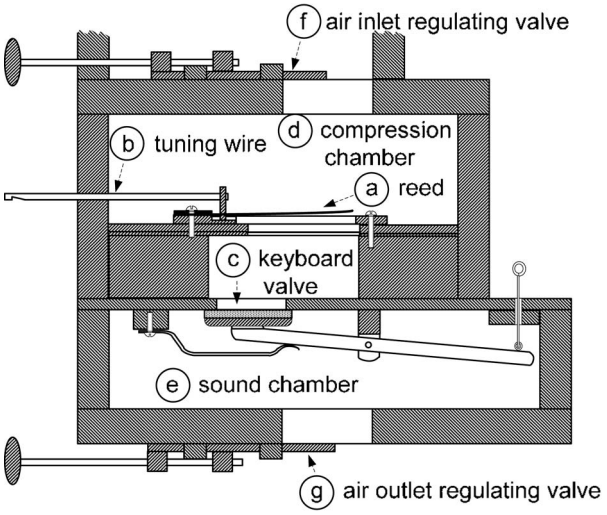


FIG. 7. Cross section through a physharmonika. The air inlet and outlet regulating valves are drawn symbolically. The typical physharmonica can be controlled either through an inlet or outlet regulating valve but not through both. Sketch adapted from Audsley (1905).

FIG. 8. Physharmonica prototype of Ulrich Averesch (Type I) with five reeds (tone c) in octaves. The scales are similar to the dimensions of reeds for a harmonium with pressure bellows.

Two further physharmonica stops have been analyzed. These instruments were recently built prototypes by Ulrich Averesch (2006), which merely consisted of c-tone reeds (see Fig. 8). Averesch is one of the very few instrument builders worldwide who builds new replacement reeds for the restoration of historic free-reed instruments and also offers a physharmonica stop for newly designed pipe organs. One of the prototypes has relatively large reeds in the tradition of reed organs with pressure bellows (Type I). The other prototype's reeds were fairly small and designed in the tradition of the American reed organ with suction bellows (Type II). In both cases, the reeds were driven with positive wind pressure provided by a pipe-organ chest (Bernard Koch). The dynamic range was determined for a wind pressure range from 34 to 108 mm H_2O. The measurements were obtained in an anechoic room in which the instruments were placed. For the frequency spectra, the sounds of the individual reeds were measured separately and then superposed in an audio editor.

The fourth instrument under examination was a French Mustel "art" reed organ. Instruments by Victor Mustel are considered to be the masterpieces in harmonium building. Although the Mustel harmonium does not belong to the family of pipe organs, it was included in this study because of the instrument's known large dynamic range achieved with a wind swell. Partly for this reason, the art of building harmoniums was still pursued for some time after the wind swell in pipe organs had been abandoned. The Mustel harmonium of this study was built in 1904 and is owned by Ulrich Averesch, who also restored the instrument.

B. Results

Figure 9 (bottom graph) shows the frequency spectrum of the Steinmeyer physharmonica (Tone C_3) for the closed wind-swell setting (minimal wind pressure). The frequency spectrum is relatively rich in higher harmonics with the maximum at the third harmonic, the twelfth. The high-

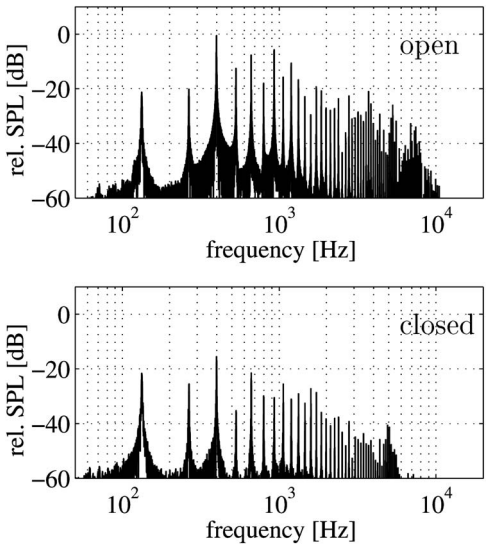


FIG. 9. Frequency spectrum of a physharmonica reed (C_3) of the Friedrich Georg Steinmeyer pipe organ in Gebsattel near Rothenburg o. d. Tauber. The top panel shows the spectrum with open swell, the bottom panel with closed swell.

frequency energy rolls off at approximately 2 kHz. The top graph shows the frequency spectrum of the same reed for the open wind-swell configuration (maximum wind pressure). While the first two partial tones are not affected much, the remaining partials are boosted by up to 15 dB. The frequency selective gain in sound pressure provides the change in tone color that is typical for the physharmonica.

In the following, the dynamic ranges of the expression devices are shown as the difference of the third-octave band filtered frequency spectra of the highest versus the lowest dynamic setting in analogy to Fig. 4. The dotted line in Fig. 10 shows the dynamic range of the Steinmeyer physharmonica. The curve was estimated from five individually measured c-reeds (C_2–C_6). As described in Fig. 9, the most effective range lies between 300 Hz and 1 kHz with values between 15 and 20 dB. The boosted frequency region above 4 kHz is less important, because the instrument does not have much acoustic energy in this frequency range. Below 300 Hz, the wind swell does not have any effect on the amplitude of the sound. In contrast, the dynamic gain of the Averesch physharmonica is more gradual than was the case for the Steinmeyer physharmonica. For frequencies between 100 and 250 Hz, the gain is about 7 dB higher than the measured gain for the Steinmeyer instrument, while the gain is less for frequencies between 300 Hz and 1 kHz. In general, the dynamic gain of the Averesch physharmonica Type I is a few decibels higher than the second Averesch physharmonica. As pointed out earlier the acoustic differences between the instruments by Averesch and Steinmeyer result most likely from differences in the mechanical construction. While the expression of the Averesch instrument was controlled by adjusting the pressure of the wind supply, the Steinmeyer instrument follows the classical physharmonica design. In the latter case, the wind exhaust of the physharmonica's enclosure is varied by a shutter, while the wind supply remains constant. Obviously, this procedure does not only affect the effective wind pressure that drives the reeds, but also attenuates the sound of the reeds in a similar manner to the swell box when the shutter is closed. Unfortunately, it was not possible to measure the wind pressure of the Steinmeyer physharmonica for a further comparison with the Averesch physharmonica. It can be assumed that the builder optimized the operating range of the wind swell between the minimum wind pressure at which the reeds responded and the maximum wind pressure the organ provides. The typical wind pressure for this type of pipe organ is around 75 mm H_2O, which is below the maximum wind pressure that was applied to the Averesch reeds. The minimum wind pressure was above the self-excitation threshold of the reeds, which is around 30 mm H_2O.

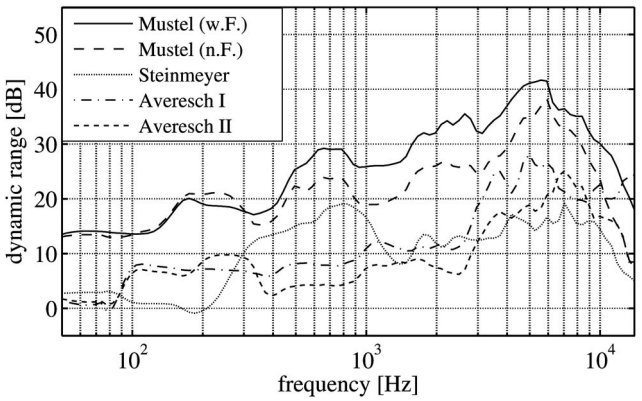


FIG. 10. Dynamic ranges for different wind swell devices. The three lowest curves show the results for the physharmonicas as indicated in the legend, while the two upper curves depict the data for the Mustel harmonium. All values are calculated from differences in the third-octave band spectra.

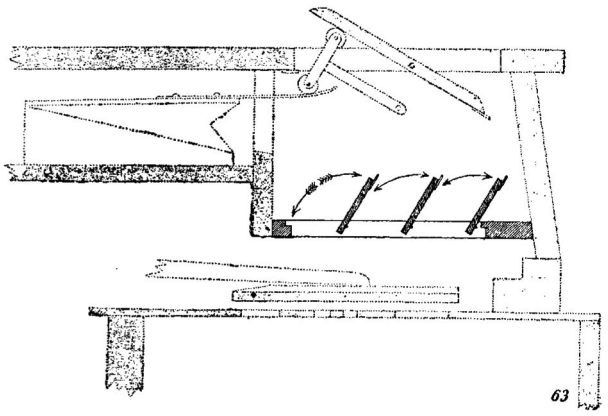


FIG. 11. Forte system of a Mustel "art" harmonium. The sketch shows the cross section through the upper part of the harmonium. The top flap shows the forte flap which can be controlled through wind pressure using the regulating bellow shown at the top left. Below the forte flap is the so-called metaphone, a very thin swell door that only affects high frequencies. The reeds are located below the metaphone.

The two curves at the top of Fig. 10 show the results for the Mustel "art" reed organ with and without activated forte flap (solid line: activated forte flap; long dashed line: deactivated forte flap). The forte flap is basically an additional swell-box (see Fig. 11). The Metaphone was kept open during the measurements. The absolute maximum level is found

FIG. 12. Percussion of a Mustel Harmonium (Harmonium Collection Averesch).

at approximately 5 kHz, but here the overall level is not very high, so that the perceived coloration is not too apparent at this frequency.

1. Percussion of a Mustel Harmonium

Part of the success of the art reed organ tradition was the introduction of new techniques to increase the dynamic range of the expression device. One reason why the dynamic range of the Mustel Harmonium is larger than those of the physharmonicas is the use of a percussion device shown in Fig. 12. The percussion device consists of a hammer for each reed that excites the reed to support the self-excitation of the wind-pressured reed when the keyboard is played. This way, the reeds respond to much lower wind pressures of 5 mm H_2O compared to the 20 mm H_2O that are needed to excite a reed without percussion. The reason why this technique never made it into pipe organ design is most likely the very low overall sound pressure level that is observed at low wind pressures of 5 mm H_2O. The fundamental frequencies of the pipes were estimated by analyzing the recorded sound files in MATLAB using the YIN algorithm (de Cheveigné, 2002). The wind pressure was determined using a water gauge. While the art reed organ is often played in small spaces, the resulting sound pressure would not be useful in a church or concert hall environment. The percussion device also has a large impact on the attack transient of the harmonium sound. It comes as no surprise that it accelerates the attack period (Fig. 13).

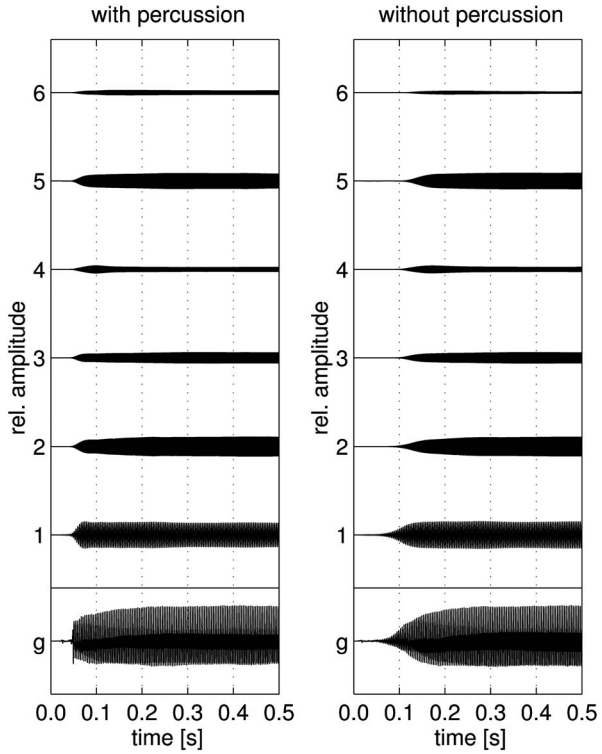


FIG. 13. Influence of the percussion on the sound of the reed C_4 of the stop *Cor anglais* 8 ft (left). The attack is accelerated and all partial tones start simultaneously compared to the same tone played without percussion (right panel). The bottom curve in each graph shows the total time signal, the curves above the total signal show filtered signals of the single partials as indicated on the *y* axis [For details see Braasch and Ahrens (2000)]. In both cases, the expression of the instrument was turned off.

2. Tuning stability of a free-reed pipe

As mentioned in Sec. I, the tuning stability of free reeds made the wind swell possible. Figure 14 demonstrates the free reed's tuning stability with varying wind pressure. While the free-reed *Klarinette* 8 ft pipe (Klais, C_4) is hardly affected by wind pressure changes from 20 to 120 mm H_2O, the frequency of the beating reed *Krummhorn* 8 ft (Bernhard Koch, C_3) varies by half a semitone within the same wind-pressure interval. A similar result for the beating read was previously published by Plitnik (2000). In case of the flue pipe *Principal* 8 ft (G_4), the frequency variation was found

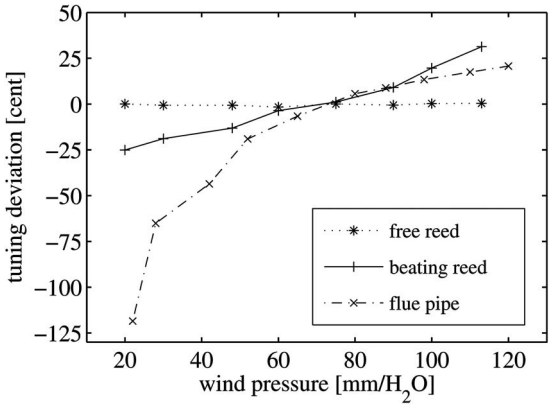


FIG. 14. Frequency variation of three organ pipes at different wind pressures.

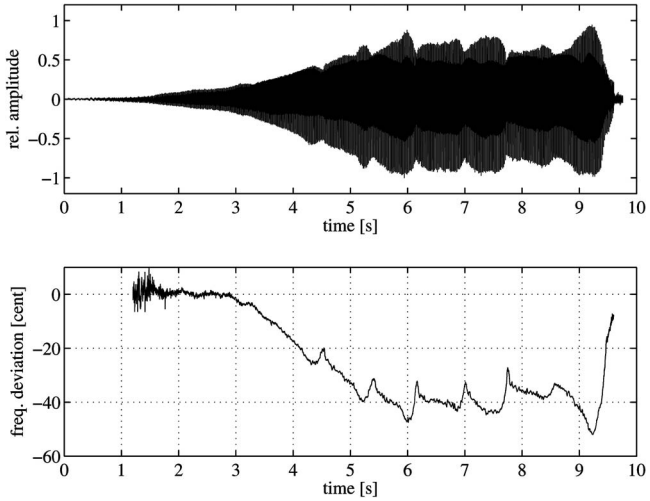


FIG. 15. The bottom graph shows how the *Basson* 8 ft C_2 reed of the Mustel Harmonium detunes with increased sound pressure (top graph). To demonstrate this effect optimally, the instrument was played with up to the maximum wind pressure.

to be more than a semitone for the same wind pressure range, but there was no linear relationship between wind pressure and frequency as was the case for the beating reed. The data are in agreement with previous published literature for the analyzed wind pressure range (Linhardt, 1960; Verge *et al.*, 1997). It should be mentioned that the fundamental-frequency dependency of flue pipes with changing wind pressure can vary with design. However, even the most complex mechanisms (see Töpfer, 1888 for details) failed to produce a flue pipe stable enough in frequency to be used in conjunction with a wind swell.

It is noteworthy that the free reed can detune, too, if driven at extremely high wind pressures (see Fig. 15). This exception is seldom stated in the literature. The top graph shows the amplitude of a crescendo played on the Mustel harmonium. The lower graph shows the corresponding frequency shift of the reed in cents. Keep in mind that this is a very extreme example for a single low note. To achieve such a tuning instability, the reed has to be exposed to wind pressures of about 200 mm H_2O. Usually, the frequency shift is less than 5 cent, since pipe organs do not operate at wind pressures above 100 mm H_2O.

Although tuning problems are typically associated with the wind swell, the two remaining expression devices, swell box and crescendo wheel, can be troubled by shift in fundamental frequency, too, if the device was not carefully designed. Although such a design flaw is very uncommon in contemporary pipe organs, it has been described in literature (Sauer, 1824; Erbslöh, 1984). The pipes in a swell box can detune through a pressure build-up for the closed swell, which reduces the effective wind pressure that drives the pipes. Here, the frequency will become lower when the swell box is closed. This is especially the case when the volume of the swell box is very small and the doors close very tightly (Erbslöh, 1984). Already in 1824, Sauer described the risk of detuned pipes, when the swell box is completely sealed up to improve the dynamic range of the swell box. Of course, the pipe organ can detune while using a crescendo wheel as well,

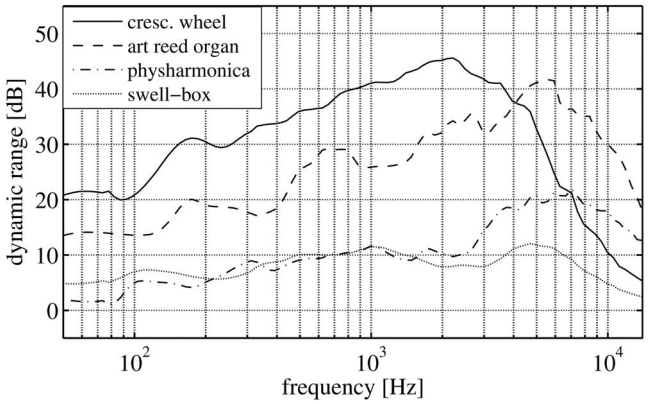


FIG. 16. Dynamic ranges of different expression devices (wind swell, swell box, and crescendo wheel). Each curve shows the average over several instruments measured as described in the text. Due to its special mechanics, the curve for the Mustel Harmonium is shown separately from the remaining instruments with wind swell. All values were calculated from differences in the third-octave band spectra.

if the wind supply is not sufficient for all pipes when the crescendo wheel is opened. However, if carefully designed, the frequency accuracy is sufficient for all three expression devices described here. The pipes' frequency shift during the swell-box activation was in the order of 1 cent for all three measured pipe organs.

C. Comparison of the wind swell to the swell box and the crescendo wheel

In Figs. 16 and 17, the wind swell is compared to the remaining expression devices. Figure 16 compares the dynamic ranges for the wind swell, the swell box, and the crescendo wheel. In all three cases, the data were averaged from all measured samples for each device. The dynamic range of the Mustel art reed organ is shown separately since its construction differs significantly from the physharmonica design, namely in percussion and forte flaps. The figure emphasizes that the dynamic range of the wind swell in physharmonicas and the swell box are of the same order—at least for frequencies below 3 kHz, while the dynamic ranges of the crescendo wheel and the Mustel Harmonium are significantly larger. Although the maximal dynamic range of the art reed organ and the crescendo wheel is in the same order, their resulting effect on the music is not comparable. While the crescendo of the art reed organ remains gradual (Fig. 17, bottom)—similar to the swell box and the physharmonica—the increase in sound pressure occurs step-wise for the crescendo wheel with the activation of each additional stop (Fig. 17 top).

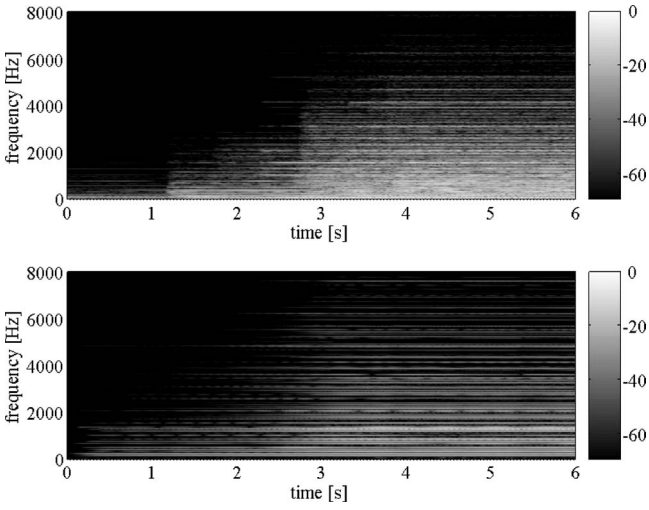


FIG. 17. The crescendo movement of the crescendo wheel (top panel, Eggert pipe organ in Wattenscheid-Höntrop, Germany) and the wind swell of a Mustel Harmonium (bottom panel). In both cases the running Fourier analysis of a c-major chord is shown. For every frame, 4096 frequency coefficients were calculated from a hanning-windowed sample of 2048 coefficients. The step rate was 1024 samples.

IV. DISCUSSION AND CONCLUSION

The main reason why swell box and crescendo wheel are typically found side by side in most larger organs is their different mode of operation (Berlioz, 1905). As Ahrens (1997, 1998, 1999) previously pointed out, the crescendo wheel simulates the type of orchestra crescendo which is reached by successively adding further musical instruments. In orchestral music this principle is often found in Wagner's music, and Locher and Dobler (1923) already recommended the use of the crescendo wheel for pipe organ transcriptions of his music in 1923. Another, quite extreme, example is Maurice Ravel's Bolero, but this orchestration technique can be found in music of the 18th century as well (e.g., Mozart, Sinf. KV 319, 1778; or Paisiello, Ouverture *Il barbiere de Sevilla*, 1784). The swell box on the other hand produces a crescendo similar to the orchestra crescendo in which the individual instruments perform a dynamic movement. Indeed, the dynamic range of the swell box of about 10–20 dB is similar to the dynamic range of the orchestral flute, which has a practical dynamic range of 15–20 dB according to Meyer (1990). While most other orchestral instruments have higher dynamic ranges, the orchestral flute is the most comparable orchestral instrument to organ pipes.

While the acoustical measurements presented here effectively demonstrate existing knowledge on the swell box and the crescendo wheel, the question remains why the wind swell disappeared at the end of the 19th century. As the measurements revealed, the crescendo effects of swell box and wind swell are very similar and the dynamic range is in the same order, too. To its disadvantage, the wind swell is restricted to operate with free-reed stops, while the swell box can be applied to a number of different pipe organ stops simultaneously. At the end of the 18th century, the simulation of an expressive solo instrument was among the major goals in pipe-organ building, but later in the 19th century a new larger pipe-organ type was developed. With the introduction of the Barker lever and the pneumatic wind chest, the pipe organ grew in number of stops, achieving a similar aesthetic to a large orchestra. While the swell box can be used to create an effect similar to the crescendo of a whole orchestral body, the wind swell is restricted to a solo instrument or small ensemble at best.

It is questionable whether constructional details of the art reed organ can be applied successfully to pipe organs to increase the dynamic range of the wind swell above that of

the swell box to bring back interest to this device. However, the large overall dynamic range of the art reed organ is reached by exciting the reeds at very low wind pressures through a hammer device. In the large environments a pipe organ is typically located in, extremely low sound pressures would not be very effective in a musical context. The same problem occurs with double wall/door swell-box constructions, which are usually designed with a general swell box enclosing the whole organ case which includes a second swell box for the swell section. A few years ago, Zacharias (2002a) proposed an interesting design of a wind swell in which the individual free-reed pipes could be played dynamically using a velocity sensitive keyboard, an idea that is impossible to pursue based on the swell-box idea. Zacharias' (2002b) design has been incorporated into a new pipe organ built by organ builder Johannes Rohlf, but it remains to be seen whether this improvement brings back more attention to the wind swell in future.

ACKNOWLEDGMENTS

Part of this work was carried out at the Ruhr-University Bochum, Germany. The author would like to thank Michael Daub and Ulrich Averesch for playing the instruments during the measurements and for sharing their expertise with him. He is also grateful to Christian Ahrens and Jens Blauert, and the church communities who own the measured pipe organs for their endorsement during this investigation. He would also like to thank two anonymous reviewers for their helpful suggestions during the review process.

Ahrens, C. (**1997**). "Die von Friedrich Weigle und J.P. Schiedmayer erbaute Parabrahm-Orgel (1908) in Eichwalde. Ein historisch bedeutsames Denkmal des Orgelbaus" ["The parabrahm pipe organ (1908) built by Friedrich Weigle and J.P. Schiedmayer. A historically important monument of pipe-organ building"], Ars Organi **45**, 195–203.

Ahrens, C. (**1998**). "*…mit Crescendo und Decrescendo zum Verwehen…* Physharmonika-Register in Orgeln des 19. und 20. Jahrhunderts" ("…Blown over with crescendo and decrescendo… Physharmonica stops in pipe organs of the 19th and 20th century"), Ars Organi **46**, 143–148.

Ahrens, C. (**1999**). "Disposition," in *Zur Geschichte der Konzertsaalorgel in Deutschland: die Klais-Orgel der Ruhr-Universität Bochum (On the History of the Concert-Hall Organ in Germany: The Klais Pipe Organ of the Ruhr-University Bochum)*, edited by C. Ahrens and J. Braasch (Main, Frankfurt) pp. 37–94.

Ahrens, C., and Braasch, J. (**2002**). "Christian Gottlieb Kratzenstein? De uitvinder van de orgelregisters met doorslaande tongen" ("Christian Gottlieb Kratzenstein, the inventor of pipe-organ stops with free reeds?"), Het Orgel **99/4**, 32–36.

Audsley, G. A. (**1905**). *The Art of Organ-Building. A Comprehensive Historical, Theoretical and Practical Treatise on the Tonal Appointment and Mechanical Construction of Concert-Room, Church and Chamber Organ* (Dodd, Mead, and Company, New York).

Averesch, U. (**2006**). "Physharmonika-Zungen für den Orgelbau" ("Physharmonica reeds for organbuilding"), ⟨http:/www.harmoniumservice.de/⟩ (last viewed 1/3/2008).

Berlioz, H. (**1905**). *Große Instrumentationslehre (The Treatise on Modern Instrumentation and Orchestration)*, edited and supplemented by R. Strauss (Leipzig, Berlin, reprint 1955).

Bicknell, S. (**1999**). *The History of the English Organ* (Cambridge University Press, New York).

Braasch, J., and Ahrens, C. (**2000**). "Attack transients of free reed pipes in comparison to striking reed pipes and diapason pipes," Acust. Acta Acust. **86**, 662–670.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.

Erbslöh, H. U. (**1984**). "Über Schwierigkeiten beim Stimmen" ("On the difficulties when tuning"), ISO Inf. **24**, 37–40.

Grave, F. K., and Grave, M. G. (**1987**). *In Praise of Harmony: The Teachings of Abbe Georg Joseph Vogler*, (University of Nebraska Press, Lincoln, London).

Jordan, W. D. (**2002**). "The first English swell-box—An open case," The Organ Yearbook **31**, 25–57.

Knecht, J. H. (**1795**). "Vom Gebrauche der Orgelregister" ("On the usage of organ stops"), in *Ausgewählte Orgelwerke. Band 1 (Selected Pipe-Organ Works, Vol. 1)*.

Kratzenstein, C. T. (**1780**). "Tentamen Resolvendi Problema ab Academia Scientarium Imperialis Petropolitana ad annum 1780 Publica Propositum" ("Document on a resolved problem for the Scientific Academy of St. Petersburg, publicly proposed in the year 1780"), Scientific Academy, St. Petersburg, Russia.

Linhardt, W. (**1960**). "Über Laden- und Traktursysteme der Orgel und ihre Einflüsse auf die Ein- und Ausschwingvorgänge der Pfeifen" ("On wind chests and tracker systems of pipe organs and their influence on the on- and offset transients of the pipes"), Dissertation Technische Hochschule Braunschweig.

Locher, C., and Dobler, J. (**1923**). *Die Orgel-Register und ihre Klangfarben (Pipe-Organ Stops and Their Tone Colors)* (Verlag von Ernst Kuhn, Bern and Biel, Switzerland, reprint Fritz Knuf, Amsterdam, 1971).

Meyer, A., and Wayne, C. W., Jr. (**1961**). "Attenuation of pipe-organ tone by swell shutters," J. Acoust. Soc. Am. **33**, 1652(A).

Meyer, J. (**1990**). "Zur Dynamik und Schalleistung von Orchesterinstrumenten" ("On the dynamics and acoustic power of orchestral instruments"), Acustica **71**, 277–286.

Peterson, R. H. (**1982**). "Swell box for hybrid pipe organ," J. Acoust. Soc. Am. **71**, 1312 (Patent Review).

Peterson, R. H., and Mornar, J. A. (**1983**). "Electronically controlled swell shutter operator for pipe organs," J. Acoust. Soc. Am. **73**, 1413 (Patent Review).

Plitnik, G. R. (**2000**). "Vibration characteristics of pipe organ reed tongues and the effect of the shallot, resonator, and reed curvature," J. Acoust. Soc. Am. **107**, 3460–3473.

Sauer, L. (**1824**). "Ueber das Crescendo in des Abt Voglers Orchestrion" ("On the crescendo in Abbé Voglers orchestrion"), Allgemeine musikalische Zeitung **23**, 370–375.

Schlimbach, G. C. F. (**1805**). "Ueber des Abts Vogler Umschaffung der Orgel zu St. Marien in Berlin, nach seinem Simplifications-System, nebst leicht ausführbaren Vorschlägen zu einigen bedeutenden Verbesserungen der Orgel" ("On Abbé Vogler's modification of this pipe organ at St. Marien in Berlin based on his simplification system, including easily accomplishable suggestions for the considerable improvement of the pipe organ"), Berlinische Musikalische Zeitung, pp. 383–386, 391–394, 404–406.

Töpfer, J. G. (**1888**). *Die Theorie und Praxis des Orgelbaus (The Theory and Practice of Organ Building)*, edited by M. Allihn (Bernhard Friedrich Voigt, Weimar).

Verge, M. P., Fabre, B., Hirschberg, A., and Wijnands, A. P. J. (**1997**). "Sound production in recorderlike instruments. I. Dimensionless amplitude of the internal acoustic field," J. Acoust. Soc. Am. **101**, 2914–2924.

Zacharias, E. (**2002a**). "Tasten-crescendo für Orgeln" ("Keyboard crescendo for pipe organs"), Instrumentenbau-Zeitschrift **7/8**, 8–9.

Zacharias, E. (**2002b**). "Die dynamische Orgel" ("The dynamic pipe organ"), Ars Organi **50/1**, 19–21.

Fast wave ultrasonic propagation in trabecular bone: Numerical study of the influence of porosity and structural anisotropy

G. Haïat[a)]
CNRS, Université Paris 7, Laboratoire de Recherches Orthopédiques, UMR CNRS 7052 B2OA, 10, avenue de Verdun, 75010 Paris, France

F. Padilla
Université Pierre et Marie Curie - Paris 6, Laboratoire d'Imagerie Paramétrique UMR 7623, 15, rue de l'école de Médecine, 75006 Paris, France and CNRS LIP UMR 7623, F-75006 Paris, France

F. Peyrin
CREATIS, UMR CNRS 5515, INSERM U630, 69621 Villeurbanne, Cedex, France and ESRF, BP 220, 38043 Grenoble, Cedex, France

P. Laugier
Université Pierre et Marie Curie - Paris 6, Laboratoire d'Imagerie Paramétrique UMR 7623, 15, rue de l'école de Médecine, 75006 Paris, France and CNRS LIP UMR 7623, F-75006 Paris, France

(Received 30 May 2007; revised 8 December 2007; accepted 15 December 2007)

Our goal is to assess the potential of computational methods as an alternative to analytical models to predict the two longitudinal wave modes observed in cancellous bone and predicted by the Biot theory. A three-dimensional (3D) finite-difference time-domain method is coupled with 34 human femoral trabecular microstructures measured using microcomputed tomography. The main trabecular alignment (MTA) and the degree of anisotropy (DA) were assessed for all samples. DA values were comprised between 1.02 and 1.9. The influence of bone volume fraction (BV/TV) between 5% and 25% on the properties of the fast and slow waves was studied using a dedicated image processing algorithm to modify the initial 3D microstructures. A heuristic method was devised to determine when both wave modes are time separated. The simulations (performed in three perpendicular directions) predicted that both waves generally overlap in time for a direction of propagation perpendicular to the MTA. When these directions are parallel, both waves are separated in time for samples with high DA and BV/TV values. A relationship was found between the least bone volume fraction required for the observation of nonoverlapping waves and the degree of anisotropy: The higher the DA, the lower the least BV/TV.
 [DOI: 10.1121/1.2832611]

PACS number(s): 43.80.Ev, 43.20.Jr, 43.58.Ta, 43.20.Gp [FD] Pages: 1694–1705

I. INTRODUCTION

Osteoporosis is a skeletal disease characterized by a decrease of bone mass and microarchitectural deterioration of bone tissue, leading to an increase of bone fragility and susceptibility to fracture. Bone mineral density measurements performed with dual x-ray absorptiometry (DXA) techniques currently represent the gold standard for fracture risk assessment. However, other bone properties, such as structural or material, contributing to bone fragility are not satisfactorily assessed with DXA. Therefore, other non-invasive modalities have been developed, such as quantitative ultrasound (QUS) of peripheral skeletal sites (Langton *et al.*, 1984; Laugier *et al.*, 1994a; Barkmann *et al.*, 200a, b; Bossy *et al.*, 2004a), which have become important modalities for the investigation of bone strength. Two ultrasonic propagation parameters are classically measured in transmission through cancellous bone: the speed of sound, and the slope of the frequency-dependent attenuation coefficient (broadband ultrasonic attenuation) (Laugier *et al.*, 1994b).

Cancellous bone is a poroelastic and biphasic medium composed of an elastic skeleton (trabecular network) filled with a viscous fluid (bone marrow). Because, the propagation characteristics are determined by several bone mechanical and structural properties, appropriate QUS measurements should reflect these bone properties. Theoretical models using the Biot theory (Biot, 1956a, b; 1962), initially developed in the context of geophysical applications, have been applied to cancellous bone with some success (Williams, 1992; Hosokawa and Otani, 1997; 1998; Kaczmarek *et al.*, 2002; Lee *et al.*, 2003; Fellah *et al.*, 2004; Wear *et al.*, 2005; Fellah *et al.*, 2006; Lee and Yoon, 2006).

Briefly, the Biot's model predicts the existence of two longitudinal waves. These waves result from the coupling of the fluid and the solid and correspond to in-phase (fast wave) and out-of-phase (slow wave) relative motion of the fluid and solid. Hosokawa and Otani (1997) first evidenced experimentally the existence of two wave modes (one fast and one

[a)]Author to whom correspondence should be addressed. Electronic mail: haiat@univ-paris12.fr

slow) propagating in bovine trabecular bone and this observation could be reproduced by others in bone specimens with marrow inside (Kaczmarek *et al.*, 2002), water (Hughes *et al.*, 1999; Kaczmarek *et al.*, 2002; Cardoso *et al.*, 2003), alcohol (Kaczmarek *et al.*, 2002) or air (Strelitzki *et al.*, 1999). Similarly, both waves have also been measured in human trabecular bone (Cardoso *et al.*, 2003; Fellah *et al.*, 2004). The measured propagation speed of the two waves were in good agreement with results obtained with Biot's theory (Williams, 1992; Hosokawa and Otani, 1997, 1998; Fellah *et al.*, 2004) or with the modified Biot–Attenborough (MBA) model (Lee *et al.*, 2003; Lee and Yoon, 2006). Biot's theory was also able to accurately predict the dependence of sound speed on porosity (Hosokawa and Otani, 1997, 1998; Wear *et al.*, 2005; Lee and Yoon, 2006). Based on this approach, a device has been developed to measure the distal radius, a site with high volume fraction of cancellous bone (Mano *et al.*, 2006). The wave forms of the fast and slow waves are analyzed in the time domain to obtain the amplitudes and the propagation time of both waves, which are then used to estimate the bone mass density and bone elasticity (Otani, 2005).

Although predicted theoretically, the observation of two clearly separated waves in the time domain is sometimes not possible due to the complex dependence of the speed of sound and amplitude of each wave on the bone volume fraction (i.e., the bone volume-to-total volume ratio, BV/TV) and on the orientation of the ultrasound beam relatively to the direction of trabeculae alignment. For example, when the propagation angle is 90° (i.e., when the trabeculae are aligned perpendicular to the direction of the beam axis as for most *in vivo* measurements conditions), the fast and slow waves completely overlap in the time domain (Hosokawa and Otani, 1997, 1998). This has been confirmed experimentally in human vertebrae in which both waves could be clearly time resolved only when the incident acoustic wave was transmitted in a direction parallel to the direction of the main trabecular alignment (Nicholson *et al.*, 1998). The acoustic anisotropy can be predicted with angle-dependent Biot and MBA models (Lee *et al.*, 2007) or with an alternative theoretical approach, proposed by Hughes *et al.* (1999), in which bone is considered as a multilayer of alternating fluid and solid layers (Hughes *et al.*, 1999). This multilayer model shows that the slow wave disappears at high refraction angles due to the fact that the energy of the slow wave is refracted from the phase propagation direction and that the signals of fast and slow waves overlap (Hughes *et al.*, 1999; Padilla and Laugier, 2000).

Thus, experimental and theoretical works provide evidence that the propagation speed of both waves is influenced not only by the bone volume fraction, but also by the bone structural anisotropy and the insonication angle. However, the combined influence of these properties on the determination of the localization in time of both waves (or on their time separability) has not been addressed so far. For this purpose, Biot's theory presents several shortcomings in that (i) it requires a large number of parameters that are not known with accuracy (Lee *et al.*, 2007) and (ii) it is limited by its inability to accurately model the complex anisotropic and heterogeneous three-dimensional (3D) bone microstructure. To overcome the technical difficulty of analytical modeling, researchers have recently turned to finite difference time domain (FDTD) computational bone models. These FDTD studies have been shown to be able to reproduce the main propagation characteristics observed experimentally (Bossy *et al.*, 2005), and interestingly to correctly predict, for an appropriate alignment of the trabecular microarchitecture in the incident ultrasonic beam, the fast and slow waves predicted by the Biot's theory (Hosokawa, 2005a).

The overall goal of the present work is to better elucidate the conditions on bone volume fraction and structural anisotropy under which the fast and slow waves are separated. Toward this end, high-resolution microcomputed tomography (μCT)-based finite difference models were developed for a cohort of human femoral specimens. Our hypothesis was that the time separability of both the slow and fast waves results from a combination of appropriate values of bone volume fraction and structural anisotropy for each insonication angle. Starting with a number of 3D reconstructed bone microstructures, studies were performed with the specific objectives of quantifying for different insonication angles the respective influence of the bone volume fraction versus structural anisotropy on the time separation of the slow and fast waves, thus providing insight into the influence of cancellous bone structure on the ultrasonic propagation characteristics.

II. MATERIAL AND METHOD

A. 3D numerical model of the trabecular structure bone

The 3D FDTD numerical computations of wave propagation were coupled to 3D binary numerical models of trabecular structures. Toward this end, cylindrical cores were machined in 34 human proximal femur specimens. The 3D microarchitecture of these samples was obtained using synchrotron radiation microcomputed tomography (SR-μCT) performed at ESRF (European Synchrotron Radiation Facility, France), with an isotropic voxel size of 10 μm. Technical details of SR-μCT and its application to bone microscanning have been published earlier (Peyrin *et al.*, 1998). Parallelepipedic domains with dimensions 5.6×5.6 mm^2 in the transverse dimensions (X, Y) and 10 mm along the Z axis were extracted from each synchrotron acquisitions. For all samples, the Z direction corresponds to the anteroposterior direction and directions (X, Y) were chosen in order to correspond approximately to the principal orientations of the trabecular network in this plane. The volumes were down sampled from their initial resolution of 10 μm, to a voxel size of 30 μm, in order to reach a compromise between memory requirementsand computation time on a standard desktop PC and a reasonably accurate representation of the trabecular structure. We checked that the microarchitectural parameters were similar when obtained with the 10 and 30 μm resolution structure.

The 34 3D SR-μCT data sets of human bone specimens were used to create binary digital 3D models of the trabecular bone specimens. For each of the 34 original bone

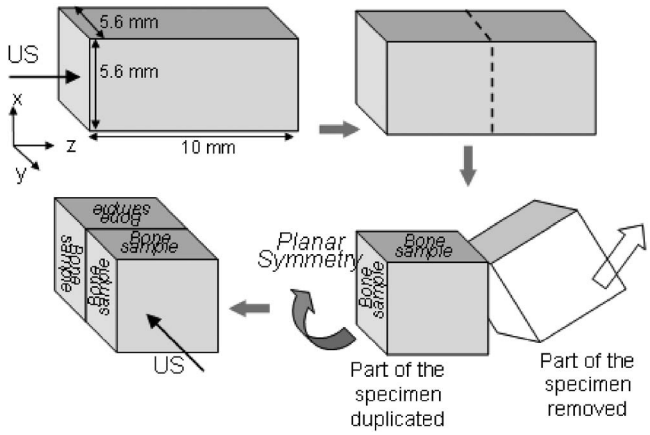


FIG. 1. Illustration of a specimen transformation for FDTD simulation along the *y* axis. The original size of the 3D bone models was 10 mm along *Z* axis and 5.6 mm along *X* and *Y* axis. FDTD simulations along *Z* axis were computed using the original models. For FDTD simulations along the two orthogonal axis (*X* or *Y*) each sample was cut by a midplane perpendicularly to the *Z* axis and transformed via a planar symmetry. A similar transformation was achieved for FDTD simulation along *X* axis.

samples, 3D data sets were deduced from the projections of the microstructure using microtomographic reconstruction techniques. Then a threshold was applied to separate bone from the background, leading to 34 3D binary data sets, similarly to what was done in (Haïat *et al.*, 2007).

In this study, ultrasonic propagation was studied in the three orthogonal directions (*X*, *Y*, and *Z*), in order to determine the effect of insonication angle on ultrasonic propagation. The bone models directly obtained from SR-μCT were used to simulate ultrasonic propagation in the *Z* direction. The same 34 bone models had to be geometrically transformed in order to simulate ultrasonic propagation in directions *X* and *Y* over a sufficiently large distance of 1 cm, so that both wave modes could potentially be distinguished. This transformation, illustrated in Fig. 1, was achieved in two steps. First, each virtual sample was cut by a plane perpendicular to the *Z* axis located approximately in the middle of the structure (at 5.6 mm from the plane $Z=0$). Second, one half of the specimen is duplicated by planar symmetry, the symmetry plane of which is perpendicular to the new direction of propagation (*X* or *Y*). The transformation results in a sample of 10 mm in length along the new propagation direction and 5.6 mm along both orthogonal directions. Each bone sample hence led to two new virtual bone models. This transformation leads to a total of 102 virtual bone models, referred to as original bone models, and corresponding to the 34 original samples and to the three orthogonal directions. Although *X* and *Y* were chosen in order to correspond approximately to the principal orientations of the trabecular network in this plane, the trabecular orientation is not always perfectly parallel to the *X* or *Y* axis. This may be a limitation of our geometrical transformation because slight discontinuities may arise in the trabecular orientation due to the procedure used to construct the samples in the *X* and *Y* directions. However, the mean structural anisotropy of all modified samples was assessed using the modified samples themselves and not the initial samples.

The 102 original bone models were used to evaluate the impact of bone volume fraction and generate new structures with thinner or thicker trabeculae. An iterative dedicated algorithm described elsewhere (Haïat *et al.*, 2006a; Barkmann *et al.*, 2007; Haïat *et al.*, 2007) was used to modify the initial microarchitecture of the trabecular network. Two scenarios of microarchitecture alteration were considered. The first one corresponds to an erosion of the trabecular network (i.e., decrease of BV/TV) and is assumed to model the thinning of the trabeculae, which is known to occur in osteoporotic patients (Luo *et al.*, 1999; Pothuaud *et al.*, 2000). The second one corresponds to the inverse scenario in which a dilation procedure is applied to the network, resulting in an increase of BV/TV. From each original bone model, a set of 4 to 13 new 3D trabecular bone structures, referred to as "simulation series," were constructed with this process. The whole process ended with 34 simulations series corresponding to a propagation direction along axis *Z* and two other sets of 34 simulations series corresponding to propagation directions along *X* and *Y*, respectively. This resulted in a total of 530 bone structures (including the original bone models which have their BV/TV ranging from 4.8% to 23%) with BV/TV ranging from 5% to 25%, which corresponds to the physiological range for human femoral trabecular bone. The simulation of wave propagation described in the next subsection was performed for each numerical bone model obtained after the application of the dilatation/erosion procedure and in the three perpendicular directions.

Each of the 102 original bone models was analyzed in term of structural anisotropy in order to evaluate the main direction of trabeculae alignment as well as their degree of anisotropy (DA). Therefore, three-dimensional morphological parameters were computed following a 3D mean intercept length (MIL) technique (Simmons and Hipp, 1997). This method is based on a three-dimensional version of the direct secant algorithm, and relies on the computation of the number of intercepts between bone/marrow interfaces and test lines with different directions in 3D space. The orientation and anisotropy of the trabecular network were estimated by fitting the MIL measurements with an ellipsoid. The DA is defined as the ratio of longest axis to shortest axis lengths. In addition, the direction of the long axis of the ellipsoid was determined. The main trabecular alignment (MTA) was given by the reference axis (*X*, *Y*, or *Z*) which was found to be the closest to the main direction of the ellipsoid.

B. Simulation of ultrasonic propagation

Simulations of ultrasonic waves propagation through trabecular bone were performed using the FDTD simulation software SIMSONIC developed by our group (Bossy *et al.*, 2004b, 2005). It is based on a resolution scheme, initially developed by de Virieux in geophysics (Graves, 1996). The value of the time step and of the cubic lattice size used in all simulations is respectively of 4.3×10^{-3} μs and of 30 μm. A numerical solution to 3D linear elastic wave propagation is computed by solving two coupled first-order differential equations. The geometry of the structure is defined by a 3D regular mapping of the mechanical properties (density, elastic coefficients) of the media.

TABLE I. Density, elastic constants, and corresponding wave propagation velocities in all simulations of ultrasonic propagation for the bone matrix.

	C_{11} (GPa)	C_{44} (GPa)	VL (m/s)	VT (m/s)	ρ (g/cm^2)
Bone tissue properties	29.6	5.994	4000	1800	1.85

This simulation tool fully takes into account reflections, refractions and mode conversions at all bone/fluid interfaces (Graves, 1996). However, our model does not account for absorption phenomena such as viscous effects occurring in the bulk or at the interface between the bone and the fluid.

The ultrasonic pressure source is defined over the whole surface at $z=0$ mm, z corresponding to the propagation axis of the considered sample. To simulate the propagation of planar waves through samples of limited transverse dimensions, symmetric boundary conditions are applied on the faces of the simulation domain parallel to the z axis. Therefore, the propagation of planar waves through trabecular structures periodized by the boundary conditions is modeled. The planar source of 5.6×5.6 mm^2 emits a broadband pressure pulse centered at 1 MHz ($\lambda=1.5$ mm in water), similar to pulses generated experimentally. Broadband pulse centered at 1 MHz was used to correspond to the experimental conditions most frequently described in literature and also to match frequencies used in clinical devices. A planar receiver is placed at the other end of the simulation mesh (at $z=10$ mm) and the signal is averaged over its surface.

For all samples, a 300 μm thick parallelepiped layer of water was positioned between the source and the trabecular network, in order to avoid any direct excitation of the trabecular network. Similarly, the same layer of water was positioned between the trabecular network and the receiver. This 300 μm water layer was set in order to impose homogeneous boundary conditions, which are given in term of displacement for the emitter. Without this layer, bone matrix as well as water could be in contact with the emitter, and the displacement in the bone matrix could affect the propagation more importantly than the displacement in water, due to mode conversion effects, thus generating artefacts in the measurements. Similarly, the water layer at $Z=10$ mm ensures that only water is in contact with the receiver.

For all simulations, the marrow is considered to be non viscous water, with density of 1 g/cm^3 and acoustic wave velocity of 1500 m/s. Bone tissue is assumed to be isotropic, non absorbing and homogeneous (i.e., same material properties are associated with all the bone voxels). Therefore, all the attenuation predicted by our model is due to scattering phenomena. Values of C_{11} (Njeh *et al.*, 1999; Turner *et al.*, 1999; Hengsberger *et al.*, 2001; Raum *et al.*, 2006), C_{44} (Wu and Cubberley, 1997), and mass density (Keaveny and Hayes, 1993; Morgan *et al.*, 2003) corresponding to human bone were used herein. Table I shows the values used in the present study, which are the same for all the simulations.

III. SIGNAL ANALYSIS

A. Position of the problem

Cancellous bone supports two compressional modes that can be observed under certain conditions (Hosokawa and Otani, 1998). However, in many experimental studies, only a single wave has been observed (Nicholson *et al.*, 1996; Strelitzki and Evans, 1996; Droin *et al.*, 1998; Wear, 2000). Likewise, both wave modes could not always be distinguished in our simulations. This situation corresponds to the "mixed modes" described by Marutyan *et al.* (2006) in a recent study where both waves overlap in time, which makes the signal analysis complicated.

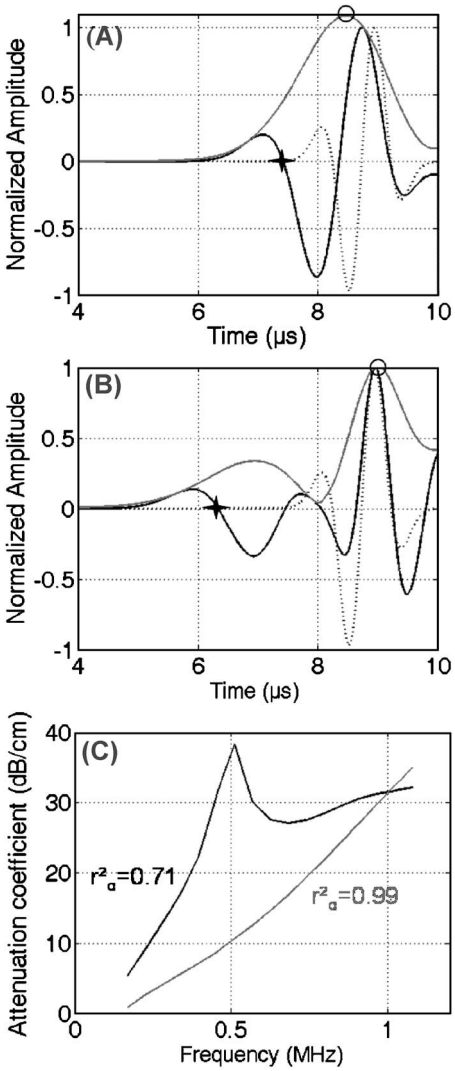


FIG. 2. Typical wave form obtained from the numerical simulation when the slow and fast wave are: (A) Overlapping and (B) separated in time. In (A) and (B), the full black line represents the rf signal transmitted through the bone specimen, the gray line its envelope and the dotted line the reference signal transmitted through water. The crosses indicate the first zero crossing and the circle the maximum of the envelope. (C) The gray line and black lines respectively correspond to the attenuation coefficient of the rf signals shown in (A) and (B). The determination coefficients (r^2) of a linear fit to the frequency-dependent attenuation coefficients is indicated in (C).

Figure 2 illustrates the difficulty of possible mixed mode situation by comparing the results obtained from the numerical simulation performed with two bone structures having similar bone volume fractions (equal to 20.6% and 20.8%). The dotted line in Figs. 2(A) and 2(B) shows the reference signal transmitted through water. The full black line of Fig. 2(A) shows a wave form obtained through a bone numerical model with a BV/TV value of 20.6%, in the direction of propagation parallel to Z. This signal, showing no visible

interference effects, may be the result from the propagation of a single mode or from the propagation of two modes that overlap in time. For each sample, the frequency-dependent attenuation coefficient $\alpha(f)$ was derived from the log-ratio of the magnitude spectrum of the signal transmitted through bone to the magnitude spectrum of the signal transmitted through water. The gray line of Fig. 2(C) shows the corresponding frequency-dependent attenuation coefficient $\alpha(f)$, which varies quasilinearly with frequency. The determination coefficient r_α^2 of the linear fit to the attenuation coefficient versus frequency computed between 200 kHz and 1.1 MHz is equal to 0.99. There is a general consensus on the quasilinear frequency dependence of attenuation in experimental studies where the signal transmitted through cancellous bone behaves as if a single wave mode propagates (Chaffai *et al.*, 2000; Wear, 2001). Moreover, using our 3D numerical simulation tool, the attenuation coefficient has also been shown to vary linearly with frequency (Bossy *et al.*, 2005; Haïat *et al.*, 2007). Therefore, we assume that including absorption phenomena in our model would not modify significantly the results. There is a general consensus on the quasilinear frequency dependence of attenuation in experimental studies where the signal transmitted through cancellous bone behaves as if a single wave mode propagates [17, 18]. Moreover, using our 3D numerical simulation tool, the attenuation coefficient has also been shown to vary linearly with frequency [1, 14]. Therefore, we assume that including absorption phenomena in our model would not modify significantly the results. In contrast, Fig. 2(B) shows the wave form obtained for another bone sample with a BV/TV of 20.8%, for which an additional early low frequency component can be observed and reveals interference effects. The black line of Fig. 2(C) displays the corresponding attenuation coefficient, which behaves far less linearly ($r_\alpha^2=0.71$) with frequency than the one derived from the wave form of Fig. 2(A). Therefore, an apparent non linear attenuation coefficient results from the interference between the fast and slow wave modes.

The nonlinear frequency dependence of the attenuation coefficient can be explained by considering the phase of the transmitted waves. When the transmission of the incident wavefront through the sample leads to two wavefronts that propagate with different velocities, the phase difference between the two wavefronts in the transmitted signal is equal to $\Delta\Phi=(k_{\text{fast}}-k_{\text{slow}})x$, where k_{fast} and k_{slow} are the wave numbers of the fast and slow waves respectively and x the sample thickness. Obviously, the phase difference $\Delta\Phi$ is a function of frequency. If the signal bandwidth contains frequencies for which the phase difference is an odd multiple of π, the signal at this specific frequency cancels out. Such an effect, generally referred to as the phase cancellation effect, results in an anomalously high value of the attenuation coefficient. Because the phase cancellation occurs at specific frequencies and because other frequency components contained in the signal are less out of phase, the resulting apparent frequency-dependent attenuation coefficient is strongly nonlinear. The final trend in the frequency dependence is thus determined by the phase and amplitude distribution of the frequency components of the pulse (i.e., by the respective speed and attenuation coefficient of the fast and slow waves). Such phase cancellation effect does not occur if a single wave front is transmitted. In this latter case, a linear frequency-dependent attenuation applies to the frequency components of the pulse and the measured attenuation coefficient is identical to the native attenuation coefficient of the sample.

As the goal of our work is to analyse the effect of BV/TV and of structural anisotropy on the separation in time of the slow and fast wave modes, we have used a heuristic method aiming at automatically detecting the presence of an interference in the received wave form.

B. Evaluation of wave separation

The criterion for the determination of the time separation of both wave modes was established based on our observations reported in Fig. 2 that the presence of two wave modes separated in time produces an interference that results in an apparently nonlinear attenuation coefficient. The determination coefficient r_α^2 being a measure of the degree of closeness to the linear relationship between attenuation and frequency, the criterion was obtained by comparing r_α^2 to a given threshold. Different thresholds have been tested and the final threshold was set at 0.97. When $r_\alpha^2>0.97$, the attenuation is quasilinear and the signal transmitted through bone is almost undistorted in comparison to the reference signal, thus corresponding either to a single pulse or to the "mixed mode" reported by Marutyan *et al.* (2006). In contrast, when $r_\alpha^2<0.97$, the transmitted signal is distorted and the two waves are considered sufficiently separated in time to produce an interference effect, which is responsible for the deviation from an approximate linearity with frequency of the attenuation coefficient.

Figures 3(A) and 3(B) show a wave form and its corresponding frequency-dependent attenuation coefficient for $r_\alpha^2=0.96$, which is just below the selected threshold, corresponding to the direction of propagation parallel to *X*. A weak additional low frequency early contribution can be observed between 6 and 8 μs in Fig. 3(A). When $r_\alpha^2<0.97$, both wave modes could always be observed in the corresponding wave form.

C. Fast and slow wave velocity assessment

Two wave velocities (first zero-crossing and envelope velocities) were determined for all radio frequency (rf) signals obtained with our simulation tool, using two different time markers. The envelope velocity was computed using as time marker the time of the maximum amplitude of the absolute value of the Hilbert transform of the corresponding analytical signal. Both wave speed estimations are illustrated in Figs. 2(A) and 2(B), where the times-of-flight obtained with the first zero-crossing and envelope techniques are indicated, respectively, by a cross and a circle in both wave forms. When both waves are observed (i.e., when $r_\alpha^2<0.97$), the envelope velocity corresponds to that of the slow wave because, in our simulation, this contribution is found to be more energetic than that of the fast wave for all the wave forms obtained. In that case, the first zero-crossing velocity corresponds to the velocity of the fast wave mode. When both wave modes overlap, these velocities cannot simply be

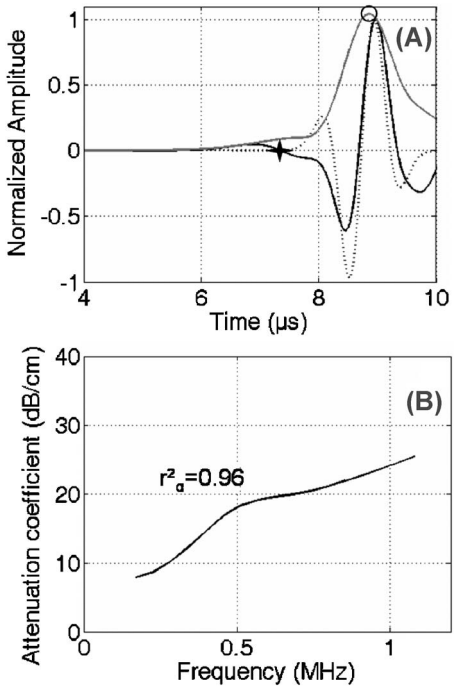


FIG. 3. Typical wave form obtained from the numerical simulation when the criteria for the determination of the time separability of both wave modes corresponds to the limit (r_α^2=0.96). In (A), the full black line represents the rf signal, the gray line its envelope and the dotted line the reference signal. The crosses indicate the first zero crossing and the circle the maximum of the envelope. (B): The black line corresponds to the attenuation coefficient of the rf signal shown in (A). The determination coefficients for both frequency-dependent attenuation coefficients vs frequency is indicated in (B).

related to either fast or slow wave mode velocity, due to the mixed mode conditions. As an illustration, the value of the envelope velocity is 1564 m/s and the value of the first zero-crossing velocity is 1704 m/s for the mixed mode of Fig. 2(A). For the interfering wave forms of Fig. 2(B), the value of the envelope velocity is 1446 m/s and the value of the first zero-crossing velocity is of 2083 m/s.

IV. RESULTS

A. Effect of bone volume fraction

The effect of changes of BV/TV (resulting from the application of the erosion/dilation procedure) is studied by considering each simulation series, corresponding to one sample and one direction of propagation. Increasing BV/TV of a bone model in a simulation series may lead to two situations only, which allows partitioning of the set of 102 simulation series into two groups. Therefore, two different groups (refered to as groups A and B) of simulation series were defined and each series was classified in one of the two groups. For simulation series of the first group (referred to as group A), increasing BV/TV does not lead to the observation of both wave modes. For all BV/TV values, the received signal always remains similar to the single pulse or the mixed mode depicted in Fig. 2(A) with a highly apparent linear attenuation coefficient. Results of the simulation for a typical simulation series of group A are illustrated in Fig. 4, in which the simulated signals are plotted for increasing BV/TV values (7.5%, 12.9%, 16.1%, and 21%) along with the determination coefficient of the attenuation coefficient versus frequency. The BV/TV value of the original bone sample corresponding to Fig. 4 is 21%. All four wave forms correspond to the direction of propagation parallel to *Z*.

A visual inspection of rf traces reveals no detectable effect of an interference between the two modes. Figure 5(A) illustrates the values of both wave velocities (first zero-crossing and envelope velocities) as a function of BV/TV for this same bone model belonging to group A. In this first set of simulation series (group A), both the zero-crossing and envelope velocities were found to increase gently with BV/TV, this increase being more important for the zero-crossing velocity. Figure 5(B) shows the values of r_α^2 for the corresponding simulation series. As expected, the attenuation coefficient varies quasilinearly with frequency and values of r_α^2 consistently remain above the threshold. In summary, for each simulation series belonging to group A, both wave modes are consistently superimposed. As a consequence, both wave modes always overlap for the sample with the lowest BV/TV of all simulation series.

For the simulation series of the second group (referred to as group B), increasing BV/TV always leads, above a certain value of BV/TV, to the observation of an interfering signal of the type described in Fig. 2(B) (i.e., distinct fast and slow wave modes), with a nonlinear apparent attenuation coefficient and therefore a value of r_α^2 below the threshold. Results of the simulation for a typical simulation series of group B are illustrated in Fig. 6, in which the simulated signals are plotted for increasing BV/TV values (7.7%, 15.2%, 18.4%, and 21.5%), along with the determination coefficient of the attenuation coefficient versus frequency. The BV/TV value of the original bone sample corresponding to Fig. 6 is 7.7%. These four wave forms correspond to the direction of propagation parallel to *Y*. No visible interference effect can be detected on the rf traces corresponding to BV/TV values of

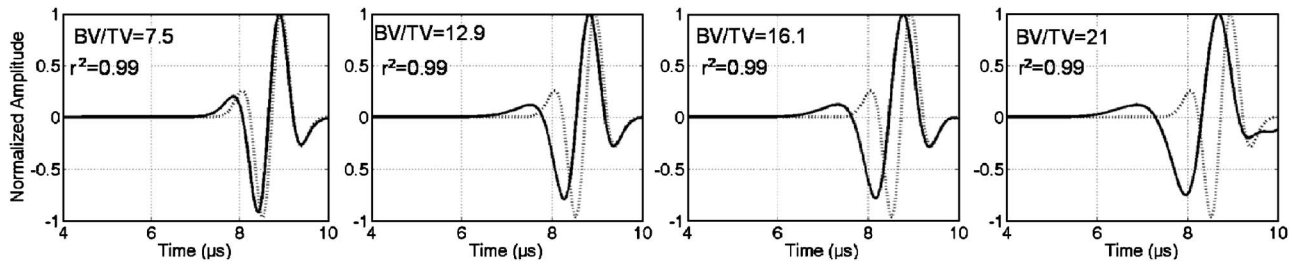


FIG. 4. Simulated radio frequency signals obtained for four bone models (BV/TV of 7.5%, 12.9%, 16.1%, and 21.0%, respectively) belonging to the same simulation series of group A. For each signal, the value of the determination coefficient of a linear fit of the attenuation coefficient versus frequency method is indicated.

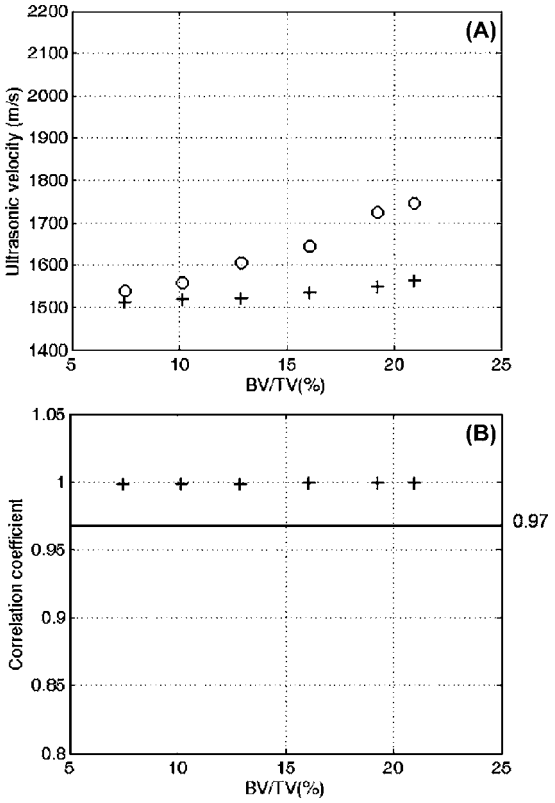


FIG. 5. (A) First zero-crossing velocity (circles) and envelope velocity (crosses) as a function of BV/TV for a simulation series of group A (both wave modes overlap in time). (B) Correlation coefficient of the attenuation coefficient vs frequency as a function of BV/TV for the same simulation series. The threshold indicating wave separation is represented by the full line.

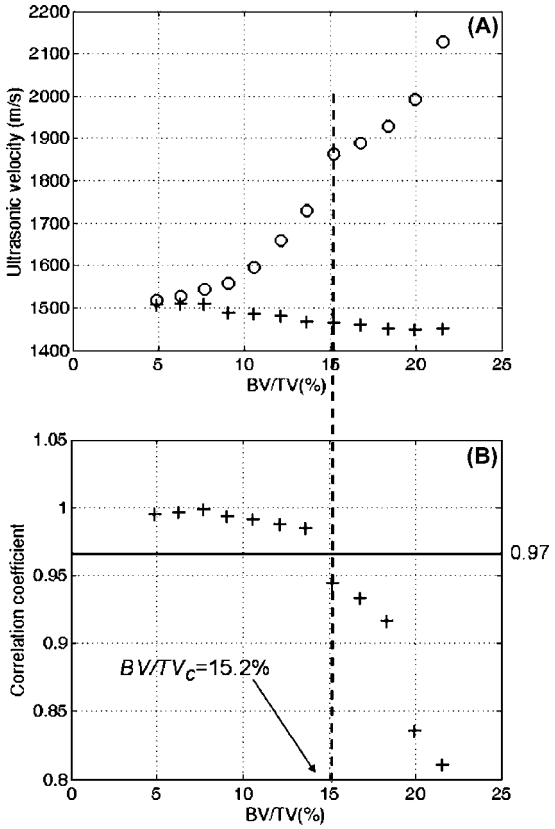


FIG. 7. (A): First zero-crossing velocity (circles) and envelope velocity (crosses) as a function of BV/TV for a simulation series of group B (both wave modes separated for high BV/TV values). (B) Correlation coefficient of the attenuation coefficient vs frequency as a function of BV/TV for the same simulation series. The threshold indicating wave separation is represented by the full line.

7.7%, whereas the rf traces corresponding to BV/TV values of 15.2% up to 21.5% reveal a composite signal resulting from the time separation between a fast and slow mode. Values of both wave velocities (first zero-crossing and envelope velocities) are plotted in Fig. 7(A) as a function of BV/TV and Fig. 7(B) shows the values of r_α^2 for this simulation series of group B. The first zero-crossing velocity is found to increase, whereas the envelope velocity (corresponding to the slow wave velocity) decreases with BV/TV. Noticeably, the increase of the zero-crossing velocity as a function of BV/TV is much more pronounced than for simulation series of group A. Moreover, as illustrated in Fig. 6, the amplitude of the fast wave mode was found to be an increasing function of BV/TV.

In what follows, $\mathrm{BV/TV}_C$ denotes the value of BV/TV for which the value of r_α^2 reaches the threshold value of 0.97. For all structures of the simulation series having a BV/TV value less than $\mathrm{BV/TV}_C$, the value of r_α^2 was always greater than 0.97, and conversely, the values of r_α^2 remain consistently below the threshold value of 0.97 for all structures of the simulation series having a BV/TV value above $\mathrm{BV/TV}_C$. The value of $\mathrm{BV/TV}_C$ is not constant but varies as a function of the simulation series. For all simulation series of group B, the value of $\mathrm{BV/TV}_C$ was determined as accurately as possible. To this end, additional bone models were constructed by varying the BV/TV in consecutive steps of about 2.5%, which gives the order of magnitude of the error on the estimation of $\mathrm{BV/TV}_C$. In the case of Fig. 7, the estimated value of $\mathrm{BV/TV}_C$ is 15.2%. In summary, for each simulation series belonging to group B with a BV/TV lower than $\mathrm{BV/TV}_C$, the corresponding rf signal is close to a single wave form,

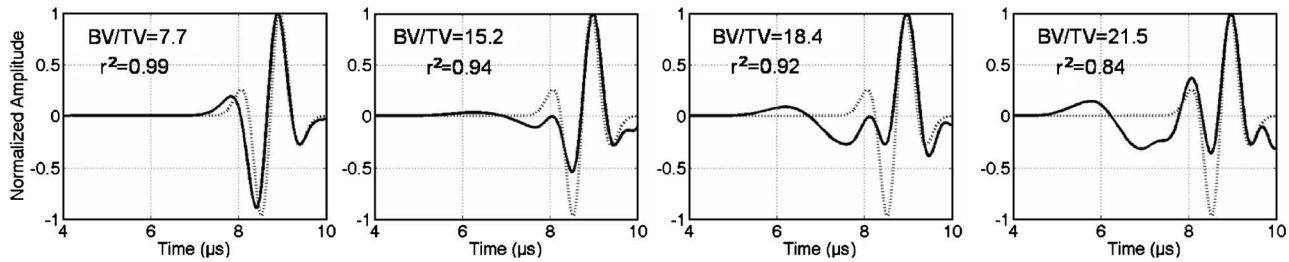


FIG. 6. Simulated radio frequency signals obtained for four bone models (BV/TV of 7.7%, 15.2%, 18.4%, and 21.5%, respectively) belonging to the same simulation series of group B. For each signal, the value of the determination coefficient of a linear fit of the attenuation coefficient vs frequency method is indicated.

TABLE II. Number of simulation series belonging to group A (single pulse or mixed mode) and group B (separation of both wave modes for high BV/TV values) for the three perpendicular directions of propagation. Italicized is the number of simulation series for which the main orientation is parallel to the ultrasonic direction of propagation. Results are given for each MTA. For each propagation direction, the total number of simulation series is 34.

Direction of propagation	MTA	Group A	Group B
Z	*X*	11	1
	Y	20	0
	Z	*1*	*1*
	All directions	32	2
X	*X*	*8*	*13*
	Y	13	0
	Z	0	0
	All directions	21	13
Y	*X*	9	2
	Y	*5*	*17*
	Z	1	0
	All directions	15	19

whereas for each simulation of all simulation series belonging to group B with a BV/TV higher than BV/TV_C, both wave modes can be distinguished.

B. Effect of the direction of propagation

Table II shows the number of simulation series belonging to groups A and B for each direction of propagation. When ultrasound propagates in direction *Z*, a single mode or a mixed mode was observed in the great majority of cases. Time separation of both waves could be observed for only two simulation series for which BV/TV_C was 18.8% and 15.1%, respectively. In contrast, for the two other perpendicular propagation directions (*X* and *Y*), it was not uncommon to observe both wave modes as 13 (respectively 19) simulation series in the direction *X* (respectively *Y*) belongs to group B. In this case, the value of BV/TV_C ranges from 9.7% to 21.85%.

Table II also shows the number of simulation series belonging to each group, for which the MTA is parallel either to *X*, *Y*, or *Z*. These data are shown for the three directions of propagation. The results demonstrate that time separation of both fast and slow waves were more frequently observed when MTA was parallel to the propagation direction. For example, for the direction of propagation *Z*, 32 out of 34 specimens had their MTA aligned with a direction perpendicular to the ultrasound beam. Among these 32 specimens with MTA perpendicular to *Z*, 31 belong to group A. Similarly, for the directions *X* and *Y*, the great majority of cases (100% and 83% for *X* and *Y* directions, respectively) with MTA perpendicular to the propagation direction were found to belong to group A. Conversely, all cases for *X* and 88% of cases for *Y* found in group B had their MTA parallel to the direction of ultrasound propagation.

As shown in Table II, the total number of sample having their main trabecular orientation in a given direction (obtained by adding the number of simulation series obtained in groups A and B) depends on the direction of propagation considered. This discrepancy is due to the fact that the trabecular orientation of the original samples is not always perfectly parallel to the *X* or *Y* axis (as described in Sec. II A). However, because the difference between the *X* or *Y* direction and the actual principal orientation of the trabecular network (mean value of 11°) is relatively weak we believe that this limitation does not significantly impact our results.

TABLE III. Mean value and standard deviation of the degree of anisotropy (DA) obtained for each direction of propagation for simulation series of each group having their main trabecular orientation parallel to the direction of propagation. The numbers in brackets indicate the number of simulation series taken into account for the statistics.

Direction of propagation and anisotropic direction	MTA	Group A	Group B
Z	*Z*	1.1 (1)	1.37 (1)
X	*X*	1.25 ± 0.09 (8)	1.45 ± 0.2 (13)
Y	*Y*	1.28 ± 0.08 (5)	1.43 ± 0.21 (17)
All data pooled		1.25 ± 0.09 (14)	1.44 ± 0.2 (33)

C. Effect of the degree of anisotropy

In order to assess the effect of the DA on the time separation of both wave modes, the mean value of DA obtained for simulation series having their MTA parallel to the direction of propagation was investigated. Table III shows the mean values and standard deviations obtained for each group and each direction of propagation. Only simulation series having the direction of propagation parallel to MTA are considered. Results are not significant for direction *Z*, due to the small number of samples. However, for directions *X* and *Y*, the mean value of DA is lower for the samples of group A than for the samples of group B, suggesting that the time separation of both wave modes is determined not only by the relative orientation of MTA with respect to ultrasound beam axis, but is also related to the DA.

In order to gain further insight into the conditions on bone volume fraction and structural anisotropy required to observe two wave modes, the relation between the DA and BV/TV_C for samples of group B was investigated in more details. For direction *X* (respectively *Y*), Fig. 8(A) [respectively Fig. 8(B)] shows a plot of DA as a function of BV/TV_C for all simulation series of group B having MTA aligned with *X* (respectively *Y*). Figure 8(C) shows the same results for the three directions pooled. Following the results of Table II, 13 and 17 simulation series are considered respectively for the directions *X* and *Y*, and 31 when all directions are pooled. For both directions *X* and *Y*, BV/TV_C is shown to decrease when DA increases with a significant negative correlation coefficient of −0.74 (respectively −0.7) in the *X* (respectively *Y*) direction. This indicates that for MTA aligned with the direction of propagation, time separation of fast and slow wave modes does not occur for specimen with low bone volume fraction and low DA. The lower the DA, the higher bone volume fraction must be for a separation of the two waves.

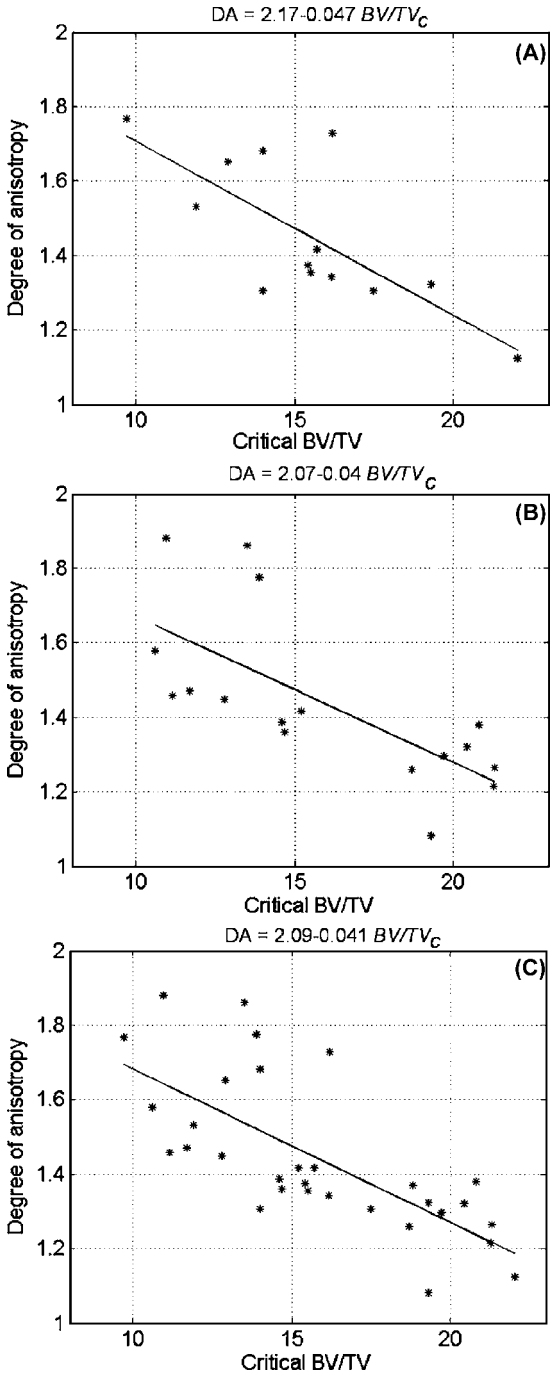


FIG. 8. Variation of the degree of anisotropy as a function of the BV/TV at which both waves start being observed for simulation series of group B where the direction of propagation is similar with the main orientation of the trabeculae. (A) Results for direction of propagation *X*, (B) for direction of propagation *Y*, and (C) all directions pooled. The equation on top corresponds to the linear fit obtained with a least-square regression.

V. DISCUSSION

As suggested by previous authors (Hosokawa and Otani, 1998; Cardoso *et al.*, 2003), the separation in time of fast and slow waves is complicated by the fact that the localization in time (or velocity) and amplitude (or attenuation) of both waves depend on several factors, including the size of the specimen, porosity level (or bone volume fraction), and direction of propagation relatively to the orientation of the trabecular network (or MTA). As a consequence, even though the fast and slow waves may be present, they are not always clearly separated and may overlap in time (Hosokawa and Otani, 1998; Cardoso *et al.*, 2003). To our knowledge, our study is the first one to assess the combined effect of bone volume fraction and structural anisotropy of trabecular bone on the time separation of slow and fast waves, whereby we demonstrate that time separation of fast and slow waves depends not only on the insonication angle and bone volume fraction but also on structural anisotropy. When the propagation direction is perpendicular to MTA (57 simulation series), wave separation occurs only in three cases. Insonication along MTA (45 simulation series) clearly favors the appearance of both wave modes, as shown in previous studies (Nicholson *et al.*, 1994; Hosokawa and Otani, 1997): out of 45 simulation series with parallel insonication, 31 ones were found in group B. Nevertheless, for parallel insonication, the separation of the waves was not always readily observable and a condition linking structural anisotropy and bone volume fraction also affects the observation of the two wave modes separated. In the case of specimens with low bone volume fraction and low DA, the two waves are not easily identified, even when the insonication is parallel to MTA. Time separation of the two waves requires a certain level of bone volume fraction (BV/TV_C) or of DA, BV/TV_C being linked to DA by an approximately linear decreasing function illustrated in Fig. 8.

The decrease of BV/TV_C when DA increases can be explained by considering the propagation in a biphasic medium. Fast waves are mostly related to a propagation mode mainly involving the solid phase, whereas slow waves are related to the fluid phase (Cardoso *et al.*, 2003). Therefore, the arrival time (and also the amplitude) of the fast wave is simply determined by the velocity times the distance of propagation (Nagatani *et al.*, 2007). This propagation length is in turn determined both by bone volume fraction and anisotropy of the structure (assuming insonation parallel to MTA). Short path length through the trabecular structure (either due to a weakly anisotropic structure or to a highly porous specimen) results in a fast wave front with low amplitude and arrival time too close to that of the slow wave. Therefore, for short path length through the trabecular structure the fast wave should remain undetectable. The detection of the fast wave requires the propagation of the wave through a sufficient amount of trabecular structure, which depends on the structural orientation of the structure and on the bone volume fraction. The result showing that BV/TV_C is negatively correlated with DA indicates that separation of both wave modes may occur at low BV/TV in highly anisotropic structures, whereas time separation of both waves in weakly anisotropic structures requires high BV/TV values. This result shows the coupling between DA, BV/TV and the time separation of both wave modes.

A limitation of this study may result from the assumption that a simulation series (which is obtained from a single sample) has similar architecture, particularly similar DA, after the erosions and dilations. The DA of all bone models of a simulation series have been computed and found to vary at maximum by an amount of $\Delta DA=0.1$, an amount which remains small compared to the difference between the aver-

age DA values of groups A and B (see Table III). Nevertheless, these small variations in microarchitecture within a simulation series may be responsible for the scatter of the results shown in Fig. 8.

A heuristic algorithm was developed to characterize the time separation of the two modes. This algorithm is based on the linearity of the frequency-dependent attenuation. To characterize the time separation of the two modes, we introduced the determination coefficient r_α^2, representing a measure of the degree of linearity between attenuation and frequency. As the value of r_α^2 decreases, time separation of the two wave modes occurs. An arbitrary threshold of $r_\alpha^2=0.97$ was chosen, above which the received signal is apparently not distorted and reveals no interference effects, which is consistent with results obtained *in vitro* (Laugier *et al.*, 1994a, 1997; Chaffai *et al.*, 2000), *in vivo* (Laugier *et al.*, 1996) and *in silico* (Bossy *et al.*, 2005; Haïat *et al.*, 2006a; Barkmann *et al.*, 2007; Haïat *et al.*, 2007) when the insonication is perpendicular to MTA. Below the threshold, time separation of both wave modes is responsible for an interference effect resulting in a complex wave form and a nonlinear apparent attenuation coefficient as shown in Fig. 2(C). The slope of the frequency-dependent attenuation coefficient is lower in the upper part of the frequency bandwidth ($f>600$ kHz), than in the lower part ($f<450$ kHz). A peak can be observed in the attenuation coefficient. These results are consistent with Fig. 6 of the experimental study of Cardoso *et al.* (2003) that shows a similar behavior of the attenuation coefficient. In their work, the authors, applying an analysis in the frequency domain to identify both waves, defined amodified frequency-dependent ultrasonic attenuation (FDUA) for the slow and fast wave contributions as the slope of the attenuation coefficient in both frequency bandwidths. The value obtained from Fig. 2(C) of the present study for the FDUA of the slow and fast waves (respectively about 80 and 15 dB/cm/MHz) span within the range obtained experimentally by Cardoso *et al.* (2003). This constitutes a further validation of the ability of our simulation method to correctly predict propagation phenomena observed in experiments.

The arbitrary threshold for r_α^2 was set empirically by inspecting all wave forms. We verified that for all values of r_α^2 lower than the threshold (0.97), both wave modes could be visually detected and Figs. 2 and 3 are given only for illustrative purposes. The approach using the correlation coefficient of the frequency-dependent coefficient finds its roots in the origin of the nonlinear trend of the attenuation coefficient. The method is far from being ideal and sometimes, a weak wave modes separation might not be detected because the associated r_α^2 is higher than 0.97. However, the method allows a first rapid and automatic classification of the signals and the identification of trends in the relationships between fast and slow wave modes separation on one hand, and BV/TV and structural anisotropy on the other hand. More sophisticated and robust signal processing techniques are required for a clear separation of wave modes. Note that the threshold value of 0.97 depends in a complicated way on the characteristics of the specimen. If absorption was included in the numerical simulation, the threshold would have probably been different. However, we do not intend to give here an absolute threshold, but simply a heuristically determined threshold value that holds for the signal generated by the current simulations.

Compared to the work of different authors (Hosokawa and Otani, 1997; Fellah *et al.*, 2004; Hosokawa, 2005b), both slow and fast waves obtained in this study were never completely separated in time [see Fig. 2(B)]. This may be caused by the short propagation distance (limited by the memory requirements to run 3D FDTD simulations), the low central frequency [1 MHz in our study compared to 2.25 MHz in Fellah *et al.* (2004)] and the small frequency bandwidth [compared to Hosokawa and Otani (1997) and Hosokawa (2005b)], leading to longer pulses. The procedure developed to relate BV/TV and structural anisotropy to the separation in time of both wave modes could be similarly applied for wave forms obtained with shorter reference signals. The results obtained for the determination of the group of the simulation series or for the values of BV/TV_C may be different but the principal trends drawn in the present study should remain the same.

Note that the heuristic approach adopted here presents limitations as it does not allow to quantify with accuracy velocity and attenuation characteristics of each wave mode. However, our primary interest was in characterizing the overlapping of both wave modes rather than recovering their true acoustical properties. Original procedures have been developed by other groups to identify interfering wave modes and obtain the estimates of the true propagation parameter values (Cardoso *et al.*, 2003; Marutyan *et al.*, 2007). In most of previous works, the determination of the velocity of both wave modes was performed by analysing the phase velocity of both contributions using a time window (Hosokawa and Otani, 1997, 1998; Hughes *et al.*, 1999; Kaczmarek *et al.*, 2002) or a frequency linear filtering (Cardoso *et al.*, 2003). Similar approaches were not possible in our study because both wave modes were never clearly separated, but rather always overlapped to some extent, which somewhat compromised velocity estimates.

Only recently, Marutyan *et al.* (2007) proposed a Bayesian approach to retrieve the true ultrasonic characteristics of overlapping fast and slow wave modes. Applying this approach to our signals might be useful to relate the true characteristics of the fast and slow wave modes to the bone properties. Nevertheless, as shown in what follows, our results appear to be fairly consistent with previous reports. First, when both wave modes overlap in time (simulation series of group A), the values of velocities are in good agreement with the results obtained experimentally *in vitro* on the same trabecular bone samples (Jenson *et al.*, 2006) measured in the direction *Z*. As shown in Fig. 5(A), both first zero-crossing and envelope velocities increase with BV/TV, but the first zero-crossing velocity increases more rapidly. Similar results were observed with experimental signals showing the first zero-crossing to be more sensitive to BV/TV than the envelope velocity (Haïat *et al.*, 2005, 2006b). Second, when both wave modes are separated in time, the fast wave velocity is found to be an increasing function of BV/TV, as reported by others (Williams, 1992; Cardoso *et al.*, 2003). Our values obtained for the fast wave velocity remain slightly below

those obtained for bovine trabecular bone. This can be explained by the higher bone volume fraction of bovine trabecular bone compared to human femoral trabecular bone. More interestingly, the slow wave velocity decreases with BV/TV, in agreement with experimental results (Hosokawa and Otani, 1997; Cardoso *et al.*, 2003) and Biot theory predictions (Williams, 1992; Hosokawa and Otani, 1997). Our values for the slow wave velocity are similar to results previously reported (Williams, 1992; Hosokawa and Otani, 1997; Cardoso *et al.*, 2003).

In this work, the fast wave exhibited a lower frequency content than the fast wave, in agreement with results reported elsewhere (Kaczmarek *et al.*, 2002; Cardoso *et al.*, 2003). In our study, the amplitude of the fast wave was always found to be lower than the amplitude of the slow wave, which is not in agreement with some of the previous experimental (Cardoso *et al.*, 2003) or theoretical findings (Fellah *et al.*, 2004; Hosokawa, 2005b), that predicted that the largest wave could in certain instances be the fast wave and in other instances be the slow wave. The amplitude of fast and slow wave may be affected by a number of factors, such as absorption (a factor which is not taken into account in the FDTD simulation), differences in experimental conditions (central frequency and frequency bandwidth, transducers placed in contact with the sample or not) and values of bone volume fraction and/or tortuosity different from values considered in the previous studies. For example, Fellah *et al.* (2004) have reported an increase of the relative amplitude of fast wave compared to the slow wave when either bone volume fraction or tortuosity increases. As shown in Fellah *et al.* (2004), our results show that the amplitude of the fast wave mode is an increasing function of BV/TV.

Currently transverse transmission QUS techniques are based on the assumption that only one wave propagates in cancellous bone. When this is not the case (two waves modes are present), the correct estimation of the frequency dependence of the acoustic parameters (attenuation and velocity) may be compromised (Hosokawa and Otani, 1998; Cardoso *et al.*, 2003). One important result is that for insonication perpendicular to the main orientation of the trabeculae, the transmitted signal apparently consists in a single pulse. Consequently, as the testing direction in clinical QUS measurements, e.g., the mediolateral direction at the heel, is perpendicular to MTA, attenuation spectral measurement or velocity measurements using adequate markers in the time domain should not be affected by limitations due to the presence of two types of waves.

VI. CONCLUSION

Our results illustrate the potential of numerical methods for the study of bone/ultrasound interaction. The numerical simulation nicely provides insights into the conditions required for the time separability of the fast and slow wave modes. Experimental results previously reported on a small number of specimens could be confirmed on a large number of virtual specimens generated by the simulations. The observation of the two transmitted compressional waves predicted by Biot's theory requires the ultrasound propagation to be parallel to that the trabecular main orientation.

Moreover, by generating specimens with controlled porosity, we could provide further insight into the interdependence between propagation of fast wave, bone volume fraction, main trabecular orientation and degree of anisotropy. These results open interesting perspectives, in particular because it elucidates the mechanism by which the two compressional waves are generated. The relation between the characteristics of the fast wave, in terms of wave speed and attenuation, to others microstructural parameters, will be useful to assess the potential of this wave mode to convey information about microstructural properties. This work hence constitutes a first step toward the resolution of the inverse problem.

ACKNOWLEDGMENT

The authors would like to acknowledge Emmanuel Bossy for his large contribution to the numerical computations.

Barkmann, R., Kantorovich, E., Singal, C., Hans, D., Genant, H., Heller, M., and Gluer, C. (**2000a**). "A new method for quantitative ultrasound measurements at multiple skeletal sites," J Clin Densitometry **3**, 1–7.

Barkmann, R., Laugier, P., Moser, U., Dencks, S., Padilla, F., Haiat, G., Heller, M., and Glüer, C.-C. (**2007**). "A method for the estimation of femoral bone density from quantitative ultrasound variables measured directly at the human femur," Bone (N.Y.) **40**, 37–44.

Barkmann, R., Lusse, S., Stampa, B., Sakata, S., Heller, M., and Gluer, C. C. (**2000b**). "Assessment of the geometry of human finger phalanges using quantitative ultrasound in vivo," Osteoporosis Int. **11**, 745–755.

Biot, M. (**1962**). "Generalized theory of acoustic propagation in porous dissipative media," J. Acoust. Soc. Am. **34**, 1254–1264.

Biot, M. A. (**1956a**). "Theory of propagation of elastic waves in a fluid-satured porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178.

Biot, M. A. (**1956b**). "Theory of propagation of elastic waves in a fluid-satured porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191.

Bossy, E., Padilla, F., Peyrin, F., and Laugier, P. (**2005**). "Three-dimensional simulation of ultrasound propagation through trabecular bone structures measured by synchrotron microtomography," Phys. Med. Biol. **50**, 5545–5556.

Bossy, E., Talmant, M., and Laugier, P. (**2004a**). "Bi-directional axial transmission can improve accuracy and precision of ultrasonic velocity measurement in cortical bone: a validation on test material," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 71–79.

Bossy, E., Talmant, M., and Laugier, P. (**2004b**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Cardoso, L., Teboul, F., Sedel, L., Oddou, C., and Meunier, A. (**2003**). "In vitro acoustic waves propagation in human and bovine cancellous bone," J. Bone Miner. Res. **18**, 1803–1812.

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

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

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

Fellah, Z. E., Fellah, M., Sebaa, N., Lauriks, W., and Depollier, C. (**2006**). "Measuring flow resistivity of porous materials at low frequencies range via acoustic transmitted waves," J. Acoust. Soc. Am. **119**, 1926–1928.

Graves, R. (**1996**). "Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences," Bull. Seismol. Soc. Am. **86**, 1091–1106.

Haïat, G., Padilla, F., Barkmann, R., Dencks, S., Moser, U., Glüer, C.-C., and Laugier, P. (**2005**). "Optimal prediction of bone mineral density with ultrasonic measurements in excised human femur," Calcif. Tissue Int. **77**, 987–996.

Haïat, G., Padilla, F., Barkmann, R., Gluer, C. C., and Laugier, P. (**2006a**). "Numerical simulation of the dependence of quantitative ultrasonic parameters on trabecular bone microarchitecture and elastic constants," Ultrasonics **44**, e289–294.

Haïat, G., Padilla, F., Cleveland, R. O., and Laugier, P. (**2006b**). "Effects of frequency-dependent attenuation and velocity dispersion on in vitro ultrasound velocity measurements in intact human femur specimens," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 39–51.

Haïat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2007**). "Variation of ultrasonic parameters with microstructure and material properties of trabecular bone: A three-dimensional model simulation," J. Bone Miner. Res. **22**, 665–674.

Hengsberger, S., Kulik, A., and Zysset, P. (**2001**). "A combined atomic force microscopy and nanoindentation technique to investigate the elastic properties of bone structural units," Eur. Cells Mater **1**, 12–17.

Hosokawa, A. (**2005a**). "Simulation of ultrasound propagation through bovine cancellous bone using elastic and Biot's finite-difference time-domain methods," J. Acoust. Soc. Am. **118**, 1782–1789.

Hosokawa, A. (**2005b**). "Simulation of ultrasound propagation through bovine cancellous bone using elastic and Biot's finite-difference time-domain methods," J. Acoust. Soc. Am. **118**, 1782–1789.

Hosokawa, A., and Otani, T. (**1997**). "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 1–5.

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

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

Jenson, F., Padilla, F., Bousson, V., Bergot, C., Laredo, J. D., and Laugier, P. (**2006**). "In vitro ultrasonic characterization of human cancellous femoral bone using transmission and backscatter measurements: relationships to bone mineral density," J. Acoust. Soc. Am. **119**, 654–663.

Kaczmarek, M., Kubik, J., and Pakula, M. (**2002**). "Short ultrasonic waves in cancellous bone," Ultrasonics **40**, 95–100.

Keaveny, T. M., and Hayes, W. C. (**1993**). "A 20-year perspective on the mechanical properties of trabecular bone," J. Biomech. Eng. **115**, 534–542.

Langton, C. M., Palmer, S. B., and Porter, S. W. (**1984**). "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91.

Laugier, P., Berger, G., Giat, P., Bonnin-Fayet, P., and Laval-Jeantet, M. (**1994a**). "Ultrasound attenuation imaging in the os calcis: an improved method," Ultrason. Imaging **16**, 65–76.

Laugier, P., Droin, P., Laval-Jeantet, A. M., and Berger, G. (**1997**). "In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and QCT," Bone (N.Y.) **20**, 157–165.

Laugier, P., Fournier, B., and Berger, G. (**1996**). "Ultrasound parametric imaging of the calcaneus: In vivo results with a new device," Calcif. Tissue Int. **58**, 326–331.

Laugier, P., Giat, P., and Berger, G. (**1994b**). "Broadband ultrasonic attenuation Imaging: A new imaging technique of the os calcis," Calcif. Tissue Int. **54**, 83–86.

Lee, K. I., Hughes, E. R., Humphrey, V. F., Leighton, T. G., and Choi, M. J. (**2007**). "Empirical angle-dependent Biot and MBA models for acoustic anisotropy in cancellous bone," Phys. Med. Biol. **52**, 59–73.

Lee, K. I., Roh, H. S., and Yoon, S. W. (**2003**). "Correlations between acoustic properties and bone density in bovine cancellous bone from 0.5 to 2 MHz," J. Acoust. Soc. Am. **113**, 2933–2938.

Lee, K. I., and Yoon, S. W. (**2006**). "Comparison of acoustic characteristics predicted by Biot's theory and the modified Biot-Attenborough model in cancellous bone," J. Biomech. **39**, 364–368.

Luo, G., Kinney, J. H., Kaufman, J. J., Haupt, D., Chiabrera, A., and Siffert, R. S. (**1999**). "Relationship between plain radiographic patterns and three-dimensional trabecular architecture in the human calcaneus," Osteoporosis Int. **9**, 339–345.

Mano, I., Horii, K., Takai, S., Suzaki, T., Nagaoka, H., and Otani, T. (**2006**). "Development of novel ultrasonic bone densitometry using acoustic parameters of cancellous bone for fast and slow waves," Jpn. J. Appl. Phys. **20**, 4700–4702.

Marutyan, K., Holland, M., and Miller, J. (**2006**). "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL55–EL61.

Marutyan, K. R., Bretthorst, G. L., and Miller, J. G. (**2007**). "Bayesian estimation of the underlying bone properties from mixed fast and slow mode ultrasonic signals," J. Acoust. Soc. Am. **121**, EL8–EL15.

Morgan, E. F., Bayraktar, H. H., and Keaveny, T. M. (**2003**). "Trabecular bone modulus-density relationships depend on anatomic site," J. Biomech. **36**, 897–904.

Nagatani, Y., Mizuno, K., Saeki, T., Matsukawe, M., and Hosoi, H. (**2007**). "Experimental and numerical study on ultrasound wave attenuation in cancellous Bone," in *2007 ICU International Congress on Ultrasonics*.

Nicholson, P. H., Muller, R., Lowet, G., Cheng, X. G., Hildebrand, T., Ruegsegger, P., van der Perre, G., Dequeker, J., and Boonen, S. (**1998**). "Do quantitative ultrasound measurements reflect structure independently of density in human vertebral cancellous bone?," Bone (N.Y.) **23**, 425–431.

Nicholson, P. H. F., Haddaway, M. J., and Davie, M. W. J. (**1994**). "The dependence of ultrasonic properties on orientation in human vertebral bone," Phys. Med. Biol. **39**, 1013–1024.

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

Njeh, C. F., Hans, D., Fuerst, T., Glüer, C.-C., and Genant, H. K. (**1999**). "Ultrasonic studies of cortical bone in vitro," in *Quantitative Ultrasound Assessment of Osteoporosis and Bone Status* (Martin Dunitz, London), p. 177.

Otani, T. (**2005**). "Quantitative estimation of bone density and bone quality using acoustic parameters of cancellous bone for fast and slow waves," Jpn. J. Appl. Phys., Part 1 **44**, 4578–4582.

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

Peyrin, F., Salome, M., Cloetens, P., Laval-Jeantet, A. M., Ritman, E., and Ruegsegger, P. (**1998**). "Micro-CT examinations of trabecular bone samples at different resolutions: 14, 7 and 2 micron level," Technol. Health Care **6**, 391–401.

Pothuaud, L., Benhamou, C. L., Porion, P., Lespessailles, E., Harba, R., and Levitz, P. (**2000**). "Fractal dimension of trabecular bone projection texture is related to three-dimensional microarchitecture," J. Bone Miner. Res. **15**, 691–699.

Raum, K., Cleveland, R. O., Peyrin, F., and Laugier, P. (**2006**). "Derivation of elastic stiffness from site-matched mineral density and acoustic impedance maps," Phys. Med. Biol. **51**, 747–758.

Simmons, C. A., and Hipp, J. A. (**1997**). "Method-based differences in the automated analysis of the three-dimensional morphology of trabecular bone," J. Bone Miner. Res. **12**, 942–947.

Strelitzki, R., and Evans, J. A. (**1996**). "On the measurement of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213.

Strelitzki, R., Paech, V., and Nicholson, P. H. (**1999**). "Measurement of airborne ultrasonic slow waves in calcaneal cancellous bone," Med. Eng. Phys. **21**, 215–223.

Turner, C. H., Rho, J., Takano, Y., Tsui, T. Y., and Pharr, G. M. (**1999**). "The elastic properties of trabecular and cortical bone tissues are similar: results from two microscopic measurement techniques," J. Biomech. **32**, 437–441.

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

Wear, K. A. (**2001**). "Ultrasonic attenuation in human calcaneus from 0.2 to 1.7 MHz," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 602–608.

Wear, K. A., Laib, A., Stuber, A. P., and Reynolds, J. C. (**2005**). "Comparison of measurements of phase velocity in human calcaneus to Biot theory," J. Acoust. Soc. Am. **117**, 3319–3324.

Williams, J. L. (**1992**). "Ultrasonic wave propagation in cancellous and cortical bone: Prediction of experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112.

Wu, J., and Cubberley, F. (**1997**). "Measurement of velocity and attenuation of shear waves in bovine compact bone using ultrasonic spectroscopy," Ultrasound Med. Biol. **23**, 129–134.

Characterization of high intensity focused ultrasound transducers using acoustic streaming

Prasanna Hariharan
Division of Solid and Fluid Mechanics, Center for Devices and Radiological Health, U. S. Food and Drug Administration, Silver Spring, Maryland 20993, USA and Mechanical, Industrial, and Nuclear Engineering Department, University of Cincinnati, Cincinnati, Ohio 45221, USA

Matthew R. Myers,[a] Ronald A. Robinson, and Subha H. Maruvada
Division of Solid and Fluid Mechanics, Center for Devices and Radiological Health, U. S. Food and Drug Administration, Silver Spring, Maryland 20993, USA

Jack Sliwa
St. Jude Medical, AF Division, 240 Santa Ana Court, Sunnyvale, California 94085, USA

Rupak K. Banerjee
Mechanical, Industrial and Nuclear Engineering Department and Biomedical Engineering Department, 598 Rhodes Hall, P.O. Box 210072, University of Cincinnati, Cincinnati, Ohio 45221, USA

(Received 3 August 2007; revised 20 December 2007; accepted 26 December 2007)

A new approach for characterizing high intensity focused ultrasound (HIFU) transducers is presented. The technique is based upon the acoustic streaming field generated by absorption of the HIFU beam in a liquid medium. The streaming field is quantified using digital particle image velocimetry, and a numerical algorithm is employed to compute the acoustic intensity field giving rise to the observed streaming field. The method as presented here is applicable to moderate intensity regimes, above the intensities which may be damaging to conventional hydrophones, but below the levels where nonlinear propagation effects are appreciable. Intensity fields and acoustic powers predicted using the streaming method were found to agree within 10% with measurements obtained using hydrophones and radiation force balances. Besides acoustic intensity fields, the streaming technique may be used to determine other important HIFU parameters, such as beam tilt angle or absorption of the propagation medium. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2835662]

PACS number(s): 43.80.Ev, 43.35.Yb, 43.25.Nm [CCC] Pages: 1706–1719

I. INTRODUCTION

Most therapeutic ultrasound procedures, including tumor ablation, hemostasis, and gene activation, rely on the ability of high intensity focused ultrasound (HIFU) to rapidly elevate tissue temperatures. In order to maximize the effectiveness of the procedure, as well as to avoid collateral tissue damage, it is desirable to predict the energy distribution of the HIFU beam within the propagation medium. An important first step in this process is the characterization of the ultrasound beam in a liquid medium. In this characterization, the acoustic intensity is determined throughout the spatial volume of interest, for transducer power levels of practical importance.

Currently, HIFU fields are often characterized using radiation force balance and hydrophone techniques, which measure the ultrasonic power and intensity distribution, respectively (Shaw and ter Haar, 2006; Harris 2005). Although these two techniques are well established and widely used, there are known limitations to both of these methods, including: (i) sensor damage due to heating and cavitation; (ii) inaccuracies due to strong focusing; and (iii) inaccurate frequency response due to generation of higher harmonics (Shaw and ter Haar, 2006). Consequently, these techniques can accurately characterize HIFU transducers only at low power. For clinically relevant high powers, there are no alternative measurement standards available to accurately characterize medical ultrasound fields generated by HIFU transducers (Shaw and ter Haar, 2006; Harris 2005).

Several new methods for measuring HIFU fields are being researched, including development of robust sensors and hydrophones (Wang *et al.*, 1999; Shaw, 2004; Schafer *et al.*, 2006; Zanelli and Howard, 2006; Shaw and ter Haar, 2006). An alternative approach to overcome the sensor-induced inaccuracies is to eliminate the use of sensors, and noninvasively measure the pressure field. One such commercially available noninvasive method is the schlieren imaging technique (Harland *et al.*, 2002; Theobald *et al.*, 2004), which utilizes changes in the optical index of refraction to qualitatively define the ultrasound field. However, for quantitative evaluation, the pressure field must be reconstructed tomographically. Other than schlieren imaging, there are no nonintrusive techniques reported in recent literature capable of measuring ultrasound field at high powers.

[a]Author to whom correspondence should be addressed. Electronic mail: matthew.myers@fda.hhs.gov

This paper describes a noninvasive method that is capable of measuring the acoustic intensity in a free field. The method incorporates acoustic streaming, the steady fluid movement generated when propagating acoustic waves are attenuated by viscosity of the fluid medium (Nyborg, 1965). Acoustic streaming arising from ultrasound absorption was first discussed by Eckart (1948), who derived an expression for the streaming velocity by applying the method of successive approximations to the Navier-Stokes equations. According to Eckart's theory, the acoustic streaming velocity is directly proportional to the square of acoustic pressure, and inversely proportional to the shear viscosity. However, Eckart's expression was derived ignoring the hydrodynamic nonlinearity term in the Navier-Stokes equation. Later, Lighthill (1978a, b) established that this steady streaming motion is due to the mean momentum flux (Reynolds stress) created by the viscous dissipation of acoustic energy in the fluid medium. Subsequently, Starritt *et al.* (1989, 1991), Tjotta and Tjotta (1993), and Kamakura *et al.* (1995) investigated, both experimentally and numerically, the effect of acoustic and hydrodynamic nonlinearity on the streaming velocity for a focused ultrasound source. Results obtained from their experiments and computations suggest that both acoustic and hydrodynamic nonlinearities can play a major role in generation of acoustic streaming.

Previous studies have exploited the relationship between the acoustic streaming field and the ultrasound intensity to various degrees. Nowicki *et al.* (1998) used both particle image velocimetry (PIV) and Doppler ultrasound to measure streaming motion generated by a weakly focused ultrasound beam in a solution of water and cornstarch. They found that similar streaming velocities were obtained for both the methods and the velocity magnitude was directly proportional to the acoustic power emitted by the transducer. Hartley *et al.* (1997) and Shi *et al.* (2002)used Doppler ultrasound to quantitatively measure acoustic streaming velocity in blood when exposed to an ultrasound source. They used streaming as a potential tool for improving hemorrhage diagnosis. Hartley *et al.* (1997) found that the streaming velocity increases with acoustic power and reduces with increase in viscosity during blood coagulation. Choi *et al.* (2004) used PIV to characterize streaming motion induced by a lithotripter. They observed that the streaming velocity correlated linearly with the peak negative pressure of the acoustic field measured at the focus. More recently, Madelin *et al.* (2006) used MRI to measure streaming velocity in a Glycerol water mixture. From the streaming velocity they tried to estimate the fluid properties such as attenuation and bulk viscosity. They used the expression derived by Eckart (1948) to calculate the acoustic field and time-averaged acoustic power of a plane ultrasound transducer from the streaming velocity data. While all of these experimental studies have observed a correlation between the acoustic intensity field (and acoustic power) and the acoustic streaming field, none of them attempted to use this correlation to accurately predict the acoustic intensity field.

Our characterization technique employs a predictor-corrector type method that back calculates the total ultrasonic power and acoustic intensity field from the streaming velocity field generated by the HIFU transducers. The acoustic streaming field set up by the HIFU beam with unknown energy distribution is measured experimentally using digital particle image velocimetry (DPIV) (Prasad, 2000). Then, an initial guess for the unknown acoustic intensity field is made. Based upon the Reynolds stress derived from this intensity, the streaming velocity is computed and compared with the experimental values. Using the error between the computed and observed velocity fields, an optimization routine is employed to refine the guess for the intensity field and the computation procedure is repeated. When the difference between the computed and measured velocity fields falls below a threshold value, the final estimate for the ultrasound intensity field is obtained. This inverse method can also be used to estimate physical properties of the fluid medium such as ultrasound absorption coefficient, or properties of the beam such as tilt angle.

The acoustic intensities of interest in this paper are those that are low enough that nonlinear propagation may be neglected, yet high enough to produce temperature increases in tissue of a few degrees per second. We label this range the "moderate" temperature regime (Hariharan *et al.*, 2007). An order-of-magnitude estimate of this regime is $100-1000\ \mathrm{W/cm^2}$, with precise values depending upon the amount of focusing, nonlinear parameter of the tissue, etc. Despite the assumption of linear propagation, intensities in the moderate regime may be damaging to conventional hydrophones, and hence good candidates for measurement via the streaming method. In terms of transducer powers, the values considered were in the range 5–30 W.

The streaming technique is described in Sec. II. In Sec. III, the intensity field and acoustic power calculated using this approach are compared with measurements made using standard measurement techniques such as hydrophone scans and radiation force balance (RFB). The applicability of the technique is summarized in Sec. IV.

II. MATERIALS AND METHODS

The transducer-characterization method utilizes an optimization algorithm, in which the difference between the experimental streaming velocity and the computed velocity is minimized as a parameter (or parameter set) characterizing the acoustic field is varied. The parameter of interest is often the total power, though quantities such as the transducer focal length, transducer effective diameter, or medium attenuation may also be determined. A flow chart of the algorithm is shown in Fig. 1. The following sections describe the steps of the flow chart.

A. Experimental measurement of streaming velocity

Figure 2 shows the experimental apparatus. Our experimental setup combines two different systems—(i) acoustic streaming generation system [Fig. 2(a)] and (ii) DPIV measurement system (Integrated Design Tools, Tallahassee, FL) [Fig. 2(b)].

In the streaming generation system, the HIFU transducer to be characterized is mounted vertically in a Plexiglas™ tank using a triaxis positioning system [Fig. 2(a)]. Sound

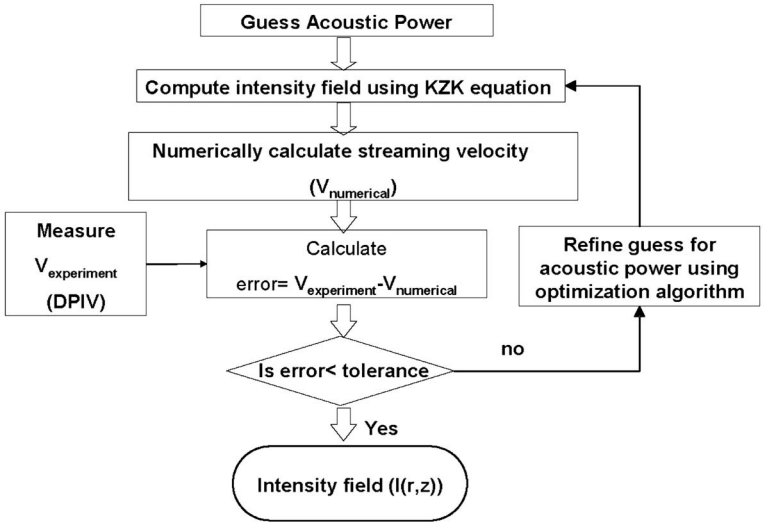


FIG. 1. (Color online) Flow chart summarizing the inverse methodology used to determine acoustic power and intensity.

sources considered in this study are single element, spherically focused, piezoceramic transducers with diameter ranging from 6.4 to 10.0 cm and focal length varying between 6.3 and 15.0 cm (Table I). The central frequency of the transducers ranges from 1.1 to 1.5 MHz. The driving signal for the transducer comes from a wave form generator (Wavetek 81, Fluke Corp., Everett, WA) in conjunction with an rf amplifier (Model 150A 100B, Amplifier Research, Souderton, PA). The output signal from the amplifier was monitored using an oscilloscope (Model 54615B, Agilent, Santa Clara, CA).

The measurement system employs a cubical Plexiglas tank ($20\times20\times20$ cm) filled with water or other appropriate fluid medium. The fluid medium was seeded with 10 μm hollow glass spheres (Dantec Dynamics Inc., Ramsey, NJ) which act as tracer particles for capturing the flow. The specific gravity of the spheres was approximately 1.1. Radiation force acting on the tracer particles was calculated using the expressions derived by King (1934) and Doinikov (1994). Using the radiation force data, the terminal velocity of hollow glass sphere particles was calculated to be negligible with respect to the measured streaming velocity (from DPIV system). This suggests that the tracer particles used in this study do not undergo motion relative to the streaming flow. In the experiments, particles were added until 5 to 10 particle pairs per interrogation area were obtained.

The DPIV system incorporates a dual pulsed 15 mJ/pulse Q-switched multimode Nd:YAG laser ($\lambda=532$ nm), with a beam diameter of 2.5 mm, repetition rate of 15 Hz, and a measured pulse width of 10 ± 1 ns, as the illumination source (New Wave Solo 1, New Wave Corp., Freemont, CA). A 100 mm focusing lens was used as laser-to-fiber coupler [Fig. 2(b)] for focusing the 2.5 mm laser on to the fiber face. The focal spot diameter at the fiber face was measured as 178 μm $\pm$ 1 μm.

TABLE I. Physical characteristics of HIFU transducers used in the experiments.

Parameters	Transducers		
	HIFU-1	HIFU-2	HIFU-3
Transducer radius	5 cm	3.8	3.2
Operating frequency	1.5 MHz	1.107 MHz	1.1 MHz
Focal distance	15 cm	11 cm	6.264 cm

Instead of solid silica-core fibers, which are not suitable for higher power delivery, this system incorporated a 700 μm core diameter hollow waveguide (HW) delivery fiber (Robinson and Ilev, 2004). This fiber was coated with cyclic olefin polymer to minimize the HW attenuation losses at the 532 nm wavelength. The output of the fiber was then directed through the microscope objective lens ($4\times$). The laser beam was then passed through the final Powell lens optics, which produced a thin laser sheet for illumination of the flow model.

Flow visualization and recording were done using an 8 bit $1K\times1K$ CCD camera (double exposure Kodak ES 1.0), which has a field of view of 2 cm$\times$2 cm. The CCD camera was connected to a computer via frame grabber to store and process the recorded images. The computer also synchronizes the camera with the laser pulse source. The time between consecutive image pairs, which is the inverse of laser pulse repetition rate, was set at 1/15 s. The pulse delay time or the time between the images was varied between 1000 and 10 000 μs depending on the streaming velocity.

Measurement of the streaming field was initiated with the positioning of the axis of the HIFU beam within the laser sheet. To perform this alignment, a hydrophone was placed in the laser sheet, which was parallel to the beam axis [z axis in Fig. 2(a)]. The transducer was then moved using a positioning system, so that the beam axis traversed a path perpendicular to the sheet [x axis in Fig. 2(a)]. When the hydrophone registered a maximum value, it was assumed that the beam axis resided in the plane of the laser sheet. As an alternative to hydrophone measurements, the transducer could be moved in the x direction until the maximum streaming velocity was observed.

After the beam axis was aligned with the laser sheet, the ultrasound was turned off and the fluid motion was allowed

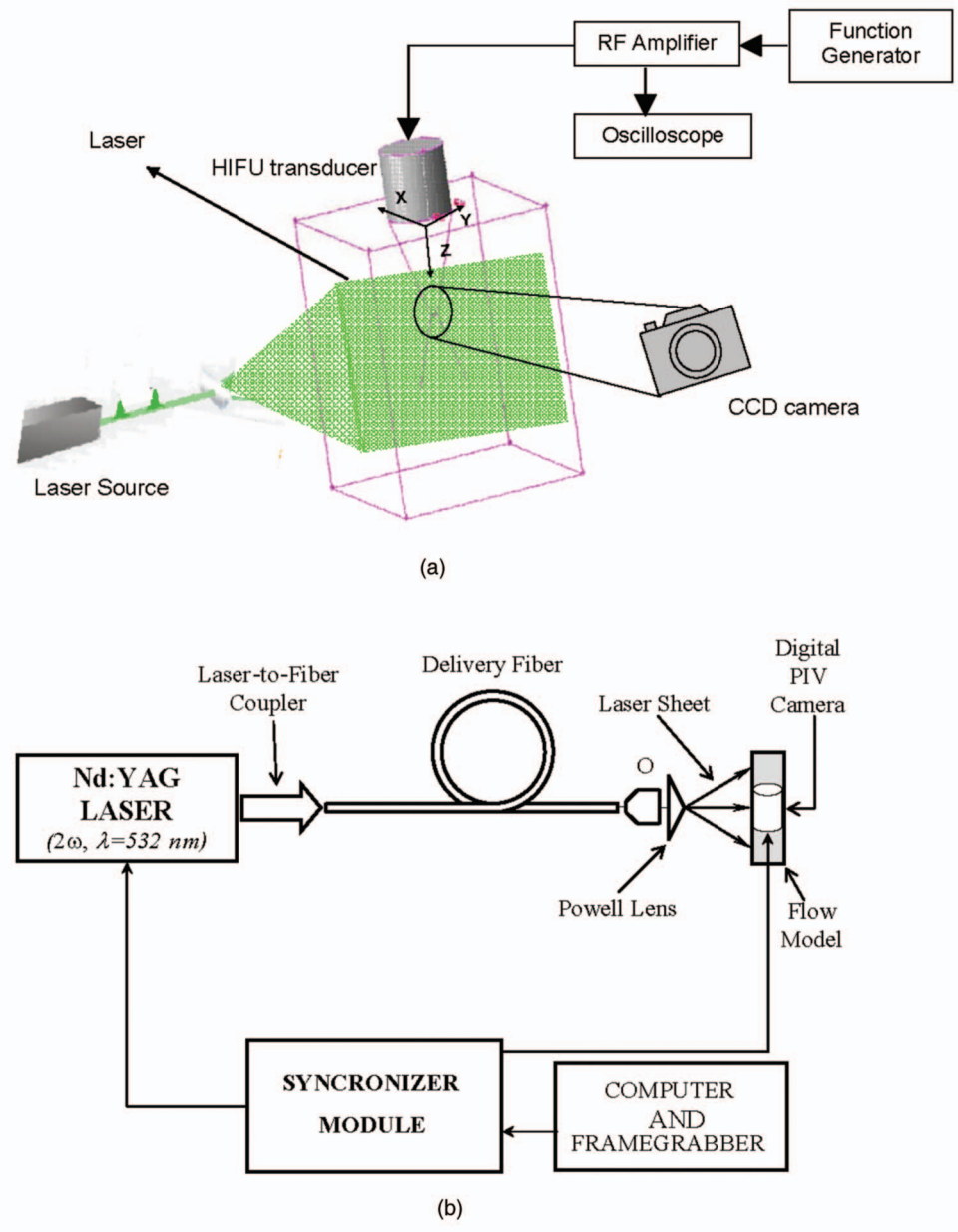


FIG. 2. (a) Schematic of the acoustic streaming generation system. (b) Block diagram of the DPIV measurement system.

to dissipate. The ultrasound was turned on and the flow field near the focus was allowed to reach steady state. Our numerical calculations predicted that flow near the focus reaches steady state in ~3 s, while outside the focus it takes longer depending on the boundary conditions. Thus, after 6.6 s the CCD camera was triggered to capture images of the flow field. Our PIV data also showed that velocity fields captured after 5, 10, and 15 s were the same.

Though a single image pair can give the instantaneous flow field, in order to avoid random errors, 50 image pairs were captured in a 5 s duration. For the 50 image pairs, the relative standard deviation in the velocity was less than 2%. All the image pairs were processed using a standard cross-correlation algorithm (IDT Provision, Tallahassee, FL) to get the streaming velocity field. Both the magnitude and direction of the velocity are provided by the algorithm. Additionally, no out-of-sheet components need to be resolved, since the flow is axisymmetric (no flow through the laser sheet) for axisymmetric transducers.

B. Numerical computation of streaming velocity

The first step in calculating the streaming velocity was to compute the momentum transferred to the test fluid by the acoustic field. Predictions of the acoustic pressure were made by solving the Khokhlov–Zabolotskaya–Kuznetsov (KZK) nonlinear parabolic wave equation. The KZK equation for an axisymmetric sound beam propagating in the z direction is (Hamilton and Morfey, 1998)

$$\frac{\partial}{\partial t'}\left[\frac{\partial p}{\partial z}+\frac{D}{2c_0^3}\frac{\partial^2 p}{\partial t'^2}+\frac{\beta}{2\rho_0 c_0^3}\frac{\partial p^2}{\partial t'}\right]=\frac{c_0}{2}\left(\frac{\partial^2 p}{\partial r^2}+\frac{1}{r}\frac{\partial p}{\partial r}\right). \quad (1)$$

Here p is the acoustic pressure, $t'=\frac{t-z}{c_0}$ is the retarded time, t is the time, c_0 is the speed of sound in the medium, $r=\sqrt{x^2+y^2}$ is the radial distance from the axis of the beam, ρ_0 is the ambient density of medium, D is the sound diffusivity of fluid medium, and β is the coefficient of nonlinearity defined by $\beta=1+B/2A$, where B/A is the nonlinearity parameter for the fluid medium. As indicated by the second tem-

poral derivative in the time-domain representation, the model employs the classic thermoviscous model of absorption, proportional to the square of the frequency in a frequency-domain representation (Hamilton and Morfey, 1998). The numerical solution is implemented using the time-domain code KZKTEXAS2 developed by Lee (1993). In the execution of the code the pressure was assumed to be constant across the transducer face, and the transducer characteristics were taken from Table I.

A steady streaming motion is generated by the absorbed acoustic energy through the spatial variation in the Reynolds stress associated with the acoustic field (Lighthill, 1978a, b). In terms of the acoustic particle velocity u_i, the Reynolds stress is given by $\overline{\rho u_i u_j}$, where the overbar denotes a time average. The force (per unit volume) associated with the Reynolds stress (Lighthill, 1978a, b) is

$$F_j = -\frac{\partial(\overline{\rho u_i u_j})}{\partial x_i}. \tag{2}$$

Here repeated indices denote summation. The streaming motion satisfies the equation of motion expressed as (Lighthill, 1978a, b),

$$\rho_0\left(\overline{u_i}\frac{\partial \overline{u_i}}{\partial x_i}\right) = F_j - \frac{\partial \bar{p}}{\partial x_j} + \mu\nabla^2\overline{u_j} \tag{3}$$

along with the conservation of mass equation

$$\rho\frac{\partial \overline{u_i}}{\partial x_i} = 0 \tag{4}$$

for an incompressible fluid.

Of primary interest in transducer characterization is the acoustic intensity in the focal region. In the focal region, the acoustic field may be modeled as a beam of locally planar waves traveling in the z direction. For this quasiplanar field, the force associated with the Reynolds stress [Eq. (2)] is in the axial direction and has the form (Nyborg, 1965)

$$F_z = \frac{2\alpha}{(\rho_0 c_0)^2}\overline{p^2} = \frac{2\alpha}{\rho_0 c_0} I, \tag{5a}$$

where α is the absorption coefficient of the medium, and

$$I = \frac{\overline{p^2}}{(\rho_0 c_0)} \tag{5b}$$

is the time averaged acoustic intensity which is calculated by solving Eq. (1). In principle, a more complex intensity relation than Eq. (5b) that accounts for the motion of the medium must be used (Morse and Ingard, 1968) in intensity calculations for a streaming medium. However, the error incurred in using Eq. (5b) is on the order of the Mach number v_0/c_0, v_0 being a measure of the streaming speed. The Mach number for the streaming jets reported on in this paper is on the order of 0.0001 (Figs. 5 and 8), and hence the use Eq. (5b) for a stationary medium expression is justified.

For an axisymmetric (e.g. spherically concave) transducer in an infinite medium, the governing equations for the streaming field [Eqs. (3) and (4)] may be rewritten as

$$\frac{\partial \overline{u_z}}{\partial z} + \frac{1}{r}\frac{\partial(r\overline{u_r})}{\partial r} = 0, \tag{6}$$

$$\overline{u_r}\frac{\partial \overline{u_z}}{\partial r} + \overline{u_z}\frac{\partial \overline{u_z}}{\partial z} = \frac{F_z}{\rho_0} - \frac{1}{\rho_0}\frac{\partial \bar{P}}{\partial z} + \nu\left[\frac{\partial^2 \overline{u_z}}{\partial z^2} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \overline{u_z}}{\partial r}\right)\right], \tag{7}$$

$$\overline{u_r}\frac{\partial \overline{u_r}}{\partial r} + \overline{u_z}\frac{\partial \overline{u_r}}{\partial z} = -\frac{1}{\rho_0}\frac{\partial \bar{P}}{\partial r} + \nu\left[\frac{\partial^2 \overline{u_r}}{\partial z^2} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \overline{u_r}}{\partial r}\right)\right]. \tag{8}$$

Equations (6)–(8) were solved using the Galerkin finite element method (Fluent Inc., 2002), in a geometry that simulated the tank of Fig. 2(a).

C. Iterative characterization

To illustrate the iterative approach, we take the unknown parameter of interest to be the acoustic power. For the first iteration, a guess is made for the power, enabling the KZK equation to be solved for the first iterate of the acoustic field. From the KZK equation, the axial component of the driving force F_z is calculated from Eq. (5). Equations (6)–(8) are then solved to obtain the first iterate of the velocity field.

Streaming velocity fields obtained from both experiment and computation are then input to a Nelder Mead multidimension optimization algorithm (Lagarias *et al.*, 1995; Mathworks, 2002). The objective function in this algorithm is

$$\text{Error}_{\text{rms}} = \sqrt{\sum_{i=1}^{n}(u_{i,\text{exp}} - u_{i,\text{num}})^2}, \tag{9}$$

where n is the number of velocity nodes in the camera's field of view. $\text{Error}_{\text{rms}}$ measures the deviation of numerical velocity profile from the experimental values. The optimization routine then attempts to reduce the error by adjusting the power estimate and recalculating the intensity field. The entire procedure is repeated until the rms error is minimized. (Further iterations produce no reduction in error.) Provided the numerical velocity field approximates the experimental one reasonably well, the acoustic power is taken to be the last iterate of the power, and the intensity field given by the most recent output of the KZK equation. The total number of iterations required for convergence of the optimization algorithm is typically 30–40, and the duration for the entire back-calculation procedure is of the order of 3 h. Computations were conducted on a 3 GHz processor with 2.0 Gbytes RAM.

D. Validation experiments

The total acoustic power and intensity field estimated from the inverse approach were compared with the measurements from radiation force balance (RFB) and hydrophone-scanning techniques. The RFB used in this study consists of a highly absorbing target suspended in a water bath (Maruvada *et al.*, 2007). The transducer to be calibrated was mounted directly above the target. When the ultrasound is turned on, the target experiences a net force arising due to the momentum associated with the ultrasound wave. The target was connected to a sensitive balance which measures the

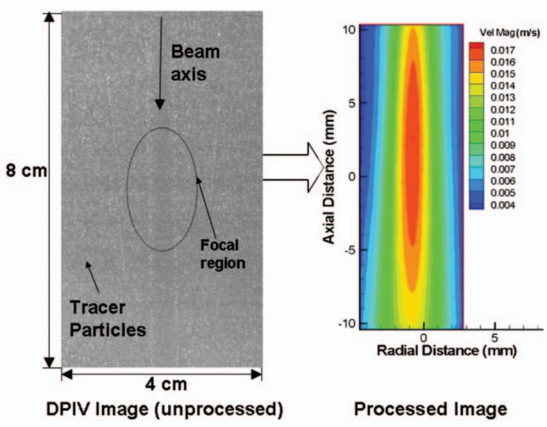


FIG. 3. Unprocessed DPIV image and streaming velocity contour, for HIFU-2 in water. Acoustic power: 3.6 W. Streaming velocity contour is obtained from the average of 50 image pairs captured in a 5 s duration. The standard deviation for 50 images at the point of maximum velocity is ~1.93%.

force acting on it. From the measured force, acoustic power was obtained by multiplying it with the sound speed, and the factor $1+(a/2d)^2$ (a=transducer radius, d=focal length) to account for focusing effects (IEC 2005). Powers determined from the RFB were estimated to have an uncertainty of less than 10%.

Pressure measurements were made using 0.6 mm piezoelectric ceramic hydrophones (Dapco Industries, Ridgefield, CT). The hydrophones were calibrated using a planar scanning technique (Herman and Harris, 1982). The uncertainty in the pressure measurements acquired using these hydrophones was approximately 20% (Herman and Harris, 1982). The transducer to be characterized was mounted horizontally in a Plexiglas tank surrounded with sound absorbing materials, and the hydrophone was moved within the HIFU beam. The movement of the hydrophone in all three coordinates was monitored and controlled using computer controlled stepper motors. A scanning step size of 0.4 and 1 mm was used along radial and axial directions, respectively. At each step, measurements of the peak positive pressure were made. Since the hydrophone scanning method is accurate only at low power levels, this validation was done at acoustic powers less than 5 W.

III. RESULTS

In Sec. III A, computed streaming velocities are compared with previously published results. Computed and measured streaming fields are also compared using a known sound source. Additionally, the stability of the streaming jet is analyzed. In Sec. III B our inverse algorithm is used to determine acoustic power and intensity in water, at low power. The inverse method is then used to determine the absorption of a more viscous medium, at moderate power. Characterization of a transducer of unknown power is then performed in the higher-viscosity medium.

A. Forward problem: Experimental and numerical streaming velocity fields

Figure 3 shows a single raw photographic image obtained from DPIV measurements and the corresponding streaming velocity contour obtained after postprocessing using a standard cross-correlation algorithm (IDT Provision, Tallahassee, FL). This velocity contour is the average of 50 image pairs captured in a 5 s duration. The standard deviation in peak velocity magnitude for 50 images, divided by the mean velocity of the 50 images, was 1.93%. This flow field was obtained in degassed water using HIFU-2 as the sound source for an input acoustic power of 3.6 W, as measured using a radiation force balance (Maruvada *et al.*, 2007). The contour origin, which is the center of field of view of the camera, does not coincide with ultrasound focus. The velocity field at focus has the cigar shape characteristic of focused ultrasound beams.

1. Validation of numerical model

Our computational model for calculating streaming velocity was validated with the data of Kamakura *et al.* (1995). Figure 4 shows the transient streaming velocity generated by a Gaussian shaded transducer with diameter 1 cm, focal

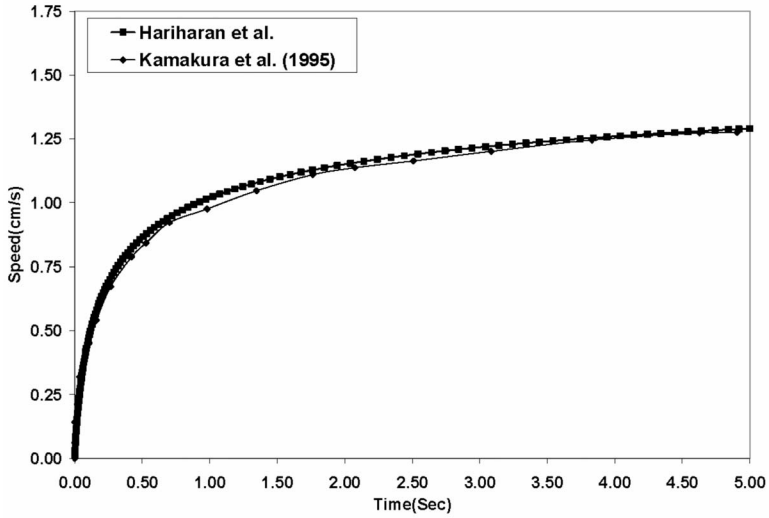


FIG. 4. (Color online) Validation of computational model with streaming-velocity data from Fig. 2 of Kamakura *et al.* (1995). Gaussian shaded transducer with diameter 1 cm, focal length 5 cm, and central frequency 5 MHz, driven at a maximum source pressure of 30 kPa.

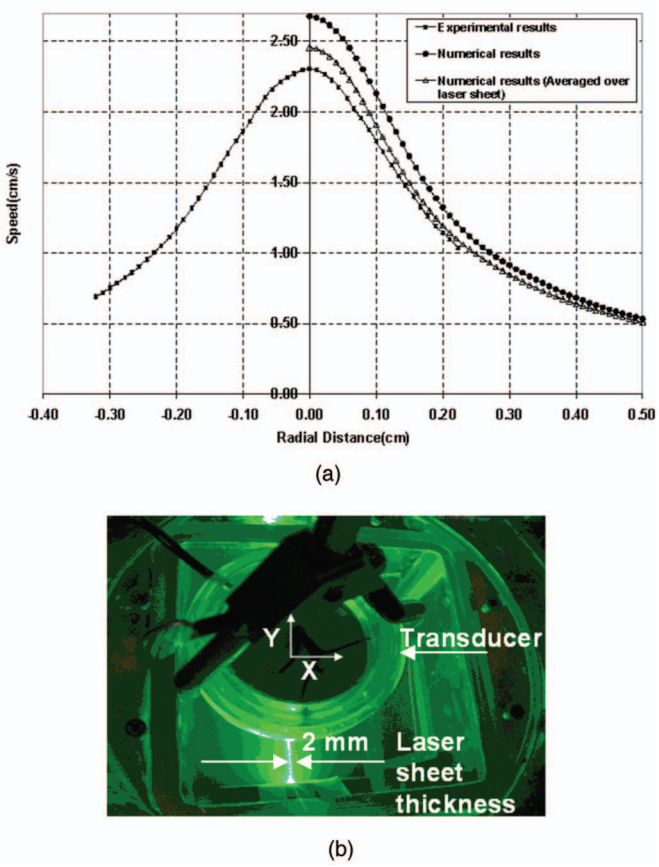


FIG. 5. (a) Radial plot of streaming velocity in water. Transducer is HIFU-2; power=5.4 W. (b) Photograph of the experimental apparatus showing the thickness of the laser sheet.

length 5 cm, and central frequency 5 MHz, driven at a maximum source pressure of 30 kPa. The transient speed calculated using the present model matches with the result published by Kamakura *et al.* (Fig. 2 of Kamakura *et al.*, 1995) to within about 1%.

2. Comparison of experimental and numerical streaming fields

Figure 5(a) compares the radial speed profiles obtained experimentally using DPIV, and computationally by solving Eqs. (6)–(8) [highest and lowest curves in Fig. 5(a), middle curve containing "averaged data" will be discussed subsequently]. The sound source used for this validation experiment is the HIFU-2 transducer (Table I), driven at a known acoustic power of 5.4 W, as measured using the RFB. The peak velocity profile calculated from the computational model is ~15% higher than the peak velocity measured using the DPIV system. A similar trend is observed for other transducers and other power levels.

The reason for this offset involves the finite thickness of the laser sheet. As mentioned earlier, the thickness of the laser sheet [Fig. 2(a)] when it exits the Powell lens is around 0.5–1 mm. However, as the laser passes through the flow medium, the sheet diverges gradually and becomes ~2 mm thick when it reaches the ultrasound beam axis [Fig. 5(b)]. Therefore, instead of visualizing a single plane (2×2 cm), the CCD camera captures particle motion in a three-dimensional space of size 2×2×0.2 cm. As a result, the streaming velocity measured at a particular point in the *YZ* plane [Fig. 2(a)] is essentially the spatial average of the velocity occurring along the thickness of 0.2 cm of the sheet in the *x* direction.

In order to account for this measurement error in our numerical algorithm, the velocity field obtained numerically was averaged over the 0.2 cm thickness of the laser sheet in the *x* direction [Fig. 5(b)]:

$$\bar{u}_{\text{averaged}} = \frac{1}{L}\int_0^L \bar{u}(r,z)dx, \quad r = \sqrt{x^2+y^2}, \tag{10}$$

where $\bar{u}$ is the velocity estimated from the computations and L is the half thickness of the laser sheet (~1 mm). This averaged velocity, $\bar{u}_{\text{averaged}}$, was then used in the optimization routine to compute $\text{Error}_{\text{rms}}$ in Eq. (9). Figure 5 (middle curve) shows the velocity profile calculated from our computations after doing the spatial averaging. The averaged velocity profile more closely matches the experimental data; the difference is approximately 6%.

3. Higher-viscosity streaming medium

DPIV experiments performed in water at higher acoustic powers (5–30 W) revealed the onset of instability in the streaming jet. This instability was manifested in the form of small puffs occurring at irregular times, just below the focus. This behavior was observed for all three HIFU transducers. As a result of the behavior, unrepeatable velocity fields were obtained.

To overcome the stability problem, the viscosity of the medium was increased (Reynolds number decreased). An alternative fluid medium composed of degassed water mixed with Natrosol-L (Hydroxy ethyl cellulose, Hercules, Wilmington, DE) in various concentrations was developed, and used in experiments where the acoustic power exceeded 5 W. The mixture was degassed using a vacuum pump prior to sonication. The viscosities of 1.4% (1.4 g/100 ml water), 2.4% and 3.4% Natrosol-L solutions were measured using a Cannon type-E viscometer to be ~5, ~12, and ~25 cp, respectively. Speed of sound for the 2.4% Natrosol solution was measured using an ultrasonic time delayed spectrometry system (Harris *et al.*, 2004) to be 1480 m/s (same as for degassed water). Finally, as discussed in Sec. III B, the attenuation coefficient of the 2.4% Natrosol solution was found using our streaming method to be roughly twice that of water.

Figure 6 shows radial speed profiles obtained using water as well as 1.4%, 2.4%, and 3.4% Natrosol solutions as the flow medium. The transducer used in this comparison study is HIFU-1 excited at an output power of 30 W. Broader, smoother, and more axisymmetric profiles can be seen with increasing Natrosol content. As the fluid viscosity increases from 1 (water) to 25 cp, the peak velocity reduces from 8.2 to 5.8 cm/s.

Figure 7 shows the fluid speed obtained in a 2.4% Natrosol solution for three different transducers driven at same voltage (42 V_{rms}). The plot shows that the velocity distribution at the focal region depends upon the focusing characteristics of the transducer. Focusing gain, 3, and 6 dB beam

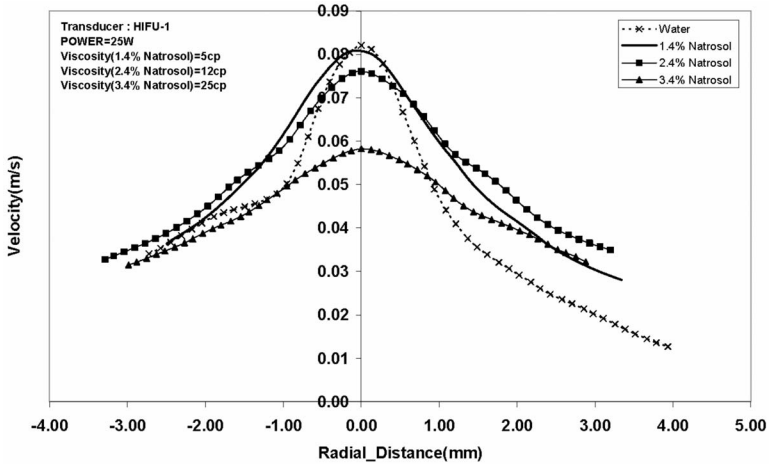


FIG. 6. (Color online) Velocity magnitude vs radial distance for (i) water, (ii) 1.4% Natrosol solution, (iii) 2.4% Natrosol solution, and (iv) 3.4% Natrosol solution.

dimensions (measured from hydrophone scans) of the transducers are listed in Table II. The focusing gain (Lee, 1993), given by the expression $\frac{\omega a^2}{2c_0 d}$ (ω being the central frequency of the transducer, a the transducer radius, c_0 the speed of sound in the medium, and d the focal length), is the highest for HIFU-1. The corresponding peak streaming velocity is maximum for this transducer. The velocity contours for HIFU-1 are relatively narrow, in accordance with the small 6 dB beam width (Table II). In contrast, the velocity contours for HIFU-3 are short in axial direction, consistent with the smaller 6 dB width in the axial direction (Table II). From the 6 dB dimensions shown in Table II, it can be seen that focal region of HIFU-2 is larger than HIFU-1 and HIFU-3. Consequently, velocity contours obtained using this transducer are elongated.

Figure 8 compares the streaming velocity profile obtained numerically and experimentally while driving HIFU-3 at an input voltage of 35 V_{rms}, in a 12 cp medium. The corresponding acoustic power measured using the RFB is 19 W. It can be seen that experimental and numerical velocity profiles match closely ($\sim$1%) in both the axial and radial directions.

The width of the velocity distribution in Fig. 8(b), based upon the distances at which the speed has dropped to half its maximum value (kinetic energy reduced to $\frac{1}{4}$ the maximum), is around 0.5 cm. By contrast, the intensity width for the same transducer (HIFU-3) is about 0.2 cm (Table II). In general, the width of the velocity distribution is larger than that of the intensity distribution, and the effect becomes more pronounced as the viscosity of the medium increases. In Fig. 6, for example, the width of the velocity profile for HIFU-1 in the 2.4% Natrosol solution (viscosity 12 cp) is about 5.4 mm. The width of the velocity profile in water (viscosity 1 cp) is between 2.4 and 3.3 mm, depending upon how the asymmetry is treated. Within the limits of linear acoustics, the width of the intensity distribution is independent of viscosity, with the value of about 1.7 mm for HIFU-1.

B. Inverse problem: Total acoustic power and HIFU beam characteristics

In the following, the optimization algorithm is used to backcalculate various parameters characterizing the transducer and the propagation medium.

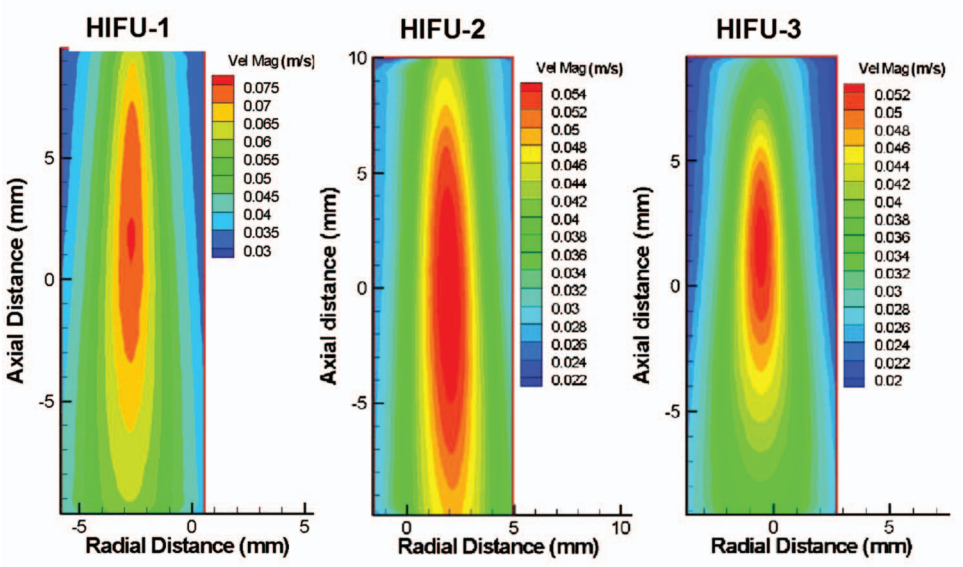


FIG. 7. Speed contours obtained in 2.4% Natrosol solution for (i) HIFU-1, (ii) HIFU-2, and (iii) HIFU-3 transducers. Transducer input voltage=42 V_{rms}.

TABLE II. Focusing characteristics of HIFU transducers used in the experiments.

Transducer	6 dB dimension (axial × radial)	Focusing gain
HIFU-1	2.4 cm × 1.7 mm	53
HIFU-2	2.95 cm × 3 mm	31
HIFU-3	1.35 cm × 2 mm	38

1. Characterizing acoustic field at low power levels (<5 W)

Since current field hydrophone mapping techniques are practical only at low power levels, characterization was first performed at low input voltages with water as the fluid medium. Figure 9 compares the source pressure ($p_0 = \sqrt{W \times 2\rho_0 c_0 / \pi a^2}$, p_0 being the peak source pressure in N/m^2, W being the acoustic power in watts, and a the transducer radius in m) obtained using RFB and streaming back-calculation methods. Using the procedure outlined in Fig. 1 and discussed in Sec. II, acoustic power was calculated from the streaming velocity for driving voltages of 15 and 18 V_{rms}. The source pressure of HIFU-2 was 0.048 and 0.058 MPa for driving voltages of 15 and 18 V_{rms}, respectively. The inverse method predicted the source pressure to be 0.046 and 0.056 MPa, which is within 5% of the values obtained from the RFB measurements. When the averaging prescribed by Eq. (10) is not done, this deviation was ~10%.

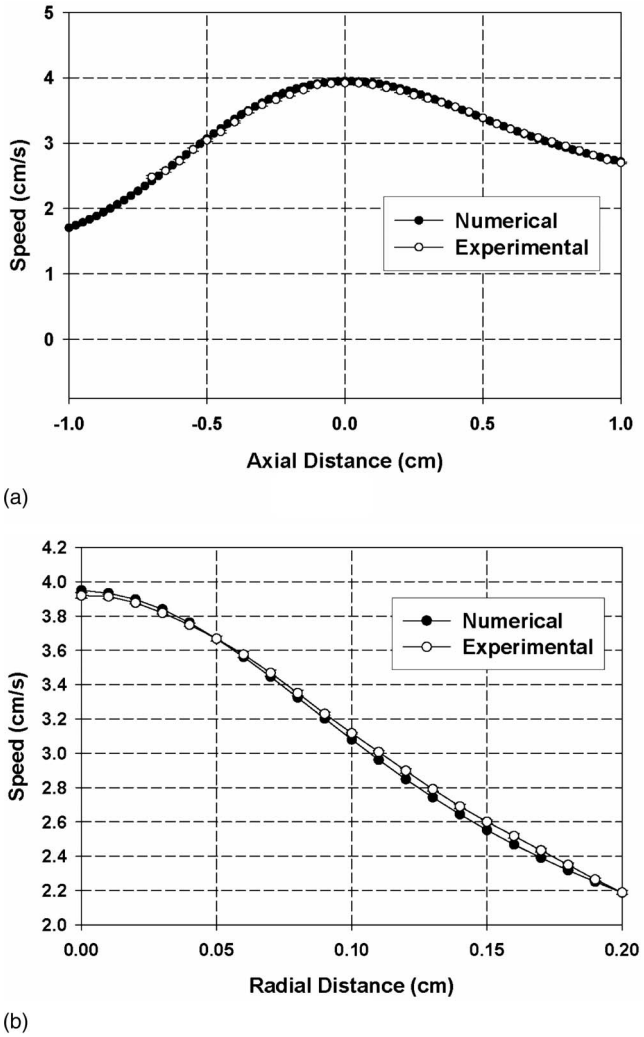


FIG. 8. (Color online) Velocity profiles obtained from computations and experiments in 2.4% Natrosol solution; (a) velocity profile along axial direction and (b) velocity profile along radial direction. Transducer is HIFU-3; input voltage=35 V_{rms} (19 W). For the experimental data, standard deviation error bars for five trials are also shown. The maximum standard deviation for five trials is 0.02 cm/s.

Figure 10 compares the axial and radial intensity fields of HIFU-2, obtained via hydrophone scanning and the inverse method for a transducer voltage of 10 V_{rms}. Peak intensity at the focus measured by the inverse method is 65 W/cm^2, approximately 4% below the peak intensity measured by hydrophone (67.8 W/cm^2). The 6 dB length and width, backcalculated from the streaming velocities, are 2.7 cm × 2.7 mm; the corresponding values obtained from hydrophone scanning are 2.95 cm × 3 mm.

2. Characterizing acoustic field at moderate power levels (<30 W)

As noted earlier, a Natrosol-based solution is the preferred medium for characterization when the acoustic power exceeds about 5 W. However, unlike water, little is known regarding the absorption properties of Natrosol solutions. Equation (5) shows that the Reynolds stress, which accounts for momentum transfer from the sound field to fluid, is proportional to the absorption coefficient of the medium. Using a transducer of known intensity, this parameter can be treated as the unknown in the optimization algorithm, in the following manner.

Figure 11(a) summarizes the procedure for calculating the acoustic absorption coefficient of the medium. Using DPIV, the streaming velocity generated by the HIFU-2 transducer in the 2.4% Natrosol-L solution was measured. First, an initial guess for the unknown α was made. In these determinations of absorption coefficient using the inverse algorithm, the attenuation coefficient of water (2.52×10^{-4} Np cm^{-1} MHz^{-2}) was taken as the first guess for α. Then, the forcing term was calculated using Eq. (5) and the streaming velocity was calculated numerically by solving Eqs. (6)–(8). The $Error_{rms}$ value was then computed from the experiment and numerical velocity profiles and input into the optimization routine. The procedure was repeated until the relative error was reduced to a value of 10^{-3}.

The absorption coefficient of 2.4% Natrosol measured using this procedure is shown in Fig. 11(b). The α values measured at the power levels 3.6, 5.8, and 11.8 W were fairly constant, approximately 5.65×10^{-4} Np/cm for the 1.1 MHz frequency considered. This value of α is used in the subsequent calculations to characterize transducer output at moderate and high power/intensity levels.

Figure 12 compares the power measurements made using the RFB and streaming in the 2.4% Natrosol medium. Transducers HIFU-1 and HIFU-3 were driven at voltages ranging from 19 V_{rms} to 42 V_{rms}. The inverse method was able to measure the acoustic power with a maximum error of about 10%. Computed intensity profiles in the Natrosol medium are shown in Fig. 13, for a power of 30 W. The 6 dB length and width, backcalculated from the streaming veloci-

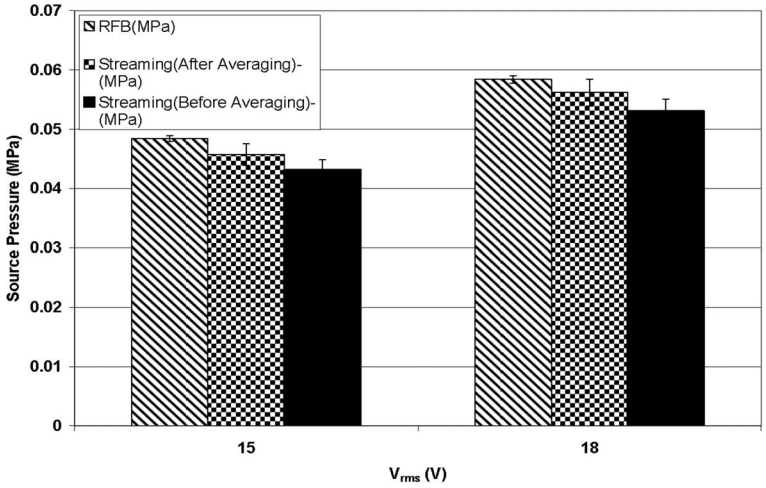


FIG. 9. Comparison of source pressure for HIFU-2 in water, obtained via the streaming method and by radiation force balance. Results based upon the streaming technique are shown with and without averaging over the 2 mm laser sheet thickness. Standard deviation error bars for three trials are also shown. For RFB measurements, the maximum standard deviation for three trials is less than 1% of the mean value. For the streaming measurements, the maximum standard deviation for three trials is 3% of the mean.

ties, are 1.2 cm × 1.85 mm. The values determined using hydrophone scanning at low power, both in water and Natrosol, are 1.35 cm × 2 mm (Table II).

IV. DISCUSSION

The streaming technique as presented in this paper is useful for characterizing HIFU transducers in the "moderate" intensity regime, roughly 100 to 1000 W/cm^2. The intensities are low enough that linear acoustic propagation is a reasonable assumption, yet high enough to possibly damage conventional hydrophones.

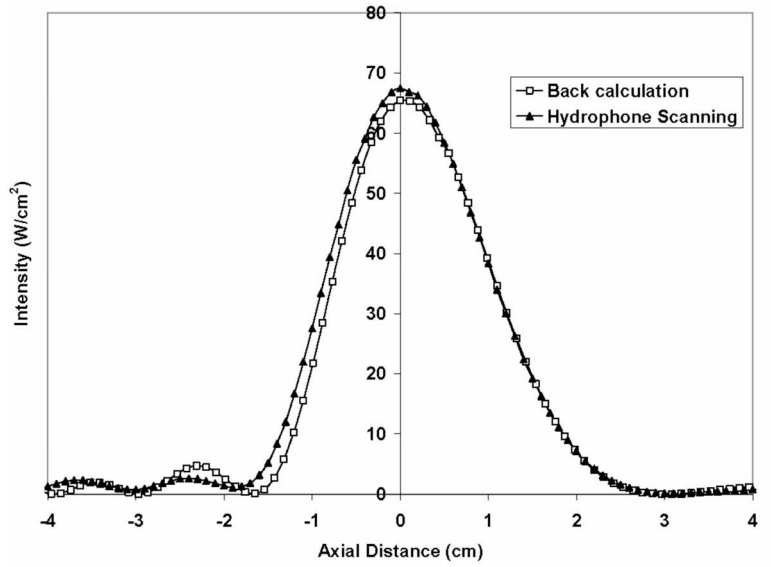
At low powers—around 3 W or 70 W/cm^2 focal intensity, close agreement with piezoceramic hydrophones (Fig. 10) was obtained. At focal intensities above about 1000 W/cm^2 (transducer powers around 20 W), divergence between the measured and predicted power levels can be seen (Fig. 12), probably due to nonlinear propagation effects. While the KZK equation can predict the generation of additional frequency components by nonlinear mechanisms, the absorption model used in our iterative method does not account for the frequency dependence of the absorption. A more comprehensive model of the absorbed acoustic energy would involve a sum of the form $\Sigma\alpha_n I_n$, where I_n is the intensity of the nth acoustic harmonic and α_n the corresponding absorption. The iterative scheme would then involve optimization over an n-dimensional space containing the intensity modes. These nonlinear-propagation effects will be treated in a future generation of the model.

The inverse method was first used to determine unknown transducer power. Because we employed the linearized version of the KZK equation, the intensity field is linearly proportional to the transducer power, and the KZK equation actually needed to be solved only once. Subsequent intensity fields required in the iteration process (iteration is still required due to the more complex dependence of velocity upon intensity in the Navier-Stokes equations) could be obtained by simply multiplying the first intensity field by a scaling factor dependent upon the current value of the acoustic power. This would not be true for other transducer-characterization parameters, such as focal length, beam tilt angle relative to the desired axis, or the absorption of the medium.

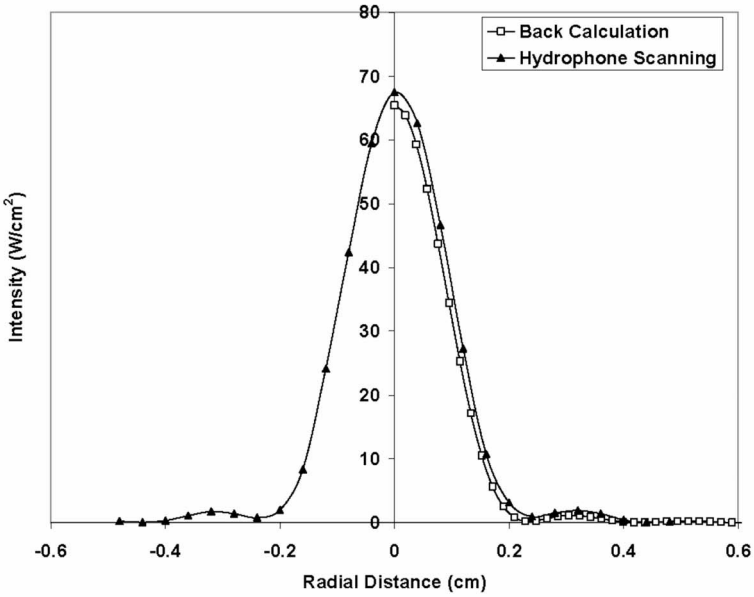
Determining an unknown absorption (Fig. 11) is a characterization problem for which the streaming technique is well suited. Conventional absorption-measurement techniques such as time delay spectroscopy (Peters and Petit, 2003) measure attenuation over a prescribed propagation distance. When the attenuation is small over the given distance, these techniques are susceptible to large error. With the streaming-based technique, however, even slight attenuations such as that of water can yield readily measurable fluid velocities. Mathematically, the forcing term in the momentum equation [Eq. (5)] contains a factor of α, producing relative changes in the fluid velocity field that are commensurate with relative changes in the absorption coefficient. This commensurate change in velocity with variation in α also allows for stable operation of the optimization algorithm; nonuniqueness issues arise when the streaming field is insensitive to changes in the parameters being determined.

The optimization technique may be used to estimate multiple parameters simultaneously, though the computation time increases significantly, and convergence can be more difficult to achieve. For example, determining the transducer intensity simultaneously with the absorption coefficient would be difficult, since the two quantities α and I_0 appear primarily as a product [Eq. (5)]. The absorption coefficient also appears in the exponential attenuating factor of the intensity but, as is the case with conventional absorption measurement techniques, this exponential factor is relatively insensitive to changes in α.

The streaming algorithm in this paper assumes an axisymmetric ultrasound beam. The technique can be extended to three-dimensional (x,y,z), nonsymmetric beams with no change in the procedures performed (Fig. 1) in the algorithm. A three-dimensional propagation code and fluid-flow simulator would be necessary, and an increase in computation time would result. The three-dimensional version of the technique would be useful in diagnosing imperfections or problems in transducer fabrication, such as a misfiring array element or a propagation axis that was skewed relative to the axis of sym-



(a)



(b)

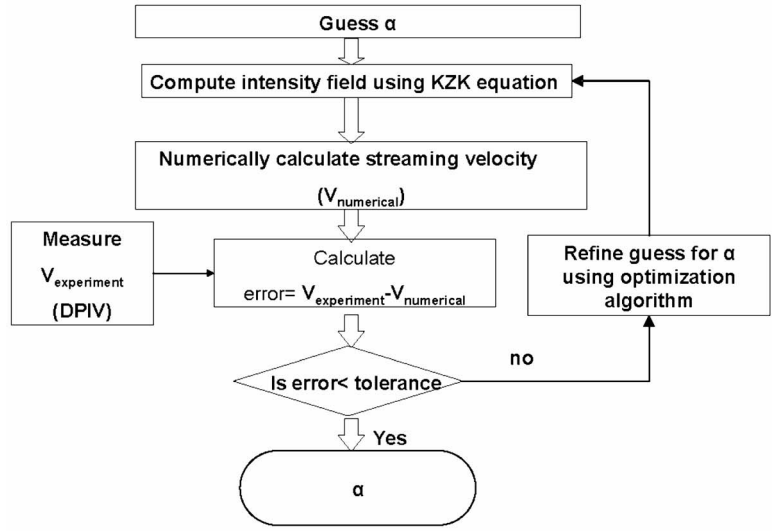
FIG. 10. (Color online) Acoustic intensity profiles as a function of (a) axial and (b) radial distances (cm), obtained from backcalculation and hydrophone scanning in water for HIFU-2 transducer. Transducer input voltage=10 V_{rms} (3.2 W).

metry of the transducer face. The problem would first be manifested in an asymmetry of the experimental streaming velocity field. The three-dimensional inverse algorithm could be used to help predict the transducer characteristics giving rise to the asymmetric velocity field that was observed. For example, the location of a nonfunctioning array element could be left as a free parameter in the optimization procedure, to be determined as part of the iterative solution.

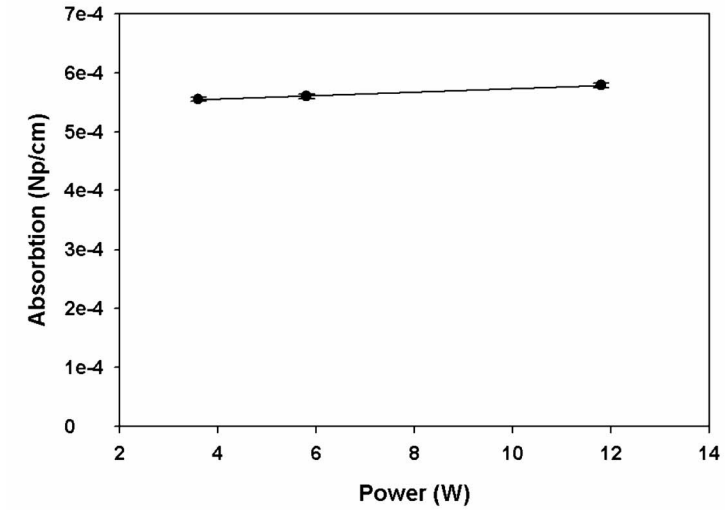
Figure 6 illustrates the complex dependence of velocity upon fluid viscosity in a streaming medium. An increase in viscosity increases acoustic absorption, i.e., produces a larger transfer of momentum from the acoustic field to the hydrodynamic field, resulting in a higher peak (on-axis) velocity. However, the higher viscosity also serves to radially diffuse momentum, resulting in a lower peak velocity. Thus, the peak velocity in Fig. 6 is essentially unchanged as the viscosity is increased by a factor of 5 from 1 to 5 cp, but drops about 25% when the viscosity is doubled from 12 to 25 cp.

The higher viscosity test fluids are essential at higher streaming velocities (higher intensities) due to stability requirements for the streaming jet. The higher viscosity test fluids also yield more symmetric velocity profiles, as can be seen by comparing the natrosol-based curves in Fig. 6 with that for water. An upper limit on the viscosity of the test fluid is imposed by the low resulting velocities resulting at high viscosities. These velocities can become difficult to quantify accurately using DPIV systems. For the intensity range considered in our studies, the agreement between experimental and computational velocity profiles was best for the 12 cp fluid (Fig. 8).

The KZK equation was used in our simulations, however, any beam propagation code can be used in the iterative algorithm outlined in Fig. 1. As noted by Tjotta and Tjotta (1993), the accuracy of the KZK degrades when the beam becomes highly focused (large aperature transducers). Tjotta and Tjotta (1993) noted that the ratio a/d of transducer ra-



(a)



(b)

FIG. 11. (Color online) (a) Flow chart of the inverse problem to estimate acoustic absorption coefficient, α (Np/cm), of the 2.4% Natrosol fluid medium. (b) Absorption coefficient of 2.4% Natrosol fluid obtained using the inverse method, for different powers using HIFU-2 transducer. Standard deviation error bars for three trials are also shown. The maximum standard deviation for three trials is less than 1% of the mean.

dius to focal length should be small in order to ensure accuracy of the KZK approach. For our transducers HIFU-1, HIFU-2, and HIFU-3, the a/d values were 0.33, 0.35, and 0.51, respectively. While the a/d value for HIFU-3 was relatively large, the agreement with experimental measurements of streaming velocity (Fig. 8) and total power (Fig. 12) indi-

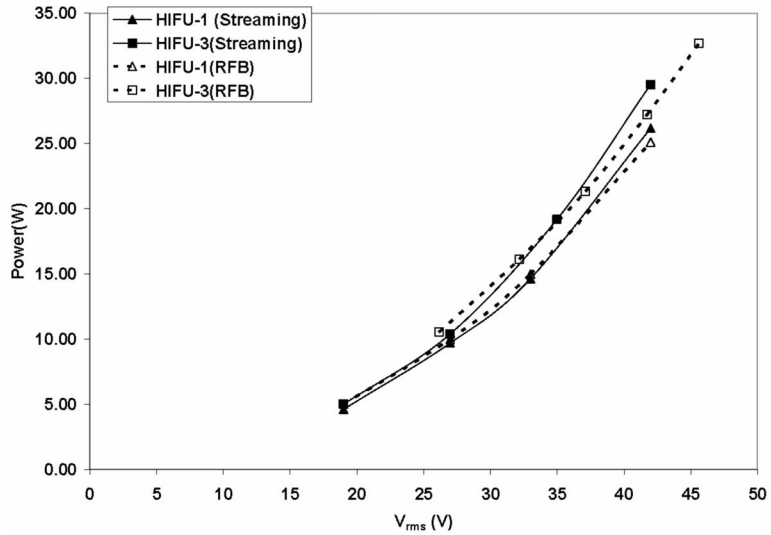


FIG. 12. (Color online) Comparison of acoustic power obtained from back-calculation and RFB in 2.4% Natrosol medium, for transducers HIFU-1 and HIFU-3.

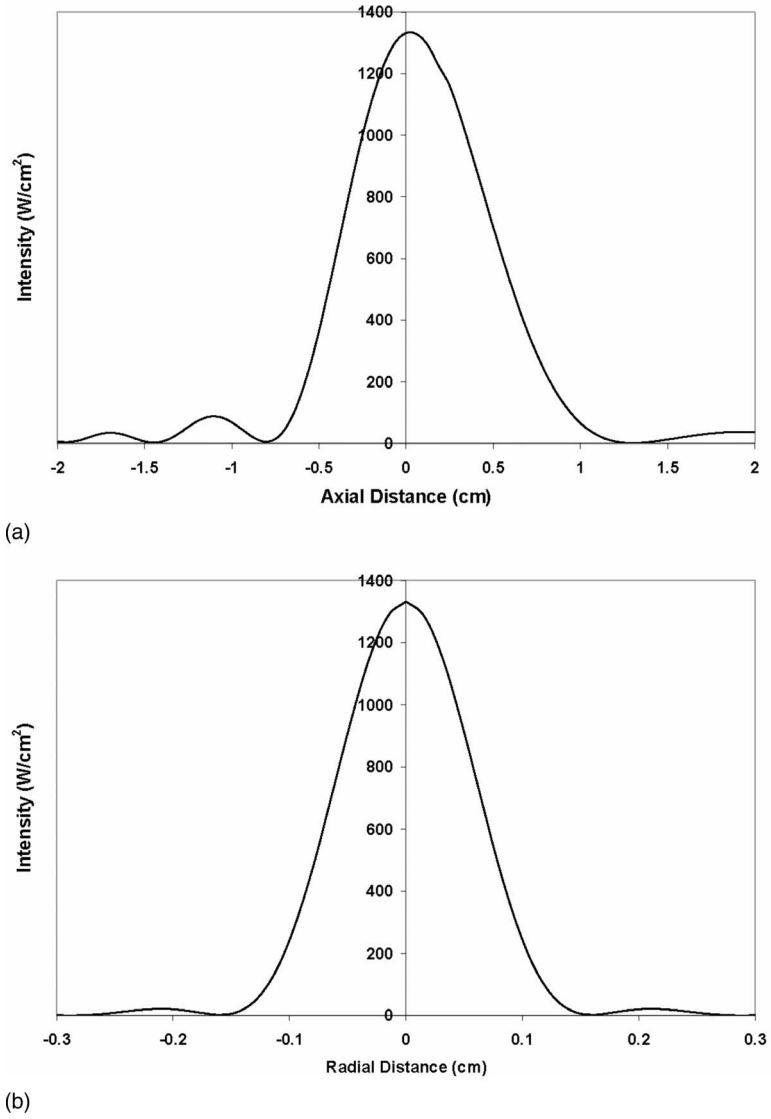


FIG. 13. (Color online) Acoustic intensity as a function of (a) axial and (b) radial distances obtained using HIFU-3 transducer in 2.4% Natrosol solution. Acoustic power=30 W.

cates that the criterion of Tjotta may be conservative. In any event, when high accuracy is required, a full-wave model such as the Helmholtz equation in the linear propagation regime or the Westervelt equation in the nonlinear regime should be employed in the iterative algorithm. The tradeoff is the increase in computation time associated with using a full-wave model compared to a parabolic equation.

A limitation of our inverse methodology is that the repetitive solution of the acoustic propagation equation and fluid-flow equations can involve significant computation time (~3 h). Our technique is also limited by the accuracy of the sound propagation code. For example, the KZK code currently employed by our model degrades in accuracy when the transducer gain becomes large. The above-mentioned drawbacks can be circumvented by developing an alternate method which determines the acoustic field directly from the experimental streaming velocity. In subsequent studies, an attempt will be made to estimate the acoustic field directly from the experimental streaming velocity data.

V. CONCLUSION

The streaming technique presented in this paper allows free-field characterization of HIFU transducers to be performed in an intensity range that may be hazardous to conventional hydrophones. The method is noninvasive and agrees well with hydrophone and radiation-force balance techniques in common ranges of applicability. In addition to predicting acoustic intensity fields, the approach may also be used to determine other important HIFU parameters, such as medium absorption, transducer focal length, or beam tilt angle.

ACKNOWLEDGMENTS

The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the U. S. Department of Health and Human

Services. Financial support from the FDA Center for Device and Radiological Health and the National Science Foundation (Grant No. 0552036) is gratefully acknowledged.

Choi, M. J., Doh, D. H., Kang, K. S., Paeng, D. G., Ko, N. H., Kim, K. S., Rim, G. H., and Coleman, A. J. (**2004**). "Visualization of acoustic streaming produced by lithotripsy field using a PIV method," J. Phys.: Conf. Ser. **1**, 217–223.

Doinikov, A. A. (**1994**). "Acoustic radiation pressure on a rigid sphere in a viscous fluid," Proc. R. Soc. London, Ser. A **447**, 447–466.

Eckart, C. (**1948**). "Vortices and streams caused by sound waves," Phys. Rev. **73**, 68–76.

Fluent Inc. (**2002**). *FIDAP User Manual ver. 8.6.2* (Fluent, Inc., Lebanon, NH).

Hamilton, M. F., and Morfey, C. L. (**1998**). "Model equations," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York), Chap. 3.

Hariharan, P., Banerjee, R. K., and Myers, M. R. (**2007**). "HIFU procedures at moderate intensities: Effect of large blood vessels," Phys. Med. Biol. **52**, 3493–3513.

Harland, A. R., Petzing, J. N., and Tyrer, J. R. (**2002**). "Non-invasive measurements of underwater pressure fields using laser Doppler velocimetry," J. Sound Vib. **252**, 169–177.

Harris, G. R. (**2005**). "Progress in medical ultrasound exposimetry," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 717–736.

Harris, G. R., Gammell, P. M., Lewin, P. A., and Radulescu, E. (**2004**). "Interlaboratory evaluation of hydrophone sensitivity calibration from 0.1 MHz to 2 MHz via time delay spectrometry," Ultrasonics **42**, 349–353.

Hartley, C. J. (**1997**). "Characteristics of acoustic streaming created and measured by pulsed Doppler ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1278–1285.

Herman, B., and Harris, G. R. (**1982**). "Calibration of miniature ultrasonic transducers using a planar scanning technique," J. Acoust. Soc. Am. **72**, 1357–1362.

International Electrotechnical Commission (IEC). (**2005**). "Power measurement—Radiation force balances and performance requirements up to 1 W in the frequency range 0.5 MHz to 25 MHz and up to 20 W in the frequency range 0.75 MHz and 5 MHz," Ultrasonics **87**, 1–44.

Kamakura, T., Kazuhisa, M., Kumamoto, Y., and Breazeale, M. A. (**1995**). "Acoustic streaming induced in focused Gaussian beams," J. Acoust. Soc. Am. **97**, 2740–2746.

King, L. V. (**1934**). "On the acoustic radiation pressure on spheres," Proc. R. Soc. London, Ser. A **147**, 212–240.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (**1995**). "Convergence properties of the Nelder-Mead algorithm in low dimensions," AT&T Bell Laboratories, Technical Report, Murray Hill, NJ.

Lee, Y. S. (**1993**). "Numerical solution of the KZK equation for pulsed finite amplitude sound beams in thermoviscous fluids," Ph.D. dissertation, University of Texas at Austin, Austin, TX.

Lighthill, J. (**1978a**). "Acoustic streaming," J. Sound Vib. **61**, 391–418.

Lighthill, J. (**1978b**). *Waves in Fluids* (Cambridge University Press, Cambridge, U.K).

Madelin, G., Grucker, D., Franconi, J. M., and Thiaudiere, E. (**2006**). "Magnetic resonance imaging of acoustic streaming: Absorption coefficient and acoustic field shape estimation," Ultrasonics **44**, 272–278.

Maruvada, S., Harris, G. R., Herman, B. A., and King, R. L. (**2007**). "Acoustic power calibration of high-intensity focused ultrasound transducers using a radiation force technique," J. Acoust. Soc. Am. **121**, 1434–1439.

Mathworks Inc. (**2002**). *Matlab Manual* (Mathworks, Inc., Natick, MA), ver. 6.5.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).

Nowicki, A., Kowalewski, T., Secomski, W., and Wojcik, J. (**1998**). "Estimation of acoustical streaming: Theoretical model, Doppler measurements and optical visualization," Eur. J. Ultrasound **7**, 73–81.

Nyborg, W. L. (**1965**). *Physical Acoustics* (Academic, New York), Vol **2B**.

Peters, F., and Petit, L. (**2003**). "A broad band spectroscopy method for ultrasound wave velocity and attenuation measurement in dispersive media," Ultrasonics **41**, 357–363.

Prasad, A. K. (**2000**). "Particle image velocimetry," Eur. Trans Telecommun. Relat. Technol. **79**, 51–60.

Robinson, R. A., and Ilev, I. K. (**2004**). "Design and optimization of a flexible high-peak-power laser-to-fiber coupled illumination system used in digital particle image velocimetry," Rev. Sci. Instrum. **75**, 4856–4863.

Schafer, M. E., Gessert, J. E., and Moore, W. (**2006**). "Development of a high intensity focused ultrasound (HIFU) hydrophone system," Proc. Int. Symp. Therapeutic Ultrasound **829**, 609–613.

Shaw, A. (**2004**). "Delivering the right dose, advanced metrology for ultrasound in medicine," J. Phys.: Conf. Ser. **1**, 174–179.

Shaw, A., and ter Haar, G. R. (**2006**). "Requirements for measurement standards in high intensity focused (HIFU) ultrasound fields," National Physical Laboratory (NPL) Report, ISSN 1744-0599.

Shi, X., Martin, R. W., Vaezy, S., and Crum, L. A. (**2002**). "Quantitative estimation of acoustic streaming in blood," J. Acoust. Soc. Am. **111**, 1110–1121.

Starritt, H. C., Duck, F. A., and Humphrey, V. F. (**1989**). "An experimental investigation of streaming in pulsed diagnostic ultrasound beams," Ultrasound Med. Biol. **15**, 363–373.

Starritt, H. C., Duck, F. A., and Humphrey, V. F. (**1991**). "Force acting in the direction of propagation in pulsed ultrasound fields," Phys. Med. Biol. **36**, 1465–1474.

Theobald, P., Thompson, A., Robinson, S., Preston, R., Lepper, P., Swift, C., Yuebing, W., and Tyrer, J. (**2004**). "Fundamental standards for acoustics based on optical methods–Phase three report for sound in water," Report No. CAIR9 ISSN 1740-0953, National Physical Laboratory (NPL), Teddington, England.

Tjotta, S., and Tjotta, N. J. (**1993**). "Acoustic streaming in ultrasound beams," in *Advances in Nonlinear Acoustics, 13th International Symposium on Nonlinear Acoustics*, edited by H. Hobaek (World Scientific, Singapore), pp. 601–606.

Wang, Z. Q., Lauxmann, P., Wurster, C., Kohler, M., Gompf, B., and Eisenmenger, W. (**1999**). "Impulse response of a fiber optic hydrophone determined with shock waves in water," J. Appl. Phys. **85**, 2514–2516.

Zanelli, C. I., and Howard, S. M. (**2006**). "A robust hydrophone for HIFU Metrology," Proc. Int. Symp. Therapeutic Ultrasound **829**, 618–622.

An ecological acoustic recorder (EAR) for long-term monitoring of biological and anthropogenic sounds on coral reefs and other marine habitats

Marc O. Lammers[a)]
Hawaii Institute of Marine Biology, University of Hawaii, 46-007 Lilipuna Road, Kaneohe, Hawaii 96744 and Joint Institute for Marine and Atmospheric Research, University of Hawaii, 1000 Pope Road, Honolulu, HI 96822

Russell E. Brainard
NOAA Fisheries, Pacific Islands Fisheries Science Center, Coral Reef Ecosystem Division, 1125 B Ala Moana Boulevard, Honolulu, Hawaii 96814

Whitlow W. L. Au and T. Aran Mooney
Hawaii Institute of Marine Biology, University of Hawaii, 46-007 Lilipuna Road, Kaneohe, Hawaii 96744

Kevin B. Wong
NOAA Fisheries, Pacific Islands Fisheries Science Center, Coral Reef Ecosystem Division, 1125 B Ala Moana Boulevard, Honolulu, Hawaii 96814

(Received 27 February 2007; revised 14 December 2007; accepted 28 December 2007)

Keeping track of long-term biological trends in many marine habitats is a challenging task that is exacerbated when the habitats in question are in remote locations. Monitoring the ambient sound field may be a useful way of assessing biological activity because many behavioral processes are accompanied by sound production. This article reports the preliminary results of an effort to develop and use an Ecological Acoustic Recorder (EAR) to monitor biological activity on coral reefs and in surrounding waters for periods of 1 year or longer. The EAR is a microprocessor-based autonomous recorder that periodically samples the ambient sound field and also automatically detects sounds that meet specific criteria. The system was used to record the sound field of coral reefs and other marine habitats on Oahu, HI. Snapping shrimp produced the dominant acoustic energy on the reefs examined and exhibited clear diel acoustic trends. Other biological sounds recorded included those produced by fish and cetaceans, which also exhibited distinct temporal variability. Motor vessel activity could also be monitored effectively with the EAR. The results indicate that acoustic monitoring may be an effective means of tracking biological and anthropogenic activity at locations where continuous monitoring by traditional survey methods is impractical.
 [DOI: 10.1121/1.2836780]

PACS number(s): 43.80.Ev, 43.80.Ka [MCH] Pages: 1720–1728

I. INTRODUCTION

A significant challenge faced by many governmental and nongovernmental management and conservation agencies is the assessment and long-term monitoring of the condition of remote marine ecosystems worldwide. The widespread distribution and isolation of many habitats found along secluded coastal areas, reefs, seamounts, and insular habitats can make monitoring logistically difficult and expensive. Research cruises often result in high ship time costs and typically allow only intermittent and limited opportunities for assessing the conditions at many sites. Moored instruments capable of measuring a wide range of environmental parameters, such as surface and subsurface temperatures, salinity, wave energy, and current flow, provide measures of environmental variability but do not obtain data directly about the biological activity taking place at a location. As a result, many significant ecological events, such as disease outbreaks, episodic infestations (e.g., harmful algal blooms, crown-of-thorn sea stars), reactions to climate change (e.g., massive coral bleaching), and the effect of storms, oil spills, and poaching often occur undetected at remote locations, complicating the interpretation of long-term monitoring data.

The application of an acoustics-based approach to monitoring may provide an important complementary method for detecting changes in the marine environment. This is because sounds present in many marine habitats can be an effective indicator of a number of biological processes, such as spawning events (Lobel, 1992; Luczkovich *et al.*, 1999; Hawkins and Amorim, 2000), courtship behaviors (Mann *et al.*, 1997), feeding (Vesluis *et al.*, 2000) competition (Johnston and Vives, 2003), and social communication among many species of fish, invertebrates, and aquatic mammals. These sounds can be detected over ranges of tens to thousands of meters, depending on the species producing them and the background ambient noise level (Mann and Lobel, 1997; Janik, 2000; Lugli and Fine, 2003; Sprague and Luczkovich, 2004). Therefore, examining the sounds occurring in remote

[a)]Electronic mail: lammers@hawaii.edu

 0001-4966/2008/123(3)/1720/9/$23.00

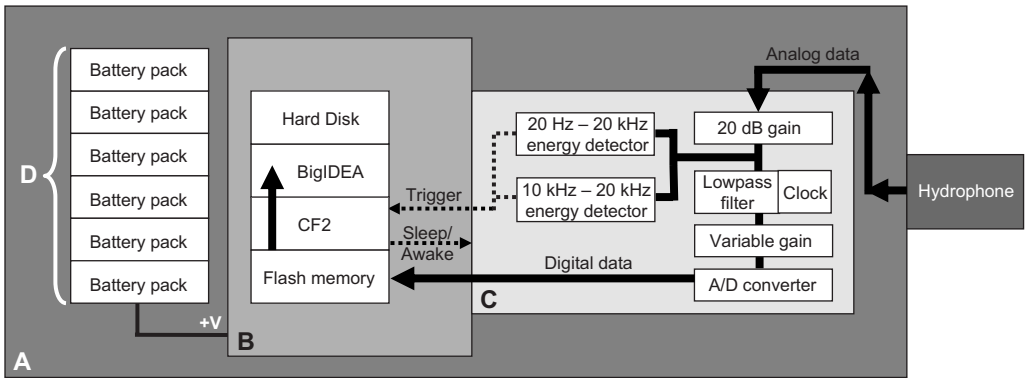


FIG. 1. A schematic representation of the EAR showing (A) the environmental interface module, (B) the central processing unit (CPU)/storage module, (C) the signal conditioning/analog-to-digital conversion (SCADC) module, and (D) the power supply module. The solid arrows (➔) represent the flow of data while the dashed arrows (⇢) describe the operational relationship between the CPU/storage and SCADC modules.

marine habitats could serve as an effective proxy for tracking certain biological processes and detecting natural and changing patterns of biological activity at many locations.

Some of the most challenging habitats to monitor are coral reefs. Coral reefs are among the most biologically diverse and complex ecosystems in the world. Their diversity supports economies around the globe through a variety of commercial activities, such as tourism, fishing, and pharmaceutical production. Despite, and partly because of, their economic and cultural value, coral reefs are rapidly being degraded in many parts of the world by a variety of stressors that include pollution, over fishing, coastal development, physical disturbance, and global climate change (Wilkinson, 2002). The difficulty of managing the health of coral reefs is compounded by the fact that they can be some of the most remote habitats in the world, making them particularly prone to experiencing unobserved changes and declines.

Acoustic monitoring of coral reef habitats is a potentially fruitful approach because many animals associated with coral reefs and nearby waters are soniferous. Numerous species of coral reef fish produce sounds (Myrberg, 1981), as do several invertebrates (Johnson *et al.*, 1947). Additionally, marine mammals that either directly or indirectly interact with the neritic environment associated with coral reefs are also quite vocal (Popper, 1980). Tracking the acoustic activity level of many of these animals is therefore a promising approach for assessing patterns of change, stability, and seasonality in biological processes occurring at different trophic levels over time. This article reports on research to develop the tools needed for monitoring the ambient sounds on remote coral reefs and other marine habitats. Several examples are provided of the types of biological and anthropogenic activities that long-term acoustic monitoring can help document and follow.

II. MATERIALS AND METHODS

Recording the ambient sound field of remote marine habitats for extended periods of several months to a year or longer presents several challenges. The recording bandwidth, power consumption, and data storage capacity of the recorder are among the factors that must be considered carefully, as these affect or involve finite system resources. The cost of production is another important consideration. A prohibitively high price that severely restricts the number of units that can be deployed reduces the system's usefulness as a monitoring tool. Finally, software algorithms that allow for the efficient processing and interpretation of the data obtained are vital to effectively deal with the large volume of acoustic recordings produced by long-term deployments made at multiple locations.

The need for long-term data and low cost preclude an approach that involves continuous recording over a wide bandwidth. Such an approach requires vast power reserves and data storage capacity, each of which reduces the affordability of the system and adds to the engineering complexity. An approach involving periodic recording with the ability to turn "on" when signals of interest occur is more desirable from both a cost and a data management standpoint.

The Ecological Acoustic Recorder (EAR) was developed to meet these requirements. The EAR is a digital recorder based on a Persistor™ CF2 microprocessor. It is a low power system that records on a programmable duty cycle, but is also capable of responding to acoustic events that meet specific criteria.

A. Hardware

Four principal components make up the EAR hardware: the environmental interface module, the signal conditioning/analog-to-digital conversion module, the central processing unit (CPU)/storage module and the power supply module. Figure 1 shows a schematic of how the different components are integrated.

The environmental interface module consists of the recording hydrophone and the package in which the electronics are housed. The system uses a Sensor Technology SQ26-01 hydrophone with a response sensitivity of −193.5 dB that is flat (±1.5 dB) from 1 Hz to 28 kHz. One of two housing packages manufactured by Sexton Photographics can be used, depending on the target deployment depth (Fig. 2). For deployments to less than 46 m, a 10.16-cm-diam by 60-cm-long by 0.64-cm-thick PVC tube is used that is enclosed on one end by a permanently sealed cap and on the other by an

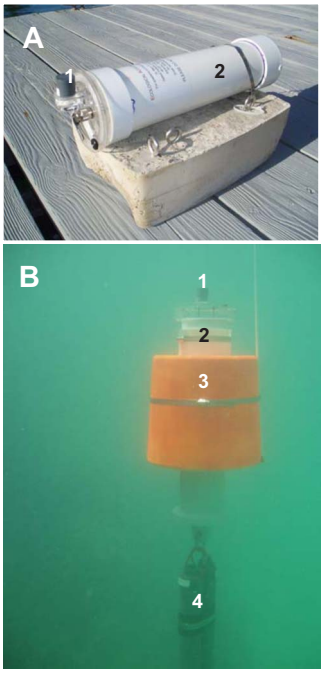


FIG. 2. (Color online) (A) The shallow water EAR showing (1) the hydrophone and (2) the PVC housing attached to a concrete anchor. (B) The deep EAR showing (1) the hydrophone, (2) the aluminum housing, (3) the syntactic foam collar, and (4) the acoustic releases attached to a sacrificial anchor (not shown).

acrylic face, on which the hydrophone is mounted. This package is typically attached to a lead or concrete anchor by divers. For deployments to 500 m, an aluminum housing is used in combination with a syntactic foam float and an acoustic release system (Sub Sea Sonics AR-60). Deep water EAR units and their releases are anchored to a sacrificial weight (typically ~70 kg of gravel/sand bags) using a 1 m stainless steel cable and released overboard. A topside acoustic release interrogator (Sub Sea Sonics ARI-60) is used to obtain range information to the unit and trigger the erosion of the release's burn wire. Two releases are typically used in parallel for redundancy.

The signal conditioning/analog-to-digital conversion (SCADC) module is a custom-designed circuit that amplifies and filters the input signal and then digitizes it. The analog signal is passed through a series of operational amplifier stages (MAX494, Maxim Integrated Products, Sunnyvale, CA) for amplification and then filtered using a switched capacitor filter (LTC1064-7, Linear Technology Corporation, Milpitas, CA). This provides an eighth-order low-pass filter with a cutoff frequency that is controlled by an input clock. That clock is provided by a programmable oscillator (MAX1077L-40, Maxim Integrated Products, Sunnyvale, CA). The cutoff frequency is set via software as a fractional percentage value of the Nyquist frequency, for example 80%. After being filtered, the signal is then digitized by a 16 bit analog-to-digital converter (Burr-Brown ADS8344, Texas Instruments, Dallas, TX). Recordings are first stored as raw binary files in flash memory and then periodically transferred to a hard drive (see the following).

In addition to conditioning the analog signal, the SCADC module also includes circuitry that monitors the input signals for specific types of acoustic events. Two event-detection circuits "listen" for energy in different frequency bands through a series of op-amps (MAX494) and differential comparator stages (LP339, National Semiconductor, Santa Clara, CA). A "wideband" event detector monitors the energy in the frequency band from 20 Hz to 20 kHz, while a "high frequency" detector monitors the energy in the band from 10 to 20 kHz. Variable resistors are used to set the signal amplitude thresholds and the energy integration period that cause each circuit to send an interrupt request to the CPU, indicating the occurrence of an acoustic event. Either (or both) event detector can be programmed to operate in parallel with duty cycled recordings.

The CPU/storage module receives input from the SCADC module and controls the recording process. It is composed of the Persistor CF2 microprocessor, a 1 Gbyte compact flash card, a Persistor BigIDEA IDE adapter, and a 120 Gbyte 2.5 in. Toshiba hard disk drive. Custom-written software on the CF2 microprocessor controls the recording duty cycle, the system's power consumption and the transfer of data from flash memory to the hard disk, and monitors the event detection circuits for interrupt requests. The CF2 is accessed via a serial connection that allows a number of variables to be modified using a terminal emulation program (e.g., HYPERTERMINAL™, MOTOCROSS™). These variables include the sampling rate, the length of individual recordings, the recording period, the recording start date and time, whether to record when an event detection circuit is triggered, and other parameters related to power monitoring and the anti-alias filter cut-off frequency.

Finally, the power supply module provides the system with a continuous voltage supply. The module consists of four battery packs wired in parallel. Each battery pack is composed of seven high capacity D-cell alkaline batteries serially wired to provide 20 500 mA h of current at 10.5 V. The modular arrangement of the battery packs allows them to easily be added or subtracted according to the power consumption needs of individual deployments. Deployments requiring more than four battery packs are realized by simply lengthening the environmental interface module.

B. Power consumption

The EAR's power consumption is regulated through the CF2 microprocessor. It controls the power supply to the SCADC module and the frequency with which the hard disk drive is accessed. To minimize the latter, the CF2 monitors the number of recordings present on the compact flash card and calculates the time it will take to transfer them to the hard disk. The disk is spun up and accessed at intervals that keep the writing process from interfering with scheduled recordings. When in standby mode between recording periods, the EAR draws 0.3 mA of current when the event detection circuits are disabled and 4.0 mA if they are enabled. When recording, the unit draws approximately 70 mA, and when writing to disk it uses between 300 and 400 mA.

C. Operation

The EAR can be programmed to begin recording either immediately when powered on, or at a future date and time. The sampling rate used will depend on the target species and/or the signals of interest. Prior experience has determined that sampling at a rate of 25 kHz (providing 12.5 kHz of bandwidth) is well suited for capturing the majority of coral reef-associated acoustic signals and the sounds of motor vessels passing nearby. If higher frequency sounds are of interest, such as those produced by cetaceans, a maximum sampling rate of 64 kHz can be selected.

The recording duty cycle and duration used for a deployment must be chosen based on several factors. These include the likelihood of capturing the signal(s) of interest, the length of the deployment, the number of recordings that can be stored on the hard disk drive, and the expected power consumption. Using a duty cycle method for recording is not well suited for capturing acoustic signals and events that are very infrequent and random, but it is effective in documenting the pattern of occurrence of regularly occurring signals typical of a specific location. The approach used to collect the data presented here was to record at a 3.3% duty cycle, or once every 15 min for 30 s. This produced approximately 4.38 Gbytes of data per month and consumed about 3.54 A when no event detection was enabled (6.25 A when event detection was enabled).

Event detection can in theory be applied to capture any acoustic event with sufficient energy and lasting long enough to meet the threshold criteria set in the SCADC circuit. This capability is especially useful for detecting the occurrence of events such as vessels transiting nearby, which generally produce sustained high levels of acoustic energy for several seconds. Calls or signals produced by animals can also be automatically detected, provided their levels are sufficiently high to distinguish them from the natural fluctuations in the background ambient noise. Thus, the efficiency of event detection will be determined primarily by the signal-to-noise ratio present at a specific site.

D. Data analysis

The volume of data produced by the EAR even for short deployments makes manual analysis of all the recordings collected unfeasible. If one considers that, at a recording duty cycle of 3.3% without event detection enabled, 2880 recordings are produced over a 30 day period, it is clear that an automated analysis approach is required if long-term data are to be used as an effective and timely monitoring tool. To this end, custom algorithms were developed in MATLAB™ and in Lockheed Martin's RIPPEN™ programming environment that process the data sets automatically. Figure 3 shows a flow chart summarizing the data reduction process. Among the variables that are extracted from each recording are: the rms sound pressure level (SPL) and its variance, the number of acoustic events detected automatically, the number of short pulses with a maximum above a defined threshold, and the frequency of occurrences of specified sounds of interest.

To establish the occurrences of specific sounds, each recorded file is searched using one or more template files representing the signal of interest. Template files are obtained by visually and aurally examining approximately 1% to 2% of the data from each site and identifying discrete, repeated biological signals. Representative examples of these sounds are then used to search the database by stepping the template files through each recording and performing a cross correlation at each step. Cross-correlation matches of 0.70 (min =0, max=1) or greater are accepted as detections of the reference signal. A manual evaluation of this approach determined that, for stereotyped signals, it is a moderately conservative method of establishing signal occurrence, with a somewhat greater tendency for producing false negative than false positive detections. This was deemed more desirable than a liberal method that results in excessive false positive detections. For signals with higher variability, the correlation threshold can be lowered to reduce the number of missed detections.

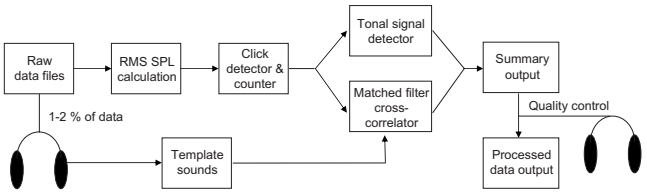


FIG. 3. A flow chart diagram summarizing the data reduction process for EAR data. A subset of data files are examined manually and template files are created of representative sounds. All data files are subjected to rms SPL calculations, a click detector and counter routine and either a matched filter analysis or a tonal signal detector or both. The summary output is checked manually to confirm the matched filter and/or tonal signal detector results.

A different method is used to detect the presence of more variable tonal sounds, such as those produced by many cetacean species and some species of fish. A short-time Fourier transform approach is applied to find consistent periods of tonal spectral peaks in each data file. These periods are then summed and reported for each recording. Data files with tonal periods in excess of 1% are manually examined to first confirm the presence of biologically related sounds and then to identify their likely origin (whale, dolphin, fish, echosounder, etc.).

E. Preliminary deployments

Test deployments of the EAR were conducted at four sites around the island of Oahu, HI. Two sites were in Kaneohe Bay (KB1, 21° 28.245 N 157° 49.585 W and KB2, 21° 26.030 157° 47.485 W), one in the Waikiki Marine Life Conservation District (WMLCD, 21° 15.94 N 157° 49.670 W) and one off Makua Beach (MB, 21° 31.75 N 158° 13.940 W). The KB1 site was monitored monthly for 10 day periods over a period of 1 year between December 2004 and November 2005. These deployments were conducted with the first generation of the EAR system and were limited to 10 day periods by the original system's higher power consumption. Subsequent deployments were conducted with the EAR system as described earlier. These lasted between 17 and 42 days. The deployment at the WMLCD site was between 25

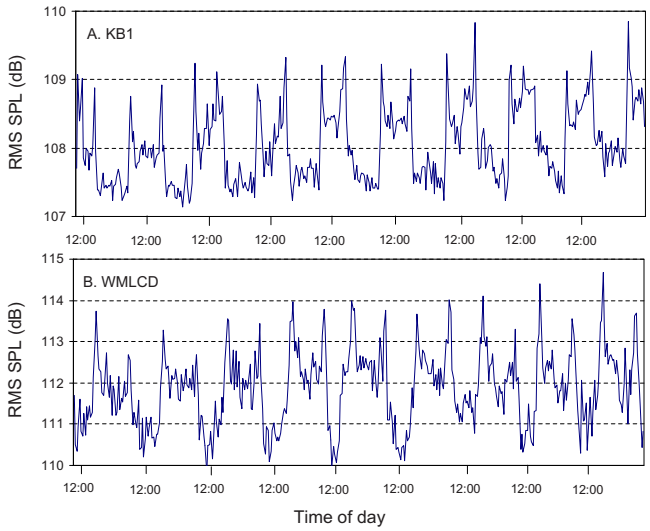


FIG. 4. (Color online) (A) Plot of snapping shrimp acoustic energy received at site KB1 from 12 to 21 December 2004. (B) Plot of snapping shrimp acoustic energy received at site WMLCD from 25 January to 2 February 2006. "12:00" represents noon on each day.

January and 10 February 2006, the deployment at MB was between 13 February and 25 March 2006 and the deployment at the KB2 site was between 21 July and 31 August 2006.

III. RESULTS

The deployments at the four sites yielded a total of 16,161 recordings. The results presented here are highlighted examples of the trends that were observed. They are not comprehensive and should be viewed only as preliminary. No specific inference relating to reef condition is intended; rather, they are presented to illustrate the utility of the EAR system as a tool to monitor temporal trends in biological and anthropogenic activity on coral reefs and other types of marine habitats.

A. KB1 deployments

Site KB1 is located on the slope of a small patch reef (200 m × 150 m) inside Kaneohe Bay at a depth of approximately 8 m. The reef is primarily dominated by large colonies of the coral species *Porites compressa*, but two other species with a significant presence are *P. lobata* and *Pocillopora meandrina*. Several species of reef fish occur at or near the site. These include members of the genera *Cheatodon*, *Dascylus*, *Scarus*, *Labroides*, *Zanclus*, *Thallasoma*, *Mulloidichthys*, *Acanthuris*, *Zebrasoma*, *Gymnothorax*, and *Abudefduf*, among others.

The predominant sound on this reef, as on all the reefs that were examined, was the "crackle" of snapping shrimp (*Alpheus sp.*). This is the summed contribution of the clicks produced by individual, benthic-dwelling shrimp rapidly closing their enlarged claw, thereby producing a collapsing cavitation bubble (Vesluis *et al.*, 2000). Figure 4(A) illustrates that this crackle has a clear, diel cyclical pattern that can be measured by plotting the averaged rms SPL of the recordings over time. This pattern exhibits two distinct features. The first is a sharp increase in activity during crepuscular periods. This is represented by a nearly 2 dB rise and then drop in rms SPL levels over a 2 h period centered on sunrise and sunset. The second is a higher level of activity during daylight hours than during the night, which was consistent throughout the 1 year period that the site was monitored.

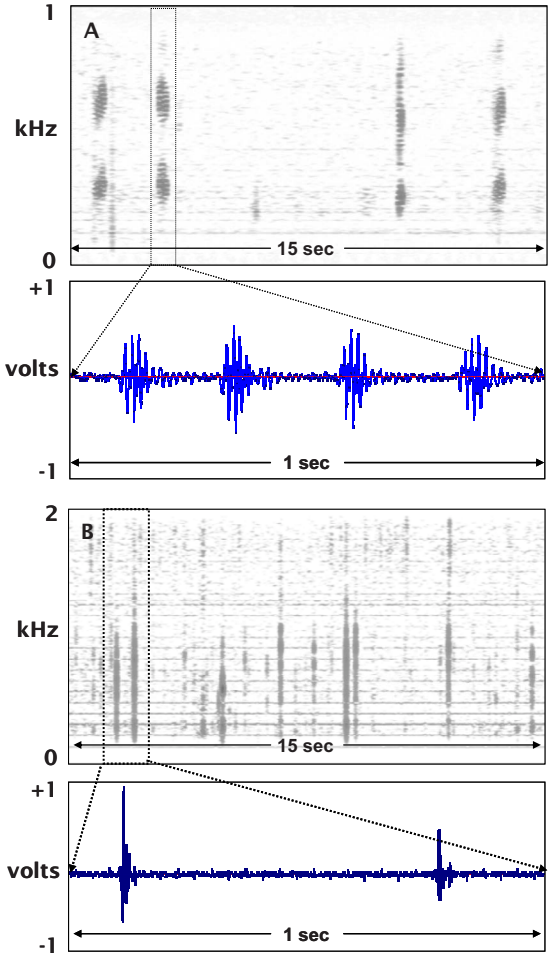


FIG. 5. (Color online) Spectrogram and wave form of signals recorded at the KB1 site from (A) *Dascyllus albisella* and (B) an unidentified species (KB1sigB).

Visual and aural examination of the data revealed that up to five sounds likely originating from fish occurred with regularity at this site. The most common of these was a pulsed train shown in Fig. 5(A) known to be produced by domino damsel fish (*Dascylus albisella*) (Mann and Lobel, 1997). This sound is thought to represent a territorial defense signal and is often heard by divers who approach a damselfish colony. It was produced significantly more during daylight hours (3.2 pulses/min) than at night (1.3 pulses/min) (One-way ANOVA, $p < 0.001$). In addition, its occurrence over the 1 year period that the site was monitored showed evidence of seasonal variability, with a sharp increase in activity noted during the month of May (Fig. 6).

The other sounds heard in the recordings could not be matched to the species producing them. The most common of these was a single, short (10 ms), pulsed signal with broadband energy between 200 and 1500 Hz, shown in Fig.

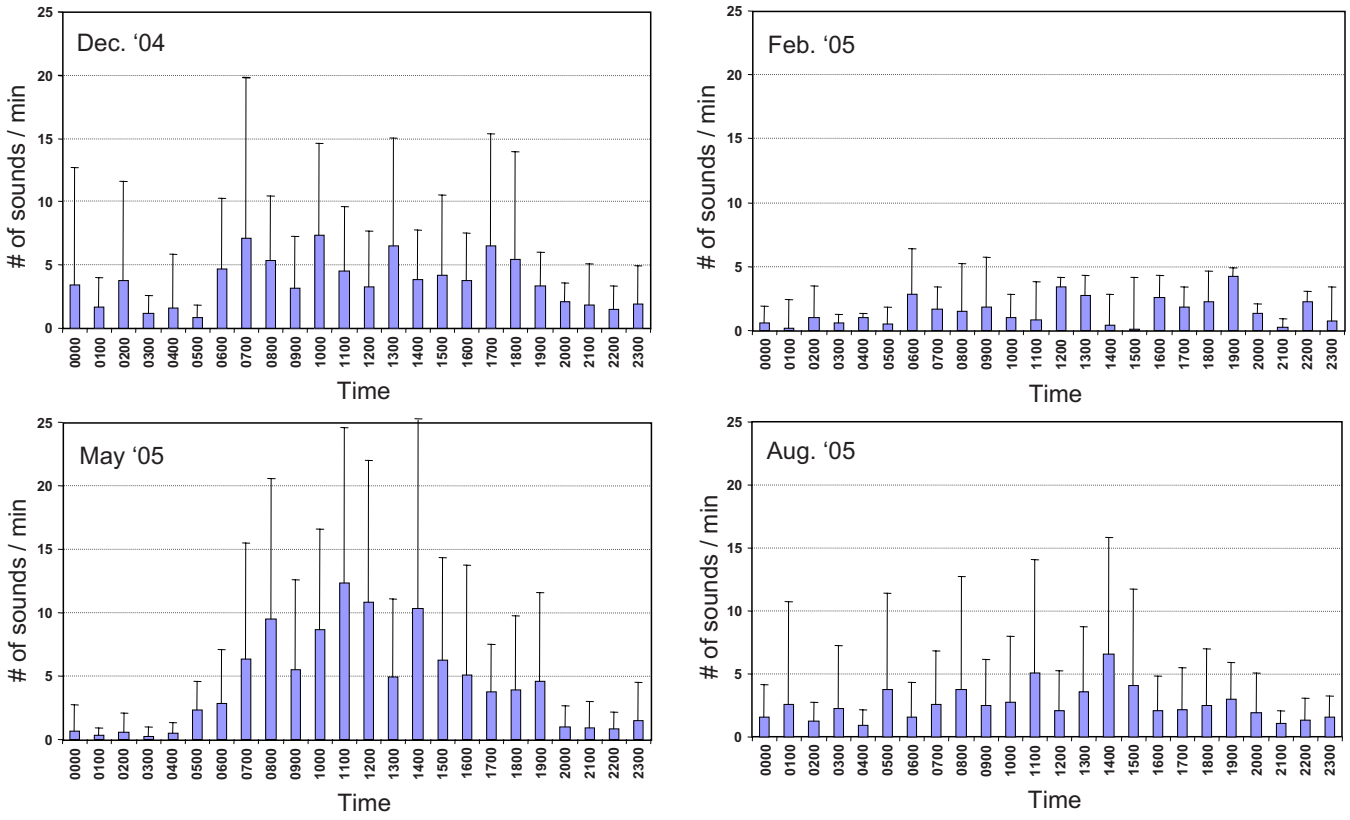


FIG. 6. (Color online) Monthly variation of *Dascyllus albisella* signals recorded at site KB1 averaged over the deployment period showing an increase in signaling during the month of May.

5(B) and termed KB1sigB. It also exhibited a distinct diel pattern of occurrence. However, unlike the damselfish pulse train, this signal occurred almost exclusively at night. More specifically, it was produced primarily about 1 h before sunrise and 1 h after sunset. Moreover, the time of maximum occurrence closely followed the change in day length between December and May (Fig. 7).

B. WMLCD deployment

The WMLCD site is characterized by mostly uncolonized basaltic rock reef. Despite limited coral cover, protection from fishing has promoted a moderately high level of fish biodiversity in the area. Many of the taxa described for the KB1 site are also found here in addition to others, such as *Rhinecantus*, which is commonly observed at the site. The EAR was deployed at a depth of approximately 6 m.

As in Kaneohe Bay, the snapping shrimp sound field exhibited a clear diel trend. A sharp increase in activity occurred both at sunrise and at sunset. This pattern was even more pronounced here than at the KB1 site, with a rise of nearly 4 dB occurring during crepuscular periods relative to daytime levels [Fig. 4(B)]. In contrast to KB1, however, overall snapping shrimp noise levels were higher during nighttime hours than during the day by approximately 2 dB. In other words, the diel trends at the two locations were reversed.

Manual examination of the data set revealed that nine distinct sounds, presumably from fish, occurred with regularity at this site. These could not be identified to the species,

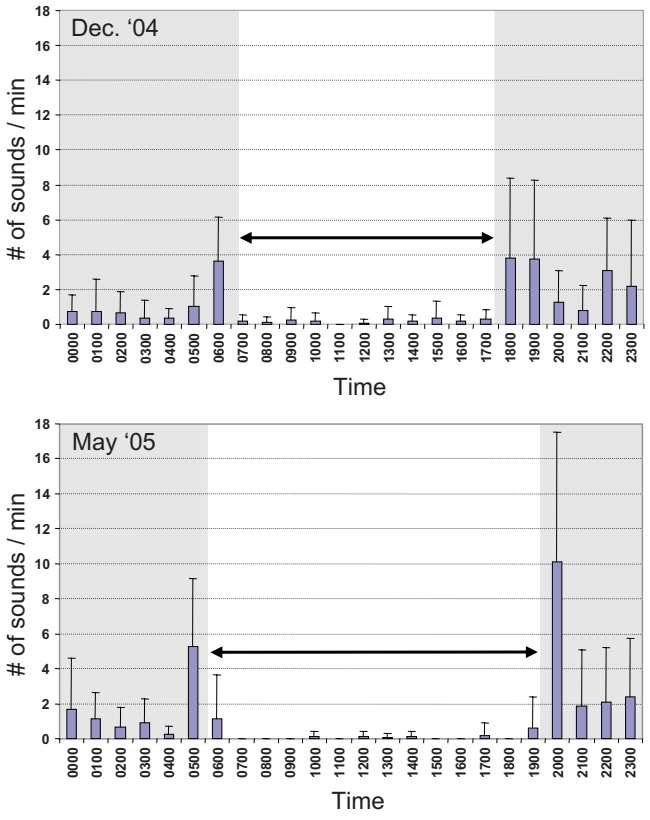


FIG. 7. (Color online) The occurrence of the KB1sigB signal as a function of time of day averaged over the deployment period. The top panel shows its occurrence in December 2004 and the lower panel shows its occurrence in May 2005. The shaded areas indicated nighttime. Note the correspondence between the shift in acoustic activity and the change in the length of day.

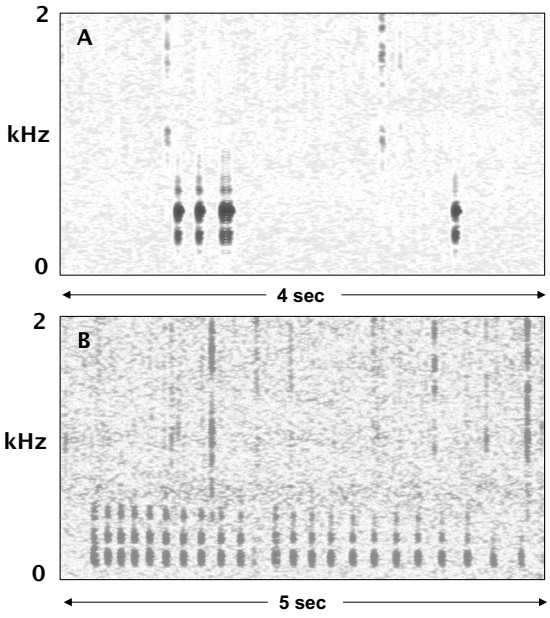


FIG. 8. Spectrograms of the two most common fish sounds recorded at the WMLCD site, (A) WMLCDsigA and (B) WMLCDsigD.

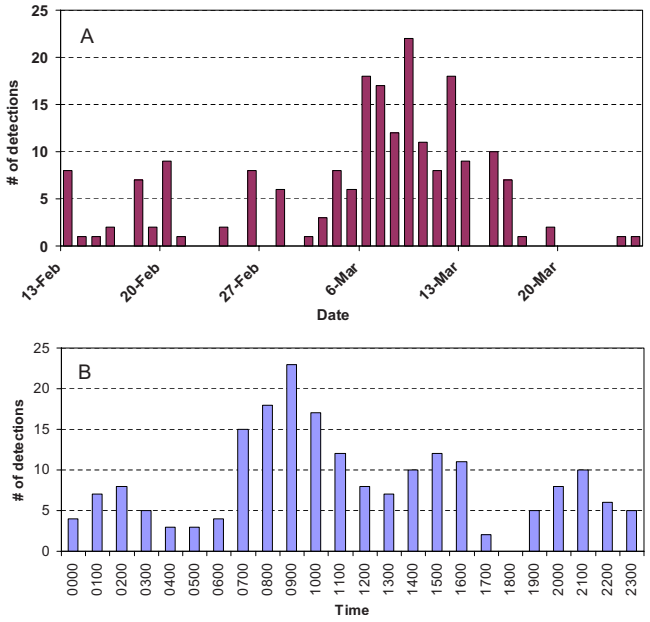


FIG. 9. (Color online) (A) The occurrence of recordings with dolphin signals at the MB site over the course of the deployment period. (B) The occurrence of recordings with dolphin signals as a function of the time of day averaged over the deployment period.

but all exhibited diel patterns of occurrence. The most common was a 30 ms pulse with a peak frequency of 500 Hz shown in Fig. 8(A) (WMLCDsigA). Although presumably produced by a different species of fish, its temporal pattern of occurrence was strikingly similar to that of the KB1sigB sound, with distinct peaks in occurrence during presunrise and postsunset periods. Another very common sound was WMLCDsigD, shown in Fig. 8(B). This 25 ms pulse with peak frequency around 170 Hz occurred rarely during daylight hours, was somewhat more common at night, and was most frequent for about 1 h just after sunset.

C. MB deployment

The MB EAR was not deployed on a coral reef but rather on a large stretch of sand at a depth of approximately 15 m. The site was chosen to gauge the effectiveness of the EAR as a tool for documenting the presence of cetaceans in an area over time. The waters off Makua Beach are well known as a daytime spinner dolphin (*Stenella longirostris*) resting area (Lammers, 2004). In addition, wintering humpback whales (*Megaptera novaengliae*) commonly occur nearby during the period between January and April. Both species are acoustically active and regularly produce social signals in the recording frequency band of the EAR (Lammers *et al.*, 2003; Au *et al.*, 2006).

Dolphin whistles and/or clicks were detected 203 times out of 3841 recordings. The presence of dolphins was recorded on 29 out of the 41 days that the EAR was deployed. The occurrence of dolphins at the Makua Beach area was not uniform throughout the deployment period. Rather, detections were few and sporadic during the initial part of the deployment, relatively consistent for a period of approximately 2 weeks during the second half of the deployment, and nearly absent during the last several days [Fig. 9(A)]. In addition, some clear diel trends were observed in the production of sounds [Fig. 9(B)]. The most active period was during the morning, with a peak at around 0900 h. Acoustic activity was considerably reduced during the middle of the day before increasing and peaking again at around 1500 h. The fewest detections occurred during the hours just prior to sunset and sunrise.

D. KB2 deployment

Site KB2 is located on the slope of the reef adjacent to Moku'O'Loe (Coconut) island in Kaneohe Bay at a depth of approximately 5 m. This site was chosen primarily because of its proximity to one of the bay's small boat channels. The primary objective of this deployment was to gauge the EAR's ability to be triggered in response to passing vessel traffic.

Five hundred seventy-four vessels were detected during the deployment, or an average14.35 vessels per day (s.d. $=5.58$). Significantly more vessels were detected on weekend days (19.58/day, s.d$=6.33$) than during weekdays (12.11/day; s.d.$=3.36$) ($P=0.002$; two-sample t-test). In addition, Fig. 10 shows that nearly all vessel traffic was detected between the hours of 0700 h and 2100 h, with the peak in traffic occurring at 1600 h.

IV. DISCUSSION

The results presented indicate that the EAR and the automated data analysis algorithms provide an effective combination of tools for documenting the temporal patterns of a wide range of acoustic signal types. The first test deployments revealed that the EAR is a useful instrument for recording long-term patterns of biological sound production on coral reefs and other marine habitats. In addition, the event

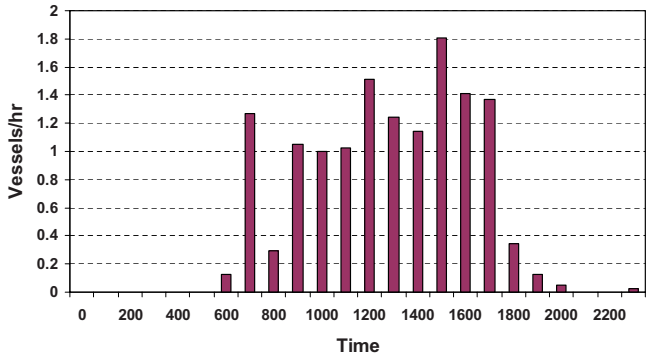


FIG. 10. (Color online) The occurrence of motor vessels at site KB2 as a function of time of day averaged over the deployment period.

detection capability of the system proved effective in documenting the occurrence of anthropogenically produced sounds, such as vessel engine noise.

It is evident from the two deployments on coral reefs that, not only are there numerous sounds present on reefs, but that their occurrences tend to follow well-defined, temporal patterns that are tied to physical variables, such as day/night cycles, changes in day length, and seasons. This relationship with the physical environment is of significance because it indicates that the ambient sound field is in fact influenced by abiotic factors experienced by the habitat. This finding lends support to the hypothesis that sounds on a reef will reflect ecologically significant changes in conditions. Some evidence for this already exists. Watanabe *et al.* (2002) provided the first indication that snapping shrimp activity is tied to changes in water temperature and dissolved oxygen concentrations. This relationship needs to be investigated further before inferences can be made on reef conditions, but the finding raises the intriguing possibility that events such as temperature-induced coral reef bleaching might be accompanied by measurable changes in the snapping shrimp sound field.

The peak in activity of damselfish sounds recorded at the KB1 site in May was coincident with the beginning of their breeding season (Randall, 1996). Whether the increase was related to spawning behavior, more territorial defense, or a change in some other activity is not presently known. However, the fact that a change was measured further validates the approach of long-term, periodic sampling of the sound field as a means of documenting trends occurring on time scales of days, months, or seasons.

The EAR's ability to document the presence of cetaceans adds to its value as an ecological monitoring tool. Cetaceans are higher trophic level consumers than most of the fish and invertebrate species occurring on coral reefs. Therefore, their presence or absence from an area has implications with respect to certain resources occurring there. Spinner dolphins in Hawaii, for example, are known to forage on a community of mesopelagic organisms that migrates vertically and horizontally toward shore each night (Benoit-Bird and Au, 2003). Not surprisingly, their signals were recorded at night, as this probably reflected their nocturnal return toward shore with their prey. Of note, however, is their rather episodic pattern of occurrence at the MB site, suggesting that prey occurrence and distribution might be heterogeneous with respect to time and space. Long-term passive acoustic monitoring of spinner dolphins over a broader area could therefore be a useful tool for gaining new insights into both spinner dolphin distribution and also the temporal and spatial occurrence of a community of prey that sustains not only cetaceans, but also tunas, billfishes, and bottomfishes (Benoit-Bird *et al.*, 2001).

Finally, the EAR's demonstrated ability to trigger in response to vessel engine noise makes it a useful tool for natural area managers concerned with monitoring the amount of legal and/or illegal vessel activity in conservation districts, parks, reserves, and sanctuaries. This is especially the case for locations where direct monitoring is either not feasible or only intermittently possible, such as many remote islands and atolls. However, it should be noted that the system's ability to detect vessels will be limited by the trigger thresholds that are set, which in turn will be determined by the natural ambient noise level. Thus, quieter environments will allow setting lower thresholds, which will improve the system's ability to detect far-off vessels and those with quieter engines. Follow-up work with the EAR system will be directed at better defining the range of vessel detection under different ambient noise conditions.

In summary, passive acoustic monitoring of coral reef and other marine habitats appears to be a fruitful means of gaining new insights into both biological and anthropogenic activities at locations of interest. Moreover, the EAR takes its place alongside other passive acoustic recorders that have been developed in recent years for monitoring marine sounds (Calupca *et al.*, 2000; Duncan *et al.*, 2000; Fox *et al.*, 2001; Wiggins, 2003; Wiggins *et al.*, 2005). Ongoing work with the EAR will continue to investigate long-term patterns of sound production on coral reefs and other marine habitats and will explore further their link with the ecosystem's condition.

ACKNOWLEDGMENTS

The authors thank Bob Herlien for his contributions to the design of the EAR circuit and software, Ken Sexton for the design of the system's environmental interface module, and Dave Lemonds and Rey Nakamura with Lockheed Martin for their contribution to the RIPPEN-based analysis software. The authors are also grateful to Sara Stieb, who assisted with processing much of the preliminary EAR data. In addition, Paul Nachtigall provided valuable logistical assistance for the work presented. Funding support was provided by the National Oceanic and Atmospheric Administration's (NOAA) Coral Reef Conservation Program as part of the Reef Assessment and Monitoring Project and also the State of Hawaii Department of Land and Natural Resources' Division of Aquatic Resources. Use or mention of any of the commercial products described in this publication does not constitute an endorsement by NOAA Fisheries or any other Federal agency. This is HIMB publication No. 1302.

Au, W. W. L., Pack, A. A., Lammers, M. O., Herman, L. M., Deakos, M., and Andrews, K. (**2006**). "Acoustic properties of humpback whale song," J. Acoust. Soc. Am. **120**, 1103–1110.

Benoit-Bird, K. J., and Au, W. W. L. (**2003**). "Prey dynamics affect foraging by a pelagic predator (Stenella longirostris) over a range of spatial and temporal scales," Behav. Ecol. Sociobiol. **53**, 364–373.
Benoit-Bird, K. J., Au, W. W. L., Brainard, R. E., and Lammers, M. O. (**2001**). "Diel horizontal migration of the Hawaiian mesopelagic boundary community observed acoustically," Mar. Ecol.: Prog. Ser. **217**, 1–14.
Calupca, T. A., Fristrup, K. M., and Clark, C. W. (**2000**). "A compact digital recording system for autonomous monitoring," J. Acoust. Soc. Am. **108**, 2582.
Duncan, J. A., Cato, D. H., Thomas, F., and McCauley, R. D. (**2000**). "The development of a compact instrument for the measurement of biological sea noise," J. Acoust. Soc. Am. **108**, 2584.
Fox, C. G., Matsumoto, H., and Lau, T. K. A. (**2001**). "Monitoring Pacific Ocean seismicity from an autonomous hydrophone array," J. Geophys. Res. **106**, 4183–4206.
Hawkins, A. D., and Amorim, M. C. P. (**2000**). "Spawning sounds of the male haddock, *Melanogrammus aeglefinus*," Environ. Biol. Fish. **59**, 29–41.
Janik, V. M. (**2000**). "Source levels and the estimated active space of bottlenose dolphin (*Tursiops truncatus*) whistles in the Moray Firth, Scotland," J. Comp., Psych. **186**, 673–680.
Johnson, M. W., Everest, A., and Young, R. W. (**1947**). "The role of snapping shrimp (Crangon and Synalpheus) in the production of underwater noise in the sea," Biol. Bull. **93**, 122–138.
Johnston, C. E., and Vives, S. P. (**2003**). "Sound production in *Codoma ornata* (Girard) (Cyprinidae)," Environ. Biol. Fish. **68**, 81–85.
Lammers, M. O. (**2004**). "Occurrence and behavior of Hawaiian spinner dolphins (*Stenella longirostris*) along Oahu's leeward and south shores," Aquat. Mamm. **30**, 237–250.
Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signaling behavior of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.
Lobel, P. S. (**1992**). "Sounds produced by spawning fish," Environ. Biol. Fish. **33**, 351–358.
Luczkovich, J. J., Sprague, M. W., Johnson, S. E., and Pullinger, R. C. (**1999**). "Delimiting spawning areas of weakfish, *Cynoscion regalis* (family Sciaenidae), in Pamlico Sound, North Carolina using passive hydroacoustic surveys," Bioacoustics **10**, 143–160.
Lugli, M., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: Ambient noise and short-range propagation in shallow streams," J. Acoust. Soc. Am. **114**, 512–521.
Mann, D. A., Bowers-Altman, J., and Rountree, R. A. (**1997**). "Sounds produced by the striped cusk-eel *Ophidion marginatum* (Ophidiidae) during courtship and spawning," Copeia **1997**, 610–612.
Mann, D. A., and Lobel, P. S. (**1997**). "Propagation of damselfish (Pomacentridae) courtship sounds," J. Acoust. Soc. Am. **101**, 3783–3791.
Myrberg, A. A., Jr. (**1981**). "Sound communication and interception in fishes," in *Hearing and Sound Communication in Fishes*, edited by A. R. Popper and R. R. Fay (Springer, Berlin), pp. 359–426.
Popper, A. N. (**1980**). "Sound emission and detection by Delphinids," in *Cetacean Behavior: Mechanisms and Functions*, edited by L. M. Herman (Wiley, New York), pp. 1–52.
Randall, J. E. (**1996**). *Shore Fishes of Hawaii* (University of Hawaii Press,), 216 pp.
Sprague, M. W., and Luczkovich, J. J. (**2004**). "Measurement of an individual silver perch (*Bairdiella chrysoura*) sound pressure level in a field recording," J. Acoust. Soc. Am. **116**, 3186–3191.
Vesluis, M., Schmitz, B., von der Heydt, A., and Lohse, D. (**2000**). "How snapping shrimp snap: Through cavitating bubbles," Science **289**, 2114–2117.
Watanabe, M., Sekine, M., Hamada, E., Ukita, M., and Imai, T. (**2002**). "Monitoring of shallow sea environment by using snapping shrimps," Water Sci. Technol. **46**, 419–424.
Wiggins, S. M. (**2003**). "Autonomous acoustic recording packages (ARPs) for long-term monitoring of whale sounds," Trans. Inst. Min. Metall., Sect. C **37**, 13–22.
Wiggins, S. M., Grasha, C., Hardy, K., and Hildebrand, J. (**2005**). "High-frequency Acoustic Recording Package (HARP) for long-term monitoring of marine mammals," J. Acoust. Soc. Am. **117**, 2525.
Wilkinson, C. (**2002**). *Status of Coral Reefs of the World: 2002* (Australian Institute of Marine Science), 378 pp.

Vocal learning in Budgerigars (*Melopsittacus undulatus*): Effects of an acoustic reference on vocal matching

Kazuchika Manabe,[a] Robert J. Dooling,[b] and Elizabeth F. Brittan-Powell
Department of Psychology, University of Maryland, College Park, Maryland 20742

(Received 24 February 2007; revised 9 November 2007; accepted 26 December 2007)

Budgerigars were trained to produce specific vocalizations (calls) using operant conditioning and food reinforcement. The bird's call was compared to a digital representation of the call stored in a computer to determine a match. Once birds were responding at a high level of precision, we measured the effect of several manipulations upon the accuracy and the intensity of call production. Also, by differentially reinforcing other aspects of vocal behavior, budgerigars were trained to produce a call that matched another bird's contact call and to alter the latency of their vocal response. Both the accuracy of vocal matching and the intensity level of vocal production increased significantly when the bird could hear the template immediately before each trial. Moreover, manipulating the delay between the presentation of an acoustic reference and the onset of vocal production did not significantly affect either vocal intensity or matching accuracy. Interestingly, the vocalizations learned and reinforced in these operant experiments were only occasionally used in more natural communicative situations, such as when birds called back and forth to one another in their home cages. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2835440]

PACS number(s): 43.80.Ka [JAS] Pages: 1729–1736

I. INTRODUCTION

The interaction between hearing and vocal production has been widely studied in birds, especially those that develop their vocalizations through learning. Much of what we know about vocal production and vocal learning in birds comes from experiments that examined song behavior following deafening (during either the nestling period or in adulthood), rearing in isolation, or selective exposure to particular types of vocalizations during development (see, for example, Kroodsma and Miller, 1982; 1996). Together, these approaches reveal considerable variation in the styles of vocal learning and vocal development in birds.

We also know from both field and laboratory studies that a number of passerine and psittacine species learn or mimic one another's calls (Dooling, 1986; Farabaugh and Dooling, 1996; Mammen and Nowicki, 1981; Mundinger, 1970; Trillmich, 1976a, b). This phenomenon is less well studied than song learning in songbirds, and there is less information on how quickly such learning can occur (but see Farabaugh *et al.*, 1994; Mammen and Nowicki, 1981). Songbirds have often been used as subjects in these experiments, but psittacids also show call learning (see the review in Farabaugh and Dooling, 1996).

Budgerigars, for example, have a complex vocal repertoire that includes learned calls and a long rambling nonstereotyped warble song. In these birds, both sexes show considerable vocal plasticity throughout life (Farabaugh and Dooling, 1996; Hile and Striedter, 2000; Hile *et al.*, 2005). The contact call is the most frequently used call in the budgerigar vocal repertoire (Brittan-Powell *et al.*, 1997; Brockway, 1964a, b, 1969; Farabaugh and Dooling, 1996; Farabaugh *et al.*, 1992, 1994; Hall *et al.*, 1997; Heaton and Brauth, 1999; Heaton *et al.*, 1999; Hile *et al.*, 2000; Hile and Striedter, 2000; Hile *et al.*, 2005; Striedter *et al.*, 2003; Wyndham, 1980). Studies have shown that birds housed together in small groups come to share contact calls with their cagemates in a matter of a few weeks (Brown *et al.*, 1988; Farabaugh *et al.*, 1994; but, see also, Hile *et al.*, 2000; Hile and Striedter, 2000). Playback experiments in the laboratory show that budgerigars call more frequently in response to a mate's call (Ali *et al.*, 1993).

The above-cited studies, as with almost all vocal learning studies in birds, relied on spontaneous behavior and vocalizations with little experimental control over the bird's motivation, calling behavior, or reinforcement contingencies. It is a relatively simple matter to direct a human to produce (or learn) a specific vocalization as a means of investigation. While it is more difficult to accomplish the strict equivalent in animals, recent studies have brought budgerigar vocal behavior under experimental control (Manabe *et al.*, 1997; Manabe and Dooling, 1997; Manabe *et al.*, 1995) by using operant conditioning with positive reinforcement. Budgerigars can be readily trained to produce specific vocalizations for food and to modify both the spectrotemporal pattern (Manabe and Dooling, 1997) and the intensity (Manabe *et al.*, 1998) of their calls by selective reinforcement. Moreover, they can be trained to do this without first being presented with an acoustic reference of the modified call they eventually come to produce. This offers an opportunity to address questions of call learning in budgerigars in a very different and much more controlled context than afforded by normal social interactive situations involving a number of birds.

Here, we train budgerigars to produce specific calls for food reinforcement (i.e., calls that match an acoustic refer-

[a]Now at Nihon University, Tokyo, Japan.
[b]Author to whom correspondence should be addressed. Electronic mail: dooling@psyc.umd.edu

ence or template) and then assess the effect of allowing the bird to hear the call (i.e., the template) it is supposed to mimic immediately prior to producing it. We also track the effects of changes in the template on call production, and we examine the effect of operant training of call production on calls the birds produce in more natural, unrestricted social situations that evoke calling behavior.

II. GENERAL METHODS

A. Subjects

The subjects in these experiments were adult budgerigars obtained from a local pet supplier and maintained in an aviary at University of Maryland. Each bird was caged separately and had free access to water and grit in their home cages. During these experiments, food was used to reinforce the vocal behavior; therefore, the birds were maintained at 90% of their free-feeding body weights. The Animal Care and Use Committee at the University of Maryland approved all animal use.

B. Apparatus

Birds were trained in a small experimental chamber (14 cm wide × 12 cm high × 17 cm deep) constructed of wire mesh and mounted in a small sound isolation chamber (Industrial Acoustic Company model AC1). A light emitting diode (LED) was affixed to each corner of a 3 cm × 3 cm square piece of sound attenuating foam. This foam was in turn mounted on the side of the cage at the level of the bird's head. An Electret condenser microphone (Sony ECM-77B), set in the middle of the foam square, detected the bird's calls. A food hopper containing hulled millet was mounted on the floor at a depth of 3 cm from the front mesh.

C. Procedure and analysis

1. Recognition of calls

The output of the Sony microphone was sent to a digital signal processing board (National Instruments AT-DSP2200). Analog data were low-pass filtered at 80 kHz before being digitized with a 16 bit analog-to-digital (A/D) converter. The digital signal was low-pass filtered at 12 kHz, using a digital filter (IIR) to reduce the frequencies in the signal and eliminate aliasing before digitizing. A/D conversion at a rate of 24 kHz was initiated when sound intensity exceeded a preset value and continued for 266 ms.

Serial power spectra were calculated using the fast Fourier transform (FFT). The FFT was performed on approximately the first 10 ms of the call while the second 10 ms of data were being sampled to RAM, and so on. For each spectrum in the series (i.e., 10 ms sample), the intensities were normalized with the peak intensity set to one (Manabe *et al.*, 1997). In all, 25 successive serial power spectra were calculated over the entire sample.

The incoming signal was classified as a contact call if the following conditions were met: (1) The duration of the entire signal was greater than 96 ms, (2) 70% of the distribution of frequencies from about 188 to 10031 Hz were between 938 and 6844 Hz, and (3) there were no call components shorter than 21 ms in the first 75 ms of the call. All signals that were classified as contact calls were stored on disk for later analysis. Call intensity (peak amplitude) was calculated on the stored digital data. All programs for experimental control and data collection were written in Microsoft C.

2. Shaping of calling behavior

The birds were habituated to the experimental chamber and then trained to eat millet from the food hopper. Once the birds reliably ate from the hopper when it was raised, manual shaping of vocalizations began. In this phase of training, typical aviary sounds were played in the test chamber to induce the birds to produce a contact call. Whenever the birds called back to the aviary tape with a contact call, the experimenter activated the hopper. When birds began to produce contact calls reliably in the absence of playback calls, the calls were reinforced automatically.

3. Generation of a call template

After several sessions in which all sounds produced by the birds met the criterion of a contact call by the call recognition software, a typical contact call produced by the bird was selected as a "template" call (see, Manabe and Dooling, 1997). The bird's call was compared to the template, with the similarity index defined as the sum of the overlapping areas of each of the 25 serial power spectra from the two calls. This index is zero if there are no overlapping areas and one if the calls were spectrally identical. In actuality, a call was compared with slight temporal variations of the onset of the other calls (e.g., five different temporal offsets of about +10, +5, 0, −5, and −10 ms to minimize the effect of slight variations in onset or intensity levels). The highest similarity index of the five temporal offsets was taken as the similarity index between the two calls. A matrix of similarity values was constructed from all calls produced in a test session. This matrix was analyzed using multidimensional scaling algorithms (MDS, Systat) and plotted in two-dimensional space. The call in the center of the largest cluster in this two-dimensional space was selected as the template call for the next phase of training.

4. Template call training

Using this template call as a model for what the bird should mimic, the next phase of training consisted of reinforcing only those calls the bird produced that were similar to the template call (i.e., above a criterion similarity index). In this phase, activating the LEDs signaled the onset of a trial and each trial was separated by a 2 s intertrial interval. Vocalizations occurring during the intertrial interval were not reinforced and delayed the onset of the next trial by 1 s. For each call produced during a trial, a similarity index was calculated in real time as described earlier with the criterion initially set very low (e.g., 0.01) so that even calls barely similar to the template were reinforced. As the bird's performance improved (i.e., more calls meeting criterion), the criterion was gradually increased to a maximum value of about 0.5 depending on the experimental conditions. All test ses-

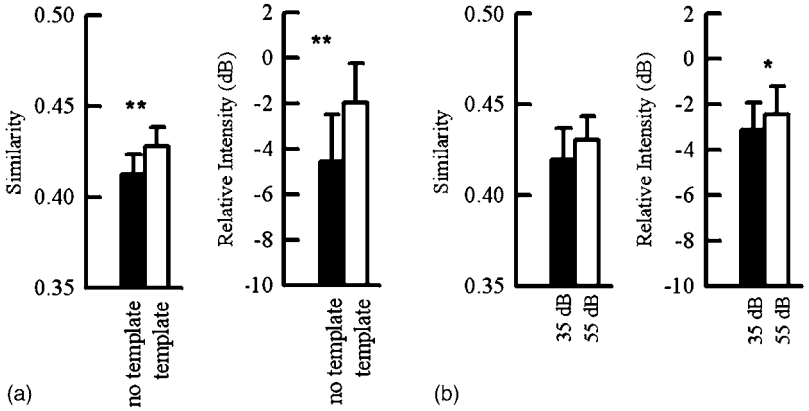


FIG. 1. (a) Mean and standard deviation of call similarity and relative intensity of contact calls from three birds. Closed bar indicates no template sound (no temp) and open bar indicates audible template (temp) presented at 55 dB. Double asterisks represents significance level of 0.001. (b) Mean and standard deviation of call similarity and relative intensity of contact calls from three birds. Closed bar indicates audible template presented at 35 dB and open bar indicates audible template presented at 55 dB. Single asterisk represents significance level of 0.05.

sions were concluded after delivery of a specific number of reinforcements (usually 48) or 20 min, whichever came first. Subjects were tested in two daily sessions, six days per week. Sessions were always separated by at least 4 h.

III. EXPERIMENT 1: EFFECTS OF TEMPLATE SOUND PRESENTATION ON THE INTENSITY AND ACCURACY OF CALL PRODUCTION

The first experiment examined the effects of presenting an acoustic reference or template (an external acoustic model of the call to be matched) to the bird and consisted of two parts. We sought to determine: (1) the effect of the presence or absence of acoustic reference on the precision of vocal accuracy and (2) the effect of the intensity level of acoustic reference on the vocal output of the test subject. We hypothesized that an acoustic reference would increase the accuracy of call production in an operant situation and that the intensity of the vocalization would depend on the intensity of the acoustic reference presented.

A. Methods

Subjects. Three male budgerigars were used for this experiment.

B. Procedure

Once the birds were trained to asymptotic levels of performance on the template-matching paradigm described earlier, a new training phase was introduced in which the template call was played to the bird before each trial. The sound pressure level of the template sound was measured with a General Radio Model 1982 Sound Level Meter and 3 m extension cable with a $\frac{1}{2}$ in. microphone placed in the position of the bird's head in the test apparatus. For the first part of the experiment, the sound pressure of the template call was set to a peak of 55 dB SPL (fast rms) at the bird's head. There were two types of trials: a nonsound template trial and a sound template trial. A total of 48 trials, 24 trials of each type, were presented randomly during a single test session. In all, each bird ran between 20 and 49 sessions for this experiment.

In the second part of the experiment, the template call was played to the bird on each trial but at one of two different intensities. The template was presented at a peak sound pressure level of either 55 dB SPL or 35 dB SPL rms at the bird's head just before the trial onset. The two different trial types occurred in a random order within a session and a total of 24 trials of each type occurred.

C. Results and discussion of experiment 1

Figure 1(a) shows the mean similarity index and the mean call intensity of three birds for each of the two trial types. The mean similarity index between all calls produced by the birds and the stored template call was significantly greater when the template was audible than in trials when the template was not presented to the bird [one-way repeated measures (RM) ANOVA; $F(2,1)=552.3$, $p<0.002$]. The mean intensity of the calls produced by the three birds in trials with an audible template was also significantly greater than in trials when the template was not presented [one-way RM ANOVA; $F(2,1)=130.6$, $p<0.008$]. This experiment shows that presenting birds with an acoustic reference results in small but significant increases in both the precision of call matching and the intensity of call production.

Figure 1(b) shows the mean similarity index and the mean call intensity of three birds. The mean similarity index between the birds' calls and the template call in high intensity trials was not significantly different from those produced in low intensity trials [one-way RM ANOVA; $F(2,1)=5.67$, $p>0.1$]. On the other hand, the mean intensity of the birds' calls in high intensity trials was significantly greater than calls produced by the birds in the low intensity trials [one-way RM ANOVA; $F(2,1)=28.1$, $p<0.03$]. These results show that while the precision of call matching remains the same for both intensities of the acoustic reference, the intensity of call production is dependent on the intensity of the

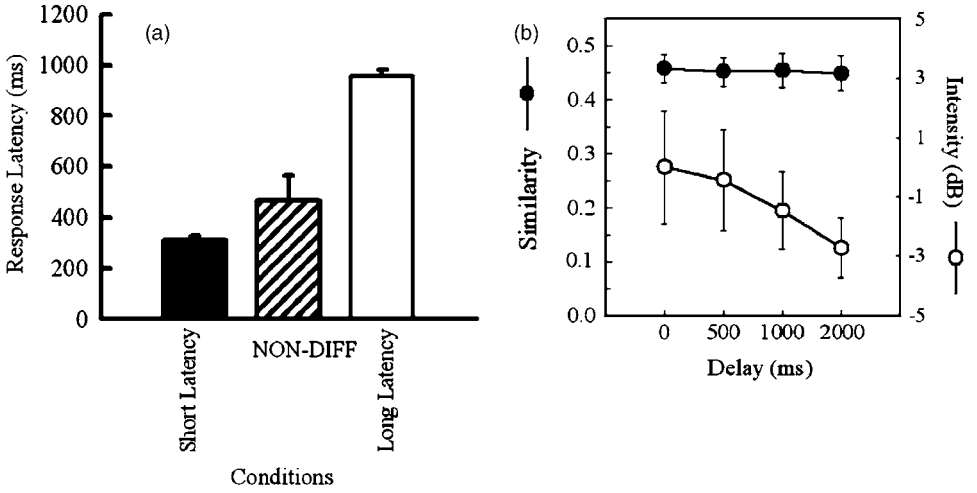


FIG. 2. (a) Mean and standard deviation of callback latency for three birds under three different conditions. Closed bar indicates the differential reinforcement of short latency, the striped bar indicates the nondifferential reinforcement condition, and the open bar indicates the differential reinforcement of long latency. (b) Mean and standard deviation of call similarity (closed circles) and intensity (open circles) across delay for three birds.

acoustic reference. Taken together, these results show that the bird's own call heard immediately prior to call production affects certain aspects of the quality of the bird's vocalization.

IV. EXPERIMENT 2: EFFECTS OF DELAY BETWEEN HEARING THE ACOUSTIC REFERENCE AND CALL PRODUCTION ON THE INTENSITY AND ACCURACY OF THE CALL

Experiment 1 showed that providing an acoustic reference before vocal production can increase both the precision of matching to the template and the intensity of vocal output. Casual observation of budgerigars in more "natural" situations, as when they are calling back and forth to one another in a large vivarium, suggests that they have control over a number of features of their calling behavior such as how quickly they respond to the call of another bird. In the following experiment, we examined whether the bird's latency to respond (vocally) to hearing a contact call was amenable to control through differential reinforcement. Moreover, if a bird's vocal response latency can be controlled by differential reinforcement, it provides an opportunity to examine other aspects of calling behavior, such as the role of memory, in guiding the precision of vocal production in matching the acoustic reference.

A. Method

Subjects. The same three birds from the previous experiments were used in this experiment.

B. Procedure

The latency with which the bird responded to the template was differentially reinforced. First, the latency of the bird's response to the template was measured in a normal test session (as described earlier). In this phase, there were no reinforcement contingencies on latency, and every call the bird produced was reinforced. Once performance reached asymptotic levels, calls were reinforced only when the latency to respond *was shorter* than the average delay value obtained in phase one. The value of the response delay criterion was gradually decreased from 1000 to 400 ms. Then, calls produced by the bird were reinforced only when the response latency *exceeded* a criterion value. Here, the criterion value was gradually increased from 400 to 700 ms.

In the second part of the experiment, a trial delay was imposed after the presentation of template sound and before the LED was lit. At first, calls matching the template were reinforced if the response occurred within 1000 ms after the LED was lit. Then, four delay intervals (0, 500, 1000, and 2000 ms) were imposed between the presentation of the template sound and the lighting of the LED. Calls produced during these delay intervals (i.e., prior to lighting the LED) were not reinforced. Furthermore, calls produced during the delay interval canceled the current trial and initiated a new trial sequence. The four delays were randomly mixed within a session, and the number of trials at each delay was set to 24. The birds were required to wait until the LED was lit to vocalize. The birds ran between 20 and 60 sessions during this experiment.

C. Results and discussion of experiment 2

Figure 2(a) shows mean response latency of three birds. These results show that differentially reinforcing response latencies was effective in both increasing and decreasing the birds' normal response latency. While the mean latency was well controlled by differential reinforcement [one-way RM ANOVA; $F(2,2)=59.9$, $p<0.002$], neither the mean similarity index [one-way RM ANOVA; $F(2,3)=1.27$, $p>0.36$] nor the mean call intensity [one-way RM ANOVA; $F(2,3)=2.54$, $p>0.10$] were significantly different across delay intervals [see Fig. 2(b)]. These results provide no evidence for a decay in auditory memory as a factor in the precision of vocal production in budgerigars. An obvious explanation for this result is that these birds, as with humans producing speech, had extensive experience with producing particular call types such that they were, in effect, "overtrained" on the

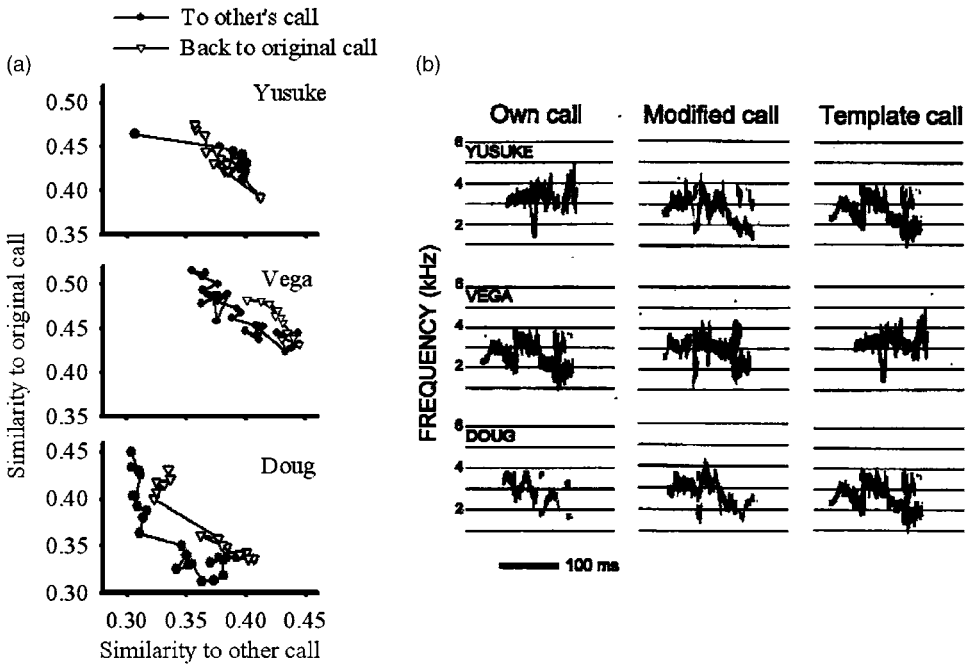


FIG. 3. (a) Mean similarities to original call and the template call for three birds. Closed circle indicates the training to make the template call. The open circle indicates training back to the original call. The line in each graph represents the average similarity between the original call from the previous training (i.e., no template sound) and the template call. (b) Sonograms of own call (original call), modified call, and template call from three birds.

motor patterns required to produce these calls. Thus, there was little lost, at least in terms of vocal precision and vocal level, from imposing a delay of up to 2 s in responding following presentation of an acoustic reference.

V. EXPERIMENT 3: EFFECTS OF CHANGING THE TEMPLATE AND ACOUSTIC REFERENCE

If providing birds with an acoustic reference or auditory template prior to vocalizing increases the precision of vocal matching, this might reflect a mechanism birds use in social situations to learn one another's contact calls. The present study examined this question by training birds to produce calls under operant control and then changed the reference sound from one of their own calls (experiment 1) to the call of another bird. The purpose of this experiment was to determine whether birds could learn new calls through a process of selective reinforcement of spectrotemporal changes in the acoustic reference that eventually converge on the contact call of another bird. In effect, this study simulates a situation that might exist when a bird is housed with a new cagemate and learns the new cagemate's contact call.

A. Method

Subjects. Two birds from the previous experiment (Yusuke and Vega) and one new bird (Doug) were used in this experiment.

B. Procedure

These birds were trained to produce a call to match an audible template call within 1000 ms after onset of LED. Once the birds' performance reached asymptotic levels, the stored template for a particular bird was changed to the stored template of one of the other two birds. In other words, the template for Yusuke was switched from his original template call to Vega's template call, the template for Vega to Yusuke's call, and the template for Doug to Vega's call, respectively. The similarity between Yusuke's original template and Vega's original template was 0.35, and the similarity between Doug's original template and Vega's original template was 0.31.

As described earlier, calls were reinforced only when the similarity between the call and the new template call met a predetermined criterion. Initially, the criterion was set low to around 0.30 and gradually increased up to a maximum of about 0.45 as the bird's performance improved. After the birds reached asymptotic levels of performance on this new template, the old template was reinstated and retraining began on the original template call. In this retraining phase, the initial criterion was set at 0.35 and again gradually increased to about 0.45 depending on the bird's performance.

C. Results and discussion of experiment 3

Figure 3(a) shows the mean similarity index between all calls produced in the two sessions and the two templates. The ordinate represents the similarity to the original call template while the abscissa represents the similarity to the new call template. Each data point is the mean similarity for all calls produced in a single daily session after the call template was switched. Changes in call similarity are significantly different across the testing sessions [ANOVA, $F(2,8)=28.14$, $p<0.05$]. In the first phase, changing the call template from "original" to "other" results in a significant decrease in similarity to the original template ($t=7.05$, $p<0.05$). Then, in a retraining phase, the template the bird had to match was switched from other back to original for all three birds.

Changing the call template back to the original template significantly decreases similarity of birds' calls to the other template ($t=-5.75$, $p<0.05$) and inceases similarity to the original template calls over a period of about 10 days.

Sonograms of each bird's original template, its new template, and the modified call resulting from changing the template are shown in Fig. 3(b). The modified call for each bird never matches the new template call perfectly but becomes more similar to the template call than to the bird's original call. Some features of the calls changed more than others. Vega's modified call, for instance, retained the constant frequency 2 kHz segment in the original call and became more like a combination of the bird's original call and the template call. This kind of recombination of segments is not unusual and may be one of several ways that budgerigars come to share contact calls in small social groups under more "natural" conditions (Farabaugh *et al.*, 1994; Hile *et al.*, 2000; Hile and Striedter, 2000).

VI. EXPERIMENT 4: CALLS IN AND OUT OF THE OPERANT CONTEXT

The previous experiments demonstrate that budgerigars can learn new contact calls under operant control using selective reinforcement and that hearing an acoustic reference prior to vocalizing affects the precision and level of vocal output. The strictly controlled operant test environment is highly unnatural so it is an open question whether vocal learning in this context is similar to what might occur in more natural social contexts. Experiment 4 examined two aspects of this problem. First, we sought to determine the extent to which calls learned in an operant context were produced in a more natural context (as when birds call back and forth to one another in their home cages). Second, we compared operant learning of contact calls in birds that had extensive learning experience in a natural context with birds that were housed in social and acoustic isolation. In all, we trained four budgerigars using the operant procedures described in the following. During the course of these experiments, the calls produced by the birds outside of the operant context were recorded on a weekly basis.

A. Method

1. Subjects

Four experimentally naive female budgerigars, all several years old, served as subjects. Two birds (There and Here) were caged separately in a large aviary that housed 50–100 other adult budgerigars. Two other birds (Bart and Maggie) were reared normally to the age of 10 months but then housed for over a year in a restricted social and acoustic environment where they could not interact visually or acoustically with other budgerigars.

2. Audio recording

During these experiments, each bird's natural vocal repertoire was recorded weekly in a seminatural situation. The two birds normally housed in the large aviary (There and Here) were moved to small, individually isolated, recording boxes ($21.5\times23\times23$ cm), which had three walls, lined with acoustic foam, a front wall of Plexiglas, and a top of wire mesh. These small boxes were housed within a larger sound isolation chamber (Industrial Acoustic Company model AC1). Calls were elicited from these birds by opening the doors to the sound isolation chambers slightly (so the birds could hear one another). The two birds (Maggie and Bart) maintained in acoustic and social isolation (i.e., already housed in similar IAC chambers) were recorded individually and stimulated with a tape of running water rather than other bird vocalizations.

In all cases, the recording boxes were fitted with an omnidirectional Realistic Electret Microphone (model 33-1063) connected to separate tracks of a four-channel Marantz tape recorder (PMD740), and 20–30 calls were recorded per bird per session. A Kay Elemetrics model 7800 Digital Kay Sonagraph was used to print spectrograms (300 Hz bandwidth; expanded mode). Contact calls are distinctive and highly stereotyped making it easy to visually categorize vocalizations into call type categories.

B. Procedure

The birds' initial contact call repertoire was sampled and classified into types (by EBP). Three of the four birds produced only one contact call type. The remaining bird (Bart) already produced several call types at the time these experiments were initiated. Bart was trained immediately on the template task described earlier in Sec. II.

The three birds that produced only one contact call were trained on a "1-back procedure" described previously (Manabe *et al.*, 1997; Manabe and Dooling, 1997). This procedure induces birds through selective reinforcement to produce more than one call type. Briefly, a call was reinforced only when it was different from the last reinforced call. Call difference, or dissimilarity, was quantified in terms of the sum of the overlapping area of 20 successive serial frequency distributions (Manabe *et al.*, 1997). Initially, all calls that varied a predetermined amount from the previous call were reinforced, even if the newly produced call differed only slightly from the last reinforced call. As the bird's performance improved, the criterion for qualifying as different became stricter. Once a bird consistently produced two, and sometimes more, acoustically distinct calls, one of the new call types was chosen to be the bird's new template call and the second phase of training began.

In this phase of the experiment, the bird was only rewarded in the operant chamber for producing the new template call. In other words, the bird's call production was only reinforced when the similarity index to the template was above a certain criterion. This criterion was increased gradually to a maximum of 0.45 as the bird's performance improved. In this experiment, the template call was not audible to the bird; in other words, there was no external acoustic reference for the bird's production. The point of this template training was to induce precise learning and reliable production of a new call type in the operant context.

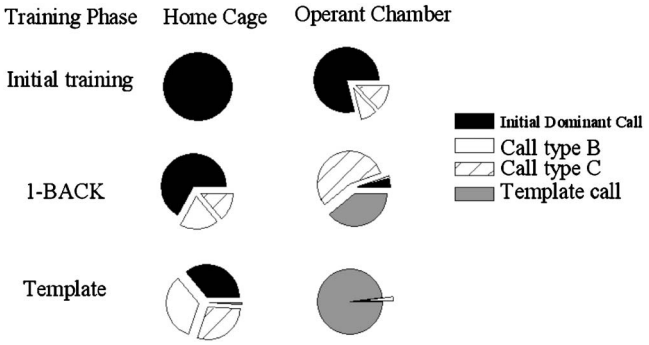


FIG. 4. Data for four birds showing percentages of calls produced (by type) in the home cage and the operant chamber. As template training concludes, the birds produce the template call the majority of the time within the operant chamber but rarely in the birds' natural repertoires.

C. Results and discussion of experiment 4

1. The effect of restricted social and acoustic conditions on learning

The two different birds, housed in small cages within sight and earshot of 50–100 birds (There and Here), readily learned both the 1-Back and template operant tasks. Surprisingly, two birds reared in restricted conditions (Bart and Maggie) failed to learn either experimental task. Specifically, these two birds did learn to associate calling with food reward but failed to reach criterion on any of the tasks. Attempts to train the birds in the operant task continued daily for approximately 8 months. The birds were then returned to a large aviary with other birds for a period of 6 months. Within a few weeks of interacting with the birds in this aviary, both birds (Bart and Maggie) produced at least two new call types in their home cage. These results show that these birds were perfectly capable of learning new call types in a more natural social situation, presumably via social reinforcement, but *not* under experimental operant conditions with food reinforcement.

2. The effect of normal social and acoustic conditions on learning

After a 6 month period in the aviary, Bart and Maggie were housed in the same room as There and Here. They were retrained in the Operant test chamber and now both learned to produce new call types in the Operant test chamber when trained on the 1-Back procedure.

All four birds learned to produce an acceptably precise match to the template call in the template task. Figure 4 summarizes the data for all four birds when housed in a normal acoustic environment. The majority of calls first produced by the birds in the Operant test chamber were the dominant contact calls (from their natural repertoire) produced in the home cage as seen in Fig. 4 (initial training). However, as the training continued, the calls in the two contexts began to differentiate until finally there was little to no overlap between calls produced in the natural and operant contexts (Fig. 4 1-Back). At the conclusion of template call training, the four birds rarely produced the template call outside of the operant context (Fig. 4, template). In other words, food reinforcement of a specific call type in the operant context did not increase its use in more natural communication contexts for these four birds. This is surprising given the demonstrable vocal plasticity of budgerigars throughout adulthood.

The overlap between call types produced in a natural calling situation and the Operant test chamber was characteristic of the earliest stages of training as shown in Fig. 4. As the criterion for reinforcement of a specific call type becomes more restrictive in the Operant test chamber, all other nonreinforced calls cease to be produced. By contrast, the number of call types produced outside of the Operant test chamber remained virtually unaffected during the last stages of this experiment (e.g., moving from the 1-Back procedure to the template training procedure). Taken together, the results from this experiment show that (1) social interactions with other birds can affect training in the Operant context and (2) training in the Operant context can influence the diversity of calls birds produce in a normal social context.

VII. GENERAL DISCUSSION

Budgerigars can be trained by food reward to modify the spectrotemporal qualities of their species-specific contact calls (Manabe *et al.*, 1995, 1997; Manabe and Dooling, 1997) and to control the intensity of their vocal output (Manabe *et al.*, 1998). The present experiments extend these findings in a number of important ways: (1) Hearing an external acoustic reference independently increases the accuracy and intensity of call matching, (2) delays of up to 2 s between template presentation and trial onset does not affect either vocal intensity or matching accuracy, (3) response (calling) latency is susceptible to modification by selective reinforcement, (4) production of other budgerigars' contact calls can be learned in an operant context through selective reinforcement, (5) calls learned and exclusively reinforced in an operant context are not used in more natural communicative contexts, and (6) birds kept in social and acoustic isolation do not readily learn new calls in an operant context.

One of the most productive areas of research in vocal learning in songbirds has been studies devoted to understanding what songs and calls birds can learn, and when, and from whom they learn. Compared to song learning, call learning in birds is much less well studied. In budgerigars, call learning, and subsequent production of contact calls, is not widely separated in time perhaps because budgerigars leave their parents and join juvenile flocks shortly after fledging (Wyndham, 1980). The development of shared vocalizations may be necessary, especially in such gregarious species, for social bonding (e.g., Farabaugh *et al.*, 1994). In most social situations, however, it is difficult to identify the key variables in vocal learning and study them separately. The present studies show that many aspects of calling behavior, including learning multiple contact call types, can be quite easily influenced through selective reinforcement procedures involving food. Presumably, in natural contexts, social reinforcement takes the place of food reinforcement. In addition, the fact that the four birds in experiment 4 came to use different contact calls in the operant environment com-

pared to their natural home environment also highlights another point—the contact call repertoire in these birds is context dependent.

The present results provide some insight into two general mechanisms by which budgerigars could learn new contact calls as adults. One, a "memory-based" process, might involve repeated presentation of the model to be learned, the subsequent formation of an auditory memory, followed by vocal-motor attempts to match that auditory memory. Another quite different mechanism, akin to "action-based" learning (e.g., Marler and Nelson, 1993), could involve selective reinforcement of call variations, through visual and social cues, which guide vocal productions so that they ultimately converge on the call type to be learned. For budgerigars, this mechanism has been suggested by previous work showing that social and visual interaction is key to contact call sharing in adult budgerigars and that acoustic interaction alone is not sufficient (Farabaugh *et al.*, 1994). The present experiments strengthen this hypothesis considerably by showing that budgerigars learn new calls in an operant context by selective reinforcement of call variation by food reinforcement eventually producing a contact call which matches another bird's contact call.

Finally, parrots almost certainly do not normally learn new vocalizations for food reinforcement in nature. Rather, they probably learn new vocalizations by mechanisms that involve various social reinforcers or other manipulations of their social environment (Brittan-Powell *et al.*, 1997; Farabaugh and Dooling, 1996; Farabaugh *et al.*, 1994; Rowley and Chapman, 1986; Treisman, 1978). But, the present experiments, by bringing a number of aspects of contact call learning under strict experimental control, do offer new opportunities to examine the biological bases of vocal learning, the operation of specific sensory and motor circuits underlying learning, and the critical role that auditory feedback plays in both the learning and maintenance of an adult vocal repertoire.

ACKNOWLEDGMENTS

We thank Peter Marvit and Michael Osmanski for comments on the manuscript Monica Burr for help running the experiments and Melonie Newman and Monica Murphy for animal care assistance. This work was supported by National Institutes of Health Grant Nos. DC00198 and DC1372 to R.J.D. and National Research Service Award No. MH10993 to E.F.B.-P.

Ali, N. J., Farabaugh, S., and Dooling, R. (**1993**). "Recognition of contact calls by the budgerigar (*Melopsittacus undulatus*)," Bull. Psychon. Soc. **31**, 468–470.

Brittan-Powell, E. F., Dooling, R. J., and Farabaugh, S. M. (**1997**). "Vocal development in budgerigars (*Melopsittacus undulatus*): Contact calls," J. Comp. Psychol. **111**, 226–241.

Brockway, B. F. (**1964a**). "Ethological studies of the budgerigar (*Melopsittacus undulatus*): Non-reproductive behavior," Behaviour **22**, 193–222.

Brockway, B. F. (**1964b**). "Ethological studies of the budgerigar (*Melopsittacus undulatus*): Reproductive behavior," Behaviour **23**, 294–324.

Brockway, B. F. (**1969**). "Roles of budgerigar vocalization in the integration of breeding behavior," in *Bird Vocalizations*, edited by R. A. Hinde (Cambridge University Press, London), pp. 131–158.

Brown, S. D., Dooling, R. J., and O'Grady, K. (**1988**). "Perceptual organization of acoustic stimuli by budgerigars (*Melopsittacus undulatus*). III. Contact calls," J. Comp. Psychol. **102**, 236–247.

Dooling, R. J. (**1986**). "Perception of vocal signals by the Budgerigars (*Melopsittacus undulatus*)," Exp. Biol. **45**, 195–218.

Farabaugh, S., and Dooling, R. J. (**1996**). "Acoustic communication in parrots: Laboratory and field studies of budgerigars, *Melopsittacus undulatus*," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University, Ithaca NY), pp. 97–118.

Farabaugh, S. M., Brown, E. D., and Dooling, R. J. (**1992**). "Analysis of warble song of the budgerigar, *Melopsittacus undulatus*," Bioacoustics **4**, 111–130.

Farabaugh, S. M., Linzenbold, A., and Dooling, R. J. (**1994**). "Vocal plasticity in budgerigars (*Melopsittacus undulatus*): Evidence for social factors in the learning of contact calls," J. Comp. Psychol. **108**, 81–92.

Hall, W. S., Cookson, K. K., Heaton, J. T., Roberts, T., Shea, S., and Brauth, S. E. (**1997**). "Audio-vocal learning in budgerigars," Ann. N.Y. Acad. Sci. **807**, 352–367.

Heaton, J. T., and Brauth, S. E. (**1999**). "Effects of deafening on the development of nestling and juvenile vocalizations in budgerigars (*Melopsittacus undulatus*)," J. Comp. Psychol. **113**, 314–320.

Heaton, J. T., Dooling, R. J., and Farabaugh, S. M. (**1999**). "Effects of deafening on the calls and warble song of adult budgerigars (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **105**, 2010–2019.

Hile, A. G., Plummer, T. K., and Striedter, G. F. (**2000**). "Male vocal imitation produces call convergence during pair bonding in budgerigars, *Melopsittacus undulatus*," Anim. Behav. **59**, 1209–1218.

Hile, A. G., and Striedter, G. F. (**2000**). "Call convergence within groups of female budgerigars (*Melopsittacus undulatus*)," Ethology **106**, 1105–1114.

Hile, A. G., Tyler Burley, N., Coopersmith, C. B., Foster, V. S., and Striedter, G. F. (**2005**). "Effects of male vocal learning on female behavior in the budgerigar, *Melopsittacus undulatus*," Ethology **111**, 901–923.

Kroodsma, D. E., and Miller, E. H. (**1982**). *Acoustic Communication in Birds* (Academic, New York).

Kroodsma, D. E., and Miller, E. H., eds., (**1996**). *Ecology and Evolution of Acoustic Communication in Birds* (Academic, New York).

Mammen, D. L., and Nowicki, S. (**1981**). "Individual-differences and within-flock convergence in chickadee calls," Behav. Ecol. Sociobiol. **9**, 179–186.

Manabe, K., Cleaveland, M. J., and Staddon, J. E. R. (**1997**). "Control of vocal repertoire by reward in budgerigars (*Melopsittacus undulatus*)," J. Comp. Psychol. **111**, 50–62.

Manabe, K., and Dooling, R. J. (**1997**). "Control of vocal production in budgerigars (*Melopsittacus undulatus*): Selective reinforcement, call differentiation, and stimulus control," Behav. Processes **41**, 117–132.

Manabe, K., Kawashima, T., and Staddon, J. E. R. (**1995**). "Differential vocalization in Budgerigars: Towards an experimental analysis of naming," J. Exp. Anal Behav. **63**, 111–126.

Manabe, K., Sadr, E. I., and Dooling, R. J. (**1998**). "Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): Differential reinforcement of vocal intensity and the Lombard effect," J. Acoust. Soc. Am. **103**, 1190–1198.

Marler, P., and Nelson, D. A. (**1993**). "Action-based learning—A new form of developmental plasticity in bird song," Neth. J. Zool. **43**, 91–103.

Mundinger, P. (**1970**). "Vocal imitation and individual recognition of finch calls," Science **168**, 480–482.

Rowley, I., and Chapman, G. (**1986**). "Cross-fostering, imprinting and learning in two sympatric species of cockatoo," Behaviour **96**, 1–16.

Striedter, G. F., Freibott, L., Hile, A. G., and Tyler Burley, N. (**2003**). "For whom the male calls: An effect of audience on contact call rate and repertoire in budgerigars, *Melopsittacus undulatus*," Anim. Behav. **65**, 875–882.

Treisman, M. (**1978**). "Bird song dialects, repertoire size, and kin association," Anim. Behav. **26**, 814–817.

Trillmich, F. (**1976a**). "Learning experiments on individual recognition in budgerigars (*Melopsittacus undulatus*)," Z. Tierpsychol **41**, 372–395.

Trillmich, F. (**1976b**). "Recognition of individual nesting box in budgerigars, Melopsittacus-undulatus shaw (Aves, Psittacidae)," Z. Tierpsychol **42**, 1–11.

Wyndham, E. (**1980**). "Diurnal cycle, behavior and social organization in the budgerigar (*Melopsittacus undulatus*)," Emu. **80**, 25–33.

Vocalizations produced by humpback whale (*Megaptera novaeangliae*) calves recorded in Hawaii

Ann M. Zoidis[a)]
Cetos Research Organization, 11 Des Isle Avenue, Bar Harbor, Maine 04609

Mari A. Smultea[b)]
Smultea Environmental Sciences LLC, 29333 SE 64th Street, Issaquah, Washington 98027 and Cetos Research Organization, 29333 SE 64th St., Issaquah, Washington 98027

Adam S. Frankel[c)]
Hawaii Marine Mammal Consortium, P.O. Box 6107, Kamuela, Hawaii, 96738-6107

Julia L. Hopkins,[d)] Andy Day,[e)] and A. Sasha McFarland[f)]
Cetos Research Organization, 11 Des Isle Ave, Bar Harbor Maine 04609

Amy D. Whitt[g)] and Dagmar Fertl[h)]
Geo-Marine, Inc., 2201 Avenue K, Suite A2, Plano, Texas 75074

(Received 17 July 2007; revised 12 December 2007; accepted 27 December 2007)

Although humpback whale (*Megaptera novaeangliae*) calves are reported to vocalize, this has not been measurably verified. During March 2006, an underwater video camera and two-element hydrophone array were used to record nonsong vocalizations from a mother–calf escort off Hawaii. Acoustic data were analyzed; measured time delays between hydrophones provided bearings to 21 distinct vocalizations produced by the male calf. Signals were pulsed (71%), frequency modulated (19%), or amplitude modulated (10%). They were of simple structure, low frequency (mean =220 Hz), brief duration (mean=170 ms), and relatively narrow bandwidth (mean=2 kHz). The calf produced three series of "grunts" when approaching the diver. During winters of the years 2001–2005 in Hawaii, nonsong vocalizations were recorded in 109 (65%) of 169 groups with a calf using an underwater video and single (omnidirectional) hydrophone. Nonsong vocalizations were most common (34 of 39) in lone mother–calf pairs. A subsample from this dataset of 60 signals assessed to be vocalizations provided strong evidence that 10 male and 18 female calves vocalized based on statistical similarity to the 21 verified calf signals, proximity to an isolated calf (27 of 28 calves), strong signal-to-noise ratio, and/or bubble emissions coincident to sound.
 [DOI: 10.1121/1.2836750]

PACS number(s): 43.80.Ka [WWA] Pages: 1737–1746

I. INTRODUCTION

Humpback whales (*Megaptera novaeangliae*) produce song, nonsong social vocalizations, and nonvocal, surface-generated percussive sounds (e.g., caused by breaches, fluke slaps, pectoral fin slaps, etc.) during migration (e.g., Kibblewhite *et al.*, 1966; Payne and McVay, 1971; Norris *et al.*, 1999; Dunlop *et al.*, 2005, 2006; Dunlop and Noad, 2007), on low-latitude winter breeding/calving grounds (e.g., Payne and McVay, 1971; Winn and Winn, 1978; Tyack, 1981, 1982; Darling, 1983; Silber, 1986; Payne, 1983; Helweg *et al.*, 1992; Au *et al.*, 2000), and on higher-latitude summer feeding grounds (e.g., Thompson *et al.*, 1977; Jurasz and Jurasz, 1979; D'Vincent *et al.*, 1985; Sharpe *et al.*, 1998). Songs are well-documented vocalizations produced by males predominantly on the wintering grounds where they appear to be part of breeding displays (e.g., Payne and McVay, 1971; Winn *et al.*, 1973; Tyack, 1981; Darling, 1983; Helweg *et al.*, 1992; Au *et al.*, 2000; Darling and Bérubé, 2001; Darling *et al.*, 2006). Singers are typically alone but may also accompany other whales including a mother and calf (e.g., Tyack, 1981; Darling, 1983; Glockner and Venus, 1983; Darling and Bérubé, 2001). Songs are characterized by continuous, repetitive, highly structured phrases and themes that contain units with harmonics above 15 kilohertz (kHz) and sounds above 24 kHz (Payne and McVay, 1971; Darling, 1983; Payne, 1983; Helweg *et al.*, 1992; Au *et al.*, 2000, 2001, 2005, 2006; Fristrup *et al.*, 2003; Potter *et al.*, 2003).

On the wintering grounds, nonsong social vocalizations have been documented among adults (Tyack, 1981; Silber, 1986). In contrast to song, adult nonsong vocalizations (i.e., "social sounds") are produced erratically (variable through time, often interrupted by silent periods), are unpredictable, do not contain the rhythmic, consistent, and continuous pat-

[a)]Electronic mail: ann@cetosresearch.org
[b)]Electronic mail: msmultea@msn.com
[c)]Electronic mail: adam.frankel@marineacoustics.com
[d)]Electronic mail: julie@cetosresearch.org
[e)]Electronic mail: andyday69@hotmail.com
[f)]Electronic mail: sasha@cetosresearch.org
[g)]Electronic mail: awhitt@geo-marine.com
[h)]Current address: Ziphius EcoServices, 8112 Springmoss Drive, Plano, Texas 75025. Electronic mail: dfertl@gmail.com

terning of song, and are known to range from 50 hertz (Hz) to over 10 kHz, with dominant frequencies below 3 kHz (Tyack, 1981; Silber, 1986). Early publications on nonsong vocalizations in surface-active groups of wintering humpbacks coined the term social sounds to describe that specific behavioral context (Payne, 1978; Tyack, 1982, 1983; Tyack and Whitehead, 1983; Silber, 1986). These social sounds have been reported to occur predominantly in groups of three or more adults composed of a mature female accompanied by a male "principal escort" and other males aggressively vying for a position near her (Tyack and Whitehead, 1983; Silber, 1986; Clapham *et al.*, 1992). Common behaviors include chasing, charging, inflated head lunges, underwater blows, aggressive contact, and fast surface traveling (e.g., Darling, 1983; Tyack and Whitehead, 1983; Baker and Herman, 1984). Although the full function of social sounds in these surface-active groups remains uncertain, they appear to invoke approach or avoidance responses from other whales, depending on sex and composition (Tyack, 1983). They also seem to signal aggression or agitation among males fighting for dominance or proximity to a female (Silber, 1986).

Despite numerous underwater and/or acoustic studies of humpback whale groups containing a mother–calf pair in Hawaiian waters (e.g., Glockner and Venus, 1983; Tyack, 1983; Silber, 1986; Glockner-Ferrari and Ferrari, 1990; Zoidis and Green, 2001; Cartwright, 2005; Zoidis *et al.*, 2005), only recently have vocalizations been attributed to individual calves based on observational evidence and underwater videoacoustic data recorded with a single omnidirectional hydrophone (Zoidis and Green, 2001; Pack *et al.*, 2005; Zoidis *et al.*, 2005). Reported evidence includes vocalizations recorded in closest proximity to a calf with no other whales nearby, strong signal-to-noise ratio, and/or vocalizations produced coincident to underwater bubble production by the calf. Although this is strong evidence that humpback calves vocalize, it cannot be ruled out that the sounds did not come from another source/whale that was nearby or not visible (i.e., the omnidirectional hydrophone used in those studies could not provide measurable evidence, such as directional bearings, linking acoustic data to a source). Published documentation that individual calves of other baleen whale species vocalize is limited to one calf from each of three different species (the first two of which were captive animals): gray (*Eschrichtius robustus*) (Wisdom *et al.*, 2001), Bryde's (*Balaenoptera edeni*) (Edds *et al.*, 1993), and North Atlantic right whales (*Eubalaena glacialis*) (Parks and Tyack, 2005; Parks and Clark, 2007).

Dunlop *et al.* (2005, 2006) and Dunlop and Noad (2007) remotely recorded nonsong vocalizations among migrating humpback mother–calf pairs and other groups with a calf off eastern Australia using an array of five separately anchored hydrophone buoys; however, it was not possible to differentiate which individual vocalized. Dunlop and Noad (2007) collectively termed these and other nonsong vocalizations recorded from any migrating humpback as "nonsong social vocalizations," although songs were also recorded. Vocalizations were further defined as being internally produced by a whale, including sounds generated from the blowhole. For the purpose of this paper and terminology consistency, we use the term "nonsong vocalizations" for sounds attributed to humpback calves that did not meet the aforementioned definition of song and were not percussively generated by body contact with the water surface.

In this paper, directional acoustic data from a two-element hydrophone array combined with underwater video empirically confirm that a male humpback whale calf produced nonsong vocalizations. Additional data indicate that both male and female calves vocalize. Descriptions of these vocalizations are presented and compared to earlier reports and vocalizations from calves of other large whale species. Contexts of some calf vocalizations are also described.

II. METHODS

Underwater humpback vocalizations and behavior were recorded during the winters of 2001–2006 off southwest Maui and Kauai, Hawaii. Data were collected by one or two snorkelers off a 6–8 meter (m) inboard or outboard vessel and by personnel above water. Divers used two systems: a Panasonic GS200 and/or GS300 digital underwater video camera equipped with a single omnidirectional hydrophone (Biomon BM 8263-3c mounted 1 m below the camera, hydrophone sensitivity to 80 kHz), and a two-element hydrophone array used one day in March 2006. The two-element array consisted of two HTI MIN-96 hydrophones (frequency response 2 Hz to 30 kHz) mounted 1.5 m apart on a bar perpendicular to the optical axis of the camera. This configuration was used to determine left–right bearings to the source of sounds produced in mother–calf groups (i.e., the offset of a sound source from the plane between the two hydrophones was measured by the time delay of the signal arriving at the two hydrophones). The unit was held parallel to the water's surface. A zero offset would indicate that the sound source was equidistant from the two hydrophones and located on a vertical plane parallel to the optical axis of the camera. Verification that sound(s) were produced by a calf for the two-element array was defined as repeated measurements of signals within a 10° angle of the plane bisecting the two hydrophones, with the calf centered in the video frame and no other whales present in that plane. Frequency response curves were obtained from the manufacturer specifications for the hydrophones and recording equipment. Hydrophones were calibrated by the manufacturer.

To analyze acoustic data, the recorded videotapes were connected to an Apple Mac Mini and raw video data were transferred to hard disk using the program IMOVIE HD. Recordings were divided into a series of video clip files. The digital video files were aurally and visually reviewed for potential acoustic signals. Visual inspection of spectrograms was useful, particularly if signals were low frequency or pulsed. All digital video clips with audible pulses were exported as audio interchange file format files. Clips downloaded from mini digital video tapes were processed using audio editing software. The sample rate of acoustic recordings was 32 or 48 kHz. Single-element recordings were downsampled by a factor of 3 to allow more detailed visualization of relatively low-frequency sounds, although the

original sampling rate data were used to determine the time delay differences.

Audio data were reviewed in RAVEN 1.2.1 (Charif *et al.*, 2004) in 30 s segments. Individual analog/digital samples were plotted and reviewed as a function of time. Detected pulses were copied and pasted into two different files to allow separate analyses of the same data. In the first file, data were low-pass filtered and decimated to allow higher resolution of low-frequency components. These data were used to create spectrograms to provide descriptive measurements of sounds. Spectrograms were calculated with a 1024 fast Fourier transform (FFT), 75% overlap, and a Hamming window. In the second file, the same original sounds recorded with the two-element array were band-pass filtered between 100 and 3000 Hz to allow more accurate measurement of time delays of each sound between the two left and right channels. Time delays were determined by hand-measuring the temporal differences between a wave form's arrival time at each sensor in milliseconds (ms). Time delays and angles were converted to bearings relative to the sound source assuming far-field propagation. Resultant bearings indicated the angular position of the sound source relative to a plane parallel to the camera's optical axis and normal to the line between the two hydrophones. These bearings are actually cones (or hyperbolic surfaces) relative to the plane parallel to the optical axis of the camera. Video from both divers was then examined to determine the position of the animal(s) in the frame at the time of sound production. The relationship between the time delay and the angle from the plane parallel to the optical axis of the camera to the sound source consists of the simple linear formula:

$$\text{angle} = \arcsin[(\Delta t \times c)/A], \tag{1}$$

where Δt=time difference between elements (s), c=speed of sound in water (m/s), and A=aperture (distance) between elements (m). Any single-measured angle could represent a source in any direction on the hyperbolic surface defined by the array geometry and the time delay differences (not necessarily the object in view of the camera, i.e., the sound source could be directly behind the camera). This was addressed by conducting both underwater and surface constant scans for animal presence and locations in relation to the recording diver. We inferred that the source animal was the one in front of the diver if no animal occurred behind the diver and no animals were visually detected on the same plane as the source animal. In addition, during calf vocalizations, no other groups were visible underwater or at the surface within 300 m.

Representative spectrograms were selected from clips to illustrate typical acoustic parameters attributable to calf vocalizations (Fig. 1). Clips were downsampled by a factor of 8, reducing the sample rate to 6 kHz. Spectrogram illustrations were made with a Hann window, a 512 FFT size, and a 98% overlap. A subsample of vocalizations with loud signal-to-noise ratios (10 dB, indicating close proximity), and data supporting the proximal position of the calf relative to other whales was selected for further analyses. Sounds with low signal-to-noise ratios, repetitive sounds similar to song, and/or sounds with more than one animal in the same plane

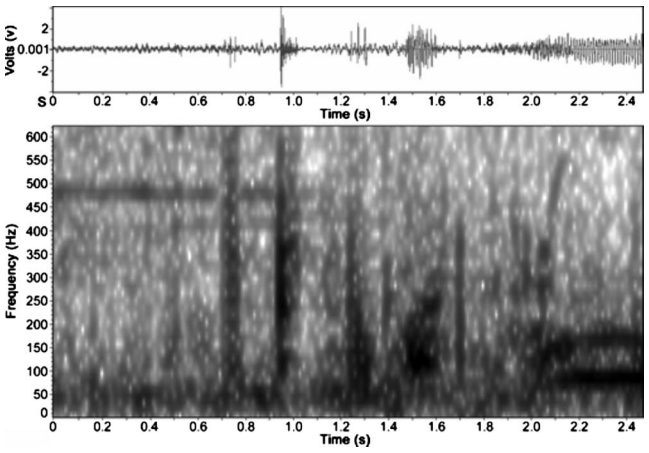


FIG. 1. Spectrograms of typical signal types attributed to a humpback whale calf based on sounds recorded with the two-element array. Five pulses are seen between 0.7 and 1.7 s. These are followed by a long tonal [or slight frequency-modulated (FM)] signal starting at 2.1 s. Above each spectrogram is a representation of the signal wave form (i.e., the recorded time series of voltages from the hydrophone shown on the *y* axis; time on the *x* axis).

based on video and field notes were discarded. Selected sounds were divided into pulsed, amplitude-modulated (AM), and frequency-modulated tonal (FM) signals. Pulsed signals had a distinct amplitude onset as opposed to the sinusoidal amplitude envelope of a typical AM signal. Duration and bandwidth, initial and ending AM and FM frequencies, number of pulses in a pulsed signal, and the first and last interpulse intervals were measured. It was noted if the interval was constant, increasing, or decreasing with time. It was not possible to calculate amplitude of sound from the source animal because we did not have the precise distance measurements necessary to estimate transmission loss with any degree of certainty.

Statistical analyses were conducted to compare vocalizations produced by the one calf recorded with the two-element hydrophone versus the selected subsample of 60 calf-attributed vocalizations recorded with the single hydrophone. Three measures of three signal types (AM, FM, and pulsed) were considered: lowest frequency, bandwidth, and duration. These measures were chosen as they are independent of each other. A mean value for each whale, measurement type, and signal type was calculated. Means were used to avoid unequal representation of one animal compared to another, as the number of vocalizations produced by each animal varied between focal sessions. Means were compared using a multivariate analysis of variance (MANOVA) , with the three dependent measures compared for the two-element hydrophone array recordings of one calf in 2006 against the single-element hydrophone recordings of 21 different calves in 2001–2005.

Underwater behavior from the recorded video of all whale groups was analyzed using a modified behavioral sampling methodology based on Mann (1999). Behavioral state, individual behaviors, and vocalizations were noted for 30 s periods that the calf was in view for at least 15 s. Distance between mother, calf, and diver and depth of animals below the water surface were estimated in mother-whale body lengths similar to Glockner and Venus (1983), assuming an

average body length of approximately 13 m for a Hawaiian mother humpback per Spitz *et al.* (2000). For example, if a calf was separated from its mother by two mother-body lengths this separation distance was estimated to be approximately 26 m; the margin of error was not measured but was believed to be approximately ±2 m based on the standard deviation and range of measured body lengths of 26 mother whales per Spitz *et al.* (2000). The mother was identified based on its closer consistent contact with the calf and genital morphology, the latter which was also used to determine sex of calves (True, 1904; Glockner, 1983).

III. RESULTS

Underwater video recordings of 170 groups containing a mother-calf humpback were filmed off Maui and Kauai from January to March for 1–3 months of each of the following years: 2001, 2002, 2004, 2005, and 2006. A total of 1007 min of videotape was recorded (mean=5.0 min per group, SD=4.9, range=<1–30 min). Nonsong vocalizations were recorded in 110 (65%) of the 170 groups with a calf. Of the total 170 calf pods, nonsong vocalizations were detected in 87% or 34 of 39 lone mother–calf pairs and 68% or 76 of 111 mother–calf escorts; no calf nonsong vocalizations were detected among the 20 mother-calf pairs with two or more escorts. Many vocalizations were of relatively low amplitude and were not detected until later aural and spectral analyses of video and acoustic data.

The two-element array system was used off Maui for one day on 9 March 2006 for a total of 51 min on three separate mother–calf groups. Quality acoustic data were obtained for one group, a male calf accompanied by a mother and escort; they were observed and recorded for 38 continuous min of which 18 min was recorded with the two-element array set up simultaneous with a second underwater video camera and single hydrophone. The remaining 109 (95%) calf groups with nonsong vocalizations were recorded using the single hydrophone-video camera set up. Nearly all (97%) of these 109 groups and videotape (96%) were recorded near Maui; the remaining 3% were near Kauai. Acoustic data and videotape were analyzed for the one calf recorded with the two-element array and for a selected subsample of 28 calves from 109 groups as described here. All 29 calves consisted of different individuals based on examination of physical attributes of both the calf and mother [e.g., fluke, pectoral fin, and body pigment patterns, body creases, sex, and/or scarring (e.g., Katona *et al.*, 1979; Glockner and Venus, 1983)]. The one vocalizing calf recorded by the two-element array was male, whereas the 28 calves recorded with the single hydrophone consisted of 10 males and 18 females. All 29 whales assessed to be mothers based on close association with the aforementioned calves were sexed via underwater video data as females.

A. Two-element array recording of one calf

Directional bearings were determined for 21 distinct biological sound signals from the one male calf based on measured time delays of recordings between the two hydrophones of the two-element array (Table I). We analyzed nine different representative video clips (separate time segments from the 18 min data series) of this encounter providing real-time visual corroboration that the calf was the only animal present in the area of the recorded sounds. The calf is clearly visible in the center of all these frames. For example, a signal identified as a grunt occurred 160.8 s into clip 15 (Fig. 2). The measured time delay between the left and right hydrophones was 0.104 ms, corresponding to an angle of −6.14°, just slightly offset from the plane between the two hydrophones. No other animals were behind the calf being videoed or behind the first diver, verifying that the calf in front of the hydrophone made the sounds. A different animal behind the first diver is the only other scenario in which this time delay difference and resulting acoustic signals could possibly have been recorded. This was ruled out based on analysis of underwater video taken by the second diver, by constant vessel observer scans confirming that no other animals were nearby,

TABLE I. Distribution of signal types recorded from the one humpback whale calf with the two-element array and attributed to 28 humpback whale calves recorded with the single hydrophone. No significant differences were found between sounds recorded with the two-element array vs the single hydrophone for the three signal types based on lowest frequency, bandwidth, and duration parameters as indicated here. AM=amplitude modulated, and FM=frequency modulated.

Signal type	Two-element array (2006) (n=1)	Single hydrophone (2001–2006) (n=28)[a]	Total n (%)	MANOVA statistics
AM	2	12	14 (17)	$F(2,1)=0.1285$, $p=0.8919$
FM	4	30	34 (42)	$F(2,4)=0.2483$, $p=0.7914$
Pulsed	15	18	33 (41)	$F(2,3)=1.7988$, $P=0.3066$
Total number of signals	n=21	n=60	n=81	

[a]Excludes duplicative sounds recorded by a single hydrophone simultaneous to the two-element array for the one calf recorded with the two-element array in 2006.

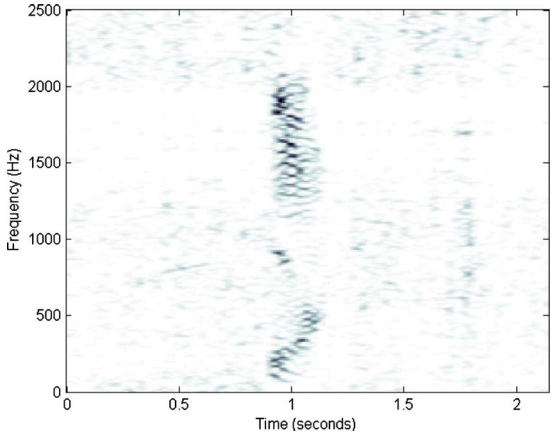


FIG. 2. (Color online) Spectrogram of an amplitude-modulated "grunt" signal produced by the humpback whale calf recorded by the two-element array.

TABLE II. Summary statistics for the amplitude-modulated (AM), frequency-modulated (FM), and pulsed signals for a male humpback whale calf recorded with the two-element (*T*), hydrophone array ($n=21$ signals), and 60 signals attributed to 28 humpback calves recorded with the single hydrophone (*S*). The mean value is provided for each measure with the standard deviation in parentheses. AM=amplitude modulated, FM=frequency modulated, Hz=hertz, and ms=milliseconds.

Signal type	Lowest frequency (Hz)	Highest frequency (Hz)	Bandwidth (Hz)[a]	Starting frequency (Hz)	Ending frequency (Hz)	Duration (ms)	Number of pulses	Interpulse interval (ms)	*n*
AM (*T*)	140 (39)	1442 (221)	1305 (182)	NA	NA	469 (276)	NA	NA	2
AM (*S*)	279 (506)	4000[b] (2482)	3519 (2608)	NA	NA	765 (614)	NA	NA	12
FM (*T*)	409 (705)	1430 (685)	1024 (395)	468 (668)	1196 (706)	253 (200)	NA	NA	4
FM (*S*)	405 (414)	936 (736)	531 (677)	531 (677)	583 (489)	582 (502)	NA	NA	30
Pulsed (*T*)	181 (315)	2536 (2115)	2359 (243)	NA	NA	164 (66)	11.3 (5.6)	14.1 (3.5)	15
Pulsed (*S*)	140 (244)	2948 (2118)	2808 (2136)	NA	NA	384 (237)	14.1 (11.4)	31.0 (21.8)	18

[a]Measured bandwidth of the AM signals is larger than that of the FM signals, as only the fundamental was measured and harmonics were excluded.
[b]Large AM value of the category highest frequency is the result of measuring all side bands of the 12 AM sounds.

and data indicating that the mother and escort were resting at deeper water depth and were not within the angle offset range.

The types and structures of the 21 vocalizations measurably attributed to the same calf via the two-element array were predominantly pulsed signals (71%), followed by 19% FM and 10% AM signals (Tables I–III). All 21 signals were relatively short in duration (<1 s; mean=208 ms, SD=147, range=89–664 ms) and of predominantly low frequency (Table II). Mean lowest frequency of the 21 signals was 220 Hz (SD=391.5, range=10–1464 Hz). Mean highest frequency was 2221 Hz (SD=1861.4, range 312–7719 Hz). Vocalizations had a relatively narrow bandwidth (2004 Hz, SD=1891, range=238–7118 Hz) with a mean maximum frequency of 2.22 kHz (range=239–7719 Hz). Table II shows summary statistics for the AM, FM, and pulsed signals recorded from this calf. "Grunt" vocalizations were the most commonly heard signal (90% of 21) from this calf based on aural analyses (Fig. 2).

The calf made a total of four close (1–5 m) approaches to the diver, once during each of four surfacing bouts. During three approaches, the calf emitted 4—8 repetitive "grunts" and looked directly at the diver at its closest point of approach. The group (consisting of the mother, calf, and adult escort) remained in the same approximate location, resting near the shallow (<30 m) bottom.

Supplemental information providing a detailed description of this encounter, including behaviors associated with calf vocalizations and types of vocalizations, is available at the Cetos Research Organization website (2007).

B. Single hydrophone recordings of calves

A subsample of 28 of the 109 calf groups with non-song vocalizations recorded only with the single hydrophone was selected for detailed analyses based on quality of videotape and acoustic signals. These 28 groups yielded a subsample of 60 sound signals considered of reasonable quality and circumstance to be attributable to a calf using methods similar to previous reports (Zoidis and Green, 2001; Pack *et al.*, 2005; Zoidis *et al.*, 2005) and data from the two-element array as follows: (1) The same 21 vocalizations made by the calf verified with the two-element array were recorded simultaneously by the single hydrophone, i.e., the omnidirectional hydrophone had coincident sounds that were verified separately as coming from the calf by the two-element array data. (2) No significant differences were found between the characteristics of three acoustic parameters of selected single-hydrophone vocalizations and those recorded and linked to the calf by the two-element array (Table II). (3) 27 of the 28 calves that produced vocalizations were alone (defined as isolated with no other whales in view within at least 15 m for at least 15 min and no humpback song sounds recorded). (4) Nonsong vocalizations were detected significantly more frequently when the calf was at or near its closest point of approach to the single hydrophone (<5 m) facing the diver/recorder whereas the mother and/or escort were resting in deeper water >10 m from the diver and calf ($n=49$, $X^2=49.0$, $p=0.002$, $df=2$). (5) The vocalizations had a relatively strong signal-to-noise ratio (10 dB). (6) Two different calves emitted bubbles concurrent with the time of the vocalizations.

Most (90%) of the above 60 signals were short in duration (<1.1 s, mean=559 ms, range=75 ms–2.5 s) and of predominantly low frequency (overall mean low frequency of all 60 was signals=306 Hz, range=10–1710 Hz). FM signals were most common (50%) followed by pulsed (30%) and AM (20%) signals (Table I). Of the 30 FM signals, eight were simple upsweeps, six were simple downsweeps, eight had near-flat frequency contours, and the remaining eight had inflections in their contours (e.g., U-shaped). Nine of the twelve AM sounds had no strong changes in AM frequency, two increased, and one had a complex pattern of modulation. The large single hydrophone AM value of the "highest frequency" (Table II) is the result of measuring all of the side bands for the 12 AM sounds. The measured bandwidth of

TABLE III. Comparison of vocalizations produced by young of large whale species. Age classifications are taken directly from referenced papers and may have been assessed by non-similar standards. Sample size provided when presented in literature. AM=amplitude modulated, FM=frequency modulated, Hz=hertz, SD =standard deviation, s=second, and MSB=mean sound bandwidth.

Species, age classification, sample size	Type: Bandwidth (Hz) MSB, SD, range[a]	Mean duration, SD, range (s)[a]	Source
Humpback whale (*Megaptera novaeangliae*) calf vocalizations linked with two-element array (2006: $n=1$ calf, 21 vocalizations)	Pulsed: MSB 2359, (SD 243) FM: MSB 1024, (SD 395) AM: MSB 1305, (SD 182) Overall range: 238–7118	Pulsed: 0.16, (SD 0.07) FM: 0.25, (SD 0.20) AM: 0.47, (SD 0.28) Overall mean: 0.21, (SD 0.15), Range: 0.01–0.66	This study
Humpback calf-attributed vocalizations recorded with single hydrophone (2001–2005: $n=28$ calves, 60 vocalizations)	Pulsed: MSB 2808, (SD 2136) FM: MSB 531, (SD 677) AM: MSB 3519, (SD 2608) Overall range: 140–4000	Pulsed: 0.38, (SD 0.24) FM: 0.58, (SD 0.50) AM: 0.77, (SD 0.61) Overall mean: 0.56 Range: 0.08—2.5	This study
Humpback calf-attributed vocalizations recorded with single hydrophone (1996–2003: $n=8$ calves, 49 vocalizations)	Constant rate pulse: MSB 794, Increasing or decreasing rate pulse: MSB 775, Upswept frequency tone: MSB 1297, Sound combination: MSB 1021, Median fundamental frequency 90, Overall range: 30–3000	Mean: 0.38, Range: 0.11–0.71	Pack *et al.*, 2005
Adult humpback groups ($n=54$)	Social sounds range: 50–10 000,[b] dominant frequencies <2000, mostly FM upsweeps	Range: 0.25– >5	Silber, 1986
Adult male humpbacks (>100)	Song range: 20– >24000	Range: 300–1980	Payne and McVay, 1971; Helweg *et al.*, 1992; Au *et al.*, 2000, 2001, 2005, 2006; Fristrup *et al.*, 2003; Potter *et al.*, 2003
Gray whale (*Eschrichtius robustus*) female calf, 1.5–7 months old ($n=1$ calf)[c] (240 h of recordings)	Ranges: Type 1a "croak" pulses[d]: 70–4000 Type 1b "pop" pulses[e]: 70–4000 Type 3 "moan"[d]: 80–2120 Type 4 "grunt"[d]: 70–5000	Type 1a: 0.039, SD 0.012 Type 1b: 0.072, SD 0.027 Type 3: 0.44, SD 0.20 Type 4: 0.34, SD 0.090	Wisdom *et al.*, 2001
North Atlantic right whale (*Eubalaena glacialis*) female calf ($n=1$ calf, 9 calls)	Warble range: 470–8410	Range: 0.43–4.77	Parks and Tyack, 2005

TABLE III. (*Continued.*)

Species, age classification, sample size	Type: Bandwidth (Hz) MSB, SD, range[a]	Mean duration, SD, range (s)[a]	Source
Bryde's whale (*Balaenoptera edeni*) juvenile ($n=1$ calf, 233 vocalizations)[c]	Discrete pulse range: 400–610 Pulsed moan range: 200–900	Discrete pulse range: 0.010,[g], Pulsed moan range: 0.5–51	Edds *et al.*, 1993
Bryde's whale calf ($n=1$ calf, 36 calls)[h]	Discrete pulse series range: 700–900	Discrete pulse range: 0.025–0.040,	Edds *et al.*, 1993

[a]Mean, MSB, median, SD, and/or range are reported if available. Studies did not always report the same units and parameters.
[b]Maximum recording equipment frequency was 10 kHz.
[c]Captive animal.
[d]Produced by 1.5 months of age.
[e]Produced at 7 months of age.
[f]1–2 year old animal
[g]10 was indicated simply as "duration" in Edds *et al.* (1993) and appears to be a range compared to other numbers presented in the original document table.
[h]Recorded with an omnidirectional hydrophone on two occasions when only first-year free-ranging calves were present at the surface, whereas an adult companion was diving. The discrete pulses were loudest when a calf was close to the hydrophone and thus may have been produced by a calf (Edds *et al.*, 1993). Discrete pulses were recorded only near these individual calves during opportunistic studies of free-ranging Bryde's whale vocalizations.

these AM signals is larger than that of the FM signals, partially because only the fundamental was measured and the harmonics were excluded from the measurement.

We recorded unusual behavioral events when calf-attributed vocalizations were recorded that to our knowledge have not been previously reported. A selective representation of these include the following: Repeated grunt series by three separate calves that increased in amplitude followed by the mother joining the calf from depth and "herding" the calf away from the diver, who was less than 5 m from the calf at the time of sound production; this type of directed movement behavior by the mother was not seen when no sounds were recorded. Two of these calves repeatedly produced grunt sounds and simultaneously created underwater bubble streams. One of these calves was recorded for a 3:04-min continuous sequence of grunts, the longest duration of any year and the farthest (>30 m) a calf was away from its mother who was in deeper water and not visible to the diver. The grunts increased in tempo and coincided with repeated bubble streams and an underwater "jaw clap."

IV. DISCUSSION

Nonsong vocalizations occur in all adult humpback group compositions on the wintering grounds, including those with a calf, as documented previously and herein (Tyack, 1983; Silber, 1986; Zoidis and Green, 2001; Pack *et al.*, 2005; Zoidis *et al.*, 2005). For the first time, measured time delays between array bearings provide measurable evidence corroborating that a male humpback calf produced 21 nonsong vocalizations as recorded with a nonstatic two-element array mounted on an underwater video camera. Previously, vocalizations have been attributed to individual humpback calves on the wintering grounds based only on observational evidence and recordings with a single-hydrophone video camera (including the 60 vocalizations from 28 other calves reported herein) (Zoidis and Green, 2001; Pack *et al.*, 2005; Zoidis *et al.*, 2005). Although convincing, the methodology used in these studies cannot unequivocally rule out that the sounds did not come from another source/whale that was nearby or not visible because one omnidirectional hydrophone cannot provide directional bearings spatially linking acoustic data to a source. However, the two-element array calf recordings combined with similarities in sound characteristics between this calf and other calf-attributed vocalizations provide further substantiation that both male and female humpback calves vocalize.

Nonsong vocalizations appear more common among wintering humpback calf pods than previously reported. Herein, nonsong vocalizations were recorded in 65% of 170 calf pods. Silber (1986) recorded nonsong social sounds in only 7% of 14 mother–calf–escort groups and in none of seven lone mother–calf pairs. Pack *et al.* (2005) attributed 49 vocalizations to eight humpback calves off Maui (the number of calf pods that did not vocalize was not reported). In addition, we found nonsong vocalizations more common in lone mother–calf pairs than among mother–calf–escort groups. Similarly, Zoidis and Green (2001) reported a significantly higher vocalization rate (10.9 vocalizations per whale per hour) among 12 mother–calf pairs compared to 50 mother–calf–escort groups (4.4), 25 surface-active groups of three or more adults with no calves (4.2), and seven adult dyads (1.7). In contrast, both Silber (1986) and Pack *et al.* (2005) reported nonsong or calf vocalizations primarily among mother–calf–escorts. Per Pack *et al.* (2005), five of eight calves that vocalized were escorted by a mother and one escort including one singing escort; the remaining three groups with such sounds were a mother–calf pair, one mother–calf with two escorts, and one lone calf (the defini-

tion, i.e., duration of how long a calf was "alone" was not indicated). Five of these calves were males and three were females (Pack *et al.*, 2005). In comparison, we attributed vocalizations to 11 male and 18 female calves.

Calf vocalizations recorded with the two-element array were simple in structure and limited in repertoire with short duration and interpulse interval, and predominantly low frequency and relatively narrow bandwidth. These characteristics were not significantly different from other vocalizations attributed to humpback calves in this study and are similar to acoustic characteristics reported in earlier reports of calf vocalizations (Zoidis and Green, 2001; Pack *et al.*, 2005; Zoidis *et al.*, 2005) (Table III). Calves appear to produce sequences of the same sound in some instances. Repeated grunt bouts were made by the two-element-array recorded calf and were also recorded near calves by the single hydrophone. Similarly, Pack *et al.* (2005) reported that four calves produced sound sequences at a mean interval of 4.4 s, the spectral characteristics of which were similar to the grunt sounds we recorded (Fig. 2). In addition, vocalizations were documented simultaneous to bubble emissions from the blowhole for both the two-element array recorded calf as well as calves recorded with the single hydrophone herein and in Pack *et al.* (2005) (Figs. 1 and 2).

Pack *et al.* (2005) found that calf-attributed vocalizations were statistically similar to "noncalf social sounds" (terminology from Pack *et al.* 2005) but were significantly different from adult song units; the former were noted to have a narrower frequency range and shorter duration (Table III). Our results for calf vocalizations based on both the two-element array and single hydrophone recordings generally corroborate these findings, as do earlier reports that song is more complex than nonsong adult and calf vocalizations (Silber, 1986; Zoidis and Green, 2001; Zoidis *et al.*, 2005). In addition, pulsed signals were frequent among calves compared to the known repertoire of pulsed sounds produced by adult humpbacks in Hawaii (e.g., Payne and McVay, 1971; Darling, 1983; Payne, 1983; Silber, 1986; Helweg *et al.*, 1992; Au *et al.*, 2000). However, the most common type of signal recorded by the single hydrophone was FM signals, as found by Silber (1986) for groups of three or more adults off Maui (Table III). Most AM signals we recorded near calves had no strong evidence of changes in AM frequency. In contrast, social sounds of surface-active adult humpback groups typically exhibit considerable modulation in the AM frequency, producing the sensation of a frequency sweep (Silber, 1986).

Underwater video documenting behaviors and surroundings concurrent with two-element array recordings were important in providing empirical real-time evidence that a calf produced vocalizations. This approach allows an animal to be monitored visually and acoustically (i.e., it remains in view) for extended periods, so that underwater behaviors can be continuously recorded concurrently with vocalizations. Previous studies (e.g., Tyack, 1981, 1983; Darling, 1983; Silber, 1986) recorded nonsong sounds among adult humpback whales by deploying a single hydrophone from a small vessel. More recently, Dunlop *et al.* (2005, 2006), and Dunlop and Noad (2007) recorded nonsong vocalizations in Australia remotely from shore using a stationary array. No directional data linking vocalizations to individuals were presented in these previous studies.

The function and biological significance of calf vocalizations are unknown. Several potential hypotheses are proposed. Some types of calf vocalizations may elicit the mother's approach. In most cases, a calf closely (<5 to 10 m) approached the diver either silently or with nonsong vocalizations during a surfacing bout, with no reaction from the mother. However, when on three occasions a calf emitted repeated grunt vocalizations, with increasing incidence and amplitude, twice with bubbles and once with an accompanying jaw clap, the mother surfaced quickly, approached, and seemingly intentionally "herded" the calf directly away. Certain sounds (e.g., FM signals, repetitive grunts, particularly those that increase in amplitude) and/or bubble streams may function as either isolation or "alarm" calls by the calf to alert and/or call the mother. Alarm or isolation calls occur among mother-young groups of other mammalian species including dolphins (reviewed in Tyack, 2000), bats (Balcombe, 1990), and primates (Robinson, 1982). The calf may also produce vocalizations when encountering a novel stimulus (i.e., a close encounter with a diver or boat), or as unidirectional contact communication from the calf to the mother [i.e., where only one individual recognizes the call of the other and not visa versa (Torianni *et al.*, 2006)]. Documentation is lacking as to whether humpback mothers or other adults use sound to communicate with calves. To date, the published evidence has not shown this.

A few studies have been able to individually link vocalizations to young of other baleen whale species including gray (Wisdom *et al.*, 2001), northern right (Parks and Tyack, 2005; Parks and Clark, 2007), and Bryde's whales (Edds *et al.*, 1993); however, some of these studies have involved captive animals and may not be representative of free-ranging animals (Table III). Although the vocalizations themselves are quite different between species and are limited to a few individuals, their overall characteristics are generally similar to humpback calf-attributed vocalizations in terms of simplicity, limited repertoire, short duration, and predominantly low and narrowband frequency, when compared to adults of the same species (Edds *et al.*, 1993; Wisdom *et al.*, 2001; Parks and Tyack, 2005; Parks and Clark, 2007) (Table III). A captive juvenile Bryde's whale produced primarily pulsed moans with amplitude and frequency modulation, and on two occasions, free-ranging isolated Bryde's calves were linked with series of 4–11 discrete pulse calls (Edds *et al.*, 1993). Series of pulsed vocalizations (i.e., pulse trains) were also commonly produced by a captive gray whale calf (Wisdom *et al.*, 2001) as well as an entrapped juvenile humpback (Beamish, 1979) and humpback calves reported herein and by Pack *et al.* (2005) (Table III). Published individual sound confirmation of Northern right whale calves is limited to one free-ranging female based on bearing data collected with an acoustic array on the feeding grounds (Parks and Tyack, 2005; Parks and Clark, 2007). This calf produced stuttered "screamlike" calls interrupted by many pauses when alone at the surface as the other adults in the group dove. The calls

were generally of higher pitch and often longer in duration than adult female "screams" used to initiate surface-active adult groups (Parks and Tyack, 2005).

The biological significance of vocalizations produced by humpback calves and how it may change depending on social structure, environmental cues, ontogeny, and behavior, as well as whether mothers vocalize, remains to be further investigated. Circumstances of calf-attributed vocalizations recorded with a single hydrophone provide strong indication that the calf is most likely the sound source rather than the mother, including in lone mother–calf pairs. Compelling observational data include vocalizations recorded coincident with bubble emissions by humpback calves during both two-element and single hydrophone recordings, similar to humpback whale calves and a juvenile reported elsewhere (Beamish, 1979; Zoidis and Green, 2001; Pack *et al.*, 2005; Zoidis *et al.*, 2005) A captive juvenile Bryde's whale emitted bubbles only when vocalizing, throughout the vocalizations, and only from the right blowhole (Edds *et al.*, 1993). Dolphins sometimes emit bubbles simultaneous to sound production (Dahlheim and Awbrey, 1982; McCowan, 1995; McCowan and Reiss, 1995; Dudzinski, 1996; Herzing, 1996; Killebrew *et al.*, 1996, 2001), although this relationship is not consistent (Caldwell *et al.*, 1990; Fripp, 2005). However, it cannot be ruled out that some sounds recorded by the single hydrophone were produced by the mother without corroboration from directional acoustic data. Northern right whale mothers vocalize to their calves when they become separated, presumably to reunite the two (Parks and Clark, 2007). Similar contact calls may occur between humpback mothers and calves but have not yet been recorded or identified, possibly due to low-amplitude source levels that are difficult to detect except at close range. The relatively low amplitude of calf vocalizations we recorded may be why calves have only recently been reported to vocalize (Zoidis and Green, 2001).

Further analyses of the acoustic characteristics and behavioral context of calf vocalizations are underway, including their role as a potential indicator of stress, with implications for management concerns in areas with elevated anthropogenic activity and underwater noise such as Hawaii.

ACKNOWLEDGMENTS

The authors thank Randy Bates, Charles Bishop, Dan DenDanto, M. Green, Carol Hart, Thea Jenson, Terry McCabe, Don Moses, Dan Shapiro, Ethan Silva, Jessica Sharman, J. Pantukhoff, and Norbert Wu for their assistance and support with field portions of this project. They thank Pete Gehring, Thomas F. Norris, Shannon Rankin, Carol Spencer, P. Stevick, and Jenelle Black for reviewing this manuscript and/or for their editorial or technical input. They thank all our donors and patrons that have funded this work so that this publication was possible. They are also grateful to E. M. Blair, J. Dennis, S. Katona, D. Kearney, P. Tadd, A. Waldron, and A. Wright for their support and encouragement over the years.

Au, W. W. L., Darling, J., and Andrews, K. (**2001**). "High-frequency harmonics and source level of humpback whale songs," J. Acoust. Soc. Am. **110**(5), Part 2, 2770.

Au, W. W. L., Lammers, M. O., Stimpert, A., and Schotten, M. (**2005**). "The temporal characteristics of humpback whale songs," J. Acoust. Soc. Am. **118**, 1940.

Au, W. W. L., Pack, A. A., Lammers, M. O., Herman, L. M., Deakos, M. H., and Andrews, K. (**2006**). "Acoustic properties of humpback whale songs," J. Acoust. Soc. Am. **120**, 1103–1110.

Au, W. W. L., Popper, A. N., and Fay, R. R., eds. (**2000**). *Hearing by Whales and Dolphins*, Springer Handbook of Auditory Research Vol. **12** (Springer, New York), 1–485.

Baker, C. S., and Herman, L. M. (**1984**). "Aggressive behavior between humpback whales (*Megaptera novaeangliae*) wintering in Hawaiian waters," Can. J. Zool. **62**, 1922–1937.

Balcombe, J. P. (**1990**). "Vocal recognition of pups by Mexican free-tailed bats, *Tadarida brasiliensis mexicana*," Anim. Behav. **39**, 960–966.

Beamish, P. (**1979**). "Behavior and significance of entrapped baleen whales," in *Behavior of Marine Animals*, Vol 3: Cetaceans, edited by H. E. Winn, and B. Olla (Plenum, New York), pp. 291–309.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). "A review of the signature whistle hypothesis for the Atlantic bottlenose dolphin," in *The Bottlenose Dolphin*, edited by S. Leatherwood, and R. R. Reeves (Academic, San Diego), pp. 199–234.

Cartwright, R. (**2005**). "A comparative study of the behaviour and dynamics of humpback whale (*Megaptera novaeangliae*) mother and calf pairs during their residence in nursery waters," Ph.D. dissertation, Manchester Metropolitan University, Manchester, U.K., pp. 1–259.

Cetos (**2007**). "Calf activity (behavioral and acoustic) during social sound production, as recorded with two-element hydrophone array," Cetos Research Organization Internet website http://www.cetosresearch.org/research/humpback_findings.htm, last accessed 17 July 2007.

Charif, R. A., Clark, C. W., and Fristrup, K. M. (**2004**). *Raven 1.2 User's Manual* (Cornell Laboratory of Ornithology, Ithaca), pp. 1–205.

Clapham, P. J., Palsbøll, P. J., Mattila, D. K., and Vasquez, O. (**1992**). "Composition and dynamics of humpback whale competitive groups in the West Indies," Behaviour **122**, 182–194.

D'Vincent, C. G., Nilson, R. M., and Hanna, R. E. (**1985**). "Vocalization and coordinated feeding behaviour of the humpback whale in southeastern Alaska," Sci. Rep. Whales Res. Inst. **36**, 41–47.

Dahlheim, M. E., and Awbrey, F. (**1982**). "A classification and comparison of sounds of captive killer whales (Orcinus orca)," J. Acoust. Soc. Am. **72**, 661–670.

Darling, J. D. (**1983**). "Migrations, abundance and behavior of Hawaiian humpback whales (*Megaptera novaeangliae*) (Borowski)," Ph.D. dissertation, University of California at Santa Cruz, Santa Cruz, CA, pp. 1–147.

Darling, J. D., and Bérubé, M. (**2001**). "Interactions of singing humpback whales with other males," Marine Mammal Sci. **17**, 570–584.

Darling, J. D., Jones, M. E., and Nicklin, C. P. (**2006**). "Humpback whale songs: Do they organize males during the breeding season?" Behaviour **143**, 1051–1101.

Dudzinski, K. M. (**1996**). "Communication and behavior in the Atlantic spotted dolphins (*Stenella frontalis*): Relationships between vocal and behavioral activities," Ph.D. dissertation, Texas A&M University, College Station, TX, pp. 1–218.

Dunlop, R. A., Noad, M. J., Cato, D. H., and Stokes, D. (**2007**). "The social vocalization repertoire of east Australian migrating humpback whales (*Megaptera novaeangliae*)," J. Acoust. Soc. Am. **122**, 2893–2905.

Dunlop, R. A., Noad, M. J., and Cato, D. H. (**2005**). "Widespread and contextual use of social communication in migrating humpback whales," *Abstracts, 16th Biennial Conference on the Biology of Marine Mammals*, San Diego, CA, p. 79.

Edds, P. L., Odell, D. K., and Tershy, B. R. (**1993**). "Vocalizations of a captive juvenile and free-ranging adult-calf pairs of Bryde's whales, *Balaenoptera edeni*," Marine Mammal Sci. **9**, 269–284.

Fripp, D. (**2005**). "Bubblestream whistles are not representative of a bottlenose dolphin's vocal repertoire," Marine Mammal Sci. **21**, 29–44.

Fristrup, K. M., Hatch, L. T., and Clark, C. W. (**2003**). "Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts," J. Acoust. Soc. Am. **113**, 3411–3424.

Glockner, D. A. (**1983**). "Determining the sex of humpback whales (*Megaptera novaeangliae*) in their natural environment," in *Communication and Behavior of Whales*, edited by R. Payne (Westview Press, Boulder), pp. 447–464.

Glockner, D. A., and Venus, S. C. (**1983**). "Identification, growth rate, and behavior of humpback whale (*Megaptera novaeangliae*) cows and calves

in the waters off Maui, Hawaii, 1977–79," in *Communication and Behavior of Whales*, edited by R. Payne (Westview Press, Boulder), pp. 223–258.
Glockner-Ferrari, D. A., and Ferrari, M. J. (**1990**). "Reproduction in the humpback whale (*Megaptera novaeangliae*) in Hawaiian waters, 1975–1988: The life history, reproductive rates and behavior of known individuals identified through surface and underwater photography," Rep. Int. Whaling Com. Special Issue Vol. **12**, 161–169.
Helweg, D. A., Frankel, A. S., Mobley, J. R., Jr., and Herman, L. M. (**1992**). "Humpback whale song: Our current understanding," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 459–483.
Herzing, D. L. (**1996**). "Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*," Aquat. Mamm. **22**, 61–79.
Jurasz, . M., and Jurasz, V. P. (**1979**). "Feeding modes of the humpback whale, *Megaptera novaeangliae*, in southeast Alaska," Sci. Rep. Whales Res. Inst. **31**, 69–83.
Katona, S., Baxter, B., Brazier, O., Kraus, S., Perkins, J. S., and Whitehead, H. (**1979**). "Identification of humpback whales by fluke photographs," in *Behavior of Marine Animals—Current Perspectives in Research, Vol. 3: Cetaceans*, edited by H. E. Winn, and B. L. Olla (Plenum, New York), pp. 33–44.
Kibblewhite, A. C., Denham, R. N., and Barnes, D. J. (**1966**). "Unusual low-frequency signals observed in New Zealand waters," J. Acoust. Soc. Am. **41**, 644–655.
Killebrew, D. A., Mercado, E. III, Pack, A. A., and Herman, L. M. (**1996**). "Burst-pulse sounds of a neonate bottlenosed dolphin," J. Acoust. Soc. Am. **100**, 2610.
Killebrew, D. A., Mercado, E., III, Herman, L. M., and Pack, A. A. (**2001**). "Sound production of a neonate bottlenosed dolphin," Aquat. Mamm., **27**, 34–44.
Mann, J. (**1999**). "Behavioural sampling methods for cetaceans: A review and critique," Marine Mammal Sci. **15**, 102–122.
McCowan, B. (**1995**). "A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, Tursiops truncatus)," Ethology **100**, 177–194.
McCowan, B., and Reiss, D. (**1995**). "Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (Delphinidae: *Tursiops truncatus*): A re-evaluation of the signature whistle hypothesis," Ethology **100**, 193–209.
Noad, M. J., Dunlop, R., Stokes, D., Miller, P., and Biassoni, N. (**2006**). "Humpback whale social sounds: Source levels and response to playback," J. Acoust. Soc. Am. **120**, 3012.
Norris, T. F., McDonald, M., and Barlow, J. (**1999**)."Acoustic detections of singing humpback whales (*Megaptera novaeangliae*) in the eastern North Pacific during their northbound migration," J. Acoust. Soc. Am. **106**, 506–514.
Pack, A. A., Herman, L. M., Deakos, M. H., Hakala, S., Craig, A. S., Olson, J. R., Spitz, S. S., Herman, E., Goetschius, K., and Lammers, M. O. (**2005**). "First report of sounds recorded from individual humpback whale calves on the Hawaiian wintering grounds," *Abstracts, 16th Biennial Conference on the Biology of Marine Mammals*, San Diego, CA, p. 216.
Parks, S. E., and Clark, C. W. (**2007**). "Acoustic communication: social sounds and the potential impacts of noise," in *The Urban Whale: North Atlantic Right Whales at the Crossroads*, edited by S. D. Kraus, and R. M. Rolland, (Harvard University Press, Cambridge, MA, pp. 310–332).
Parks, S. E., and Tyack, P. L. (**2005**). "Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups," J. Acoust. Soc. Am. **117**, 3297–3306.
Payne, R. S. (**1978**). "Behavior and vocalizations of humpback whales (*Megaptera* sp.)," in *Report on a Workshop on Problems Related to Humpback Whales (Megaptera novaeangliae) in Hawaii*, edited by K. S. Norris and R. R. Reeves, Report prepared for the U.S. Marine Mammal Commission, Washington, D.C., NTIS PB-280 **794**, pp. 56–78.
Payne, R. S., editor (**1983**). *Communication and Behavior of Whales* (Westview Press, Boulder).
Payne, R. S., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 585–597.
Potter, J. R., Deakos, M. H., Koay, T. B., Durville, C., and Pack, A. A. (**2003**). "Up close and personal: Recording humpback whale song at close ranges (10–50 m)," *Proceedings, Oceans 2003 MTS/IEEE Conference*, San Diego, CA, p. 472.
Robinson, J. G. (**1982**). "Vocal systems regulating within-group spacing," in *Primate Communication*, edited by C. T. Snowdon, C. H. Brown, and M. R. Petersen (Cambridge University Press, Cambridge), pp. 94–116.
Sharpe, F. A., Dill, L. M., Beaver, V., and Spellman, B. (**1998**). "Killing me softly: Feeding calls of the Alaskan humpback whale," *Abstracts, World Marine Mammal Science Conference*, Monaco, pp. 123.
Silber, G. K. (**1986**). "The relationship of social vocalizations to surface behavior and aggression in the Hawaiian humpback whale (*Megaptera novaeangliae*)," Can. J. Zool. **64**, 2075–2080.
Spitz, S., Herman, L. M., and Pack, A. A. (**2000**). "Measuring sizes of humpback whales (*Megaptera novaeangliae*) by underwater videogrammetry," Marine Mammal Sci. **16**, 664–676.
Thompson, P. O. W., Cummings, W. C., and Kennison, S. J. (**1977**). "Sound production of humpback whales, *Megaptera novaeangliae*, in Alaskan waters" J. Acoust. Soc. Am. **62**, 1–182.
Torriani, V. G., Vannoni, E., and McElligott, A. G. (**2006**). "Mother-young recognition in an ungulate hider species: A unidirectional process," Am. Nat. **168**, 412–420.
True, F. W. (**1904**). "The whalebone whales of the western North Atlantic compared with those occurring in European waters with some observations on the species of the North Pacific," Smith. Contrib. Know. **33**, 1–332.
Tyack, P. (**1981**). "Interactions between singing Hawaiian humpback whales and conspecifics nearby," Behav. Ecol. Sociobiol. **8**, 105–116.
Tyack, P. (**1983**). "Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds," Behav. Ecol. Sociobiol. **13**, 49–55.
Tyack, P., and Whitehead, H. (**1983**). "Male competition in large groups of wintering humpback whales," Behaviour **83**, 132–154.
Tyack, P. L. (**1982**). "Humpback whales respond to sounds of their neighbors" Ph.D. dissertation, Rockefeller University, New York, NY, pp. 1–207.
Tyack, P. L. (**2000**). "Functional aspects of cetacean communication," in *Cetacean Societies: Field Studies of Dolphins and Whales*, edited by J. Mann, R. C. Connor, P. L. Tyack, and H. Whitehead (University of Chicago Press, Chicago), pp. 270–307.
Winn, H. E., Bischoff, W. L., and Taruski, A. G. (**1973**). "Cytological sexing of cetaceans," Mar. Biol. (Berlin) **23**, 343–346.
Winn, H. E., and Winn, L. K. (**1978**). "The song of the humpback whale (*Megaptera novaeangliae*) in the West Indies," Mar. Biol. (Berlin) **47**, 97–114.
Wisdom, S., Bowles, A. E., and Anderson, K. E. (**2001**). "Development of behavior and sound repertoire of a rehabilitating gray whale calf," Aquat. Mamm. **27**, 239–255.
Zoidis, A. M., and Green, M. (**2001**). "Relationship of social vocalizations to pod size, composition and behavior in the Hawaiian humpback whale," *Abstracts, 15th Biennial Conference of the Biology of Marine Mammals*, Vancouver, B.C., http://www.cetosresearch.org/main.htm last accessed 21 September **2007**.
Zoidis, A. M., Smultea, M. A., Fertl, D., Day, A. J., DenDanto, D., Ertl, A. S., Hayes, J., and Whitt, A. D. (**2005**). "Can you hear me now? Social sounds and underwater behavior of Hawaiian humpback whale (*Megaptera novaeangliae*) calves," *Abstracts, 16th Biennial Conference of the Biology of Marine Mammals*, San Diego.

Variation in call pitch among killer whale ecotypes

Andrew D. Foote[a)]
School of Biological & Biomedical Sciences, University of Durham, Durham, DH1 3LE, United Kingdom

Jeffrey A. Nystuen
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Box 355640, Seattle, Washington 98105

(Received 9 August 2007; revised 13 December 2007; accepted 27 December 2007)

Vocal structure can vary between populations due to variation in ecology-dependent selection pressures, such as masking by background noise and the presence of eavesdroppers. Signalers can overcome these obstacles to effective communication by avoiding frequencies that overlap with background noise or the audible range of eavesdroppers. In the Northeastern Pacific three "ecotypes" of killer whale coexist in sympatry, but differ from one another in their diet and habitat use. The minimum frequency (F_{min}) and the frequency containing the peak energy between 0 and 10 kHz (F_{peak}) of a random sample of calls produced by a population of each ecotype was measured. The offshore ecotype produced calls with a significantly higher F_{min} than the other ecotypes, which could be a strategy to avoid masking by low frequency chronic bandlimited wind noise found in the offshore environment. The resident ecotype produced calls with a significantly higher F_{min} and F_{peak} than the transient ecotype. This could be to reduce detection by their salmonid prey, which has a narrow band, low frequency auditory range.
 [DOI: 10.1121/1.2836752]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.66.Dc, 43.80.Nd [WWA] Pages: 1747–1752

I. INTRODUCTION

Variation in vocalization structure can occur between populations or groups of the same species due to ecological differences such as background noise levels or habitat-dependent transmission properties, or the auditory range of eavesdropping competitors, predators, or prey (Hunter and Krebs, 1979; Rydell and Arlettaz, 1994; Slabbekoorn and Peet, 2003). Vocalizations may then undergo functional selection as the acoustic properties of an environment may favor vocalizations with particular characteristics, e.g., adjusting the frequency range of a signal so that it does not overlap with the frequency band of background noise or the auditory range of eavesdroppers (Rydell and Arlettaz, 1994; Slabbekoorn and Peet, 2003). In this study we compare the frequency parameters of calls produced by three sympatric killer whale ecotypes that differ in prey and habitat preference.

Three ecotypes of killer whale occur in sympatry in the Northeastern Pacific. The resident ecotype specializes in foraging on salmonid species; the transient ecotype specializes in foraging for marine mammals (Ford *et al.*, 1998; Saulitis *et al.*, 2000; Herman *et al.*, 2005). The majority of sightings of these two ecotypes are predominantly in coastal waters, however sightings are biased toward the summer months and less is known about their winter distribution (Ford *et al.*, 2000). Lastly the offshore ecotype, thought to be primarily pisciverous (Herman *et al.*, 2005; Krahn *et al.*, 2007), is most commonly sighted 15 km or more offshore and rarely found in coastal waters (Ford *et al.*, 2000). Each ecotype is distinct from the other two in both genotype (Hoelzel *et al.*, 1998; Barrett-Lennard, 2000) and phenotype (Ford *et al.*, 2000).

Killer whales maintain contact with conspecifics using stereotyped broadband calls (Ford, 1989). Call types are distinguishable by the frequency modulation contours (Ford, 1989; Deecke *et al.*, 1999; Yurk *et al.*, 2002). Calls have the potential to convey information on signaler's sex, matriline and pod identity, behavioral state, direction of travel, and distance relative to the receiver (Ford, 1989, 1991; Deecke *et al.*, 2000; Miller and Bain, 2000; Miller, 2002; Nousek *et al.*, 2006; Miller *et al.*, 2007). There is strong evidence that call types are learned (Deecke *et al.*, 2000; Yurk *et al.*, 2002; Foote *et al.*, 2006). Call type dialects are found both within and between populations (Ford, 1991).

Previous studies have investigated variation in the acoustic behavior of resident and transient killer whales due to differences in the eavesdropping ability of their preferred prey species (Barrett-Lennard *et al.*, 1996; Deecke *et al.*, 2005). Background noise is another potential source of ecology-dependent variation in killer whale call structure. The character of ambient noise is likely to vary between the coastal and offshore habitats as it is highly correlated with wind speed (Knudsen *et al.*, 1948; Wenz, 1962; Wille and Geyer, 1984). In this study we compare the minimum frequency and the frequency of peak energy between 0 and 10 kHz of calls produced by pods from each ecotype and discuss how variation may relate to ecological variables such as background noise or the presence of eavesdroppers.

[a)]Current affiliation: School of Biological Sciences, University of Aberdeen, Lighthouse Field Station, George Street, Cromarty, Ross-shire, IV11 8YJ, United Kingdom. Electronic mail: a.d.foote@abdn.ac.uk

II. METHODS

A. Killer whale recordings

Recordings were collected from three populations of killer whale from the eastern North Pacific, one of each ecotype (transient, resident, and offshore). Recordings were made using a variety of recording devices, but all with a flat response between 0.1 and 10 kHz, the range in which all measured minimum frequencies fell.

The Southern Resident population is commonly sighted in the coastal waters of Washington State and British Columbia during the summer months, their winter distribution is known to range between Monterey Bay, CA and the Queen Charlotte Islands, BC. The population consists of three pods, J, K, and L pods, each of which has a distinctive call type repertoire (Hoelzel and Osborne, 1986; Ford, 1987, 1991). They share no call types in common with any other killer whale population in this study (Ford, 1987, 1991). A database of 278 recordings was analyzed from Haro Strait, WA and the neighboring waterways. Individual calls were isolated and saved as separate sound files. Thirty soundfiles of individual calls were then selected randomly, these were checked to confirm that the quality of the recording and signal-to-noise ratio was sufficient for measurements to be made and where this was not the case replacement soundfiles were chosen, also at random.

The West Coast Transient population is found from California to Southeast Alaska (Ford and Ellis, 1999). The West Coast Transient population has a single call type repertoire; however there are regional variations in call type usage (Deecke, 2003). They share no call types in common with any other killer whale populations in this study (Ford, 1987). Our recordings of the West Coast Transient population were also from Haro Strait, WA and neighboring waterways, a database of six recordings were analyzed and calls were isolated and selected in the same way as for the Southern Resident Population.

The offshore ecotype also has a large home range, and individuals sighted off the California coast have been resighted in the Bering Sea (Krahn *et al.*, 2007). A database of four recordings, from nonconsecutive encounters, totaling 81 min from relatively infrequent encounters with this ecotype in the coastal waters of Johnstone Strait, BC were analyzed. Three recordings were made in the same week, although on separate days, and were thus likely to be of the same group. Calls were isolated and selected as noted earlier. As recordings of all three ecotypes were made in the coastal waters around Vancouver Island during summer months, background noise levels should be similar in frequency characteristics and the suitability of calls for analysis based on their signal-to-noise ratio should be the same for all three ecotypes.

B. Ecological variables

As the offshore killer whales were recorded during rare incursions into coastal waters, it was inappropriate to take the background noise levels from these recordings. Ambient noise levels and spectral profiles were measured from a coastal site (Haro Strait; 48°30′N, 123°08′W) and an offshore site (Bering Sea; 56°51′N, 164°03′W) using PAL (Passive Aquatic Listeners) recorders (Ma and Nystuen, 2005). The Haro Strait PAL was deployed from May to September 2005 in approximately 30 m of water. This location is a critical habitat for the Southern Resident population of killer whales (Ford *et al.*, 2000), and the whales were regularly sighted during the deployment period. This area is also commonly used by West Coast Transients during this period (Baird and Dill, 1995). The Bering Sea PAL was deployed May–September 2004 in approximately 70 m of water, offshores have been sighted in this area at this time of year (Zerbini *et al.*, 2007). Each PAL consisted of a low-noise wideband hydrophone (either an ITC-8263 or a Hi-Tech-92WB), signal preamplifier, and a recording computer (Tattletale-8). The nominal sensitivity of these instruments was −160 dB relative to 1 V/μPa and the equivalent oceanic background noise level of the preamplifier system is about 28 dB relative to 1 $\mu Pa^2\ Hz^{-1}$. Bandpass filters were present to reduce saturation from low frequency sound (high pass at 300 Hz) and aliasing from above 50 kHz (low pass at 40 kHz). The PAL sampled at fixed intervals resulting in a time series of spectral level at 2 kHz, to coincide with the omnidirectional component of killer whale calls.

Audiograms of key prey species were taken from the literature and compared with the mean minimum frequency from the 30 calls measured. The current published data on the predominant species that constitute the diet of offshore killer whales are not as detailed as for the other two ecotypes. Therefore we focused only on transient and resident killer whales. Resident killer whale calls were compared with an audiogram for Atlantic Salmon *Salmo salar* (Hawkins and Johnstone, 1978) as a literature search found no audiogram for Chinook salmon *Oncorhynchus tshawytscha*, the preferred prey species of the Southern Resident population based on observations, prey fragment sampling, and stomach contents analyses (Ford *et al.*, 1998; Ford and Ellis, 2006). Transient killer whale calls were compared with audiograms for the harbor porpoise *Phocoena phocoena* (Kastelein *et al.*, 2002), and harbor seal *Phoca vitulina* (Kastak and Schusterman, 1998), their two most common prey species based on observation and stomach contents analyses (Baird and Dill, 1996; Ford *et al.*, 1998).

C. Acoustic and statistical analysis

Recordings were digitized using Soundforge software at 44.1 kHz and analyzed using Canary 1.2.4. Spectrograms and energy spectra were produced with a filter bandwidth of 88.24, FFT size 1024, and 87.5% overlap. We measured F_{min}: the lowest frequency on the spectrogram, and F_{peak}: the frequency with highest amplitude between 0 and 10 kHz. Using the measurement panel in Canary, we were able to measure frequency parameters to a resolution of 0.01 kHz. We used analysis of variance (ANOVA) tests to compare the F_{min} and F_{peak} of the call samples from each population (Fig. 1).

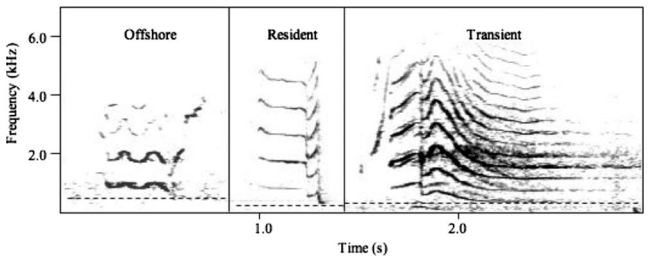


FIG. 1. Spectrograms of examples of calls of each ecotype of NE Pacific killer whale (filter bandwidth of 88.24, FFT size 1024, 87.5% overlap). Dashed lines indicate the F_{min} variable measured.

III. RESULTS

A. Frequency variation between ecotypes

Our selection of 30 resident calls was extracted from 28 recordings from nonconsecutive encounters totaling 512 min, and calls from all 3 pods (J, K, and L) were included. The selection of 30 transient calls came from 6 recordings, each from nonconsecutive encounters totaling 219 min. Last, 30 offshore calls were selected from 4 recordings totaling 80 min. There was significant variation between the three ecotypes in mean F_{min} (ANOVA: $F_{2,87}=38.23$, $p<0.0001$; Fig. 2). A post hoc t-test indicated the offshore ecotype produced calls with a significantly higher minimum frequency ($\bar{x}\pm\mathrm{SE}=0.89\pm0.24$ kHz, $N=30$) than the resident ($\bar{x}\pm\mathrm{SE}=0.55\pm0.33$ kHz, $N=30$; $t=4.738$, $p<0.0001$) and transient ecotypes ($\bar{x}\pm\mathrm{SE}=0.34\pm0.06$ kHz, $N=30$; $t=12.141$, $p<0.0001$). The difference between the resident and transient mean F_{min} was also significant ($t=3.014$, $p=0.0038$). There was also significant variation in the F_{peak} between ecotypes ($F_{2,87}=4.511$, $p=0.014$; Fig. 2). A post hoc t-test showed that the resident ecotype mean F_{peak} ($\bar{x}\pm\mathrm{SE}=1.88\pm1.02$ kHz) was significantly higher than the mean F_{peak} of the transient ecotype ($\bar{x}\pm\mathrm{SE}=1.28\pm0.42$ kHz; $t=2.919$, $p=0.0042$) but not the offshore ecotype mean F_{peak} ($\bar{x}\pm\mathrm{SE}=1.55\pm0.76$ kHz; $t=1.421$, $p=0.161$). There was also no significant difference between the transient and offshore mean F_{peak} ($t=1.703$, $p=0.0939$).

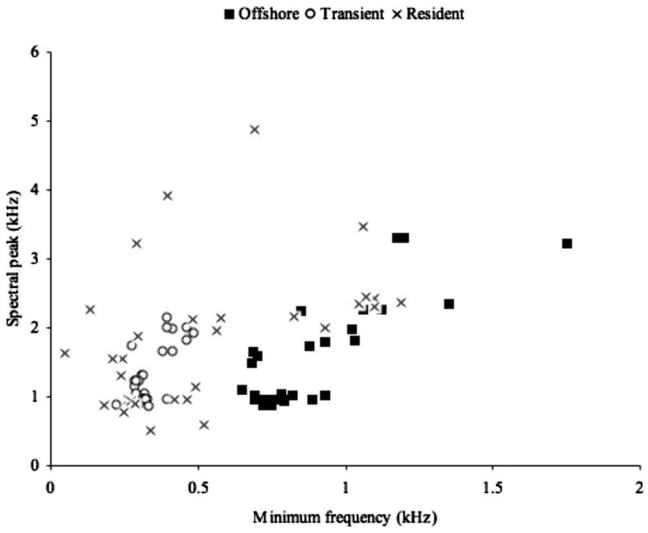


FIG. 2. F_{min} (the lowest frequency detectable on the spectrogram) plotted against F_{peak} (the frequency with highest amplitude between 0 and 10 kHz) of a random sample of 30 calls each recorded from resident, transient, and offshore Pacific killer whale ecotypes.

TABLE I. Sound budgets from nearshore and offshore sites. Peaks represent short transient sounds such as whale calls.

	Dominant sound source (% of time present)			
	Ships	Wind	Rain	Peaks
Nearshore Pacific (Haro Strait)	23	53	8	12
Offshore Pacific (Bering Sea)	1	85	6	7

	Sound pressure density spectrum levels 2 kHz (db relative to 1 $\mu Pa^2\,Hz^{-1}$)				
	Ships	Wind	Rain	Peaks	Mean
Nearshore Pacific (Haro Strait)	63 ± 7	55 ± 6	58 ± 8	59 ± 8	58 ± 8
Offshore Pacific (Bering Sea)	72 ± 5	60 ± 7	64 ± 6	62 ± 7	60 ± 8

B. Ecological correlates

By examining the spectral characteristics of the sound, it was possible to identify the source (see Ma and Nystuen, 2005) and produce a sound budget for each location, including the percentage of time that a particular source is present and the loudness of that source (Table I). The mean ambient noise levels were higher overall at the offshore site and wind noise was the main contributor and almost constant. Wind noise was only present at the coastal site approximately half of the time and then was lower in sound level than the offshore site. The noise from nearby ships were the loudest events detected and broadband in frequency range, this loud, broadband noise source was present more of the time in the coastal site, but was temporally variable. The higher F_{min} of the offshore ecotype would therefore be consistent with the hypothesis that persistent low frequency ambient noise will select for higher pitch calls. However this is not the case for our measurements of F_{peak}.

Although the mean minimum frequency of resident killer whale calls was above the auditory range of salmonids, the F_{min} ranged to as low as 0.13 kHz and a number of calls overlapped with the peak hearing range of salmonids (Hawkins and Johnstone, 1978; Fig. 3). However even the minimum F_{peak} (0.508 kHz) measured from the 30 resident calls was above the auditory range of salmonids (0.03–0.40 kHz; Hawkins and Johnstone, 1978). All the measured F_{min} and F_{peak} values for transient killer whale calls overlapped with the hearing range of harbor seal and harbor porpoise, their preferred marine mammal prey (Fig. 3).

IV. DISCUSSION

Interpopulation variation in F_{min} could have resulted from a number of factors and could be functional or an epiphenomenon of the learning process (Slater, 1986, 1989). For example this variation could facilitate population identity

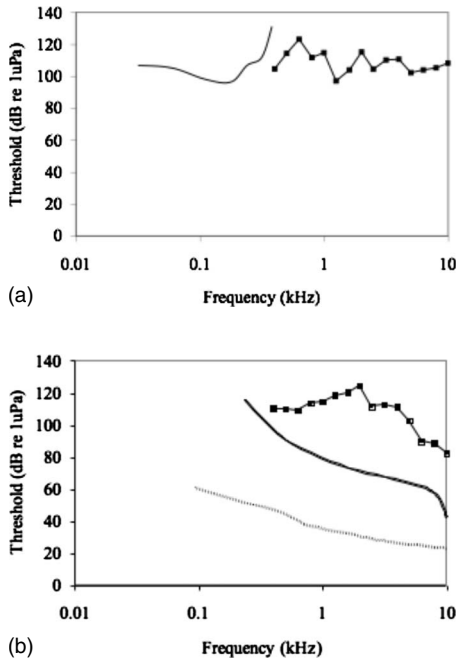


FIG. 3. (a) Audiogram of the Atlantic salmon *Salmo salar* (Hawkins and Johnstone, 1978) and one-third octave received levels (dB) up to 10 kHz of the most common resident call type (SI) from our sample, estimated to have been recorded <500 m from source based on surface observations, and on-axis based on the number of visible harmonics (see Miller, 2002). (b) Audiogram up to 10 kHz of the harbor porpoise *Phocoena phocoena* (solid line; Kastelein *et al.*, 2002), and harbor seal *Phoca vitulina* (dashed line; Kastak and Schusterman, 1998), and one third octave received levels (dB) of the most common transient call (WCT07) from our sample, estimated to have been recorded <500 m from source based on surface observations and on-axis based on the number of visible harmonics (see Miller, 2002).

(Yurk, 2005). However, the frequency modulations of call types within pod specific call type repertoires achieve this regardless of the frequency range of the fundamental (Yurk *et al.*, 2002). The structure of killer whale call types is known to shift randomly over time by a drift mechanism (Deecke *et al.*, 2000), which could lead to interpopulation variation in $F_{\min}$. However, the range of $F_{\min}$ from published call type repertoire catalogues of other resident ecotype populations is very consistent in their $F_{\min}$, e.g., Kamchatka Residents 0.5 ± 0.2 kHz (Filatova *et al.*, 2004), Northern Residents 0.4 ± 0.3 kHz, Southern Residents 0.5 ± 0.3 kHz (Ford, 1987) and similar to our random 30 calls from the Southern Resident population. This suggests that some common ecological or contextual factor may have led to directional convergence in $F_{\min}$.

Although ambient noise is likely to vary greatly spatially and temporally due to local conditions and our two sites are not expected to be representative, generic coastal and offshore sites, our measurements do conform to the general prediction that offshore waters will have higher levels of low frequency noise due to wind noise (Knudsen *et al.*, 1948; Wenz, 1962; Urick, 1983; Wille and Geyer, 1984). Wind noise is frequency dependent, decreasing 5 dB per octave with increasing frequency between 0.5 and 5.0 kHz (Wenz, 1962), therefore the offshore into ecotype's use of high pitch calls, in terms of $F_{\min}$, may avoid masking by low frequency wind noise. A previous comparison of the calls recorded from killer whales in the high ambient noise waters of the Ross Sea (Urick, 1983) found they were higher in minimum frequency than Pacific Resident calls (Awbrey *et al.*, 1982) and that Ross Sea killer whales seasonally use an "acoustical niche" outside the frequency range of leopard seal vocalizations (Mossbridge *et al.*, 1999). Inter and intraspecies comparisons of other odontocetes have also found that offshore species or populations have higher pitched whistles and concluded that this may be to reduce masking by ambient noise (Ding *et al.*, 1995a, b). However, this does not explain the significant difference in $F_{\min}$ between the resident and transient ecotypes, which are both found in coastal waters such as Haro Strait. Additionally, the full range of each ecotype is not fully known and resident and transient ecotypes may spend more time in offshore waters during the winter when there are fewer sightings in nearshore waters (Ford *et al.*, 2000).

Ship noise was the loudest sound source in both environments and was more persistent in the coastal environment (Table I). However, ship noise, although most intense at low frequencies, can be broadband between 0 and 10 kHz and could mask the entire omnidirectional component of killer whale calls (Bain and Dahlheim, 1994). A change in frequency would therefore not compensate for such broadband masking. Foote *et al.* (2004) reported an antimasking strategy of increasing call duration in the presence of boat noise by killer whales, which had developed within a decade concurrent to an increase in vessel traffic. They suggested that this would decrease call detection and perception thresholds through increased redundancy. Any vocal response may therefore depend upon both the temporal (e.g., transient or chronic) and spectral properties (e.g., narrowband or broadband, low or high frequency) of noise.

Deecke *et al.* (2005) reviewed the costs and limitations to transient killer whales of producing call types outside the broadband auditory range of their main prey species, which included the increased attenuation of higher frequencies and the reduced directionality of low frequencies. The first of these costs would be negligible if resident killer whales produced calls not overlapping with the low-frequency narrowband hearing range of salmonids (Hawkins and Johnstone, 1978), and may explain the difference in $F_{\min}$ between the resident and transient calls. Some resident killer whale calls did contain energy below 400 Hz, which would therefore be detectable by salmonids, at least at close range, but calling while foraging does not appear to be constrained in resident killer whales (Deecke *et al.*, 2005). Frequency is just one factor when considering the active space of a signal, the hearing threshold of the receiver is also important. Salmonids have poor hearing even at the peak of their audible range (Hawkins and Johnstone, 1978) and the distance over which salmonids can detect killer whale calls would be considerably less than marine mammals. As our recordings were not made with a calibrated hydrophone and the exact distance between the signaler and hydrophone were not known, it is not possible to precisely quantify the distance that salmon could detect a resident killer whale calls. However, Fig. 3 indicates that within approximately 500 m, detection

of resident killer whale calls would be limited by the hearing threshold in salmonids but easily detectable by harbor porpoise *Phocoena phocoena* and harbor seal *Phoca vitulina*. The higher F_{peak} of resident calls compared to transient calls may also be a strategy to reduce detection by salmonid prey. However the ratio of energy in the first two harmonics can be sex-dependent (Miller *et al.*, 2007), therefore, differences between ecotypes in F_{peak} in our samples could be due to a bias in recording predominantly calls produced by adult male residents and female transients (see Miller *et al.*, 2007).

Differences between habitats in wind noise or the selection pressure from eavesdropping prey would be consistent over time scales that would allow genotypic selection of call frequency. However, the plasticity of killer whale vocal behavior (Foote *et al.*, 2004, 2006) would enable call structure to be changed over ontogenetic time scales. Active selection of the vocal repertoire could be rapidly achieved by the dropping of call types with lower frequencies and increasing the usage of call types with higher frequencies, or by a spectral shift of existing call types.

Future recordings using multihydrophone arrays to localize the individual calling will help to control for potentially confounding variables such as sex, body size, and context, which we were unable to account for. Further within-population comparisons between recordings with different background noise levels may also reveal short-term changes to call structure in response to temporal variations in ambient noise, such as that noted by Foote *et al.* (2004), to boat noise. Combining recordings of killer whales with known dietary specialization and measurements of ambient noise, as we have done in this study, from new locations would help further understand the role of ecology in shaping killer whale call structure.

ACKNOWLEDGMENTS

The authors would like to acknowledge Kenneth Balcomb and Candice Emmons of the Center for Whale Research, Rich Osborne of the Whale Museum, Paul Spong and Helena Symonds of Orcalab for the use of acoustic recordings. Guen Jones, Nicholas Meidenger, Erin O'Connell, Jennifer Snowball, Anne-Marie Van Dijk, Val Veirs, and Monika Weiland provided valuable assistance in the field. We thank Whitlow Au, Rob Barton, Marilyn Dahlheim, Volker Deecke, Rus Hoelzel, Bridget Senior, Ari Shapiro, Peter Slater, and three anonymous reviewers for their helpful comments. Partial funding was provided by the ACS Puget Sound Chapter, Northwest Fisheries Science Center, PADI Aware, The Whale Museum, and the Whale and Dolphin Conservation Society. Field work was carried out in full compliance with state and federal laws.

Awbrey, F., Evans, W., Jehl, J., Thomas, J., and Leatherwood, S. (**1982**). "Comparison of Ross Sea and Pacific Northwest killer whale vocalizations," Rep. Int. Whal. Comm. **32**, 667–670.

Bain, D. E., and Dahlheim, M. E. (**1994**). "Effects of masking noise on detection thresholds of killer whales," in *Marine Mammals and the Exxon Valdez*, edited by T. R. Loughlin (Academic, San Diego), pp. 243–256.

Baird, R. W., and Dill, L. M. (**1995**). "Occurrence and behaviour of transient killer whales: Seasonal and pod-specific variability, foraging behaviour and prey handling," Can. J. Zool. **73**, 1300–1311.

Baird, R. W., and Dill, L. M. (**1996**). "Ecological and social determinants of group size in transient killer whales," Behav. Ecol. Sociobiol. **7**, 408–416.

Barrett-Lennard, L. G. (**2000**). "Population structure and mating systems of North-eastern Pacific killer whales," Ph.D. dissertation, University of British Columbia, Vancouver, BC.

Barrett-Lennard, L. G., Ford, J. K. B., and Heise, K. A. (**1996**). "The mixed blessing of echolocation: Differences in sonar use by fish-eating and mammal-eating killer whales," Anim. Behav. **51**, 553–565.

Deecke, V. B. (**2003**). "The vocal behaviour of transient killer whales (*Orcinus orca*): Communicating with costly calls," Ph.D. dissertation, University of St. Andrews, St. Andrews, Scotland.

Deecke, V. B., Ford, J. K. B., and Slater, P. J. B. (**2005**). "The vocal behaviour of mammal-eating killer whales: Communicating with costly calls," Anim. Behav. **69**, 395–405.

Deecke, V. B., Ford, J. K. B., and Spong, P. (**1999**). "Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects," J. Acoust. Soc. Am. **105**, 2499–2507.

Deecke, V. B., Ford, J. K. B., and Spong, P. (**2000**). "Dialect change in resident killer whales: Implications for vocal learning and cultural transmission," Anim. Behav. **60**, 629–639.

Ding, W., Wursig, B., and Evans, W. E. (**1995a**). "Whistles of bottlenose dolphins: Comparisons among populations," Aquat. Mamm. **21**, 65–77.

Ding, W., Wursig, B., and Evans, W. E. (**1995b**). "Comparisons of whistles among seven odontocete species," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 299–325.

Filatova, O. A., Burdin, A. M., Hoyt, E., and Sato, H. (**2004**). "A catalogue of discrete calls of resident killer whales (*Orcinus orca*) from the Avacha Gulf of Kamchatka peninsula," Zoolo. J. **83**, 1169–1180.

Foote, A. D., Griffin, R. M., Howitt, D., Larsson, L., Miller, P. J. O., and Hoelzel, A. R. (**2006**). "Killer whales are capable of vocal learning," Biol. Lett. **2**, 509–512.

Foote, A. D., Osborne, R. W., and Hoelzel, A. R. (**2004**). "Whale-call response to masking boat noise," Nature (London) **428**, 910.

Ford, J. K. B. (**1987**). "A catalogue of underwater calls produced by killer whales (*Orcinus orca*) in British Columbia," Can. Data Rep. Fish. Aquat. Sci. **633**, 165.

Ford, J. K. B. (**1989**). "Acoustic behaviour of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia," Can. J. Zool. **67**, 727–745.

Ford, J. K. B. (**1991**). "Vocal traditions among resident killer whales (Orcinus orca) in coastal waters of British Columbia," Can. J. Zool. **69**, 1454–1483.

Ford, J. K. B., and Ellis, G. M. (**1999**). *Transients: Mammal-Hunting Killer Whales of British Columbia, Washington, and Southeastern Alaska* (University of British Columbia Press, Vancouver).

Ford, J. K. B., and Ellis, G. M. (**2006**). "Selective foraging by fish-eating killer whales *Orcinus orca* in British Columbia," Mar. Ecol.: Prog. Ser. **316**, 185–199.

Ford, J. K. B., Ellis, G. M., and Balcomb, K. C. (**2000**). *Killer Whales; The Natural History and Genealogy of Orcinus orca in British Columbia and Washington* (UBC press, Vancouver).

Ford, J. K. B., Ellis, G. M., Barrett-Lennard, L. G., Morton, A. B., Palm, R. S., and Balcomb, III, K. C. (**1998**). "Dietary specialisation in two sympatric populations of killer whales (*Orcinus orca*) in coastal British Columbia and adjacent waters," Can. J. Zool. **76**, 1456–1471.

Hawkins, A. D., and Johnstone, A. D. F. (**1978**). "The hearing of the Atlantic salmon, *Salmo salar*," J. Fish Biol. **13**, 655–674.

Herman, D. P., Burrows, D. G., Wade, P. R., Durban, J. W., Matkin, C. O., Leduc, R. G., Barrett-Lennard, L. G., and Krahn, M. M. (**2005**). "Feeding ecology of eastern North Pacific killer whales from fatty acid, stable isotope, and organochlorine analysis of blubber biopsies," Mar. Ecol. Prog. Ser. **302**, 275–291.

Hoelzel, A. R., Dahlheim, M., and Stern, S. J. (**1998**). "Low genetic variation among killer whales (*Orcinus orca*) in the Eastern North Pacific and genetic differentiation between foraging specialists," J. Hered **89**, 121–128.

Hoelzel, A. R., and Osborne, R. W. (**1986**). "Killer whale call characteristics: Implications for cooperative foraging strategies," in *Behavioral Biology of Killer Whales*, edited by B. Kirkevold and J. S. Lockard (Liss, New York), pp. 373–403.

Hunter, M. L., and Krebs, J. R. (**1979**). "Geographical variation in the song of the great tit (*Parus major*) in relation to ecological factors," J. Anim. Ecol. **48**, 759–785.

Kastak, D., and Schusterman, R. J. (**1998**). "Low frequency amphibious

hearing in pinnipeds: Methods, measurements, noise, and ecology," J. Acoust. Soc. Am. **103**, 2216–2228.
Kastelein, R. A., Bunskoek, P., Hagedoorn, M., Au, W. W. L., and de Haan, D. (**2002**). "Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 334–344.
Knudsen, V. O., Alford, R. S., and Emling, J. W. (**1948**). "Underwater ambient noise," J. Mar. Res. **7**, 410–429.
Krahn, M. M., Herman, D. P., Matkin, C. O., Durban, J. W., Barrett-Lennard, L., Burrows, D. G., Dahlheim, M. E., Black, N., LeDuc, R. G., and Wade, P. R. (**2007**). "Use of chemical tracers in assessing the diet and foraging regions of eastern North Pacific killer whales," Mar. Environ. Res. **63**, 91–114.
Ma, B. B., and Nystuen, J. A. (**2005**). "Passive acoustic detection and measurement of rainfall at sea," J. Atmos. Ocean. Technol. **22**, 1225–1248.
Miller, P. J. O. (**2002**). "Mixed-directionality of killer whale stereotyped calls: A direction of movement cue?" Behav. Ecol. Sociobiol. **52**, 262–270.
Miller, P. J. O., and Bain, D. E. (**2000**). "Within-pod variation in the sound production of a pod of killer whales, *Orcinus orca*," Anim. Behav. **60**, 617–628.
Miller, P. J. O., Samarra, F. I. P., and Perthuison, A. D. (**2007**). "Caller sex and orientation influence spectral characteristics of 'two-voice' stereotyped calls produced by free-ranging killer whales," J. Acoust. Soc. Am. **121**, 3932–3937.
Mossbridge, J. A., Shedd, J. G., and Thomas, J. A. (**1999**). "An 'acoustic niche' for antarctic killer whale and leopard seal sounds," Marine Mammal Sci. **15**, 1351–1357.
Nousek, A. E., Slater, P. J. B., Wang, C., and Miller, P. J. O. (**2006**). "The influence of social affiliation on individual vocal signatures of Northern Resident killer whales (*Orcinus Orca*)," Biol. Lett. **2**, 481–484.
Rydell, J., and Arlettaz, R. (**1994**). "Low-frequency echolocation enables the bat *Tadarida teniotis* to feed on tympanate insects," Proc. R. Soc. London, Ser. B **257**, 175–178.
Saulitis, E. L., Matkin, C. O., Barrett-Lennard, L. G., Heise, K. A., and Ellis, G. M. (**2000**). "Foraging strategies of sympatric killer whale (*Orcinus orca*) populations in Prince William Sound," Marine Mammal Sci. **16**, 94–109.
Slabbekoorn, H., and Peet, M. (**2003**). "Birds sing at a higher pitch in urban noise," Nature (London) **424**, 267.
Slater, P. J. B. (**1986**). "The cultural transmission of bird song," Trends Ecol. Evol. **1**, 94–97.
Slater, P. J. B. (**1989**). "Bird song learning: Causes and consequences," Ethol. Ecol. Evol. **1**, 19–46.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York).
Wenz, G. M. (**1962**). "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956.
Wille, P. C., and Geyer, D. (**1984**). "Measurements on the origin of the wind-dependent ambient noise variability in shallow water," J. Acoust. Soc. Am. **75**, 173–185.
Yurk, H. (**2005**). "The evolutionary history of resident killer whale clans in the northeastern Pacific using vocal dialects," Ph.D. dissertation, University of British Columbia, Vancouver.
Yurk, H., Barrett-Lennard, L., Ford, J. K. B., and Matkin, C. O. (**2002**). "Cultural transmission within maternal lineages: Vocal clans in resident killer whales in southern Alaska," Anim. Behav. **63**, 1103–1119.
Zerbini, A. N., Waite, J. M., Durban, J. W., LeDuc, R., Dahlheim, M. E., and Wade, P. R. (**2007**). "Estimating abundance of killer whales in the nearshore waters of the Gulf of Alaska and Aleutian Islands using line-transect sampling," Mar. Biol. (Berlin) **150**, 1033–1045.

Classification of broadband echoes from prey of a foraging Blainville's beaked whale

Benjamin A. Jones,[a] Timothy K. Stanton, Andone C. Lavery, and Mark P. Johnson
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1053

Peter T. Madsen
Department of Biological Sciences—Zoophysiology, University of Aarhus, Denmark

Peter L. Tyack
Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1053

(Received 5 September 2007; revised 22 November 2007; accepted 4 December 2007)

Blainville's beaked whales (*Mesoplodon densirostris*) use broadband, ultrasonic echolocation signals with a −10 dB bandwidth from 26 to 51 kHz to search for, localize, and approach prey that generally consist of mid-water and deep-water fishes and squid. Although it is well known that the spectral characteristics of broadband echoes from marine organisms vary as a function of size, shape, orientation, and anatomical group, there is little evidence as to whether or not free-ranging toothed whales use spectral cues in discriminating between prey and nonprey. In order to study the prey-classification process, a stereo acoustic tag was deployed on a Blainville's beaked whale so that emitted clicks and the corresponding echoes from targets in the water could be recorded. A comparison of echoes from targets apparently selected by the whale and those from a sample of scatterers that were not selected suggests that spectral features of the echoes, target strengths, or both may have been used by the whale to discriminate between echoes. Specifically, the whale appears to favor targets with one or more nulls in the echo spectra and to seek prey with higher target strengths at deeper depths. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2828210]

PACS number(s): 43.80.Lb, 43.30.Sf, 43.30.Pc, 43.80.Ev [WWA] Pages: 1753–1762

I. INTRODUCTION

Toothed whales emit broadband, ultrasonic signals to navigate and locate prey (Au, 1993). Although these uses of sound have proven difficult to study in the wild, considerable progress has been made with various species of dolphins in captivity. It has been shown that trained dolphins can discriminate between targets with subtle differences in size, shape, and material composition using information contained in the broadband echoes (Nachtigall, 1980). Identifying the specific features of acoustic signals that dolphins use to discriminate between targets is the subject of ongoing research. Findings thus far have led to the speculation that dolphins use spectral information to discriminate between artificial objects (Vel'min and Dubrovskiy, 1976; Dubrovskiy, 1989; Au and Pawloski, 1989; Fuzessery *et al.*, 2004).

In independent studies, research on acoustic scattering from individual marine organisms has shown that different anatomical groups of animals can be classified by the frequency spectra of their backscattered signals (Martin *et al.*, 1996; Stanton *et al.*, 1998; Martin Traykovski *et al.*, 1998). One goal of this research has been to develop acoustic classification schemes based on the frequency spectra of the scattered signals. Constructive and destructive interference of the sound wave, scattered by anatomical features such as tissue interfaces, gas inclusions, and bone, create a frequency-dependent interference pattern. This pattern is specific to the size, shape, orientation, material properties (i.e. sound speed and density), and internal morphology of the scatterer (review by Simmonds and Maclennan, 2005). Studies of broadband acoustic scattering from individual zooplankton (e.g., Stanton *et al.*, 1998) and fish (e.g., Reeder *et al.*, 2004) have shown that structure in the spectra of high frequency, acoustic backscattering is, indeed, due to anatomical features and is highly affected by orientation. In addition, active broadband acoustic systems have been used in efforts to achieve reliable species recognition based on the spectral signature of backscatter from both individual fish (Au and Benoit-Bird, 2003) and aggregations of fish (Simmonds *et al.*, 1996; Zakharia *et al.*, 1996).

These findings, combined with the evidence for spectral selection in captive dolphins, suggests that it is possible that free-ranging toothed whales use spectral features of broadband, acoustic backscattering to aid in the classification of prey. A few studies have shown prey selection by toothed whales in the wild (MacLeod *et al.*, 2006; Whitehead *et al.*, 2003), but the basis for such selection is not known. Observing acoustically based prey selection in free-ranging toothed whales poses a significant challenge: while sounds emitted by echolocating whales can be recorded with surface-deployed hydrophones, the lower amplitude echoes, scattered by prey, are considerably more difficult to measure. Recently, this problem has been overcome, for some species of beaked whales, by using acoustic recording tags (DTAGs)

[a]Author to whom correspondence should be addressed. Electronic mail: ben.jones@navy.mil; current affiliation: U.S. Naval Maritime Forecast Center/Joint Typhoon Warning Center, Pearl Harbor, HI 96860.

affixed directly to echolocating whales (Johnson and Tyack, 2003; Zimmer *et al.*, 2005; Madsen *et al.*, 2005; Johnson *et al.*, 2006). Using these devices, both the whale's emitted signals and echoes from scatterers in the water column are recorded, providing a detailed view of the echolocation process in the wild.

In one such study, Madsen *et al.*, (2005) reported data from a DTAG attached to a *Mesoplodon densirostris*. This species of beaked whale uses broadband signals, or *clicks*, to search for and localize prey consisting of mesopelagic fishes and squid (Johnson *et al.*, 2004; Pauly *et al.*, 1998). Numerous echoes from scatterers in the water column were recorded including some that were associated with series of rapid clicks, called buzzes, emitted by the whale. These distinctive sounds, also observed in other toothed whales and bats, have been associated with prey capture attempts (Madsen *et al.*, 2002; Miller *et al.*, 2004; Griffin, 1958; Johnson *et al.*, 2006). Although the whale ensonified a large number of scatterers on each dive, it selected only a small percentage of them for capture. This suggests that the animal is actively selecting certain types of prey (Madsen *et al.*, 2005); however, the criteria for such selection were not examined in that study.

In this paper a data set from a related study (Johnson *et al.*, 2006) is examined in order to compare the acoustic scattering signatures of the targets selected by the whale with those not selected. Both the spectral characteristics and relative target strengths of individually resolved echoes are examined. Significant differences were observed with respect to spectral content of selected and nonselected targets. Moreover, differences in the echoes from prey selected from two different depth strata suggest that at least two distinct prey types may be targeted by the whale. Finally, the results are qualitatively compared with previous laboratory and modeling studies of broadband scattering from marine organisms to provide context for echo-to-echo variation.

II. MATERIALS AND METHODS

A. Signals emitted by the Blainville's beaked whale

M. densirostris emit two different signals associated with different phases of echolocation-based foraging (Johnson *et al.*, 2004, 2006). Distinctive frequency modulated (FM) clicks are produced at inter-click intervals (ICIs) of approximately 0.4 s while the whale searches and approaches prey. When the whale closes on a selected prey, it emits a rapid series of unmodulated clicks called a buzz. These clicks can be readily distinguished from the slower FM clicks by a lower output level and an ICI well below 0.1 s (Madsen *et al.*, 2005). The FM clicks are relatively long with a duration of about 270 μs (Fig. 1(a)). The frequency spectrum of these clicks is within 10 dB of the maximum level between approximately 26 and 51 kHz with very little energy at lower frequencies (Johnson *et al.*, 2006). When observed close to the acoustic axis, FM clicks appear to have a consistently broad and relatively featureless spectrum, i.e., free of peaks and nulls (Fig. 1(b)), while off-axis clicks, generally, have more complex spectra (Johnson *et al.*, 2006). Although the beamwidth of the *M. densirostris* FM clicks is not known, Zimmer *et al.* (2005) has estimated the −3 dB beam width of another larger beaked whale with similar click characteristics, *Ziphius cavirostris*, at about 6°.

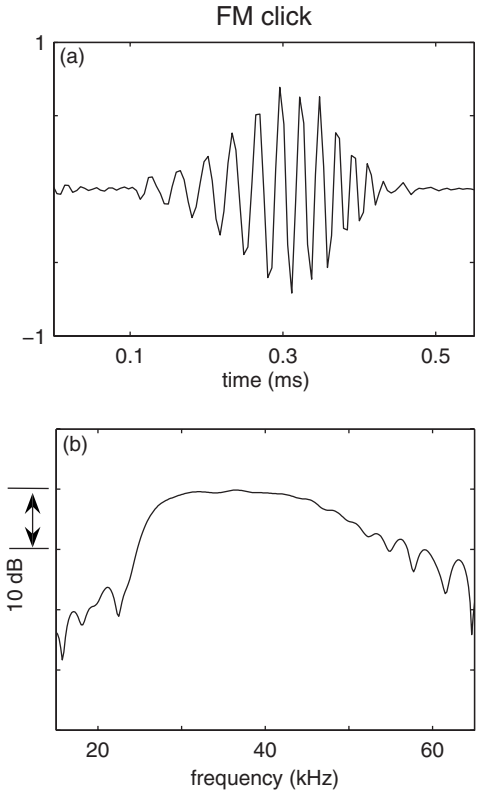


FIG. 1. Echolocation signal (FM click) of *M. densirostris*, observed during this study. The whale recorded was a conspecific foraging with the whale discussed in this study. (a) Time series (arbitrary vertical scale), (b) frequency spectrum.

Although echoes from scatterers ensonified by FM clicks are frequently detectable in tag acoustic recordings, significantly fewer echoes are observed during buzzes. This is likely due to the lower output level of buzz clicks (Madsen *et al.*, 2005; Johnson *et al.*, 2006). For this reason, only echoes from targets ensonified by FM clicks are considered in the spectral and target strength analyses presented in this paper. Buzzes are used as an indication that the whale has selected a prey item and is approaching to capture it (Madsen *et al.*, 2005).

B. Data acquisition

1. Instrumentation

Data reported here were collected with a miniature acoustic and orientation recording tag, the DTAG (Johnson and Tyack, 2003). The tag attaches to the surface of a whale with four suction cups and automatically releases after a programmed period of time. The DTAG contains two hydrophones spaced 2.5 cm apart which are sampled synchronously at a rate of 192 kHz per channel. The hydrophones were calibrated in a pressure test facility to a depth of 500 m. The overall frequency response is flat within 3 dB between 0.5 and 67 kHz with the upper cutoff resulting from an antialias filter in the tag. The tag also samples a three-axis ac-

celerometer and a three-axis magnetometer at 50 Hz for measuring orientation, and a pressure sensor for extracting depth.

2. Field measurements

An adult *M. densirostris*, tagged in October of 2004 near the Canary Islands, provided 9.5 h of acoustic data. The position of the tag during the dives was determined using photo documentation of the whale when surfacing between dives. The tag was initially placed on the right side of the whale but slid to a dorsal position approximately 1 m posterior of the blowhole after about 1 h. This position minimized shading of the hydrophones by the whale's body providing favorable conditions to record echoes from scatterers ensonified by the whale. The tag remained near the dorsal ridge for the following three foraging dives. It is the data recorded during these dives that are examined in this study.

C. Analysis methods

Use of the whale's biosonar as a sound source and the animal's body as the platform for the receiver introduces several complexities in studying acoustic backscattering. First, it is not possible to make far-field measurements of the sounds emitted by the whale on which the tag is mounted. This means that the level and spectrum of the outgoing signal are not precisely known. In order to estimate the frequency response of scatterers, we restrict the analysis to the frequency range over which echolocation signals, recorded from other conspecifics, have been observed to be essentially featureless (Johnson *et al.*, 2006). Furthermore, to estimate the relative target strengths of scatterers we use the near field recording of the whale's emitted signal as a proxy for output level and compensate for click to click variation in the amplitude of the emitted signal. A second complexity is that the position of the scatterer with respect to the whale's sonar beam can have a significant influence on spectral content. In order to reduce uncertainties associated with this effect, only echoes with echo-to-noise ratio (ENR) above 15 dB are chosen. Since scatterers located near the axis of the sonar beam will tend to produce the highest echo levels, selecting high ENR echoes ensures that a preponderance of the echoes examined are nearly on axis.

The approach taken here is to identify series of echoes associated with FM clicks, that appear to represent individual targets. Several parameters are computed for each echo (e.g., the number of nulls in the echo spectrum and the relative target strength) in the series. The parameters of the echoes in each series of echoes, or *echo train*, are compiled into a single statistic which can then be compared across trains. The objective is to determine if there are consistent differences in the parameters of echoes from targets apparently selected by the whale as compared to those not selected. Variation in the echo parameters will also be related to the depth and the overall density of acoustic scatterers in the environment.

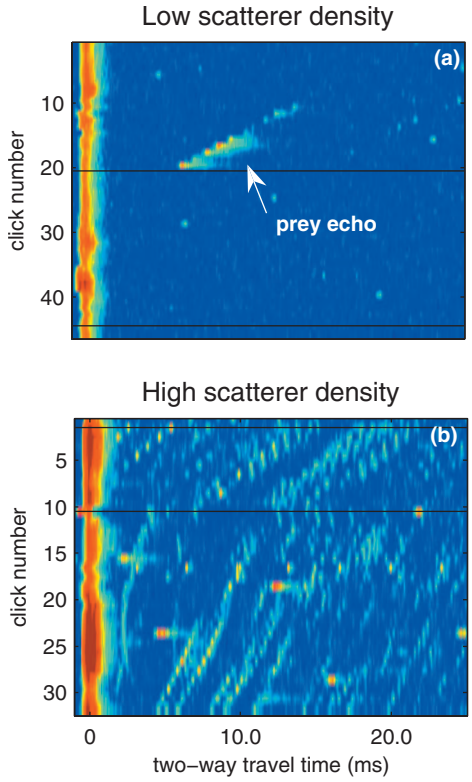


FIG. 2. Two representative examples of echograms displaying scattered FM clicks of a foraging beaked whale which are used to identify echo trains: (a) low density of scatterers, (b) high density of scatterers. Click number is shown on the vertical axis and the horizontal axis shows time since last emitted click. Echo strength is indicated by color with red corresponding to higher sound pressure and blue corresponding to lower sound pressure. Plot (a) shows a single echo train as the whale approaches a target. The horizontal black line represents a period of greater than 1 s between outgoing clicks that meet a preset threshold level which, in this case, indicates the whale's switch to the lower amplitude buzz.

1. Identification of buzzes, echoes and echo trains

This analysis was restricted to the section of the tag recording made during the three foraging dives in which the tag was dorsally located. A supervised, click detection algorithm was used to identify FM clicks and buzzes (Johnson *et al.*, 2006). Echograms were then constructed from the sequences of FM clicks produced by the tagged whale (see Fig. 2) by aligning the envelopes of 25 ms sections of the tag recording starting at each click (Johnson, 2004). The duration of the sections restricts the maximum two-way travel time (TWTT) of echoes to 25 ms corresponding to a maximum range from source to scatterer of 19 m. This TWTT limit ensures adequate echo-to-noise ratio, ENR, for later processing. Echo trains from a single scatterer were readily identifiable in the echograms as series of echoes with consistent, slowly varying TWTT (Fig. 2(a)).

Segments of the tag sound recording with a duration of 400 μs, centered on each echo from an individual target, were extracted for analysis. This window size was selected to accommodate the increased duration of echoes of FM clicks caused by scattering from a finite-sized object. Another 400 μs segment was taken 600 μs prior to each echo as a contemporary sample of the noise levels. The signal and noise segments were filtered with a 12-pole (six low-pass

and six high-pass) Butterworth bandpass filter with cutoff frequencies of 20 and 70 kHz to reduce out-of-band noise. The echo level, *EL*, and the noise level, *NL*, were estimated as $10\log_{10}(.)$ of the signal power in each echo and noise segment, respectively. The ENR was then estimated from the difference *EL*−*NL*. The magnitude of the spectrum of each echo was estimated by calculating a 512-bin fast Fourier transform (FFT).

2. Whale-selected and nonselected targets

Echo trains were grouped according to whether they occurred in the proximity of a buzz, indicating that the whale tried to capture the target, or not. Echo trains that terminated within 5 s of a buzz were considered to represent a whale-selected target (Madsen *et al.*, 2005). If multiple echo trains were detectable prior to a buzz, indicating multiple scatterers, the echogram was expanded in the vicinity of the buzz to look for echoes associated with low-level buzz clicks. If an echo train associated with the buzz was detected, then the echo train immediately prior to the buzz that most closely matched the TWTT of the echoes associated with the buzz was chosen as the whale-selected target. However, if no echo was visible from the buzz clicks, or two echo trains were very close together in time prior to the buzz, that sequence was not considered in the analysis. An echo observed during the buzz, though, was not a criterion for a whale-selected target if there was only one echo train leading up to the buzz. Echo trains occurring more than 5 s prior to a buzz were considered to represent targets not selected by the whale. A subset of these echo trains was randomly chosen for comparison using the same maximum TWTT criteria as for whale-selected echo trains.

3. Spectral classification

Acoustic scattering spectra of marine organisms are characterized by interference patterns specific to the size, shape, material properties, and orientation of the organism. This information should be contained in the echoes recorded by the DTAG; however, the echo spectrum will also be influenced by the spectrum of the sound source and any interfering noise. While the effects of noise can be minimized by selecting echoes with a high ENR, the spectrum of the sound source will depend on the unknown angle of the scatterer to the axis of the outgoing sonar click. Given the broad, smooth spectrum of on-axis FM clicks, targets close to the acoustic axis should give rise to echoes with spectral features associated only with the scattering characteristics of the target. In contrast, off-axis targets will produce echoes with features due to both the target and off-axis distortion.

Assuming *M. densirostris* have a beam pattern similar to *Z. cavirostris* and delphinids, the output energy of the FM clicks will decrease rapidly with increasing off-axis angle outside of the main beam (Zimmer *et al.*, 2005). Thus, by placing a minimum ENR criterion on echoes chosen for analysis a set can be obtained with a preponderance of echoes from on-axis targets.

Given the limited ENR of echoes available in this study, three robust parameters were chosen to characterize the echo spectra: (1) number of nulls, (2) frequency of the first null, and (3) frequency interval between the first two nulls (Fig. 3). Nulls were defined as any dip in the spectral magnitude with greater than 4 dB difference between the local minimum and the average of the two adjacent maxima. To avoid detecting small ripples in the spectrum, possibly due to noise, nulls were rejected if either adjacent peak was less than 1 dB above the null. The frequency band between 27 and 50 kHz was selected for analysis. This range falls within the −10 dB end points of the FM click signal and is also within the range of frequencies over which on-axis FM clicks appear to be largely featureless (Fig. 1(b)). To verify that on-axis clicks from *M. densirostris* are indeed featureless in this band, we counted the number of spectral nulls in a set of 144 likely on-axis clicks from a conspecific measured by Johnson *et al.* (2006). According to the definition given above, only six (<5%) of these clicks had one or more nulls supporting the notion that on-axis FM clicks are essentially featureless.

A Monte Carlo simulation, using 1000 trials, was performed by combining an on-axis *M. densirostris* click with artificial, band-limited (20–70 kHz) noise to determine the likelihood that noise in the simulated echo signal would give rise to spectral nulls. With an in-band ENR of 15 dB, fewer than 2% of the trials (14 of 1000) resulted in a spectral null of at least 4 dB in depth in the frequency range of interest (27–50 kHz). No simulated echo had more than one spectral

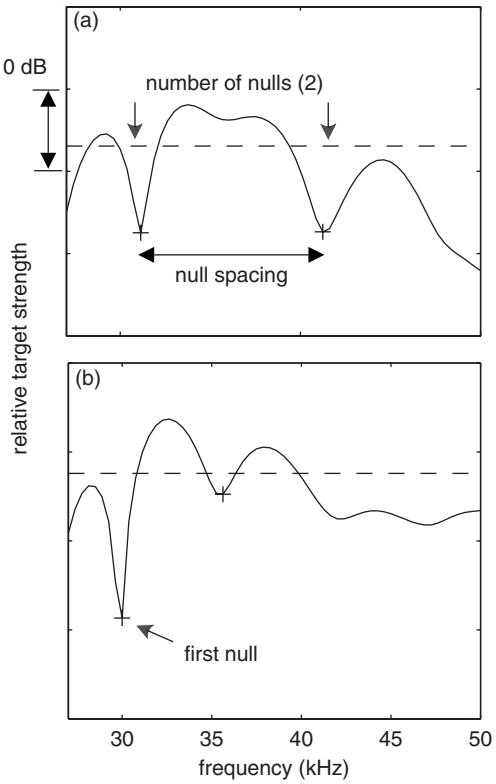


FIG. 3. Example of automated structure analysis results. Plus signs indicate nulls and horizontal dashed lines indicate mean relative target strength of frequency band displayed. An example of the two criteria for a null are illustrated in the plot (b): (1) The dip near 42 kHz does not meet the criterion of being more than 1 dB lower than both adjacent peaks; (2) the dip at 47 kHz does not meet the criterion of being more than 4 dB lower than the average of the two adjacent peaks.

null. Thus, by selecting echoes with ENR > 15 dB, nulls in the resulting echo spectra are most likely due to the scattering characteristics of the target, as opposed to off-axis distortion or noise.

4. Target strength estimation

Target strength, *TS*, can be calculated from the active sonar equation (Urick, 1983):

$$TS = EL + 2TL - SL, \quad (1)$$

where *TL* and *SL* are the one-way transmission loss and source level, respectively, in decibels. The *TL* for spherical spreading is given by $TL = 20 \log r + \alpha r$, where r is the one-way distance from the source to the scatterer and α is the absorption loss in dB/m. Absorption can be neglected as its effect on *TL* for short target distances is small (<0.5 dB) in the frequency range considered here. All relative target strengths presented in this study are band averaged, in which the mean is calculated prior to logarithmic conversion.

Although *EL* and *TL* can be estimated from the tag recordings, the *SL* of *M. densirostris* is not precisely known and may vary from click to click. The location of the tag behind the whale's head, in the near field and off axis with respect to the sound source, prevents measurements of *SL*. However, measurements of the outgoing FM clicks by the tag provide a proxy for changes in the power output of the on-axis emissions from click to click (Madsen *et al.*, 2005). Thus, the relative *TS* (i.e., the *TS* with respect to an unknown source level) can be estimated by replacing *SL* by S_h, the received level in decibels of the outgoing click as recorded by the tag. The relative *TS*, although not useful in an analysis requiring absolute measures of target strength, can be used in quantitative comparisons between echoes observed at a given tag position on the same whale.

5. Statistical comparison of echo trains

A statistical comparison of echo parameters across individual echoes is complicated by the fact that echoes within the same echo train represent the same target, and so are not independent samples. To overcome this problem, we use echo trains as the unit of analysis and combine the parameters of echoes within the train into a summary statistic. For each parameter (e.g., number of nulls), a set of discrete levels or bins was selected. The proportion of the echoes in each echo train with a parameter value falling in each bin was calculated producing a histogram of that parameter for each echo train. This reduces the variable length echo trains into fixed-length vectors suitable for statistical analysis. All histograms were normalized by the number of echoes in each echo train. As an example of this technique, consider an echo train containing four echoes with 3, 2, 0, and 3 spectral nulls in the echoes. If the null-count parameter was represented by bins of 0, 1, 2, … nulls, the histogram for this train would be [0.25 0 0.25 0.5 0…0]. As previously described, echo trains were differentiated according to whether they occurred close to a buzz or not (i.e., whether the whale appeared to select the target as a prey item or not). Thus, two sets of histograms were generated summarizing the echoes in whale-selected and nonselected targets.

6. Density of scatterers

In addition to the relative TS and spectral content of echoes, the depth and density of scatterers were also estimated from the data. Scatterer density was inferred from the number of echoes visually detected in the echogram within a 25 ms window following each FM click (Fig. 2). The number of echoes per click, in the proximity of buzzes, had a strongly bimodal distribution with either few (1–5) echoes or many (>10) echoes being excited by each click. Only three of the 47 buzzes analyzed occurred in an environment of intermediate scatterer density (i.e., with between five and ten echoes recorded for each FM click prior to the buzz). Based on this, we define intervals in the data in which there were more than five echoes per click as having high scatterer density and the converse for low scatterer density.

III. RESULTS

During dives 2 through 4, the dives during which the tag was located near the dorsal ridge, the whale dove to depths between 600 and 1320 m. FM clicks were only observed at depths greater than 440 m. As indicated by buzzes observed during these dives, the whale hunted at various meso- and bentho-pelagic depths. During dive 3 the whale hunted near the seafloor at depths between 580 and 680 m. Foraging during dives 2 and 4 appeared to be spread out in depth over several hundred meters; although, in both cases the majority of buzzes were concentrated in depth ranges of less than 200 m.

A. Echoes selected for analysis

A total of 89 buzzes were observed during the three dives (34, 24, and 31 in dives 2, 3, and 4, respectively) indicating prey capture attempts. Of these, 47 were preceded by unambiguous echo trains that are assumed to correspond to scattering from the selected prey. The remaining buzzes were either associated with irresolvable echo trains due to a cluttered acoustic environment or not associated with echo trains with sufficient ENR. In dive 3, a high percentage of buzzes could not be correlated with echo trains due to reverberation from the nearby seafloor. In general, when an echo train was associated with a buzz, the first echo of the train was discernable from background noise at a distance from the tag to the scatterer of between 5 and 15 m. In each case the echo train terminated shortly before the start of the buzz at a distance of 3–5 m.

The 47 echo trains identified as corresponding to whale-selected prey contained a total of 426 discrete echoes. Of that total, 132 echoes, in 37 echo trains, met the minimum ENR criterion. In order to accumulate a sufficient number of echoes for comparison, 92 echo trains from scatterers not selected by the whale were chosen at random. Of these, 42 trains containing 120 echoes remained after applying the same ENR criterion. The number of echoes analyzed from each dive is shown in Table I. Many more nonselected echo

TABLE I. Distribution of echo trains (bold) and echoes (parenthesis) from whale-selected and nonselected scatterers observed in three dives examined. All echoes selected for analysis (bottom half of table) had an echo-to-noise ratio, ENR, of at least 15 dB.

	Whale selected	Nonselected
Total identified		
Dive 2	**18** (139)	**36** (276)
Dive 3	**9** (77)	**18** (130)
Dive 4	**20** (209)	**38** (177)
Total	**47** (425)	**92** (583)
Selected for analysis		
Dive 2	**14** (43)	**22** (67)
Dive 3	**5** (18)	**12** (37)
Dive 4	**18** (71)	**8** (16)
Total	**37** (132)	**42** (120)

trains were required to obtain a sample size similar to that of the whale-selected echo trains as they generally had fewer echoes in each train. This is likely due to the shorter length of time that the whale maintained these nonselected scatterers within its acoustic beam.

B. Echo classification

1. Spectral characteristics

The spectra of most individual echoes from prey selected by the whale contained one or more nulls within the frequency band examined (e.g., Fig. 4). Additionally, echo-to-echo variability was observed in the spectra of some echo trains (e.g., Figs. 4(a) and 4(b)). Echoes from whale-selected targets had more spectral nulls than did echoes from scatterers not selected by the whale (median: whale-selected: 2, nonselected:1; Wilcoxon rank-sum $p=0$, $n=79$). Notably, less than 12% of the echoes within whale-selected echo trains were characterized by featureless echoes (i.e., echoes with no nulls meeting the criteria defined above), whereas more than 48% of the echoes within nonselected echo trains were composed of such echoes (Fig. 5(a)).

A comparison of other spectral features between the whale-selected and nonselected echo trains show less notable differences. No significant differences were found in the medians of the first null locations and the spacings between the first two nulls (Figs. 5(b) and 5(c)) between whale-selected and nonselected targets [(Wilcoxon rank-sum $p=0.11$, $n=64$ (null location), $p=0.14$, $n=45$ (null spacing)].

2. Target strengths

Echo trains of targets selected by the whale were composed of echoes with significantly higher relative target strengths [median difference of 12 dB, Wilcoxon rank-sum: $p=0$, $n=79$ (Fig. 5(d))] than nonselected targets. Echo trains from whale-selected prey also showed relatively wide echo-to-echo variation in relative target strengths. For echo trains containing at least two echoes, 73% contained echoes that varied by at least 3 dB and 40% contained echoes varying by at least 6 dB. Echo trains of targets not selected by the whale had less variability with 55% of echo trains having at least a 3 dB variation and none varying by more than 6 dB. This difference in variability between echo trains of targets selected by the whale and those not selected may be due, in part, to the fact that there are fewer echoes, on average, in the echo trains of nonselected targets.

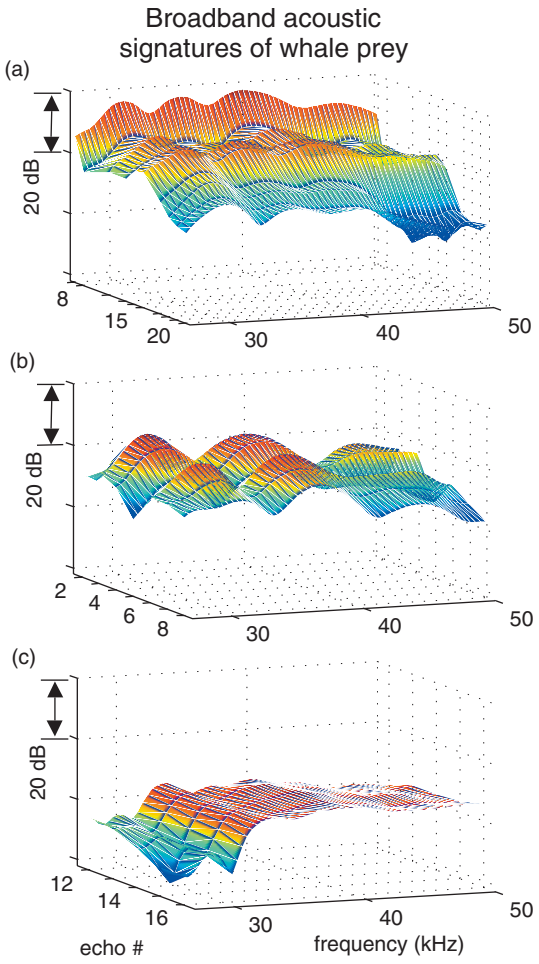


FIG. 4. Broadband acoustic signatures of three prey selected by the whale. Each plot is comprised of frequency spectra of a series of echoes that make up one echo train. Only echoes meeting the ENR criterion are included, hence not all echoes in each echo train are displayed. The echo number of displayed echoes are indicated by the presence of a tick mark. Echo values are interpolated across the gaps associated with echoes not chosen such as between echoes 15 and 17 in plot (a). Plots (a) and (b) show examples of high target strength prey observed at deep depths (below 700 m) in environments with low scatterer density. Plot (c) shows an example of lower target strength prey observed in shallower water (above 700 m) in environments with high scatterer density.

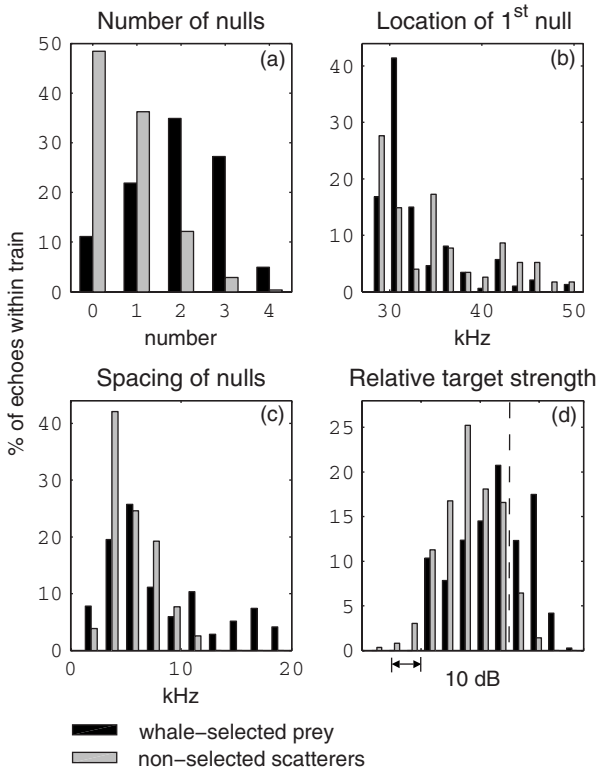


FIG. 5. Normalized distributions of various characteristics of echoes from whale-selected prey and randomly chosen scatterers not selected by the whale (black and gray bars, respectively). The characteristics are: (a) number of nulls, (b) location of first null, (c) spacing between the first two nulls, and (d) relative target strength. Each characteristic of an individual echo is weighted by the fraction of the total number of echoes within a train that the echo represents. Distributions are normalized by the total number of echo trains (whale selected or nonselected). Vertical dashed line in target strength plot (d) represents absolute target strength value of −65 dB based on an arbitrary source level of 200 dB.

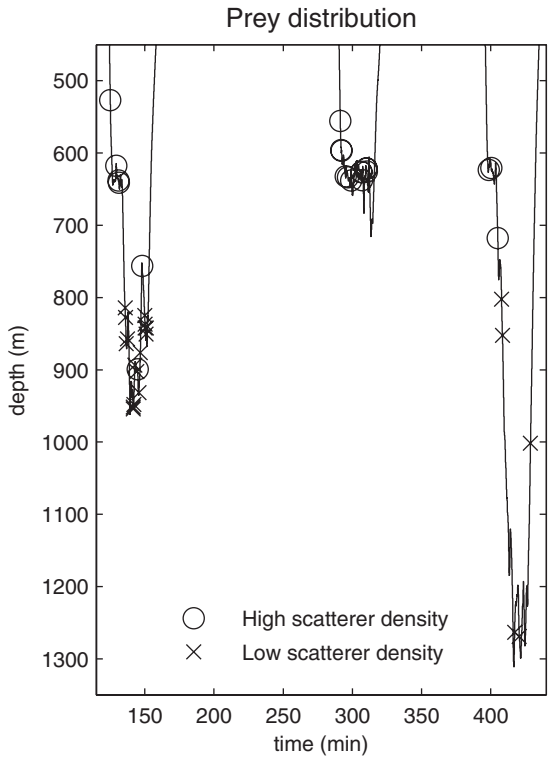


FIG. 6. Depth distribution of prey categorized by scatterer density. Depth is truncated above 450 m as no predation events were observed at shallower depths.

C. Comparison of echo characteristics between two groups of prey

Scatterer density in regions where the whale hunted varied widely within the three dives (Fig. 2) and two distinct regimes can be identified. Scatterer density was low during much of dives 2 and 4 while high scatterer density was observed throughout the shallower and benthic dive 3 (Fig. 6). The two density regimes were separated spatially in the water column with the high density group found, generally, in water shallower than 700 m and the lower density group found at greater depths. A comparison was made between these two groups in which the three outliers (i.e., scatterers in high density environments, deeper than 700 m) were discounted.

In both regimes, the targets selected by the whale were characterized by more highly structured echoes than the group of nonselected scatterers. In the low scatterer density case (Fig. 7(b)) the median number of nulls for whale-selected and nonselected targets were two and one, respectively (Wilcoxon rank-sum $p=0$, $n=52$). Similarly, in the high scatterer density case (Fig. 7(a)), the median values were one and zero, respectively (Wilcoxon rank-sum $p=0.04$, $n=27$); although, the result in the high density regime is not significant if Bonferroni correction is taken into account on the division of data. As in the combined results, a relatively small percentage of echoes in whale-selected echo trains were composed of featureless echoes in both the shallow, high density group (whale selected: 14%, versus nonselected: 57%) and the deep, low density group (whale selected: 11%, nonselected: 40%). Finally, a comparison between the two environments (high and low scatterer density) of only whale-selected targets showed no significant difference in the number of nulls.

The distribution of relative target strengths of the two populations provided further information about these two scattering groups. The target strength distribution of whale-selected and nonselected echoes within each environment (i.e., shallow, high scatterer density and deep, low scatterer density) are overlapping (Fig. 8) with no significant difference noted in the median target strengths in the shallow, high scatterer density case (Wilcoxon rank-sum $p=0.36$, $n=27$). A difference of 6 dB in the medians was observed in the deep, low scatterer density case with whale-selected targets having higher target strengths ($p=0.02$, $n=52$). Furthermore, a comparison between the two environments (i.e., high and low echo density regimes) of only the whale-selected targets reveals a 15 dB difference in the median target strengths with the targets in the deep, low density regime having higher values ($p=0$, $n=37$).

In order to explore the implication of this difference in target strengths, a scattering model was used to estimate the possible difference in average scatterer size. Using a simple, finite cylinder model for randomly oriented scatterers of length, L, target strength averaged over orientation varies as $10\log(L^2)$ in the geometric scattering region (Stanton *et al.*, 1993). This first-order approximation can be used to estimate

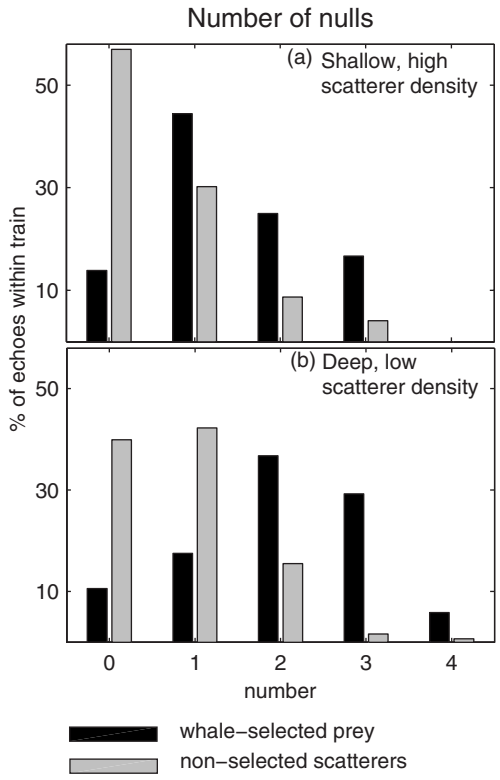


FIG. 7. Normalized distributions of number of nulls in frequency responses of selected and nonselected scatterers (black and gray bars, respectively) in (a) shallow, high density aggregations and (b) deep, low density aggregations. Weighting and normalization are identical to Fig. 5.

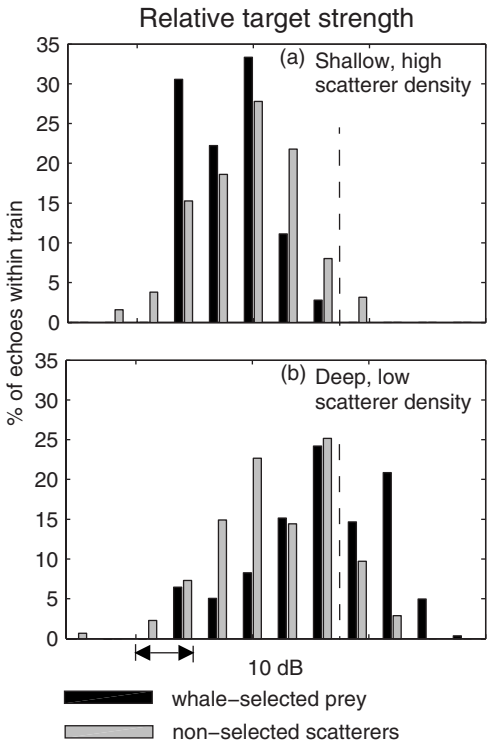


FIG. 8. Normalized distributions of relative target strengths of whale-selected prey and nonselected scatterers (black and gray bars; respectively) in: (a) shallow, high density aggregations, and (b) deep, low density aggregations. Vertical dashed line represents absolute target strength value of −65 dB based on an arbitrary source level of 200 dB. Weighting and normalization are identical to Fig. 5.

a ratio of the lengths of prey from the difference observed in average target strengths. For scattering from two organisms of the same shape and material composition, a 15 dB difference in target strengths correlates to a length factor of approximately 5.5.

IV. DISCUSSION AND CONCLUSIONS

The characteristics of echoes from targets ensonified by broadband echolocation signals have been studied through data obtained, *in situ*, by a recording device mounted on a foraging Blainville's beaked whale. By setting stringent criteria on the echoes analyzed and, in part, due to the opportunely smooth spectrum of the whale's emitted signal, the spectral characteristics of backscattering by prey have been analyzed and relative target strengths have been estimated.

Significant structure, resembling the type of interference patterns observed when marine organisms scatter broadband sound in laboratory experiments exist in the frequency spectra of echoes measured in this study. Additionally, echo-to-echo variability in terms of both spectral structure and relative target strengths is observed within some echo trains selected by the whale. Although both spectral distortion of the emitted signal, related to the prey location within the acoustic beam, and noise can induce spectral nulls in the received echoes, these effects have been minimized by selecting only high ENR echoes. Therefore, variations in echo characteristics within echo trains, arguably, relate to the orientation of the scatterer with respect to the incident sound beam rather than location in the beam. Other modeling studies and laboratory experiments of the broadband spectra of echoes from fish and squid show a qualitatively similar variability (Reeder *et al.*, 2004; Jones, 2006). In these studies of scattering from individual organisms, changes in orientation as small as 5° have been shown to dramatically change the backscattering spectra.

Comparing all groups of echoes from scatterers selected by the whale to other randomly chosen, nonselected scatterers, we have shown that the targets preyed upon by the whale have a higher degree of structure within their frequency responses. Conversely, the whale was much less likely to prey upon targets characterized by featureless echoes. This result is contrary to what would be expected if structure in the echoes resulted primarily from the location of the scatterer with respect to the axis of the sonar beam. There is no behavioral reason to suppose that whale-selected targets would tend to be ensonified further off-axis than nonselected targets. In fact, the contrary ought to be the case as is seen in bats (Moss and Ghose, 2003), i.e., the whale likely orients itself so as to ensonify selected targets close to its acoustic axis. Thus, neglecting the frequency response of the scatterer, one would expect to see more spectral nulls in the nonselected targets than in those selected for capture attempts. Here we see a significant trend in the opposite direction indicating that the observed spectral features are most likely due to the frequency-dependent target strength of the scatterer.

Given our lack of knowledge of the type, size, and orientation of the ensonified scatterers, no definite statement

can be made regarding the type or size of organisms that might produce the structured versus featureless echoes, or the reason for the whale preferring more structured returns. It is even possible that the whale's decision process is completed at a further range than that of the targets considered in this study. It must be assumed that the whale's range of detection is greater than that of the tag-mounted hydrophones. The combination of directivity of the whale's aural receiving system, as shown by Au and Moore (1984) in other toothed whales, and poorer ENR of the hydrophone outputs due to flow noise over the tag should provide the whale with a significantly better ENR. Nonetheless, it is clear from this study that there is a significant difference in the echo characteristics of prey selected by the whale and those not selected when within the range of our sensor. Possible explanations for increased structure in the selected scatterers are that they are a different species, different size, and/or in a different orientation when ensonified. Stomach content data, net samples, and acoustic scattering studies of the species present in the areas and depths at which Blainville's beaked whales forage are needed to narrow down the type and size of organism giving rise to the echoes, and to ultimately allow the results to be interpreted at an ecological level.

Two depth-stratified regimes were identified which exhibited differences in scatterer density and prey target strength. The whale hunted prey with lower target strengths in high density aggregations between depths of 600 and 650 m. Prey with higher target strengths were selected by the whale at depths below 700 m in environments with low scatterer density. In the shallow, high scatterer density environment the target strength distributions of selected and nonselected targets showed no significant difference suggesting that the relative target strength of a scatterer is not the whale's sole means of discriminating between prey and nonprey targets. In the low scatterer density environment it is clear that the whale favored targets with a significantly higher degree of structure in their echo spectra. Combined, these results provide some evidence of spectral-based prey selection, though further data will be required to determine whether this is the whale's primary method of discrimination.

Finally, an explanation has been proposed as to why this large marine predator chooses to dive to significantly deeper depths, and thus expend greater amounts of energy (Tyack *et al.*, 2006), to hunt less dense populations of organisms. A significant difference in the relative target strengths for whale-selected targets was observed between the shallow, high density aggregation and the deep, low density aggregation. While several parameters can give rise to a difference in target strengths (e.g., size, shape, tissue composition and gas inclusions), we have shown that, by allowing length to dominate, the higher target strengths observed at deeper depths could be explained by a prey size some 5–6 times that of prey hunted in shallower water. This, in turn, could give the whale an incentive to expend the energy required to dive to greater depths.

Little is known about the potential prey communities in the deep ocean due to the difficulty and expense of studying these habitats. The method presented here, using a deep-diving, echolocating predator as a natural source of sound and analyzing the scattered signals, is a novel way to explore this challenging environment. It is hoped that future studies will provide data which can be used to further analyze these whales' behavior and characterize the prey field where they hunt.

ACKNOWLEDGMENTS

Thanks to Natacha Aguilar de Soto and the El Hierro field team of the University of La Laguna for data collection. Thanks also to Tom Hurst and Alex Shorter for DTAG construction. Field work was supported by the U.S. National Oceanographic Partnership Program, the U.S. Office of Naval Research, and the Canary Islands government. Research was conducted under a permit from the government of the Canary Islands to the University of La Laguna and was approved by the WHOI Institutional Animal Care and Use Committee. Analysis of the data was supported by the Office of the Oceanographer of the U.S. Navy, The Academic Programs Office at the Woods Hole Oceanographic Institution and the Danish Natural Science Research Council through a Steno scholarship to Peter T. Madsen.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). "Acoustic backscattering by Hawaiian lutjanid snappers. II. Broadband temporal and spectral structure," J. Acoust. Soc. Am. **114**, 2767–2774.

Au, W. W. L., and Moore, P. W. B. (**1984**). "Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin *Tursiops truncatus*," J. Acoust. Soc. Am. **75**, 255–262.

Au, W. W. L., and Pawloski, J. L. (**1989**). "Detection of noise with rippled spectra by the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **86**, 591–596.

Dubrovskiy, N. (**1989**). "On the two auditory subsystems in dolphins," in *Proceedings of a NATO Advanced Workshop and Symposium of the Fifth International Theriological Congress on Sensory Abilities of Cetaceans*, edited by R. Thomas and J. A. Kastelein, NATO ASI Series (Plenum, Rome, **1990**), pp. 233–254.

Fuzessery, Z., Feng, A., and Supin, A. (**2004**). "Central auditory processing of temporal information in bats and dolphins," in *Echolocation in Bats and Dolphins*, edited by C. F. V. M. Thomas and J. A. Moss (University of Chicago Press, Chicago), pp. 115–122.

Griffin, D. R. (**1958**). *Listening in the Dark; the Acoustic Orientation of Bats and Men* (Yale University Press, New Haven).

Johnson, M., Madsen, P., Zimmer, W., Aquilar-Soto, N., and Tyack, P. (**2006**). "Foraging Blainville's beaked whales (*Mesoplodon densirostris*) produce distinct click types matched to different phases of echolocation," J. Exp. Biol. **209**, 5038–5050.

Johnson, M., Madsen, P. T., Zimmer, W. M. X., de Soto, N. A., and Tyack, P. L. (**2004**). "Beaked whales echolocate on prey," Proc. R. Soc. London, Ser. B **271**, S383-S386.

Johnson, M. P., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng., **28**, No. 1, 3–12.

Jones, B. A. (**2006**). "Acoustic scattering of broadband echolocation signals from prey of Blainville's beaked whales: Modeling and analysis," Master of Science, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution.

MacLeod, C. D., Santos, M. B., Lopez, A., and Pierce, G. J. (**2006**). "Relative prey size consumption in toothed whales: Implications for prey selection and level of specialization," Mar. Ecol.: Prog. Ser. **326**, 295–307.

Madsen, P. T., Johnson, M., de soto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.

Madsen, P. T., Wahlberg, M., and Mohl, B. (**2002**). "Male sperm whale (*Physeter macrocephalus*) acoustics in a high-latitude habitat: Implications for echolocation and communication," Behav. Ecol. Sociobiol. **53**, 31–41.

Martin, L. V., Stanton, T. K., Wiebe, P. H., and Lynch, J. F. (**1996**). "Acous-

tic classification of zooplankton," ICES J. Mar. Sci. **53**, 217–224.
Martin Traykovski, L. V., Stanton, T. K., Wiebe, P. H., and Lynch, J. F. (**1998**). "Model-based covariance mean variance classification techniques: Algorithm development and application to the acoustic classification of zooplankton," IEEE J. Ocean. Eng. **23**, 344–363.
Miller, P. J. O., Johnson, M., and Tyack, P. L. (**2004**). "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture," Proc. R. Soc. London, Ser. B **271**, 2239–2247.
Moss, C. F., and Ghose, K. (**2003**). "The sonar beam pattern of a flying bat as it tracks tethered insects," J. Acoust. Soc. Am. **114**, 1120–1131.
Nachtigall, P. E. (**1980**). "Odontocete echolocation performance on object size, shape, and material: in animal sonar systems," in *Proceedings of the Second International Interdisciplinary Symposium on Animal Sonar Systems*, edited by J. Busnel and R. G. Fish (Plenum, Jersey, Channel Islands), pp. 71–95.
Pauly, D., Trites, A. W., Capuli, E., and Christensen, V. (**1998**). "Diet composition and trophic levels of marine mammals," ICES J. Mar. Sci. **55**, 467–481.
Reeder, D. B., Jech, J. M., and Stanton, T. K. (**2004**). "Broadband acoustic backscatter and high-resolution morphology of fish: Measurement and modeling," J. Acoust. Soc. Am. **116**, 747–761.
Simmonds, E. J., Armstrong, F., and Copland, P. J. (**1996**). "Species identification using wideband backscatter with neural network and discriminant analysis," ICES J. Mar. Sci. **53**, 189–195.
Simmonds, E. J., and Maclennan, D. N. (**2005**). *Fisheries Acoustics*, Fish and Fisheries (Blackwell Science Ltd., Oxford).
Stanton, T. K., Chu, D., Wiebe, P. H., and Clay, C. S. (**1993**). "Average echoes from randomly oriented random-length finite cylinders: Zooplankton models," J. Acoust. Soc. Am. **94**, 3463–3472.
Stanton, T. K., Chu, D., Wiebe, P. H., Martin, L. V., and Eastwood, R. L. (**1998**). "Sound scattering by several zooplankton groups. I: Experimental determination of dominant scattering mechanisms," J. Acoust. Soc. Am. **103**, 225–235.
Tyack, P., Johnson, M., Aquilar-Soto, N., Sturlese, A., and Madsen, P. T. (**2006**). "Extreme diving of beaked whales," J. Exp. Biol. **209**, 4238–4253.
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd edition (McGraw–Hill, New York).
Vel'min, V. A., and Dubrovskiy, N. A. (**1975**). "Auditory analysis of sound pulses in dolphins (in Russian)," Proceedings of the Academy of Sciences of USSR **225**, 470–473.
Whitehead, H., MacLeod, C. D., and Rodhouse, P. (**2003**). "Differences in niche breadth among some teuthivorous mesopelagic marine mammals," Marine Mammal Sci. **19**, 400–406.
Zakharia, M. E., Magand, F., Hetroit, F., and Diner, N. (**1996**). "Wideband sounder for fish species identification at sea," ICES J. Mar. Sci. **53**, 203–208.
Zimmer, W. M. X., Johnson, M. P., Madsen, P. T., and Tyack, P. L. (**2005**). "Echolocation clicks of free-ranging Cuvier's beaked whales (*Ziphius cavirostris*)," J. Acoust. Soc. Am. **117**, 3919–3927.

Methods for automatically analyzing humpback song units

Peter Rickwood[a)]
Building 6, University of Technology, Sydney (UTS), P. O. Box 123, Broadway, NSW 2007, Sydney, Australia

Andrew Taylor[b)]
University of New South Wales, School of Computer Science and Engineering, Sydney 2052, Australia

(Received 22 April 2007; revised 3 December 2007; accepted 27 December 2007)

This paper presents mathematical techniques for automatically extracting and analyzing bioacoustic signals. Automatic techniques are described for isolation of target signals from background noise, extraction of features from target signals and unsupervised classification (clustering) of the target signals based on these features. The only user-provided inputs, other than raw sound, is an initial set of signal processing and control parameters. Of particular note is that the number of signal categories is determined automatically. The techniques, applied to hydrophone recordings of humpback whales (*Megaptera novaeangliae*), produce promising initial results, suggesting that they may be of use in automated analysis of not only humpbacks, but possibly also in other bioacoustic settings where automated analysis is desirable.
[DOI: 10.1121/1.2836748]

PACS number(s): 43.80.Lb, 43.60.Np, 43.58.Ta [WWA] Pages: 1763–1772

I. INTRODUCTION

Almost all analysis of bioacoustic signals is done with substantial human guidance. This can be extremely time consuming. It can also raise serious concerns regarding subjectivity. Lack of quantification can make replication or comparison of manual analyses difficult. The availability of personal computers able to store many hours of bioacoustic signals and to process these signals faster than real time makes automatic analysis attractive. This paper describes methods for performing such automatic analysis, and presents results from the application of these methods to the songs of humpback whales (*Megaptera novaeangliae*).

There are two principal tasks in the detection and characterization of distinct bioacoustic signals, such as animal vocalizations:

1. **Signal detection:** The detection of the vocalization, originating from one or more target sources, from the remainder of the signal (the background). The target sources may be a single individual, multiple conspecific individuals or multiple individuals from multiple species. The signal background may include: bioacoustic signals from nontarget individuals of the same species; bioacoustic signals from other species; and nonbioacoustic sounds.
2. **Signal characterization:** The characterization of vocalizations from the target individual(s) isolated in the previous stage. This characterization will involve reduction of the signal into a more useful and typically much more compact form.

This paper describes automatic techniques for both tasks and presents the results of applying these methods to characterize a noisy set of complex bioacoustic signals—humpback whale *Megaptera novaeangliae* vocalizations. An open-source software implementation of the techniques described in this paper is available from the authors.

II. TARGET SIGNAL DETECTION

Signal detection is a problem common to many domains. For example, it has been well studied for human speech recognition.[1,2] Domain-specific characteristics presumably explain why these techniques have not been transferred to bioacoustic applications, where researchers typically isolate bioacoustic signals manually,[3,4] or avoid the issue altogether.[5]

For a general technique to detect target signal to be effective, there can be no strong assumptions about the amplitude, duration, or frequency characteristics of the target signal, as these can vary greatly between target species.

The approach taken in this paper is energy based. It assumes fast Fourier transform (FFT) frames containing target signals will contain more energy within a certain frequency range than other frames. Further assumptions might produce better detection of target components but risk the reduction of generality.

A. Signal detection, step 1: Feature vector extraction

The first step in the extraction process is to partition the recording into fixed-length intervals. Background noise is assumed to vary sufficiently slowly that it is almost constant within each interval. For estimation of background noise, it is assumed that target calls occupy at most 70% of each interval. The technique thus depends on target signals being interspersed by sufficient periods of background noise, which is typical of bioacoustic environments. The exact

[a)]Electronic mail: peter.rickwood@gmail.com
[b)]Electronic mail: andrewt@cse.unsw.edu.au

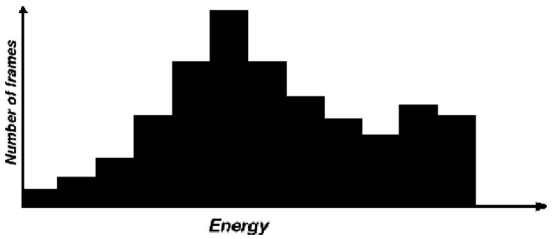


FIG. 1. Combined energy histogram of target and noise signal is usually a combination of a low energy noise distribution (assumed to be symmetric here), and a potentially complex distribution covering signal energy.

choice of interval length is relatively unimportant if it significantly exceeds the duration of calls. Frame length is a manually chosen parameter—we have not explored automatic determination of interval length but this may be possible. The power spectrum of the interval is obtained from the square of the magnitude of a short-time FFT. The user must also choose parameters for FFT window size, FFT step size and a frequency range appropriate for the target signals.

For the humpback recordings used to validate the technique, sound is digitized at 32 kHz, we use an interval length of 20 s, a Hann-windowed 1024 point FFT with a 1/3 window step and a 90–5600 Hz frequency range. These choices are based on extensive prior research on humpback recordings.[6–12] We exclude the 0–90 Hz range because ambient noise typically dominates this range. As a consequence of these choices, 1876 FFT frames are calculated for each 20 second interval, and the band-limited (90–5600 Hz) energy contained in each of these FFT frames is calculated, resulting in 1876 values—each value being the band-limited energy of a single 0.032 s (overlapping) portion of the signal.

B. Signal detection, step 2: Differentiating signal and noise

The initial labeling of frames is based on a histogram of the band-limited energy values calculated from each FFT frame in the interval. Figure 1 shows an illustrative histogram. A threshold T is estimated and all frames with band-limited energy greater than this are labeled as target signal. The remainder are labeled as noise.

The threshold energy T is estimated by assuming that frames can be classified as either containing target signal, or not. Those that do not contain target signal are termed "noise" frames. It is assumed that the band-limited energy of noise frames is distributed according to a distribution f, and for "target signal" frames it is distributed according to g. The energy of all observed frames within an interval will then follow the distribution $(f+g)$. We have found the following heuristics useful. They clearly will not hold in practice, but a method based on them for estimating a threshold energy has proven useful and robust enough to provide useful results.

1. Windows drawn from g will in general have greater energy than those drawn from f, i.e., $E(g(X)) > E(f(X))$;
2. f is approximately symmetric;
3. The probability of an observation drawn from g having less energy than the mode of f is 0. (i.e., $E(g(X) < \text{mode}(f) = 0)$, and
4. The mode of $(f+g)$ is the same as the mode of f.

The procedure is then as follows:

1. Estimate the mode of $(f+g)$, which also gives us an estimate of the mode of f (from assumption 4).
2. Reconstruct an estimate of f from the estimated mode. This can be done using assumptions 1, 2, and 3.
3. For all possible thresholds, estimate the number of misclassified frames if all frames with band-limited energy below that threshold are deemed noise and all above deemed signal. As we have estimates for $(f+g)$ and f, we also have an estimate for g, which means this is a trivial calculation. See Fig. 2.

Reliable estimation of the mode from a frequency histogram is nontrivial, and there is an extensive literature on this.[13,14] The mode estimation method employed in this application consisted of a weighted average of the three largest peaks in the histogram. Other methods could be employed, but we have not investigated this. Steps 2 and 3 in the previous list perhaps require a little more explanation. Because it is assumed that f is symmetric, and f and g do not overlap below the mode of f, an estimate of the lower half of f can be obtained directly. Knowing the lower half of a symmetric distribution is clearly the same as knowing the whole distribution, so f is known entirely, and this allows the calculation of the separate components of the observed distribution $(f+g)$ (see Fig. 2). For any given threshold energy E, the sum of f to the left of E is assumed to be the number of correctly classified noise frames, and the sum of g to the right of E is assumed to be the number of correctly classified signal frames. It should be clear that an optimal threshold exists that maximizes the number of correctly classified frames, given the stated assumptions. Figure 2 is a simple example

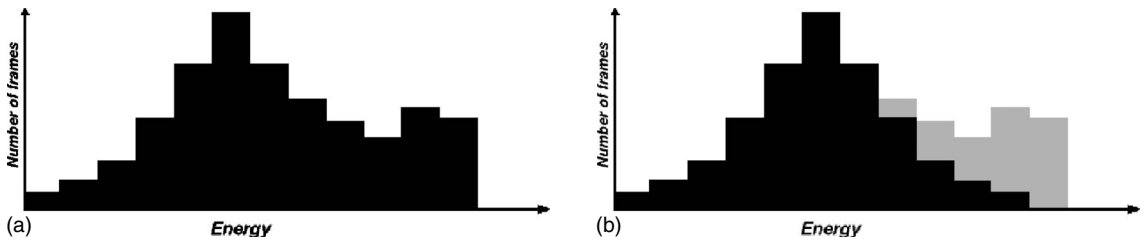


FIG. 2. An illustration of how it is possible, given an energy histogram, to separate signal and noise distributions by assuming that the distribution covering noise is symmetric. For each possible energy threshold, the black area to the right of that threshold is proportional to the expected number of incorrectly labeled noise frames, and the gray area to the left of the threshold is proportional to the number of incorrectly labeled target frames.

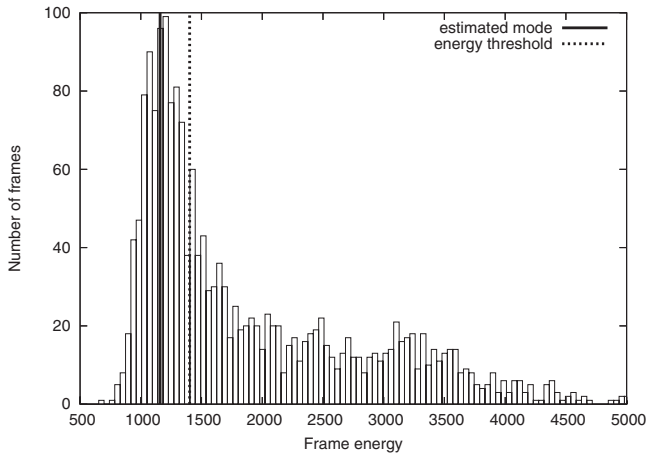


FIG. 3. An actual histogram taken from the recording, with the estimated mode and threshold energies shown. All frames with band-limited energy greater than the threshold would be labeled as signal.

for illustrative purposes. Figure 3 shows an actual energy histogram, with the estimated mode, and the estimated threshold value.

Although the assumptions required for the initial labeling of frames will clearly not hold exactly, we have found them to be useful heuristics and robust in practice.

Once a threshold energy is calculated, each FFT frame in the interval can be tentatively labeled as either target (if it contains more energy than the threshold), or background (if less). Detection is improved by making a further assumption:

- Both noise and target signals have a minimum length of d seconds. That is, both noise and target frames do not occur in isolation, but as part of a sequence of other frames that have the same noise/signal classification.

The assumption allows us to apply a temporal smoothing procedure, where each frame is classified as target only if the majority of other frames in a d second neighbourhood around it are also classified as target. Applying this smoothing criterion results in the (re)classification of frames, and can be applied repeatedly until no frames are reclassified. The separation period can be small relative to the expected signal length—for humpbacks, we use $d=0.05$—despite the fact that most humpback song units are much longer than this.

The histogram analysis, followed by temporal smoothing, is performed for each interval in the recording, so that there is a complete classification of each interval in the recording into either target or noise. Each contiguous sequence of frames labeled as target is regarded as a single distinct call.

Figure 4 shows this technique as applied to an interval containing a series of humpback song units. The initial threshold-based classification is shown in the left-hand panel of Fig. 4 (unsmoothed). The temporal smoothing procedure just described transforms this to the right-hand panel of Fig. 4. The differences are minor, with the smoothing technique acting principally to reclassify short periods labeled as signal as noise, and short periods labeled noise, but surrounded by signal frames, as signal. All 15 song units are successfully detected, with only 1 false positive (the seventh interval from the left), and some artificial elongation of the ninth and sixteenth intervals. The technique successfully excludes most ambient noise, as well as the fainter song units of a second (more distant) whale, which are circled in the right-hand panel of Fig. 4. Although the effectiveness of the technique will clearly decline as the signal/noise ratio increases, it has proven effective when employed over the 11 h of ocean hydrophone recording available for this study.

III. UNSUPERVISED CLASSIFICATION OF SOUNDS

There has been a variety of work on (supervised) classification of bioacoustic signals into (predetermined) categories, (e.g., Refs. 4, 5, and 15). This paper does not address supervised classification. It presents techniques for unsupervised classification (clustering)—classification without predetermined categories.

Applications of unsupervised classification to bioacoustic signals have been limited, (e.g., Refs. 3, 16, and 17), and have typically employed a highly target-specific approach. Unsupervised classification can be viewed as constructing a

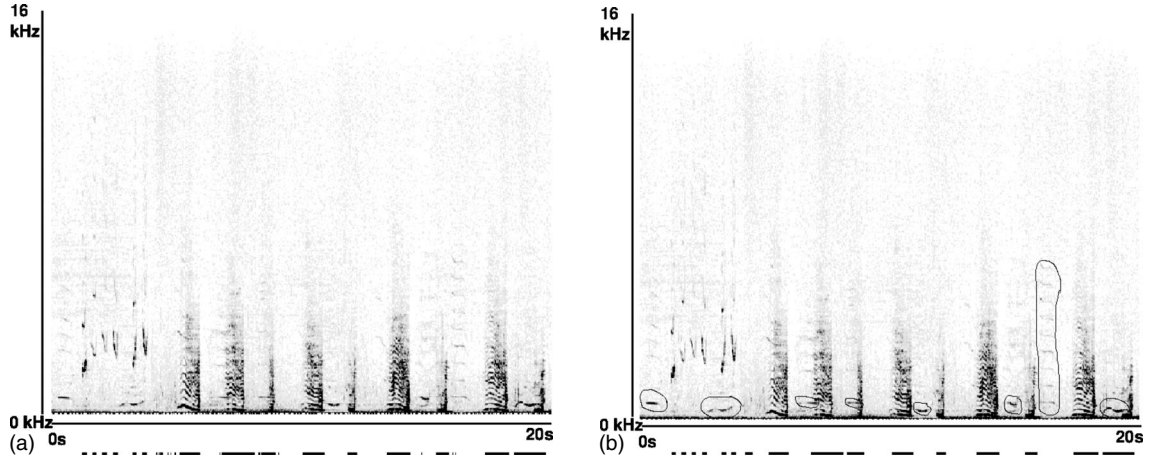


FIG. 4. The black line under each spectrogram indicates which frames have been classified as target. (Left) Based on energy alone and (right) shows the effect of smoothing on this classification. The song units from a second (more distant) whale are circled.

model for a set of signals. Although for some purposes, measurement of several signal parameters may sufficiently model the signal (see Refs. 5, 8, 18, and 19), analysis of complex bioacoustic signals requires more complex models. This paper details general methods to automatically build a class of such models. These methods robustly handle errors in signal detection, allowing automatic signal detection to be employed and hence fully automatic signal analysis.

Unsupervised classification techniques for sound are not well established, because the temporal nature of a sound is difficult to capture with standard attribute-value clustering techniques such as AUTOCLASS[20] or SNOB.[21] This has resulted in the use of a large number of more exotic techniques being used, such as Kohonen networks,[22] adaptive resonance theory networks,[23,24] several less well known clustering techniques (see Refs. 3 and 16), and many others. Often, the user must specify suitable values for a large number of parameters to obtain good performance from these classifiers. The technique that we describe avoids the need for extensive parameter tweaking by using a penalty function (inspired by information theory) which finds a trade-off between model complexity and model fit to data.

A. Feature extraction

After target signals have been isolated, we extract features capturing time varying frequency and amplitude characteristics of the signal. We do not extract features from the raw sound wave form. Instead we take the power spectrum vector $\mathbf{p}=[p_1 \cdots p_n]$ computed previously for signal detection and reduce its dimensionality by computing a feature vector $\mathbf{q}=[q_1 \cdots q_m]$, where each q_i is the sum of one or more of the elements of $\mathbf{p}$.

The size of the power spectrum vector(n), the size of the feature vector(m) and the computation of the feature vector elements (q_i) are user specified. For humpbacks, we use $n=512$ and $m=36$. That is, we use a 1024 point FFT, which produces a 512 element power spectrum vector, and this is then reduced to a 36 element vector, as we now describe. Each q_i is the simple sum of a number of elements in the power spectrum vector. The number of elements in the spectrum to be summed in the computation of each q_i is given by the series $[2, \lfloor 2 \times 1.05 \rfloor, \lfloor 2 \times 1.05^2 \rfloor, \lfloor 2 \times 1.05^3 \rfloor, \ldots]$. That is, q_1 is the sum of the first two values, q_2 is the sum of the next $\lfloor 2 \times 1.05 \rfloor$ values, q_3 is the sum of the next $\lfloor 2 \times 1.05^2 \rfloor$ values, and so on. For example, $q_5=p_{12}+p_{13}$ (375–406 Hz) and $q_{33}=p_{140}+p_{141}+\cdots+p_{148}$ (4375–4625 Hz). We discard the first 3 (low frequency) elements and the last 333 (high frequency) elements of the power spectrum vector.

This process transforms every sequence of FFT frames labeled as target signal into a sequence of feature vectors describing the time-varying power spectra of the recording over the time period covered by those frames. Henceforth, we term each such sequence of feature vectors "call." The sequence of feature vectors describing each call are the basis for all further analysis. More sophisticated feature than those described here are commonly used in speech recognition (see Refs. 25 and 26), and more complex feature vectors might improve performance with humpbacks, for example, but at the cost of simplicity and perhaps generality.

B. Vector quantization

Each target signal O (detected by the technique in Sec. II), consists of a sequence of feature vectors (i.e., $O=[\mathbf{q}_1 \cdots \mathbf{q}_n]$, with each $\mathbf{q}_i$ being a feature vector $\mathbf{q}_i=[q_{i_1} \cdots q_{i_m}]$), extracted as described earlier. Each such sequence of vectors is reduced to a sequence of integers through the process of *vector quantization*.[27–29] Vector quantization essentially clusters vectors in a multidimensional space, and replaces each vector with the index of the cluster centroid that the vector is closest to, under some distance metric (e.g., squared Euclidean distance). This process allows each call to be described as a sequence of integers $[i_1 \cdots i_n]$ rather than a sequence of feature vectors. The vector quantization/clustering required to do this is performed on all feature vectors extracted (i.e., on every feature vector of every target sound detected in the target detection step).

In any application of vector quantization, the codebook size (i.e., number of cluster centroids) is specified by the user. For humpbacks, we used 512. It is possible that this value could be determined by an adaptive/automatic approach, but this was not investigated.

C. Hidden Markov modeling of sound

Hidden Markov models (HMMs)[30,31] are well known as a representation for supervised classification of sound and have been applied to speech,[29,32–34] music[35,36] and bioacoustics.[19,37–41] However, the ability of HMMs to model the time-varying nature of sound signals also makes them suitable for unsupervised classification, provided that a suitable unsupervised clustering technique for HMMs can be devised.

We present a method of using HMMs to perform unsupervised classification, which is not computationally expensive, and which determines the number and form of classes automatically. For completeness, we give a very brief overview of HMMs, for a full treatment see Ref. 42.

A HMM consists of a number of states, and an implied set of observable symbols. Given this, an n-state HMM is describable by two probability matrices:

- *The transition matrix* M is an $n \times n$ matrix where entry $M[i][j]$ is the probability of a transition from state i to state j.
- *The emission matrix* E is an $n \times p$ matrix where entry $E[i][s]$ is the probability of observing symbol s in state i.

HMMs are well studied, and several important algorithms have been developed:

(1) **Baum–Welch expectation maximization training:** The *Baum–Welch* algorithm finds estimates for the HMM transition and emission probabilities that are locally maximal for some set of observation sequences.
(2) **Viterbi alignment:** The *Viterbi* algorithm provides a way of determining the most likely state sequence, given

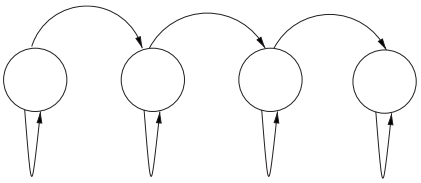


FIG. 5. A Bakis hidden Markov model.

a particular symbol sequence, and a HMM.

(3) **Calculation of** $P(O|M)$: $P(O|M)$ is the probability of observing a particular sequence of symbols O, given a particular HMM, M.

Traditionally, Bakis HMMs (HMMs with transitions only allowed from left to right, see Fig. 5) have been used to model time-varying sounds.

Vector quantization has reduced each call to a sequence of integers. We will term this a *call sequence*. The integers become the alphabet of our HMM. As each detected call is described as a sequence of integers (in our case, in the range [1–512]), it should be clear that any set of calls can be modeled using HMMs with an alphabet of 512 symbols. Given a set of integer sequences $\{O_1 \cdots O_r\}$ (where each O_i is some integer sequence of symbols $[i_1 \cdots i_n]$), each of which describes a single call through time, we can calculate the following:

(1) $P(\{O_1 \cdots O_r\}|M) = P(O_1|M)P(O_2|M) \cdots P(O_r|M)$: The conditional probability of observing the call sequences, given a particular Hidden Markov Model M.
(2) Calculate M_{opt}, the HMM that locally maximizes $P(\{O_1 \cdots O_r\}|M)$ (using Baum–Welch reestimation).

For now, assume there are k different types of call in the data set—a method for determining k will be presented later. k HMMs $(\{M_1 \cdots M_k\})$ modeling these types of calls can be constructed using this algorithm:

(1) Arbitrarily partition all observed call sequences into k disjoint sets. Call these $\{C_1 \cdots C_k\}$.
(2) For each $C_i \in \{C_1 \cdots C_k\}$, calculate M_{opt}, the HMM that locally maximizes $P(C_i|M)$. This HMM is said to describe the calls in C_i. In this way, each HMM is associated with a disjoint subset of call sequences.
(3) For every call sequence (i.e., each O_i), calculate $P(O_i|M)$ for each HMM $M \in \{M_1 \cdots M_k\}$. O_i is then assigned to the subset C_i associated with the HMM that maximizes $P(O_i|M)$.
(4) If any call sequences changed their subset membership, go to step 2.

In practice, waiting for complete termination (i.e., no change in subset membership) is inefficient, and, indeed, convergence is not guaranteed, so the iterations cease when only a small number ($\approx 1\%$) of call sequences have an altered subset membership.

We are now then in a situation where, if k (the number of distinct "types" of calls) is known, then k separate HMMs $\{M_1 \cdots M_k\}$, each of which models a different type of call can be constructed. Further, given any particular call sequence O_i, it is possible to determine its type by calculating which of these HMMs maximizes $P(O_i|M)$. In other words this allows time varying bioacoustic signals to be grouped into distinct call types without any prior labeling of calls or specification of call types.

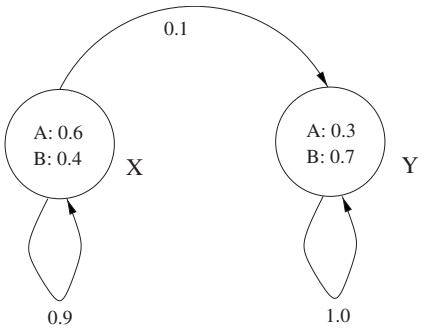


FIG. 6. A simple two-state HMM.

The previous algorithm crucially relies on the value of k. The following section describes how this is determined using an information-theory derived penalty function.

D. MML based classification

Minimum message length (MML) encoding[21] is an information-theoretic approach to balancing a model's fitting of observed data against its complexity (see also the similar Minimum Description Length (MDL) encoding[43,44]). In essence MML encoding states that, for a set of models $\{m_1, m_2, \ldots, m_n\}$, each of which is an alternative way of describing a probability distribution over observed data, one can assess the quality of any particular model m_i by measuring the length of the binary string required to describe the model and the observed data. The best model is the one that minimizes the length of this description string. MML encoding has been successfully employed in widely used supervised[45] and unsupervised[21] classification systems.

We wish to choose a value for k, the number of call types, that minimizes the number of bits, usually termed the *transmission cost*, required to encode the calls (the observed data), and the models that describe those calls. This requires an encoding scheme for HMMs.

Suppose we have a simpler two-state HMM, M, shown in Fig. 6, with observable symbols $\{A,B\}$ and a single observed sequence: $[A,A,A,B,B,A,B]$.

The transmission cost consists of two parts: The model cost $C(M)$, which is the number of bits to encode the model; plus the message cost $C([A,A,A,B,B,A,B]|M)$, which is the number of bits to encode the observed symbols $[A,A,A,B,B,A,B]$, using the model M.

Communicating the model consists of communicating the transition and emission probabilities for each state of the model. The optimal encoding for an event with probability p is $-\log_2 p$ bits.[46] For the first state of the model shown in Fig. 6, these are:

Symbol	Probability	Codeword length
A	0.6	0.74 bits
B	0.4	1.32 bits

Transition	Probability	Codeword length
$X \rightarrow X$	0.9	0.15 bits
$X \rightarrow Y$	0.1	3.32 bits

The cost of communicating the transition probabilities for this state is 3.47 bits, and the cost of communicating the emission probabilities is 2.06 bits. So the total cost for that state is 5.53 bits. This process is repeated for each state to get the total cost of specifying the entire model. For this example, it is 7.78 bits, as state 2 takes 2.25 bits.

It is straightforward to calculate the probability $P(O|M)$ of observing some sequence O for a particular HMM M (Ref. 47) allowing us to us to calculate the message cost $-\log_2 P(O|M)$, which is the number of bits required to communicate the observed data given the specified model.

IV. DETERMINING THE NUMBER OF CLUSTERS

The ability to calculate the transmission cost of a HMM and the data it describes allows us to compare alternative ways of classifying observations. For example, one could initially set k to 1 (i.e., have a single category of vocalization, modeled by a single HMM M_1^1), and calculate $C(M_1^1)$ and then calculate $\Sigma_{i=1}^{r} C(O_i|M_1^1)$ for all observations $O_1 \cdots O_r$. The sum $C(M_1^1)+\Sigma_{i=1}^{r} C(O_i|M_1^1)$ is the total transmission cost for model and data. We could then repeat for increasing values of k, and see how the total transmission cost varies. Choosing the value of k that minimizes total transmission cost would give an estimate of k. Although this approach is feasible, it can be computationally expensive, and so we now describe a more efficient method of achieving the same end—that is, estimating a value of k that has good fit to data, but is sufficiently small so that the categories are intelligible to a human viewing the results. Note that, in a purely statistical sense, simpler models are not always preferable[48,49] but they do have the (nonstatistical) advantage of being easier to interpret.

(1) Initially, a "guess" is made for the initial number of classes, k_{guess}. Calls are assigned to classes at random. A HMM is then generated for each class, which is then trained (using the Baum–Welch reestimation) on the members of the class. This will give k_{guess} HMMs, each of which models some disjoint subset of the instances that are being classified. This procedure has already been described in detail in Sec. III C.

(2) Iterative expectation-maximization[50] is now performed, as described in Sec. III C, until class membership is (nearly) stable:

(a) M: Train each HMM using the members that belong to its class. This is done with Baum–Welch reestimation.

(b) E: Reestimate class membership for each call, based on the newly trained HMMs by assigning each call to the class whose HMM maximizes $P(O|M)$.

(c) Go to step (1) if any reassignments took place in step (2).

(3) For each of the k_{guess} HMMs (call them $\{M_1 \cdots M_{k_{\text{guess}}}\}$), calculate whether removing that HMM results in a reduction in total transmission cost.

(4) Test to see if splitting any classes results in an information gain (i.e., a reduced number of bits). That is, for every HMM M, modeling a set of observations O, partition O into two random subsets (O', O''), generate a HMM for each subset (M', M''), and perform HMM training and reestimation (as in 2), until subset membership is stable. Once this is done the transmission cost of the original grouping [i.e., $C(M)+C(O|M)$] is compared with that of the proposed split $[C(M')+C(O'|M')+C(M'')+C(O''|M'')]$. Whichever has the lower transmission cost is retained. That is, if $C(M')+C(O'|M')+C(M'')+C(O''|M'') < C(M)+C(O|M)$, M is replaced with M' and M'', otherwise $M$is retained and M' and M'' are discarded. If any of these "splits" do occur, the splitting test is applied recursively to each new subset, until there is no information gain in splitting. After this process is complete, let the remaining number of classes be k''. Figure 7 demonstrates the recursive splitting of a single class (covering four observations $o2$, $o4$, $o7$, $o9$) into three separate classes.

(5) Consider there now to be k'' remaining classes/HMMs, with each still representing a disjoint subset of instances. At this point, the algorithm repeatedly performs the expectation/maximization described in step 2, until no changes to class membership takes place. This second reestimation step helps to move out of undesirable locally optimal solutions produced by the greedy split/join procedure.

(6) Perform one final time the procedure described in step 3—removing HMMs as there is an information gain in doing so. Let the remaining number of clusters be k'''.

(7) This completes the classifying procedure, with k''' being the number of classes "decided" on by the algorithm.

One objection that may be raised is that the technique for automatic grouping of calls requires an initial guess, by the user, of the number of distinct types of vocalization. However, our initial experiments with humpback song units seem to indicate that this initial guess is unimportant, playing little role in the final number of classes determined by the technique. In other words, the technique is not sensitive to this initial guess. Figure 8 illustrates.

We have thus far avoided specifying how we determine the number of states in the HMM that models each call category. Although this could be determined through some automatic scheme, we choose the simpler and faster method where the number of states is proportional to the mean length of the calls modeled by the HMM. As the number of states is proportional to call length, the user must specify a parameter relating call length to the number of underlying HMM states.

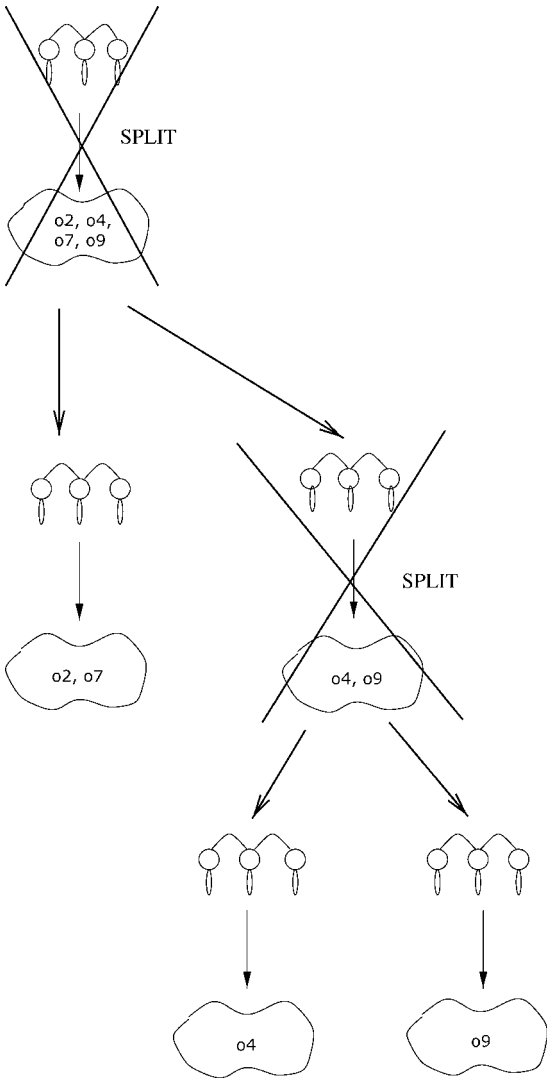


FIG. 7. Recursive splitting of a single class.

V. RESULTS

The techniques described (signal detection, feature extraction, clustering) were applied to 11 h of humpback recordings. The recordings consisted of nine separate recordings, each of between 54 and 90 min, taken on separate days via hydrophone buoy off the coast of North Stradbroke Island in Queensland, Australia—a known migration path of humpbacks. Each recording captures the vocalizations of an individual humpback as it passes near the buoy. The depth and orientation of the whale to the buoy vary through time. Wave and wind noise also vary throughout the recordings. Boat motors are recorded, as are other (more distant) whales. Noise levels vary by over 10 dB through the 11 h of recordings, but in any 20 s interval, variation in background noise is generally less than 3 dB. We performed no filtering or other preprocessing to minimize the effect of noise sources. At no stage did we intervene in the process of signal detection, extraction, or classification. Thus, we did not vet the results of the detection algorithm before running the feature extraction and clustering. The steps thus form a single processing pipeline, requiring ≈4 h of processing time on a 3 GHz desktop to run the detection, feature extraction, and clustering programs on the 11 h of recordings.

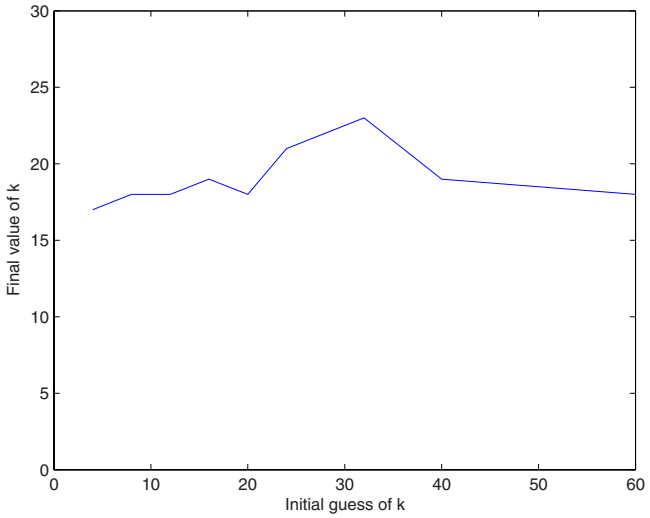


FIG. 8. Final number of clusters against initial number of seed clusters.

Some indication of the accuracy of the segmentation technique is given by Fig. 4. More convincing evidence of the ability of the segmentation technique to do a reasonable job is provided by the results of the signal classification algorithm—as the classification/clustering procedure is built on top of the segmentation and extraction procedures, it cannot perform well if the segmentation procedure does not. However, given the unsupervised nature of the processing, we cannot report detailed quantitative results on classification accuracy.

The unsupervised classification technique described produced 19 distinct song unit types after processing ≈11 h of humpback recordings. As noted, the detection technique described in Sec. II is not perfect, and a cursory inspection of the 19 song unit types identifies 9 of them as describing noise clusters (i.e., pertaining to false positives in the detection procedure), rather than actual humpback song units—leaving 10 distinct types of humpback song units identified. Despite the fact that the signal detection procedure produces a number of false positives, the ability of the algorithm to group mistakenly isolated noise into distinct categories based on their spectral characteristics is pleasing, and is an indication of the robustness of the technique. Figs. 9(a)–9(h) show randomly chosen extracts from the 11 h of recording with labels attached to indicate the "category" to which each song unit has been assigned. To keep Figs. 9(a)–9(h) clear, noise categories mentioned earlier are excluded. A few representative samples of each signal category are circled, with the rest simply having their category listed above the song unit. Some song units that a human might consider similar are grouped separately (e.g., 10 and 9). The groupings produced by the automatic technique were considered plausible by a domain expert but we cannot provide any quantitative evaluation. Even domain experts could not provide a precise manual categorization of the all song units present in the recordings.

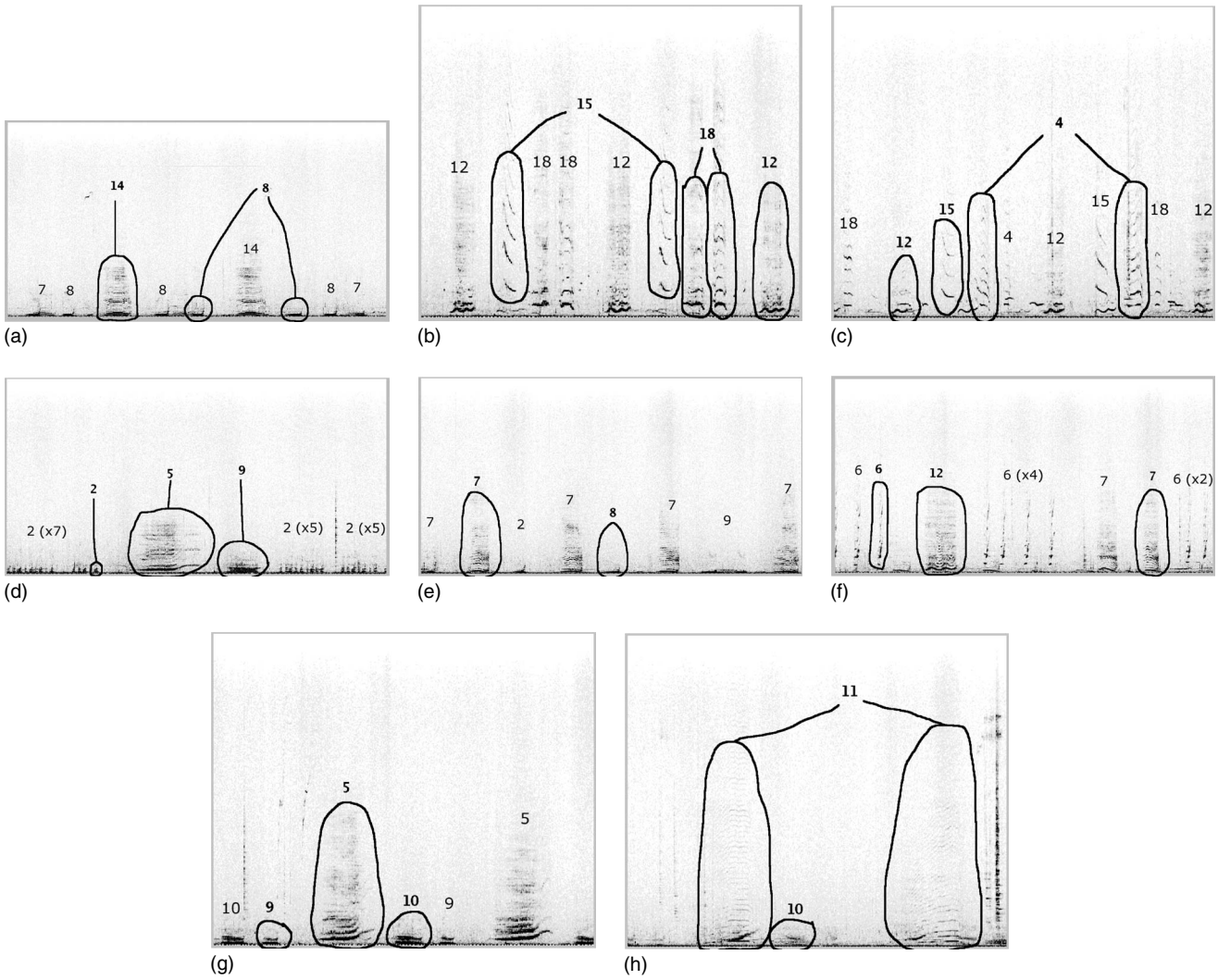


FIG. 9. 20 s frames with song unit categories (as determined by the algorithm) labeled. Intervals belonging to obvious noise categories excluded.

VI. DISCUSSION

The fact that results produced by a fully automatic analysis were deemed broadly consistent with the groupings likely to be chosen by a domain expert is the strongest evidence of the effectiveness of the automatic techniques described in this paper. The vocalizations of humpbacks are particularly complex, and the ones analyzed in this paper were all recorded under uncontrolled conditions. However, the recording conditions were generally benign, in that humpback vocalizations can be distinguished based primarily on band-limited energy, and there is only limited interference from nearby humpbacks. The clustering technique succeeded in grouping similar song units together despite significant variations in signal levels, interference from other oceanic sounds, and recording equipment fluctuations.

The techniques do require some initial manual configuration, but this is unavoidable given the wide range of target signals the techniques can, in principle, be applied to. Table I lists all user-specified parameters. Although this list is not small, the required parameters are generally easily understood in terms of the underlying signal being analyzed (with the exception of the "VQ codebook size" parameter). There are, for example, no obscure parameters required for the sole purpose of mathematical tuning (such as the "vigilance" parameter used in ART networks), as this is avoided through the use of the MML-based penalty function. Essentially the user needs to specify the basic signal processing parameters to be used, and several important parameters describing the broad character of the target signal to be analyzed. Little effort was made to tailor the technique specifically for hump-

TABLE I. The complete list of user-supplied parameters required for end-to-end processing of raw sound data to clustered song units.

Stage	Parameter
Call detection	FFT size and step
Call detection	Signal frequency range
Call detection	Detection segment size
Call detection	Smoothing interval
Feature extraction	Feature extraction scheme parameters
Clustering	VQ codebook size
Clustering	Number of initial categories
Clustering	HMM state duration constant

back vocalizations, and yet the achieved results are promising, compared to previous specialized attempts in that domain (e.g., Refs. 3 and 16).

We see many areas where the technique could be improved. The greatest improvement would be gained, we believe, by incorporating human feedback into the process. The detection procedure, for example, makes no attempt to use spectral features to differentiate noise and signal. Although this has the advantage of keeping the technique simple and general, it does degrade performance. The ideal procedure, we believe, is to begin by making few assumptions about the spectral characteristics of the target signal, and then, after human input, rerun the algorithm with the benefit of this human expertise. For example, it is clear to any humpback expert (and indeed any nonexpert who has spent some time listening to humpbacks) that 9 of the 19 "classes" produced by the classification algorithm are in fact ambient noise. If an expert can indicate this, and inform the program which classes are constituted of noise, the remaining classes, confirmed as containing actual humpback song units, could be used to rerun the detection procedure, but this time, both positive and negative examples (i.e., of noise and of signal) would be available. By looking for the distinguishing properties between the two, it should be possible to do substantially better than in the initial run, where no positive or negative examples are available. In effect, with only the brief intervention of an expert, it would be possible to "bootstrap" the whole procedure from an unsupervised task to a supervised one, and yet this could be done with much less effort than in a wholly supervised procedure. The manual detection and classification of song units by an expert would be reduced to the brief vetting of results produced by the initial unsupervised run.

Another area of improvement lies in allowing a human to vary the penalty function, producing either more, or fewer categories. We do not claim that the MML-based penalty function employed produces the "correct" number of clusters—indeed, we believe such a notion makes no sense. However, in this application, the technique does group song units in a way remarkably similar to human experts, who also typically recognize from around a half dozen to a dozen distinct types of song units (see Refs. 3, 9–12, and 51). The effect of the penalty function applied is to concentrate computational effort in looking for simpler models that fit the data. This is done not because simpler models are more probable (i.e., in the Bayesian prior sense), but because they are more preferable, as they are more comprehensible. If the aim is to maximize human intelligibility, then it makes sense for a human to have some input to the process. A human could do this quite simply by indicating if too many (or too few) classes were produced, and the penalty function could be adjusted accordingly.

ACKNOWLEDGMENTS

The humpback recording were kindly supplied by Fiona MacKnight and Mike Noad (Uni of Qld.). They and Doug Cato (Defence Science and Technology Organisation) discussed humpback vocalizations at length with us. The authors are also indebted to the comments of three anonymous reviewers, each of whom provided valuable detailed feedback and suggestions.

[1] S. Das and M. Picheny, "Issues in practical large vocabulary isolated word recognition: the IBM Tangora system," in *Automatic Speech and Speaker Recognition*, edited by C. Lee, F. Soong, and K. Paliwal (Kluwer Academic, New York, 1996), pp. 457–479.
[2] L. Rabiner and M. Sambur, "An algorithm for determining endpoints of isolated utterances," Bell Syst. Tech. J. **54**, 297–315 (1975).
[3] D. Helweg, D. Cato, P. Jenkins, C. Garrigue, and R. McCauley, "Geographic variation in south Pacific humpback whale songs," Behavior **135**(1), 1–27 (1998).
[4] J. Potter, D. Mellinger, and C. Clark, "Marine mammal call discrimination using artificial neural networks," J. Acoust. Soc. Am. **96**, 1255–1262 (1994).
[5] G. Grigg, H. McCallum, A. Taylor, and G. Watson, "Monitoring frog communities: an application of machine learning," *Eighth Innovative Applications of Artificial Intelligence Conference*, 1996.
[6] L. Frazer and E. Mercado, "A sonar model for humpback whale song," IEEE J. Ocean. Eng. **25**, 160–182 (2000).
[7] P. Frumhoff, "Aberrant songs of humpback whales (*Megaptera novaeangliae*): clues to the structure of humpback songs," in *Communication and Behavior of Whales*, edited by R. Payne (Westview Press, Boulder, Colorado, 1983), pp. 81–128.
[8] G. Hafner, C. Hamilton, W. Steiner, T. Thompson, and H. Winn, "Signature information in the song of the humpback whale," J. Acoust. Soc. Am. **66**, 1–6 (1979).
[9] K. Payne, P. Tyack, and R. Payne, "Progressive changes in the songs of humpback whales (*Megaptera novaeangliae*): A detailed analysis of two seasons in Hawaii," in *Communication and Behavior of Whales*, edited by R. Payne (Westview, Press, Boulder, Colorado, 1983), pp. 9–58.
[10] R. Payne and S. McVay, "Songs of humpback whales," Science **173**, 585–597 (1971).
[11] P. Thompson, W. Cummings, and S. Ha, "Sounds, source levels, and associated behavior of humpback whales, southeast Alaska," J. Acoust. Soc. Am. **80**, 735–740 (1986).
[12] H. Winn and L. Winn, "The song of the humpback whale *Megaptera novaeangliae* in the West Indies," Mar. Biol. (Berlin) **47**, 97–114 (1978).
[13] D. R. Bickel, "Robust estimators of the mode and skewness of continuous data," Comput. Stat. Data Anal. **39**, 153–163 (2002).
[14] U. Grenander, "Some direct estimates of the mode," Ann. Math. Stat. **36**, 131–138 (1965).
[15] A. Taylor, "Bird flight call discrimination using machine learning," J. Acoust. Soc. Am. **97**, 3370–3371 (1995).
[16] D. Chabot, "A quantitative technique to compare and classify humpback whale sounds," Ethology **77**, 89–102 (1988).
[17] E. Mercado and A. Kuh, "Classification of humpback whale vocalizations using a self-organizing neural network," *Proceedings of IJCNN'98*, 1998, pp. 1584–1589.
[18] J. Buck and P. Tyack, "A quantitative measure of similarity for *Tursiops truncatus* signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506 (1993).
[19] J. Kogan and D. Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," J. Acoust. Soc. Am. **103**, 2185–2196 (1998).
[20] P. Cheeseman, J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman, "AutoClass: A Bayesian classification system," in *Proceedings of the Fifth International Conference on Machine Learning* (Morgan Kaufmann Publishers, San Francisco, CA, 1988), pp. 54–64.
[21] C. Wallace and D. Boulton, "An information measure for classification," Comput. J. **11**, 185–194 (1968).
[22] T. Kohonen, "The self-organizing map," Proc. IEEE **78**, 1464–1480 (1990).
[23] G. Carpentera and S. Grossberg, "ART 2: self-organization of stable category recognition codes for analog input patterns," Appl. Opt. **26**, 4919–4930 (1987).
[24] V. B. Deecke and V. M. Janik, "Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls," J. Acoust. Soc. Am. **119**, 645–653 (2006).
[25] S. Furui, "Cepstral analysis technique for automatic speaker verification," IEEE Trans. Acoust., Speech, Signal Process. **ASSP–29**, 254–272 (1981).
[26] S. Furui, "Speaker-independent isolated word recognition using dynamic

features of speech spectrum," IEEE Trans. Acoust., Speech, Signal Process. **ASSP–34**, 52–59 (1986).
[27]D. Burton, "Text-dependent speaker verification using vector quantization source coding," IEEE Trans. Acoust., Speech, Signal Process. **35**, 133–143 (1987).
[28]J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," Proc. IEEE **73**, 1551–1558 (1985).
[29]L. Rabiner, S. Levinson, and M. Sondhi, "On the application of vector quantization and hidden Markov models to speaker-independent, isolated word recognition," Bell Syst. Tech. J. **62**, 1075–1106 (1983).
[30]B. Juang, "On the hidden Markov model and dynamic time warping for speech recognition—A unified view," AT&T Bell Laboratories Technical Journal **63**, 1213–1238 (1984).
[31]L. R. Rabiner, "A tutorial on hidden Markov models and selected applications," in *Readings in Speech Recognition*, edited by A. Weibel and K. Lee (Morgan Kaufmann, San Francisco, CA, 1990), pp. 267–296.
[32]S. Levinson, "Structural methods in automatic speech recognition," Proc. IEEE **73**, 1625–1648 (1985).
[33]L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE **77**, 257–285 (1989).
[34]L. Rabiner, B. Juang, and C. Lee, "An overview of automatic speech recognition," in *Automatic Speech and Speaker Recognition*, edited by C. Lee, F. Soong, and K. Paliwal (Kluwer Academic, New York, 1996), pp. 1–30.
[35]M. Eichner, M. Wolff, and R. Hoffmann, "Instrument classification using hidden Markov models," in *Proceedings of the ISMIR 2006, 7th International Conference on Music Information Retrieval*, Victoria, Canada, 8–12 October 2006, 349–350.
[36]K. Lee and M. Slaney, "Automatic chord recognition from audio using a HMM with supervised learning," in *International Conference on Music Information Retrieval*, 2006, pp. 133–137.
[37]P. Clemins, M. Johnson, K. Leong, and A. Savage, "Automatic classification and speaker identification of African elephant (loxodonta africana) vocalizations," J. Acoust. Soc. Am. **117**, 956–63 (2005).
[38]D. Curless, M. Roch, and S. Madhusudhana, "Classification of wolf call types using remote sensor technology," J. Acoust. Soc. Am. **121**, 3106–3106 (2007).
[39]R. Suzuki, J. Buck, and P. Tyack, "Information entropy of humpback whale songs," J. Acoust. Soc. Am. **119**, 1849–1866 (2006).
[40]M. Trawicki and M. Johnson, "Automatic song-type classification and speaker identification of Norwegian ortolan bunting (*Emberiza Hortulana*)," 2005.
[41]B. Weisburn, S. Mitchell, C. Clark, and T. Parks, "Isolating biological acoustic transient signals," in *International Conference on Acoustics, Speech, and Signal Processing* (IEEE Computer Society, Los Alamitos, CA, 1993), pp. 269–272.
[42]L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall, Upper Saddle River, NJ, 1993).
[43]P. Grunwald, "A tutorial introduction to the minimum description length principle," in *Advances in Minimum Description Length: Theory and Applications* (MIT Press, Boston, MA, 2004).
[44]J. Rissanen, "Stochastic complexity," J. R. Stat. Soc. Ser. B (Methodol.) **49**, 223–239 (1987).
[45]J. Quinlan, *C4-5: Programs for Machine Learning* (Morgan Kaufmann, San Mateo, CA, 1993).
[46]C. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379-423 and 623-656, July and October, (1948).
[47]L. Rabiner and B. Juang, "An introduction to hidden Markov models," IEEE ASSP Mag., January, 4-16 (1986).
[48]C. Schaffer, "Overfitting avoidance as bias," Mach. Learn. **10**, 153–178 (1993).
[49]P. Domingos, "The role of Occam's Razor in knowledge discovery," Data Min. Knowl. Discov. **3**, 409–425 (1999).
[50]A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. Ser. B (Methodol.) **29**, 1–38 (1977).
[51]D. Cato, "Songs of humpback whales: the Australian perspective," Memoirs of the Queensland Museum **30**, 277–290 (1991).

Feasibility of ultrasound phase contrast for heating localization

Caleb H. Farny[a] and Greg T. Clement
Department of Radiology, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts 02115

(Received 8 August 2007; revised 20 December 2007; accepted 22 December 2007)

Ultrasound-based methods for temperature monitoring could greatly assist focused ultrasound visualization and treatment planning based on sound speed-induced change in phase as a function of temperature. A method is presented that uses reflex transmission integration, planar projection, and tomographic reconstruction techniques to visualize phase contrast by measuring the sound field before and after heat deposition. Results from experiments and numerical simulations employing a through-transmission setup are presented to demonstrate feasibility of using phase contrast methods for identifying temperature change. A 1.088-MHz focused transducer was used to interrogate a medium with a phase contrast feature, following measurement of the baseline reference field with a hydrophone. A thermal plume in water and a tissue phantom with multiple water columns was used in separate experiments to produce a phase contrast. The reference and phase contrast field scans were numerically backprojected and the phase difference correctly identified the position and orientation of the features. The peak temperature reconstructed from the phase shift was within 0.2 °C of the measured temperature in the plume. Simulated results were in good agreement with experimental results. Finally, employment of reflex transmission imaging techniques for adopting a pulse-echo arrangement was simulated, and its future experimental application is discussed.
 [DOI: 10.1121/1.2835438]

PACS number(s): 43.80.Qf, 43.35.Bf, 43.80.Ev, 43.35.Yb [FD] Pages: 1773–1783

I. INTRODUCTION

Noninvasive, inexpensive, and prompt visualization of temperature rise remains a significant barrier to optimizing thermal therapies for cancer tissue ablation such as high-intensity focused ultrasound (HIFU), radio-frequency (rf) ablation, and cryoablation. Variability in tissue structure between patients and target movement during treatment makes it difficult to predict *a priori* where and how much energy will be deposited in the treatment region, so treatment time and outcome can suffer greatly without adequate visualization and guidance. For HIFU treatment, measurement of the proton resonant frequency shift with magnetic resonance (MR) imaging is the current gold standard[1,2] for monitoring thermal damage but there is a clear motivation for a cheaper, portable method with higher temporal and spatial resolution. Each of these requirements contributes to an ongoing investigation of ultrasound-based thermographic methods. A number of other applications ranging from nondestructive testing[3,4] to meteorology[5,6] have a similar need for noninvasive, portable methods of temperature measurement.

Nearly all ultrasound methods for measuring temperature rise in tissues are based on empirical relationships for sound speed and temperature for various tissue types. Similar to water, the sound speed for most tissue types increases approximately quadratically with temperature but exhibits an inflection point near 70 °C and decreases for higher temperatures; the sound speed of some tissues types, such as fat, instead decreases with increasing temperature.[7] Based on these relationships, knowledge of local sound speed will yield the local temperature. A drawback to this approach occurs when multiple tissue types having varying dependencies on temperature are present in the measurement volume, resulting in an approximation of the actual temperature. However, although a direct measurement of the *in vivo* temperature remains an important goal for treatment guidance, an intermediate goal of simply localizing the beam focus relative to the target would greatly improve treatment efficacy. A method designed to locate the focus rather than quantify temperature rise justifies use of a homogeneous approximation for the tissue structure, allowing a single relationship between sound speed and temperature. A short, low-power sonication has been employed as a "citing shot"[8,9] before the therapeutic sonication is applied, for the sole purpose of determining the position of the focus *in vivo*.

A number of ultrasound-based techniques take advantage of standard ultrasound diagnostic imaging platforms for measuring sound speed change. Commonly, the spatial dimensions on which a diagnostic image is based assume that the medium corresponds to a uniform 1540 m/s sound speed, that of normal body temperature. Under this assumption, an increase in temperature will shift the image, and by comparing the apparent spatial shift with an initial (unheated) image, the sound speed change can be determined.[13–19] A second phenomenon which will occur is thermal expansion, which results in a physical shift in the tissue.[9,11–13,19]

Seip and Ebbini[10] used the spectrum of the measured backscattered rf signal to track heating. Based on the change in the average tissue scatterer spacing from thermal expansion and the sound speed dependence on temperature, it was found that the shift in the resonance peak of the rf signal was

[a]Electronic mail: cfarny@bwh.harvard.edu

linearly proportional with the change in temperature. The technique was shown to agreeably estimate temperature rises for *in vitro* tissue samples up to 10 °C for waterbath heating and up to 4 °C for the cooling curve in *in vivo* ultrasound-heated tissue samples. However, it was theorized that higher temperature changes and resulting tissue damage would disrupt the scatterer structure on which the technique was based. Maass-Moreno and Damianou[11] and Maass-Moreno *et al.*[12] interlaced diagnostic and therapeutic ultrasound pulses to track heating in excised turkey breast. Cross-correlation techniques were used to determine the time change in the backscattered diagnostic signal. Linear approximation of the sound speed with temperature produced accurate temperature measurements for elevations up to 10 °C but estimations of higher temperature rises suffered.

Simon *et al.*[13] correlated the backscattered ultrasound images as the temperature in tissue phantoms was increased, and found that the axial derivative of the successive echo shifts was proportional to temperature rise. The results reliably estimated temperature rises up to ~4 °C but it was found that spatial ripples introduced by a thermoacoustic lensing effect degraded the image. Two-dimensional low-pass filtering reduced the ripples but reduced spatial resolution as well. A number of other groups[14–17] have successfully applied similar approaches, but primarily only in *in vitro* samples experiencing low (<10 °C) temperature rises. Varghese *et al.*[18,19] showed success in measuring temperature change for *in vitro* and *in vivo* tissue samples for temperature elevations up to 100 °C from rf ablation therapy. The group cross-correlated rf ultrasound echo signals to determine the apparent time shift, despite applying a linear approximation for the sound speed relationship with temperature. It was theorized that the success of the resulting two-dimensional (2D) temperature map even at high temperatures was partly due to thermal expansion effects which were not explicitly accounted for in the analysis.

King *et al.*[20] used a different approach for measuring thermal effects. An acoustic camera equipped with a 128 ×128 polyvinylidence difluoride (PVDF) sensing array was positioned in a through-transmission arrangement with a planar 10 MHz transducer operated in pulse mode. *Ex vivo* rabbit liver and bovine fat tissue samples were instrumented with a thermocouple and sonicated with a high-power focused ultrasound source in order to induce temperature rises up to 100 °C. It was found that image intensity from the acoustic camera correlated well with the cooling curve in the sample following therapeutic sonication, and the lesion size measured from the intensity image correlated well with lesion sizes determined from histology. The authors theorized that the local change in sound speed caused a sufficient change in phase, altering the intensity on which the image was based. The technique offered a real-time measurement of temperature change but due to its insensitivity to different trends in the sound speed–temperature relationship of different tissue types, the resulting image was thought to be best suited for localization of the focus.

Based on the findings by King *et al.*,[20] it is theorized that by measuring phase shift, the change in sound speed and thus, change in temperature can be determined. Comparing the change in phase angle between an unheated reference and subsequent heated wave can determine a change in sound speed, which is contained in the phase angle of the received sound wave. A similar approach was taken by Mizutani *et al.*[21] and Ishikawa *et al.*,[22] who applied ultrasound computed tomography (CT) techniques to measure the local time of flight at multiple positions through a region heated approximately 5–15 °C in air. A through-transmission setup was linearly and rotationally scanned around the target region so that the change in sound speed at multiple locations was known. Using an empirical relationship to determine temperature from sound speed, a filtered backprojection method was used to reconstruct the temperature rise to within 1 °C of the thermocouple-measured temperature.

It is proposed that reflex transmission imaging (RTI), planar projection, and tomographic methods can be applied to reconstruct a 2D map of sound speed and corresponding temperature rise from the change in phase angle, in a method that would not depend on speckle-tracking and correlation methods. RTI is a technique which has been used to improve backscattered *B*-mode image quality of attenuating media such as kidney stones embedded in soft tissue[23,24] and skin tumors.[25] Typically, a diagnostic ultrasound imager is modified so that the signal reflected from a region beyond the focus is integrated. The principle is based on the fact that the integrated signal will depend on the characteristics of the return path length, which includes the focal zone, and can be treated as a virtual source, emulating a through-transmission setup. It is desirable to use a pulse-echo arrangement as opposed to through-transmission due to the limited number of acoustic windows in the body. Planar backward projection techniques in combination with tomographic reconstruction have been employed in transmission imaging simulations to visualize 2D temperature maps from phase contrast measurements.[26] It is noted that phase measurements often feature a higher signal-to-noise ratio than traditional backscattered ultrasound.

Interest here lies in understanding whether phase contrast from backprojection of the beam measured through an unheated and subsequent heated medium can be used to locate and measure the region of heating. In this feasibility study, a through-transmission arrangement was used and planar backprojection and tomographic reconstruction methods were employed to experimentally and numerically localize temperature rise and/or sound speed contrast from the change in phase angle. To apply this technique, it is necessary to measure the amplitude and phase of the sound field throughout a 2D plane with sufficient spatial resolution to properly backproject the beam. Therefore, two separate experimental arrangements featuring quasi-steady-state phase contrasts were employed due to the lengthy measurement time. The first arrangement used a thermal plume to produce a phase contrast, whereas the second arrangement consisted of a tissue phantom with multiple water columns. Finally, extension of the through-transmission approach to RTI was simulated in a pulse–echo arrangement, and future experimental employment is discussed.

II. THEORY

A. Phase angle

The relationship between sound speed and temperature is well-known for water and has been measured for a number of tissue types.[7] The heating in the experiments was conducted in water, and a sixth-order fit of the measured relationship in water[27] was used to determine the corresponding sound speed. The sound speed is contained in the complex argument of the pressure wave, so by determining sound speed, the temperature can be found. If the reference sound speed c_0 is known, the change in sound speed can be determined from the difference in phase $\Delta\Phi$, where

$$\Delta\Phi = \omega z\left(\frac{1}{c_h} - \frac{1}{c_0}\right), \tag{1}$$

where ω is the angular frequency, z is the distance over which the contrast occurs, and c_h is the sound speed in the heated region.

B. Planar projection

Forward and backward planar projection techniques[28] are a computationally efficient method for determining the sound field at a different plane provided the amplitude and phase of the field and medium characteristics are adequately known. Each point of the field is treated as an individual plane wave and a propagation operator in 2D wave vector space can be used to model the phase at the new point in space, thereby projecting the entire field to the new plane. The relationship between the initial field P_i and projected field P_0 in wave vector space is

$$P_0(k_x,k_y,z) = \frac{1}{2\pi}\int\int P_i(k_x,k_y,z_i)\exp\left(i(z_0 - z_i) \times \sqrt{\frac{\omega^2}{c^2} - k_x^2 - k_y^2}\right)dk_x dk_y, \tag{2}$$

where k_x and k_y are the wave numbers and the exponential term is the transfer function propagation operator. The effects of evanescent waves, which are small for sufficiently large distances ($dz > \lambda$), must be set to zero to avoid buildup of numerical error for back projection.[29]

C. Tomographic reconstruction

The experimental arrangement used here permitted for only one angle of interrogation of the medium, as opposed to rotation of the transmitter and receiver around the target commonly employed in CT imaging. In order to comply with this restriction, the measurements specifically used axially symmetric targets, allowing application of the Fourier slice theorem to reconstruct the distribution of heat throughout the target region. The 2D Fourier transform of an object $p(x,y)$ is defined as

$$P(k_x,k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\exp(i(k_x x + k_y y))dk_x dk_y. \tag{3}$$

The Fourier slice theorem states that a Fourier transform of a projection along an angle θ,

$$P_\theta(x) = \int_{-\infty}^{\infty} p(x,y)dy, \tag{4}$$

is equivalent to a slice through the origin in wave vector space.[30] For a radially symmetric object such as a plume, the phase is no longer dependent on y, reducing the integral to

$$P(k_x,0) = \int_{-\infty}^{\infty} P_{\theta=0}(x)\exp(ik_x x)dk_x. \tag{5}$$

This relationship reconstructs the 2D temperature distribution by taking the one-dimensional Fourier transform of the phase difference across the plume, rotating it about the origin in wave vector space and taking the 2D inverse transform.

III. MATERIALS AND METHODS

A. Experimental setup

Experiments were conducted in a tank lined with absorbing rubber pads and filled with deionized water. Interrogation of the medium was conducted by a focused transducer driven at its 1.088 MHz center frequency (50 mm diameter, 100 mm focal length) and the field at the focus was measured with a needle hydrophone and preamplifier (0.2 mm diameter, Precision Acoustics, Dorchester, UK). A function generator (396, Fluke Corp., Everett, WA) supplied the pulse train (35 cycles, 2 kHz pulse repetition frequency) which was amplified (240L, E & I Ltd., Rochester, NY) before passing through a matching network to the transducer. The hydrophone was scanned in a raster pattern with a 0.2 mm step size by a computer-controlled positioning system (VP9000, Velmex, Inc., Bloomfield, NY). The signal was digitized with an 8-bit oscilloscope (TDS 210, Tektronix, Inc., Richardson, TX), averaged, and transferred via GPIB for storage on the local computer. Data acquisition and analysis was performed with MATLAB (The MathWorks, Natick, MA). Due to the large measurement field and averaging time, the scan duration was several hours. Each measurement consisted of a reference scan of the homogeneous medium and a "heated" scan that contained the phase contrast in the sound path. The temperature of the water in the reference scans was monitored before and immediately following the scan and the sound speed in the analysis and corresponding simulation was based on the measured ambient water temperature.

Two separate setups were employed to produce a phase-contrast: a thermal plume and a tissue phantom. In order to simulate a thermal effect from a HIFU source but also comply with the steady-state time restriction, a heater was encased in thermal insulation with an open column in the top to produce a thermal plume, shown in Fig. 1(a). The reference scan consisted of measuring the field with the heater off and the heated scan was performed with the heater turned on. The thermal plume was positioned approximately 24 mm in

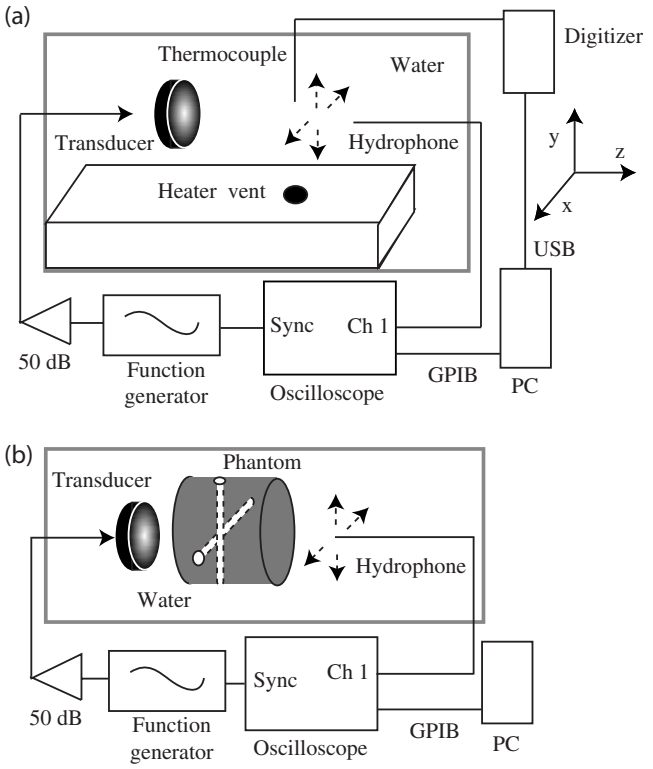


FIG. 1. Experimental arrangement for the phase contrast from (a) a thermal plume and (b) a tissue phantom intersected by water columns.

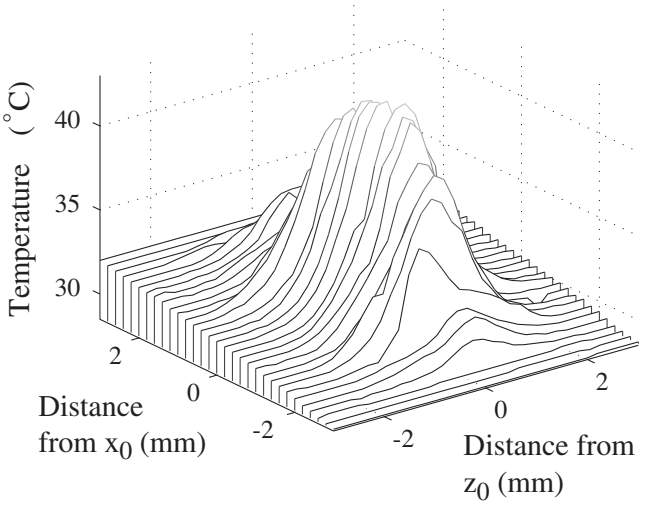


FIG. 2. Temperature profile across the $x-z$ plane in the thermal plume measured with a thermocouple.

front of the focus and a thermocouple (type T, 50-μm wire diameter) was positioned in line with the hydrophone in the center of the plume to monitor the temperature in the plume along the sound path. At each position in the scan three thermocouple measurements were digitized (24 bit, NI-9211A, National Instruments Corp., Austin, TX) and averaged following transfer to the computer via USB. The thermocouple tip was positioned approximately 5 mm above the hydrophone $x-z$ plane to minimize interference with the sound propagation. The sound field was measured with and without the thermocouple present and its presence did not exhibit a significant impact on the field. Maximum steady-state temperatures in the plume ranged from 5 to 15 °C above the reference temperature. Radial symmetry of the temperature distribution in the plume was determined by scanning the thermocouple across the acoustic (z) and lateral (x) plane, shown in Fig. 2.

Insulation around the heater helped minimize heating of the surrounding water but the water outside of the plume inevitably warmed up, as is evident in the temperature along the x axis in Fig. 2. Typical scans were 20×20 mm, resulting in overall temperature rises of ~4 °C in the water surrounding the plume. This gradual temperature rise had an effect on the phase measurement and, left uncorrected, manifested alignment errors in the beam pattern following numerical back projection. The corresponding phase decrease due to the ambient temperature rise was corrected by determining the slope of the phase decrease and adding the incremental change at each position in the phase measurement of the heated scan. Following this procedure, the backpropagated beam position compared favorably with the backpropagated beam position from the reference scan. The ambient temperature rise also prevented use of a higher spatial resolution for the scan, as the lengthy scan duration would have resulted in a significantly lower difference between the plume and ambient water temperature. The impact of the 0.2 mm step size was evaluated by comparing the sound field with that measured with a 0.1 mm step size. The back-projected fields for both scans were nearly identical and showed no signs of spatial aliasing, demonstrating that the larger step size was sufficient.

The second phase-contrast setup is shown in Fig. 1(b). A cylindrical tissue phantom was constructed (63 mm length, 50 mm diameter) so that it would possess a higher sound speed than that of the surrounding water. In this arrangement, the phantom was positioned with the front face 18 mm from the transducer and the reference scan at the focus consisted of measuring the field through the homogeneous phantom with a short water path on either side. Two phase contrast scans were performed following excision of a water column in the phantom. The column was constructed by inserting a thin-walled brass pipe (3.18 mm outer diameter) into the phantom and water from the tank was flushed through the resulting cavity with a syringe. The first scan was conducted following creation of a horizontal column and the second scan was conducted following placement of a second, vertically oriented column, so that a progression of features could be visualized following analysis of the multiple scans. The phase contrast in this arrangement was due to the lower sound speed of the water in the column relative to the surrounding phantom matrix. In addition to placing multiple identifiable features in the phantom, the reduced sound speed in the column was expected to result in a phase column opposite to that of the thermal plume.

The phantom construction was based on a commonly used recipe[31] with the sole difference of omitting graphite, as matching the attenuation was not of interest in this study. The phantom sound speed was measured to be $c_p = 1511$ m/s, using a time of flight method.[32] Given the long scan times, it was anticipated that swelling of the phantom would be an issue; the phantom holder restricted the side of

the phantom but swelling of the top and bottom walls of the phantom towards the transducer and the hydrophone could impact the phase measurement over time. The phantom surface was treated in glutaraldehyde prior to measurement to minimize swelling[33] and multiple reference scans were conducted to investigate the potential impact of swelling on the phase. The phase at the focus was found to shift approximately 0.49 rad uniformly across the field. The impact of such a shift has the effect of uniformly increasing the phase angle contrast but as the reconstruction of an absolute temperature was not the goal of the phantom measurements, this small shift in phase was neglected. These two reference scans were also used as a control case before the water columns were introduced into the phantom to examine whether a significant phase contrast was revealed following back projection. No identifiable phase contrast feature was apparent and the phase difference across the image was negligible, with a standard deviation of 0.027 rad.

A second phase contrast orientation was also tested in the phantom measurements. The orientation exhibited in Fig. 1 sought to mimic an arrangement where the diagnostic transducer was oriented perpendicular to the therapeutic beam axis, so that the heated region encountered by the diagnostic beam was radially symmetric about the therapeutic beam axis. In a clinical setting where acoustic windows are limited it may be difficult to readily achieve such an arrangement; an ideal situation would instead have the diagnostic transducer positioned inline with the therapeutic transducer. Such an arrangement would eliminate registration issues regarding transducer alignment but would require certain assumptions about the geometry of the heated region in order to properly reconstruct the temperature rise. This arrangement was tested by changing the orientation of the water column so that its axis of symmetry was aligned with the transducer acoustic axis. A fresh, homogeneous phantom was used and a similar reference scan was conducted, followed by a scan with an on-axis water column placed throughout the cylindrical length of the phantom.

B. Simulation

The experiments described in the previous section were simulated to provide a practical means for testing the impact of the spatial dependence of the phase contrast feature. The phase contrast was simulated by creating an initial sound field and forward projecting it via Eq. (2) to the focus through a homogeneous reference medium with a sound speed of either water, for the thermal plume, or the measured phantom sound speed. The spatial distribution for the phase shift from the simulated plume was based on the measured temperature distribution of the peak temperature gradient shown in Fig. 2, whereas the phase shift for the simulated water column was based on the sound speed from the measured temperature of the tank water and the column diameter. The phase contrast scenario was tested by projecting the field along the axis of propagation to the position of the phase contrast and adding a phase shift equivalent to the experimental shift (this step was omitted for the reference field projection). The fields were then projected to the focus. Both

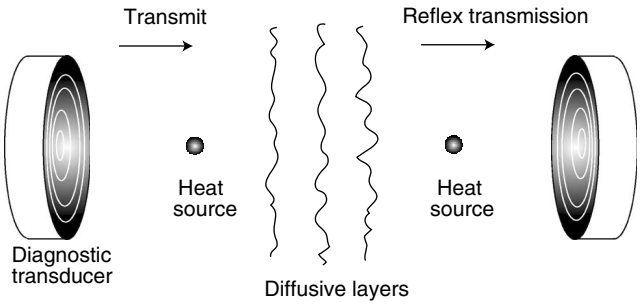


FIG. 3. Simulation arrangement for the phase contrast from a heat source with tissue effects present. The transducer used in the simulation was a 14-ring focused transducer with emission only from rings 8–11. The first case tested the arrangement shown here, whereas the second case did not have the heat source present upon transmit and the third case did not have the heat source present upon reflex transmission.

projected fields were backprojected to the position of the contrast region and the phase difference was tomographically reconstructed. The simulation was performed on a 78 $\times$ 78 mm grid with a 0.25-mm stepsize. The initial sound field was created by setting the amplitude on the 50-mm diameter transducer face to a constant value and the radius of curvature was simulated by incrementally shifting the initial phase depending on the radial position on the grid.

The feasibility of using RTI in a pulse–echo arrangement instead of through transmission was also simulated. The setup for this arrangement is shown in Fig. 3. The field was projected to the focus as described earlier, for the homogeneous reference case and with a phase contrast layer introduced along the path for the heated case. Near the focus, a diffusive phase shift layer was introduced to both fields in order to simulate the tissue structure which is expected to be responsible for scattering and reflecting the transmitted pulse back to the transducer. The field was then "reflected" back to the transducer by taking the complex conjugate of the phase and forward projecting the field to the equivalent position of the transducer, with a phase contrast layer again present for the heated field. The two fields were then back projected to the position of the contrast layer and reconstructed as before.

The phase shift of the diffusive layer was constructed by generating a random distribution for the phase. Physiologically the diffusion is not expected to necessarily follow a random distribution but rather a more structured pattern based on the tissue structure. The magnitude of the shift was based on measurements of the phase shift introduced by reflection off a tissue sample. Bovine muscle ($\sim$50 mm thick, store bought pot roast) was positioned at a 45° angle to the beam path approximately 50 mm from the transducer and the field was scanned with the hydrophone. The received pulse train at each position in the scan was analyzed to judge the extent of the phase shift from the expected phase of the transducer frequency. The reflected wave is a result of the initial reflection from the front tissue surface superimposed with energy reflected from a lateral part of the beam having transmitted into the tissue and subsequently scattered off the tissue structure towards the hydrophone. Multiple scans ex-

hibited a mean phase shift of ± 1 rad, so the random phase distribution for the simulated diffusive layer fell within ± 1 rad.

A potential drawback to using RTI in a pulse–echo arrangement is that the pulse is subject to heating on transmission and reflection, inducing a higher phase shift. Here it was hypothesized that a ring transducer may be employed such that the transmit pulse does not appreciably pass through the heated region but diffusive effects in the tissue sufficiently alter the beam pattern so that the pulse passes through the heated region only upon reflection. Such an arrangement would prevent the heated scan from experiencing up to double the phase shift but would require proper selection of the rings chosen for transduction. Interrogation of the heated region in the simulation was based on a 1.54 MHz concentric 14-ring focused transducer[34] (100-mm diameter, $f=1$), which will be available for future experimental study. Here, the heat source consisted of a spherical heat distribution rather than a linear plume distribution in order to more closely match the heat distribution resulting from a focused sound beam. This geometry resulted in a Gaussian distribution for the temperature radially away from the origin, rather than in just the lateral direction as in the plume arrangement.

A number of variables exist for simulating heating in this arrangement, such as number and position of transmit rings, position of the plume, number and position of the diffusive layers in the field. These variables were examined to evaluate the diffusion effect of tissue and three scenarios of heating were compared. Rings 8–11 were used for transmit and it was assumed that the entire returning sound field could be measured. Three diffusive layers were positioned 10 mm apart and 10 mm from the heat source, which was positioned 50 mm from the transducer. The heat source featured a 15 °C temperature rise relative to the rest of the medium, where $T_0=24$ °C. The sound speed dependence was modeled as that of water, corresponding to $c_0=1494$ m/s and $c_{\text{heat,max}}=1527$ m/s. The first scenario tested reconstruction of the temperature rise following propagation through the heat source and first diffusive layer on transmit, "reflection" off the second layer, and subsequent return through the third layer and heat source. The diffusion effects of the tissue for disrupting the beam focus was investigated in the second and third scenarios, where heating was applied only on the return or transmit paths, respectively. These two scenarios represent nonphysical situations but are included to demonstrate the feasibility of using a ring transducer and RTI for measuring temperature rise in tissue. The reference field for all scenarios differed only by omitting the phase contrast of the heat source, so that the same random phase distribution for the diffusive layer was used for both the reference and heated fields.

IV. RESULTS

A. Thermal plume

The temperature of the plume measured with the thermocouple is shown in Fig. 4(a). The positioning system scanned the thermocouple and hydrophone along the vertical y axis from −10 to 10 mm and marched across the lateral x axis from −10 to 10 mm. The temperature rise of the surrounding tank water rose approximately 3.5 °C over the ~4 h scan and the peak temperature rise of the plume was approximately 5 °C. Inspecting the temperature along the lateral (x) axis at any given vertical position revealed an approximately linear ambient temperature rise. Figure 4(b) shows the phase difference near the plume following back-projection of the two measured scans. The plume is identified by the slightly wavy decrease in the phase difference in the middle of the lateral axis. The difference between $\Delta\Phi$ in the plume ($x\sim 0$ mm) compared to the background $\Delta\Phi$ ($x\sim 3$ mm) was approximately 0.25 rad. The width of the measured plume was in good agreement with the width of the phase contrast plume.

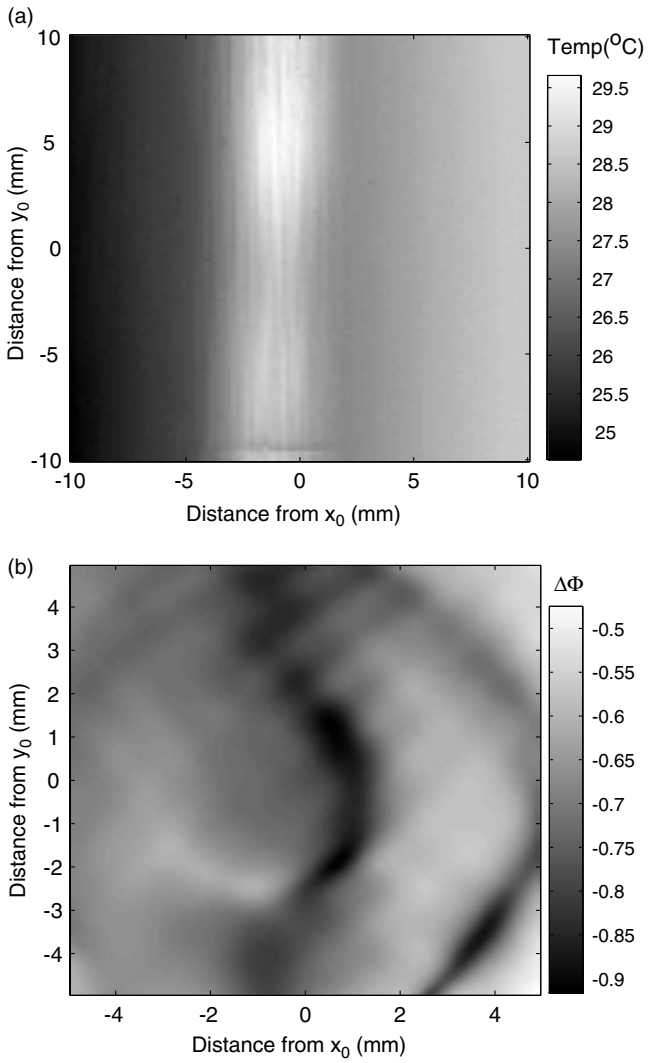


FIG. 4. (a) Temperature in the plume measured across the $x-y$ plane with a thermocouple. The scan started at $x=y=-10$ mm and moved in 0.2 mm increments, scanning along the y axis before returning to $y=-10$ mm and moving to the next position on the x axis. (b) Phase difference between the heated and reference fields following back projection close to the plume.

The results from the tomographic reconstruction of the simulation and experimentally measured phase contrast are shown in Figs. 5(a) and 5(b), respectively. Figure 5(a) compares the temperature distribution on which the simulated phase contrast was based to the reconstructed temperature

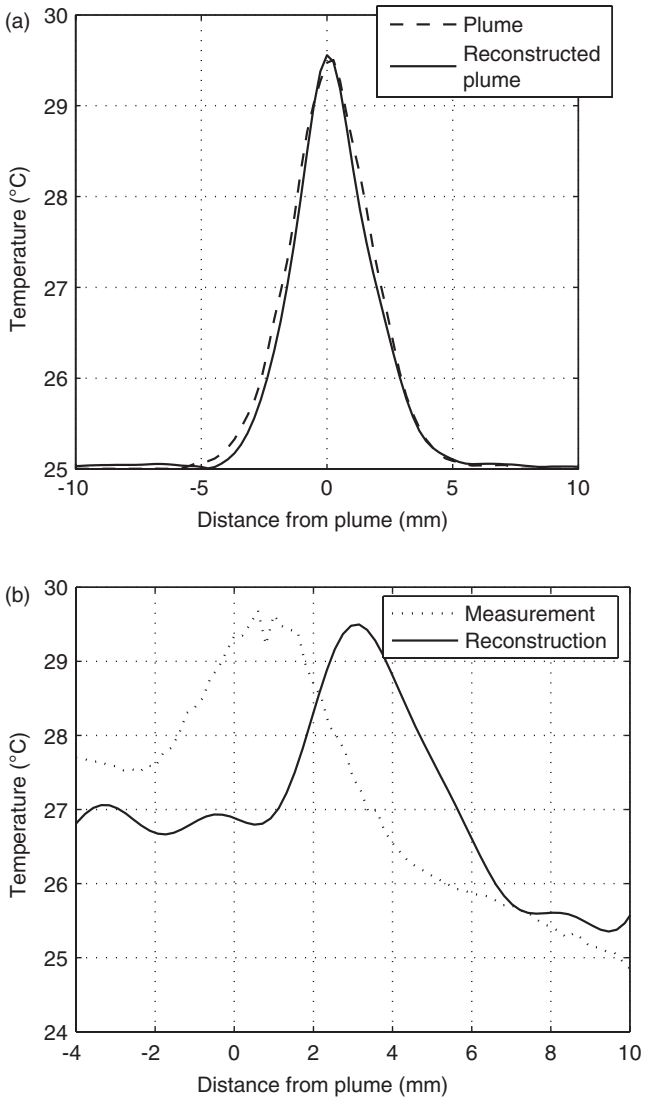


FIG. 5. (a) Comparison of the simulated plume temperature with the reconstructed temperature. (b) Comparison of the measured and estimated temperature following tomographic reconstruction.

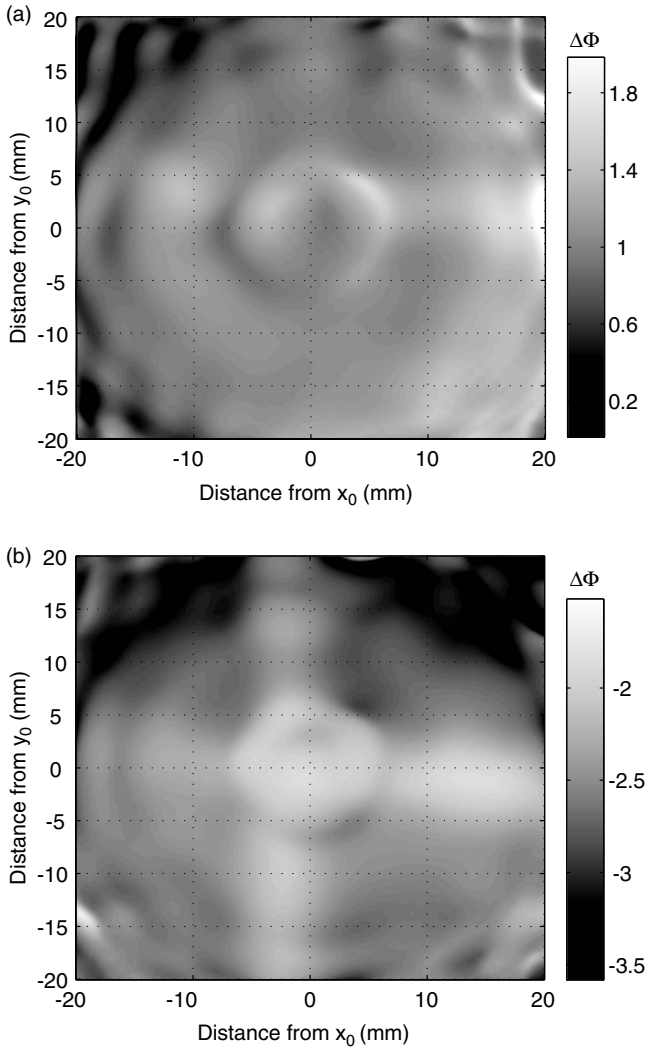


FIG. 6. (a) Phase difference between the phase contrast and reference fields following backprojection close to the horizontal water column in the phantom. (b) Phase difference following backprojection of the vertical and horizontal water column.

distribution. The two curves agree well, with a peak temperature difference of 0.06 °C. Figure 5(b) shows the temperature measured with the thermocouple compared to the temperature following reconstruction of the phase contrast measured with the hydrophone. The two curves are shifted slightly along the lateral axis, most likely due to alignment errors between the thermocouple and hydrophone. Despite the lateral shift, the width of each curve agrees well and the peak temperature differs by only 0.2 °C.

B. Phantom

The phase difference of the water columns present in the phantom measurement is displayed in Fig. 6. The horizontal water column is evident in Fig. 6(a) as a horizontal increase in $\Delta\Phi$ positioned just above the focus. The water column had to be carefully excised so as to keep the phantom position faithful to its position in the reference scan, so the column had a slight angle compared to the lateral x axis and was angled slightly out of the $x-y$ plane. In contrast to the thermal plume measurement where the phase decreased in the plume, the phase was higher in the water column due to the lower sound speed of the water compared to that of the phantom. The phase contrast of the horizontal column is not immediately clear in some parts of the image, where the image instead has a circular ripple effect. The addition of the vertical water column is visible along with the horizontal column in Fig. 6(b). Positioning errors in creating the vertical column were again evident as the two columns did not intersect, creating a region near the origin where the water path length with respect to the acoustic axis (z) was effectively doubled. Inspection of the phantom following measurement revealed that the two columns were separated by 7 mm along the z axis. In the non-overlapping portions of the columns the phase difference is approximately 0.4–0.6 rad but the phase contrast is higher near the origin where the water path was longer.

Of note in the phantom results are the concentric ripples in the phase difference and the higher phase difference seen in the positive side of the horizontal column compared to the

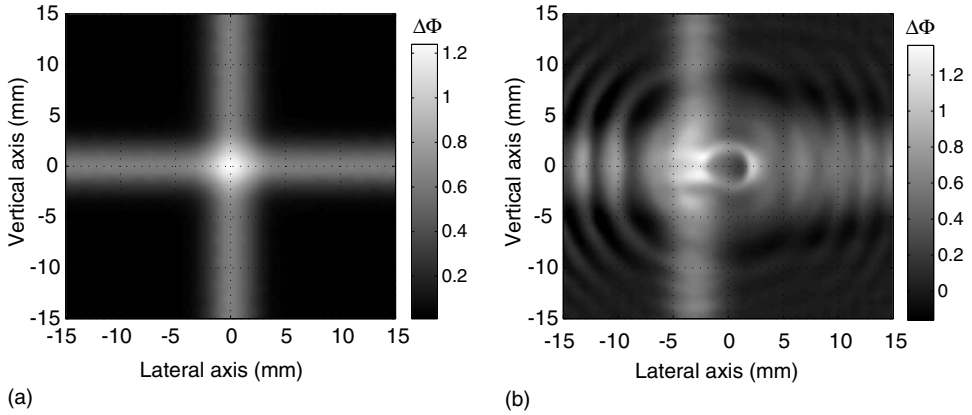


FIG. 7. (a) Backprojected phase difference between the simulated phase contrast and reference fields. Two overlapping water columns were used to produce the phase contrast. (b) Backprojected phase difference after separating the two columns by 7 mm and offsetting the vertical column from the origin of the lateral axis.

negative lateral side in both panels of Fig. 6. These features were examined in a simulation of the water columns present in the phantom. Figure 7(a) shows the simulated phase contrast for the water columns after back projection of the reference and phase contrast fields. Here the phase contrast of each water column was introduced so that the columns intersected in the $x-y$ plane but with double the phase shift at the region of intersection near the origin (i.e., the phase shift from each column was added separately into the simulation). The resulting phase difference reveals straight, perpendicular columns with a maximum phase difference of 0.6 rad outside of the origin and approximately 1.2 rad at the origin. The simulation presented in Fig. 7(b) sought to more closely replicate the experiment, where the columns were slightly misaligned. In the experiment, the columns were not only positioned 7 mm apart from each other along the acoustic z axis but were also set at a small angle out of the $x-y$ plane. This particular arrangement was simulated by positioning the columns 7 mm apart from one another along the z axis and the vertical column was offset from the lateral origin by 4 mm. The angle of each column out of the $x-y$ plane was replicated by separating each column at the origin and positioning the two halves of the column 2 mm apart along the acoustic z axis. Figure 7(b) exhibits a ripple effect similar to Fig. 6 and the phase difference is no longer symmetric, with a higher contrast on the negative side of the horizontal column compared to the positive side. In addition, the phase difference in the offset case was similar to the experimental phase difference, ranging between 0.4 and 0.6 rad in the nonoverlapping regions.

The final experimental arrangement tested the detection of a water column placed along the acoustic axis. The phase difference from back projection of this measurement is shown in Fig. 8. The water column is immediately evident by the circular region centered about the lateral origin and $y=7$ mm where the phase difference is highest. In contrast to the phase difference in the perpendicularly oriented water columns the phase difference here exhibits a sharp increase. The plateau effect seen in this phase contrast was most likely due to the geometry of the column in comparison to the perpendicular columns; the distance along the z axis over which the pulse traveled for the two media at any given position was the same, whereas the length of the path through the water column in the perpendicular orientation changed smoothly across the column, steadily shifting the phase.

C. RTI simulation

Figure 9 shows results from the RTI simulation where tissue diffusion effects were examined. The temperature, following projection through the medium and reconstruction of the phase contrast, is plotted for the case of the diagnostic pulse experiencing heating on transmission and reflection and for the cases of propagation through the heated region solely on the return or transmit paths. Compared to the actual temperature rise of 1527 m/s, the realistic situation where roundtrip heating occurs resulted in a maximum reconstructed sound speed of 1534 m/s. The case of idealized heating only on the return trip resulted in a slightly more accurate maximum sound speed of 1531 m/s. The overall width of the heated region is slightly wider in the roundtrip heating case compared to the return-path heating but is nonetheless similar to the simulated heat geometry. Restricting heat deposition to the transmit path resulted in negligible phase contrast throughout the entire 2D field, as contrasted

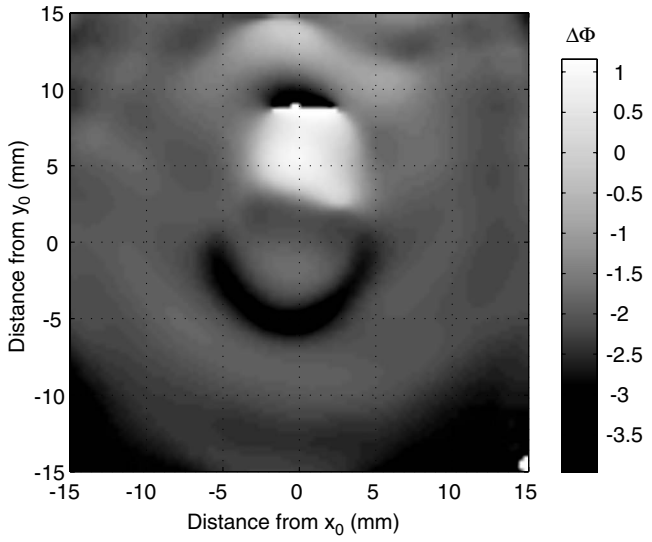


FIG. 8. Phase difference between an on-axis water column and reference field following backprojection.

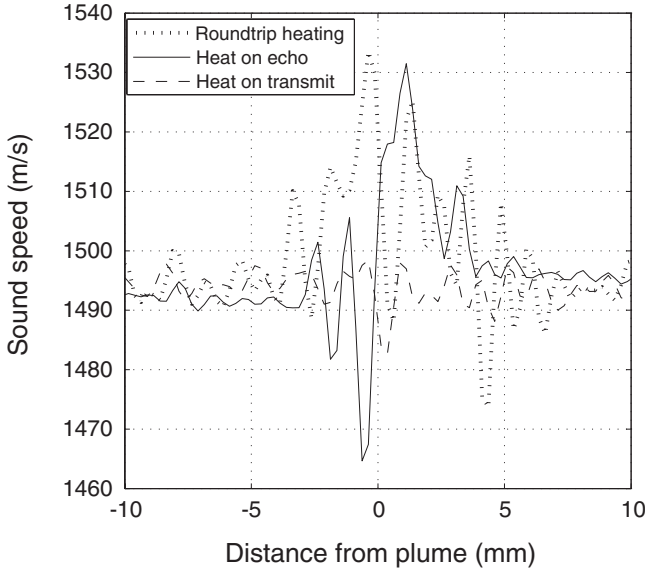


FIG. 9. Simulated comparison of the estimated sound speed for multiple heating arrangements in a pulse–echo setup. The case of heating on transmission and echo (dotted) is compared with heating only on echo (solid) and heating only on transmit (dashed). The simulated maximum sound speed was 1527 m/s.

with the roundtrip heating simulation in Fig. 10. The phase contrast following heating on transmission and reflection is clearly visible near the origin of Fig. 10(a), whereas no such contrast can be seen in Fig. 10(b).

V. DISCUSSION

A thermal plume and a phantom intersected by water columns were separately used to produce a small phase change in the field, and analysis of the sound fields revealed accurate identification of the geometry and location of the phase contrast feature. The phase difference following back-projection and subtraction of the reference field matched the expected change, where the phase decreased for the temperature rise in the plume but increased for the lower sound speed in the water column. In addition, the temperature from the field reconstruction in Fig. 5 showed excellent agreement with the measured plume temperature. Both measurements featured an offset in the overall phase difference due to a shift in the field, either from bulk heating of the water or from swelling of the phantom. These problems underscore the difficulty of replicating heating from a HIFU source for a measurement that extends over several hours, and the need for a faster method of acquiring the field data. In the thermal case, the background water temperature rise was addressed by assuming a linear increase in temperature and the phase was adjusted accordingly. Accounting for swelling in the phantom was not as straightforward and reveals the issue that phase contrast measurements are susceptible to error due to target movement, an important factor to address in future work. However, phase measurements are attractive due to their low noise sensitivity, and combining the therapeutic and diagnostic transducers could assist in reducing problems stemming from target movement. The on-axis water column

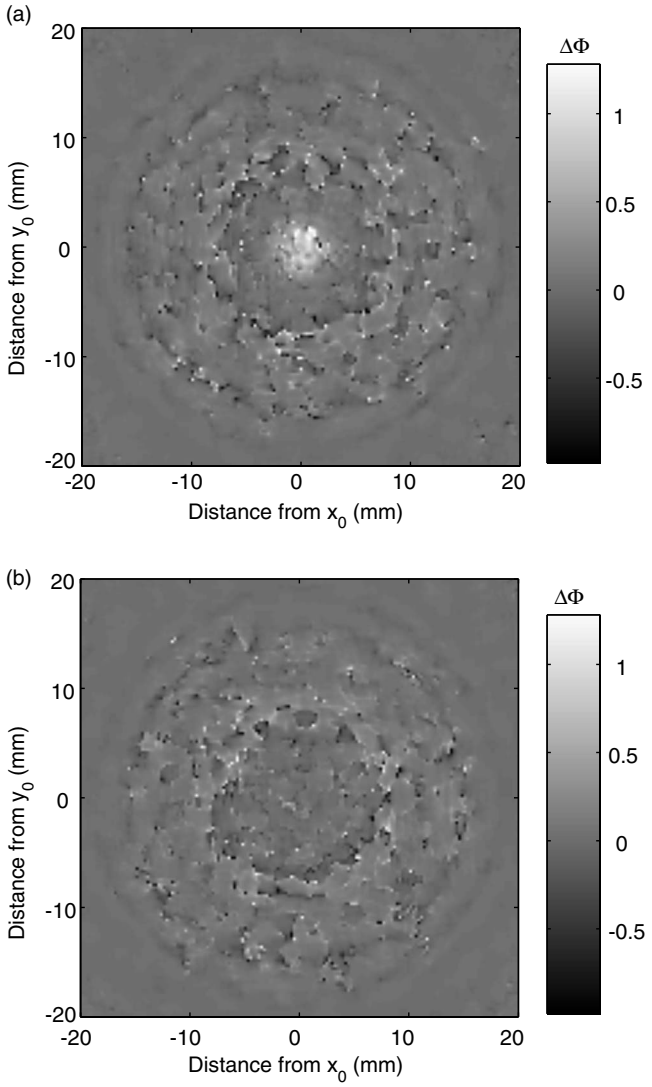


FIG. 10. Comparison of the simulated 2D phase difference for different heating arrangements in a pulse–echo setup when the outer rings of the transducer are used. Phase differences when heating is simulated (a) on transmission and reflex and (b) only on transmission.

measurement demonstrated that a co-axial arrangement of the diagnostic and therapeutic transducers is also feasible for detecting a phase change.

A couple of solutions exist for reducing the time required to measure the echo field, as a mechanical scan with a hydrophone is a lengthy process which would not be clinically feasible. An obvious, although hardware-intensive, solution is to use an array for the diagnostic transducer and acquire the rf signal from each element. This approach would require analog-to-digital conversion for each element to sufficiently reconstruct the field, and assumes that the entire sound field is contained within the geometrical footprint of the transducer so that the field can be properly back-projected. A second solution would be to use a phase-contrast filter as described by Clement and Hynynen.[26] This method would still require an array transducer in order to obtain sufficient spatial resolution but would not require a high sampling rate.

Simulation of the experimental setup revealed good agreement with the experimental results and helped explain some of the trends observed in the experimental data. The phase contrast in Fig. 6(b) is not uniform throughout the entire water column, as the contrast is greater for the right side of the horizontal column, as well as for the bottom portion (negative) of the vertical column, and the image shows ripples in the phase difference. Separating the simulated water columns along the acoustic z axis and offsetting the vertical column revealed a similar change in the phase contrast symmetry and emergence of concentric ripples in Fig. 7(b). These ripple artifacts have been well-documented in the ultrasound temperature imaging literature[13–15] and are attributed to a refractive thermal lensing effect. Deviation of the phase difference away from the expected uniform, straight distribution demonstrates that alignment of the contrast feature relative to the sound beam can alter the perceived geometry of the feature, as distortion from refraction becomes increasingly pronounced when the contrast is positioned out-of-plane.

The present experimental configuration did not permit RTI, but simulation of such an arrangement with an approximation of the tissue effects showed that reflex transmission might be feasible in not only localizing the heat source in a 2D plane but measuring the temperature as well. The phase shift produced from multiple tissue layers was simulated by shifting the phase in a random distribution. Coupled with using only the outer rings of a concentric ring transducer, Fig. 10 demonstrates that the simulated tissue diffusion adequately served to alter the beam so that sufficient energy passed through the heated region only upon reflection. Although the simulation demonstrates that a ring-shaped beam pattern can be made sensitive to heating solely on the return path, the reconstructed sound speeds displayed in Fig. 10 drop below the reference speed at certain points. This is a nonphysical result which is believed to be an artifact from the simulation, due to error in the phase that occurs when the beam ring pattern amplitude is low. When all the rings are employed and the beam amplitude is sufficiently high throughout the entire heated region, heating is detected on both the transmission and reflection paths. Although this results in a higher sound speed, the sound speed is always above the reference speed, as the beam pattern does not contain nulls due to a ring pattern and consequent phase errors. Until the sound field reflected from the tissue structure following heat deposition is measured it will be difficult to know exactly how the tissue structure will affect the beam geometry along the entire pathlength. Nonetheless, if heating sufficiently alters the sound field on both the transmission and return paths at the very least a pulse–echo RTI setup should be effective at visualizing the 2D heat deposition.

As with most ultrasound thermographic methods, an additional factor in judging the success of this method is the underlying assumed relationship between temperature and sound speed. Multiple tissue layers of varying composition will affect not only the final temperature reconstruction but the backward projection as well. The use of ultrasound phase-contrast transmission imaging for identifying multiple tissue types has been previously investigated[26] and remains an ongoing area of investigation.

The current method of tomographic reconstruction assumes radial symmetry, as integration of the phase difference is performed along a single direction. This assumption is fine for the current experimental arrangement but future setups where the heated region may be oblong rather than circular will require certain assumptions regarding the temperature distribution or a different approach for integrating the field. Left unaddressed, asymmetry in the temperature distribution would almost certainly result in miscalculation of the spatial extent and amplitude of the temperature. Once an array-based measurement system is developed, the acquisition time will decrease dramatically, which will allow the effect of alignment and lesion geometry on the reconstruction algorithm to be more easily investigated. Ultimately, the result from this system and analysis will be compared to MR thermometry measurements.

Other unknown factors include effects of cavitation on the propagation of the signal through a bubbly region. The sound speed is not expected to change appreciably due to the likely-low number of bubbles present but it is possible that scattering effects may preclude the diagnostic signal from propagating past and back through the bubbly region. However, a significant number of bubbles present in the field may suggest that boiling has already occurred, at which point the temperature will be significant and possibly above the target temperature.

Based on the sound speed dependence of water on temperature, the experimental and simulated investigations presented here were intended to show feasibility that the phase contrast obtained after backward projection and tomographic reconstruction can localize and measure a two-dimensional heat distribution. Experimental restrictions on measuring the sound field limited us to adopting a through-transmission setup, so although using an ultrasound through-transmission setup to measure temperature rise is not a novel endeavor, the techniques outlined and results obtained here incorporate a new approach that should translate well to a pulse–echo arrangement.

Many previous techniques[13–19] rely on a beam-formed amplitude contrast image and correlation methods to determine the temperature change. These techniques can be easily corrupted by speckle noise, an effect which does not directly represent a physical feature in the region of interest. The approach described here gives a map of the sound field at the location of the feature which is of interest and may provide a more accurate description. Another advantage of this approach is that the temperature change is based on the phase information, which is less susceptible to noise and attenuation effects. Additionally, the computational time required to numerically project the field is relatively fast, promising a near-real-time imaging method. A number of issues still remain, such as motion and image registration, double-valued sound speeds at high temperatures, and thermal lensing effects, and will be addressed in future investigations. Many of these problems will be more easily addressed with an array-based acquisition system. The backprojection approach gives

some flexibility in observing how the field develops as it propagates, which may provide insight into accounting for thermal lensing effects. One solution to distinguishing between double-valued sound speeds at high temperatures would be to track the temperature change continuously by interleaving the diagnostic and therapeutic sonications. If the temperature change is sufficiently smooth and continuous, the direction of the sound speed change should indicate the direction of the temperature change.

VI. CONCLUSION

The ultrasound phase-contrast method described here obtained estimates in good agreement with the measured and observed phase contrast in the medium. The temperature contrast estimate was within 0.2 °C of the measured temperature and the water columns were clearly identified in the phantom measurements. Phase contrast was produced using both a thermal and sound speed change method, demonstrating that the technique is not limited solely to thermal features. Future application of these methods to a pulse–echo arrangement will allow the technique to be better suited for clinical adoption where acoustic access is limited, either with a perpendicular or parallel orientation to the therapeutic acoustic axis. The fast reconstruction time in conjunction with an array-based diagnostic transducer and phase contrast filter may allow for intermittent interrogation of the field while the therapeutic field is off, providing a continuous update of the temperature during treatment.

ACKNOWLEDGMENTS

Financial support for this research was provided by the NIH via Award No. U41 RR19703.

[1]B. Quesson, J. A. de Zwart, and C. T. Moonen, "Magnetic resonance temperature imaging for guidance of thermotherapy," J. Magn. Reson. Imaging **12**, 525–533 (2000).
[2]N. McDannold, "Quantitative MRI-based temperature mapping based on the proton resonant frequency shift: Review of validation studies," Int. J. Hyperthermia **21**, 533–546 (2005).
[3]A. Kristensen and J. Dalen, "Acoustic estimation of size distribution and abundance of zooplankton," J. Acoust. Soc. Am. **80**, 601–611 (1986).
[4]M. Keidar, I. D. Boyd, and I. I. Beilis, "Electrical discharge in the Teflon cavity of a coaxial pulsed plasma thruster," IEEE Trans. Plasma Sci. **28**, 376–385 (2000).
[5]W. M. L. Morawitz, P. J. Sutton, P. F. Worcester, B. D. Cornuelle, J. F. Lynch, and R. Pawlowicz, "Three-dimensional observations of a deep convective chimney in the Greenland Sea during winter 1988/89," J. Phys. Oceanogr. **26**, 2316–2343 (1996).
[6]D. D. Iorio, D. Lemon, and R. Chave, "A self-contained acoustic scintillation instrument for path-averaged measurements of flow and turbulence with application to hydrothermal and bottom boundary layer dynamics," J. Atmos. Ocean. Technol. **22**, 1602–1617 (2005).
[7]F. A. Duck, *Physical Properties of Tissue* (Academic, London, 1990).
[8]K. Hynynen, N. I. Vykhodtseva, A. H. Chung, V. Sorrentino, V. Colucci, and F. A. Jolesz, "Thermal effects of focused ultrasound on the brain: Determination with MR imaging," Radiology **204**, 247–253 (1997).
[9]N. McDannold, C. M. Tempany, F. M. Fennessy, M. J. So, F. J. Rybicki, E. A. Stewart, F. A. Jolesz, and K. Hynynen, "Uterine leiomyomas: MR imaging-based thermometry and thermal dosimetry during focused ultrasound thermal ablation," Radiology **240**, 263–272 (2006).
[10]R. Seip and E. S. Ebbini, "Noninvasive estimation of tissue temperature response to heating fields using diagnostic ultrasound," IEEE Trans. Biomed. Eng. **42**, 828–839 (1995).
[11]R. Maass-Moreno and C. A. Damianou, "Noninvasive temperature estimation in tissue via ultrasound echo-shifts. Part I. Analytical model," J. Acoust. Soc. Am. **100**, 2514–2521 (1996).
[12]R. Maass-Moreno, C. A. Damianou, and N. T. Sanghvi, "Noninvasive temperature estimation in tissue via ultrasound echo-shifts. Part II. In vitro study," J. Acoust. Soc. Am. **100**, 2522–2530 (1996).
[13]C. Simon, P. VanBaren, and E. S. Ebbini, "Two-dimensional temperature estimation using diagnostic ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1088–1099 (1998).
[14]N. R. Miller, J. C. Bamber, and G. R. ter Haar, "Imaging of temperature-induced echo strain: Preliminary *in vitro* study to assess feasibility for guiding focused ultrasound surgery," Ultrasound Med. Biol. **30**, 345–356 (2004).
[15]M. Pernot, M. Tanter, J. Bercoff, K. R. Waters, and M. Fink, "Temperature estimation using ultrasonic spatial compound imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 606–615 (2004).
[16]M. D. Abolhassani, A. Norouzy, A. Takavar, and H. Ghanaati, "Noninvasive temperature estimation using digital images," J. Ultrasound Med. **26**, 215–222 (2007).
[17]A. Anand, D. Savery, and C. Hall, "Three-dimensional spatial and temporal temperature imaging in gel phantoms using backscattered ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 23–31 (2007).
[18]T. Varghese, J. A. Zagzebski, Q. Chen, U. Techavipoo, G. Frank, C. Johnson, A. Wright, and F. T. Lee, "Ultrasound monitoring of temperature change during radiofrequency ablation: Preliminary *in-vivo* results," Ultrasound Med. Biol. **28**, 321–329 (2002).
[19]T. Varghese and M. J. Daniels, "Real-time calibration of temperature estimates during radiofrequency ablation," Ultrason. Imaging **26**, 185–200 (2004).
[20]R. L. King, G. T. Clement, S. Maruvada, and K. Hynynen, "Preliminary results using ultrasound transmission for image-guided thermal therapy," Ultrasound Med. Biol. **29**, 293–299 (2003).
[21]K. Mizutani, K. Nishizaki, K. Nagai, and K. Harakawa, "Measurement of temperature distribution in space using ultrasound computerized tomography," Jpn. J. Appl. Phys., Part 1 **36**, 3176–3177 (1997).
[22]E. Ishikawa, K. Mizutani, and K. Nagai, "Fast method for visualization of temperature distribution using acoustic computerized tomography," Jpn. J. Appl. Phys., Part 1 **40**, 5446–5449 (2001).
[23]P. S. Green and M. Arditi, "Ultrasonic reflex imaging," Ultrason. Imaging **7**, 201–214 (1985).
[24]P. S. Green, J. S. Ostrem, and T. K. Whitehurst, "Combined reflection and transmission ultrasound imaging," Ultrasound Med. Biol. **17**, 283–289 (1991).
[25]D. Rallan, N. L. Bush, J. C. Bamber, and C. C. Harland, "Quantitative discrimination of pigmented lesions using three-dimensional high-resolution ultrasound reflex transmission imaging," J. Invest. Dermatol. **127**, 189–195 (2007).
[26]G. T. Clement and K. Hynynen, "Ultrasound phase-contrast transmission imaging of localized thermal variation and the identification of fat/tissue boundaries," Phys. Med. Biol. **50**, 1585–1600 (2005).
[27]V. A. Del Grosso, and C. W. Mader, "Speed of sound in pure water," J. Acoust. Soc. Am. **52**, 1442–1446 (1972).
[28]P. R. Stepanishen and K. C. Benjamin, "Forward and backward projection of acoustic fields using FFT methods," J. Acoust. Soc. Am. **71**, 803–812 (1982).
[29]D. Liu and R. Waag, "Propagation and backpropagation for ultrasonic wavefront design," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1–13 (1997).
[30]A. C. Kak, and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE Press, Piscataway, NJ, 1988), Chap. 3, pp. 49–60.
[31]M. M. Burlew, E. L. Madsen, J. A. Zagzebski, and R. A. Banjavic, "A new ultrasound tissue-equivalent material," Radiology **134**, 517–520 (1980).
[32]J. C. Bamber, "Ultrasonic properties of tissues," in *Ultrasound in Medicine*, edited by F. A. Duck, A. C. Baker, and H. C. Starritt, (IOP Publishing, Philadephia, 1998), pp. 68–69.
[33]L. K. Ryan, and F. S. Foster, "Tissue equivalent vessel phantoms for intravascular ultrasound," Ultrasound Med. Biol. **23**, 261–273 (1997).
[34]T. Fjield, X. Fan, and K. Hynynen, "A parametric study of the concentric-ring transducer design for MRI guided ultrasound surgery," J. Acoust. Soc. Am. **100**, 1220–1230 (1996).

Interaction of microbubbles with high intensity pulsed ultrasound

Siew Wan Fong and Evert Klaseboer
Institute of High Performance Computing, 1 Science Park Road, #01-01 The Capricorn, Singapore Science Park II, Singapore 117528

Boo Cheong Khoo[a)]
Department of Mechanical Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260

(Received 3 September 2007; revised 17 December 2007; accepted 26 December 2007)

High intensity pulsed ultrasound, interacting with microbubble contrast agents, is potentially useful for drug delivery, cancer treatment, and tissue ablation, among other applications. To establish the fundamental understanding on the interaction of a microbubble (in an infinite volume of water) with an ultrasound pressure field, a numerical study is performed using the boundary element method. The response of the bubble, in terms of its shape at different times, the maximum bubble radius obtained, the oscillation time, the jet velocity, and its translational movement, is studied. The effect of ultrasound intensity and initial bubble size is examined as well. One important outcome is the determination of the conditions under which a clear jet will be formed in a microbubble in its interaction with a specific sound wave. The high speed jet is crucial for the aforementioned intended applications. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2836746]

PACS number(s): 43.80.Sh, 43.80.Gx, 43.35.Ei [DLM] Pages: 1784–1793

I. INTRODUCTION

Encapsulated microbubbles are often used as contrast agents in nondestructive medical imaging. The acoustic impedance difference between bodily fluid (such as blood) and the surrounding tissues is low. Therefore the introduction of microbubbles in the blood stream enhances the amount of backscattering and allows imaging of the blood flow, both spatially and temporally. Currently, these bubbles provide a noninvasive method to visualize a specific part of the body for diagnostic purposes (Feinstein, 2004; Lepper *et al.*, 2004).

Recently, the use of microbubble contrast agents has been proposed to be extended to therapeutic procedures (Pitt *et al.* 2004). A few notable examples in development are: thrombosis and vascular plaque treatments (Unger *et al.*, 1998; Tachibana and Tachibana, 1995; Tsutsui *et al.*, 2006) and drug and gene deliveries (Taniyama *et al.*, 2002; Li *et al.*, 2003; Bekeredjian *et al.*, 2005). More significantly, if these bubbles are coated with ligands (antibodies and peptides), they can attach to specific targeted cells (Unger *et al.*, 1998; Klibanov, 1999; Dayton *et al.*, 2004; Lanza *et al.*, 1996). This technique can be used to enhance the reflected ultrasonic signal from the diseased area, such as cancerous tissue, for a better image in diagnostics, because a large number of microbubbles will attach to the respective tissues. Microbubbles not only enhance image contrast but may also be utilized for therapeutic procedures. One such possibility is the destruction of the attached diseased cells when the bubbles are acoustically forced to collapse.

It is known that when a shock wave or pulsed ultrasound of sufficiently high pressure interacts with a bubble, the latter will collapse violently and an extremely high speed jet can develop in the bubble (up to 2.6 km/s as reported in Bourne and Field, 1999). This microjet traverses the bubble and is directed in the propagation direction of the wave (Leighton, 1994). Similar jets are observed for a bubble collapsing near a solid boundary (toward the boundary) or near a free surface (away from the surface) without the presence of a sound wave. Nevertheless the jet formed in bubbles collapsing near such surfaces is generally an order of magnitude lower in speed (about 100 m/s) when compared to jets seen in shock wave–bubble interaction. High speed jets in bubble–sound wave interactions are difficult to observe, although bubble wall velocities exceeding 700 m/s have been observed experimentally (May *et al.*, 2002).

The impacting jet of an attached bubble interacting with a high amplitude acoustic wave will create high shear stress on the cell membranes causing the membranes to deform or even rupture. Possible applications are drug delivery via sonoporation (Miller *et al.*, 2002; Ohl *et al.*, 2006; Le Gac *et al.*, 2007) or cell destruction (Prentice *et al.*, 2005). Shock wave–bubble interaction has been previously studied, among others, by Ding and Gracewski (1996), Turangan *et al.* (2007), and Klaseboer *et al.* (2006, 2007). Recently, the use of pulsed ultrasound has found renewed interest, as a possible replacement for extracorporeal shock wave lithotripsy (i.e. kidney stone removal therapy using shock waves). The use of pulsed ultrasound is expected to reduce collateral damage on nearby tissue caused by strong shock waves (Ikeda *et al.*, 2006).

The basic phenomena concerning the interaction of a microbubble with a pulsed ultrasound wave are provided in

[a)]Also affiliated with: Singapore-MIT alliance, 4 Engineering Drive 3, Singapore 117576. Author to whom correspondence should be addressed. Electronic mail: mpekbc@nus.edu.sg

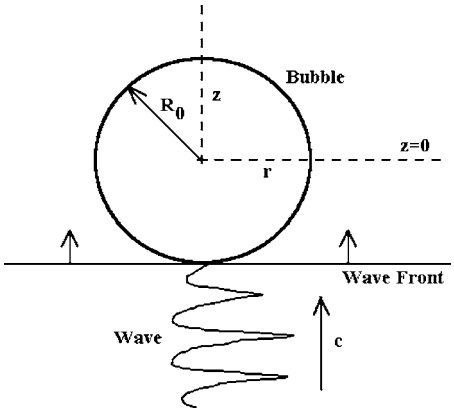


FIG. 1. A bubble hit by an ultrasound wave consisting of pulsed ultrasound with three peaks in this case. The wave propagates from below and travels with the speed of sound, c, in an upwards direction. At $t=0$, the wave front, indicated by a horizontal line, first makes contact with the bottom part of the bubble. The bubble will start to interact with the sound wave after this instant. The bubble has an initial radius R_0. The wave is not drawn to the same scale as the bubble.

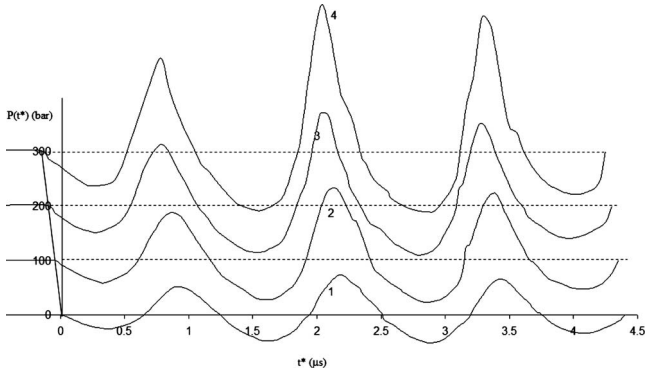


FIG. 2. Pulsed ultrasound with increasing intensities of 1000 (pulse 1), 3000 (pulse 2), 5000 (pulse 3), and 9000 W/cm^2 (pulse 4) as used by Xu *et al.* (2005), see also Table I. The number indicated relates to these four pulses (1–4). All four pulses start off with a tensile component that will cause the bubbles to expand before they are forced to collapse by one of the compressive components of the pulse. Each pulse has four tensile parts and three compressive parts. The time scale t^* is relative, due to the finite value of the velocity of sound. The wave arrives earlier at certain parts of the bubble and then progressively moves across the whole bubble. The plot is staggered, in order to show the four pulses in one graph. The plots consist of the sum of the acoustic and the atmospheric pressure: thus all pulses have $P(t^*)=1$ bar for $t^* <=0$. In the simulations, every bubble collapsed at times smaller than 4 μs. The pulses have a main frequency of 0.8 MHz.

this work. The emphasis of this article is to obtain a better understanding of the underlying physics via numerical simulations. Experimental results are very difficult to obtain, due to the very small length and time scales (see, e.g., May *et al.*, 2002) and numerical simulations seem to be the ideal tool to investigate such an interaction. The findings could lead to future improvements in the design and application of clinical treatments involving bubble–pulsed ultrasound interaction.

A microbubble is basically a small gas bubble with a thin membrane made of albumin, galactose, lipid, or polymers (Lindner, 2004) to prevent it from dissolving too quickly. The size of the microbubble contrast agents currently available ranges from 1 to 8 μm. This small size allows them to follow the blood flow and to be transported to the site of interest. Hence, due to this typical range of size of the microbubbles, our simulations are performed by focusing predominantly on bubbles between 1 and 10 μm: a few larger sized bubbles are also investigated.

A quiescent bubble, with initial radius R_0, may be hit by a strong sound wave, such as high intensity pulsed ultrasound or a shock wave, and as a result the bubble will start to interact with the wave (Fig. 1). Some typical sound waves were used in this article, taken from experiments done by Xu *et al.* (2005). These focused pulsed ultrasonic waves were used to mechanically remove tissue in a localized, controlled manner for clinical treatments. Acoustic peaks in the observed signals, detected using a hydrophone, were found when effective tissue erosion occurred. This leads to the contention that collapsing cavitation bubbles are responsible for the desired tissue removal effect. Four different sound wave intensities are used (see Fig. 2 and Table I), namely 1000, 3000, 5000, and 9000 W/cm^2, which are named pulses 1, 2, 3, and 4, respectively. The main frequency of the pulses is 0.8 MHz. These ultrasound waves are generated using a single-element transducer. Each pulse is made up of three cycles. The first cycle of pulse 1 obtains a minimum of −25 bar, Pulse 2 −39 bar, Pulse 3 −48 bar, and Pulse 4 −62 bar. The first cycle's maxima were 52, 86, 112, and 169 bar for pulses 1–4, respectively, and all occurred at about 0.9 μs after the initial instance of the wave front. The second maxima were 74, 134, 177, and 270 bar for pulses 1, 2, 3, and 4, respectively. All these second maxima occurred 2.2 μs after the initial wave front. It is noted that all four pulses start with a negative pressure component. This will turn out to be essential because it will make the bubble grow first, before it collapses. The profiles of Xu *et al.*, 2005 are used here, as they represent a typical set of sound waves as used for clinical tests. The jets from the collapsing bubbles can be utilized for various purposes such as drug delivery or cell lysing.

A model to simulate the interaction of pulsed ultrasound with a microbubble is presented in Sec. II using a boundary element method (BEM). The results in Sec. III show the interaction between bubbles of various sizes and the four different pulsed ultrasonic waves. The focus of this study is to investigate when a clear high speed jet is formed in the bubble during the interaction with the sound wave. A discussion is given in Sec. IV followed by conclusions in Sec. V.

II. PULSED ULTRASOUND–BUBBLE INTERACTION MODELED WITH THE BOUNDARY ELEMENT METHOD

The bubble will show considerable volume variation, due to the high pressures involved. Consequently, the bubble behavior will be highly nonlinear. The model of pulsed ultrasound–bubble interaction given here follows closely the work of Klaseboer *et al.* (2006), who used a BEM. The primary difference is that in the current article a pulsed ultrasound wave form (Fig. 2) instead of a compressive shock wave with constant amplitude is used. Also, the initial size of the bubbles considered is much smaller here. A brief description of the theory is given in the following for completeness.

TABLE I. Intensity and peak pressures, both negative and positive, of the first cycle of the pulsed ultrasound waves, pulses 1–4. Their effects on the collapse time t_c, the maximum radius of the microbubbles, R_{max} and the translation of the bubble center at the moment of jet impact for initial bubble radii $R_0=1$ and 10 μm. All other bubbles investigated, ranging from $R_0=2-9$ μm, obtain values between these two extremes.

Pulse	Intensity (W/cm^2)	Peak pressure of the first cycle (bar)		t_c(μs)		R_{max} (μm)		Translation (μm)	
		Negative	Positive	1 μm	10 μm	1 μm	10 μm	1 μm	10 μm
1	1000	−25	52	1.23	1.37	25.1	29.8	8.1	8.2
2	3000	−39	86	1.21	1.31	31.2	35.7	11.1	11.8
3	5000	−48	112	1.20	1.30	34.8	39.3	12.5	13.5
4	9000	−62	169	1.23	1.32	40.5	44.7	15.2	17.1

The flow around a bubble can be described with a velocity potential, Φ, which satisfies the Laplace equation $\nabla^2\Phi=0$. The velocity vector $\mathbf{u}$ may be obtained by taking the gradient of the potential, $\mathbf{u}=\nabla\Phi$. The pressure, p, anywhere in the fluid surrounding the bubble can be found from the unsteady Bernoulli equation

$$p = P_{\text{ref}} - \rho\frac{D\Phi}{Dt} + \frac{1}{2}\rho|\mathbf{u}|^2, \tag{1}$$

where the reference pressure is indicated by P_{ref}, t represents time and ρ is the density of the fluid. The material derivative $D/Dt=\partial/\partial t+\mathbf{u}\cdot\nabla$ is used in Eq. (1), as the bubble surface will move with the flow. Gravity has been neglected in Eq. (1), due to the very small sizes of the bubbles. Usually, P_{ref} is a constant, e.g., the hydrostatic pressure. However, as in Klaseboer *et al.* (2006, 2007) and Calvisi *et al.* (2007), the influence of the sound wave is incorporated through the introduction of the acoustic and atmospheric pressure in the reference pressure. This wave is both time and space dependent as it travels at the speed of sound. Further, it is assumed that the wave is planar and thus

$$P_{\text{ref}} = P(t,z) = P(t^*), \tag{2}$$

where

$$t^* = t - (z + R_0)/c \tag{3}$$

and c represents the speed of sound, which is assumed to be $c=1500$ m/s. For $t^*<0$ P_{ref} is the hydrostatic pressure P_{ATM} ($P_{\text{ATM}}=1$ bar). The choice of Eq. (3) assumes that at $t=0$, the wave front has just reached the bottom of the bubble ($z=-R_0$) as shown in Fig. 1. An axisymmetric configuration is assumed, where the r-axis is parallel to the advancing wave front and the wave is traveling in the positive z direction. The pressure profile of each of the wave forms of Fig. 2 is discretized into 52 points, from which the value of $P(t^*)$ is obtained using linear interpolation. The initial center of the bubble is located at $(r,z)=(0,0)$. The gas inside the bubble is considered an ideal gas as a first approximation. The pressure inside the bubble, p_b, is assumed to behave adiabatically, i.e., $p_b=p_0(V_0/V)^\gamma$, where $p_0=P_{\text{ATM}}$ and γ is the ratio of specific heats of the bubble contents assumed to be $\gamma=1.4$ for an ideal gas in this work. The initial volume V_0 corresponds to the initial bubble radius R_0 as $V_0=4\pi R_0^3/3$. It is assumed that the effects of surface tension and possible membrane stiffness are small compared to the very large pressures induced by the strong sound wave. As such, it is possible to set the pressure in the fluid just outside the bubble equal to the pressure inside the bubble, thus $p=p_b$ in Eq. (1). However, the effect of surface or membrane tension can be easily incorporated; some results with the effect of surface tension included are given in Sec. IV. Thus finally,

$$\rho\frac{D\Phi}{Dt} = P(t^*) - p_0\left(\frac{V_0}{V}\right)^\gamma + \frac{1}{2}\rho|\mathbf{u}|^2. \tag{4}$$

The pressure $P(t^*)$ is set equal to the pressure caused by the sound wave, by substituting the z-value of the bubble surface for each point on the bubble surface at a given time. Thus a different pressure is present on different parts of the bubble surface. Note that the reference pressure $P(t^*)$ is not the pressure at the bubble wall, but rather it is the pressure due to the wave that would have occurred at a certain location and time, if the bubble had not been present. Thus the pressure caused by the ultrasound wave in the fluid is incorporated through the reference pressure with a time delay. The boundary element method can now be used to obtain a relationship between the potential on the bubble surface and the normal velocity $\partial\Phi/\partial n=\mathbf{n}\cdot\nabla\Phi$ on the surface of the bubble, i.e.

$$s(\mathbf{x})\Phi(\mathbf{x}) + \int_S \Phi(\mathbf{y})\frac{\partial G(\mathbf{y},\mathbf{x})}{\partial n}dS = \int_S G(\mathbf{y},\mathbf{x})\frac{\partial\Phi(\mathbf{y})}{\partial n}dS, \tag{5}$$

with s the solid angle, $G(\mathbf{y},\mathbf{x})=1/|\mathbf{x}-\mathbf{y}|$ represents the free space Green's function, and $\partial G/\partial n$ its normal derivative. As the passing of the wave imposes different conditions at different locations through Eq. (4), the potential and normal velocity are different at different positions on the bubble surface. The surface of the bubble is divided into 51 nodes and 50 linear elements (one element less due to the axial symmetry of the problem). Equation (4) is used to update the potential at each node and the normal velocities for each node can be obtained with Eq. (5). Once the velocity of each node is calculated, the surface of the bubble can be updated and the cycle can be repeated for the next time step. Equation (5) has the advantage that only a single integration on the surface of the bubble is required. It is not necessary to solve the whole fluid field, this is done indirectly via the boundary element method. The numerical method was tested thoroughly in the past and was used to predict underwater explosions (Wang *et al.*, 1996), and the behavior of oscillat-

ing bubbles near elastic materials (Fong *et al.*, 2006; Klaseboer and Khoo, 2004). The time step is constant during the simulations and taken to be $5\times10^{-5}R_0$. The validity of the model/approach, i.e., using potential theory to describe the interaction of a strong sound wave with a bubble, was discussed in detail in Klaseboer *et al.* (2006) and further confirmed by experimental data in Klaseboer *et al.* (2007). It is believed that the flow field around the bubble is mainly controlled by inertial effects and the compressibility in the fluid due to the passage of a rectangular shaped shock wave as in Klaseboer *et al.* (2006), or a lithotripter shock wave as in Klaseboer *et al.* (2007) can be effected through the reference pressure in Eq. (4). As the jet impact velocity is of interest, rather than the flow resulting from this impact, the BEM simulations were stopped when the jet hits the opposite surface of the bubble. Theoretically, the simulations could have been continued by introducing a toroidal bubble with a vortex ring in the interior of the bubble: this could be the subject for a future study. It is also apparent from Klaseboer *et al.* (2007) that the possibility of shock wave reflection (not modeled here) in the interior of the bubble before jet impact does not greatly affect the general motion of the flow. As such, the employment of potential flow is deemed directly applicable in the present simulation of microbubbles interacting with strong sound waves, in particular the pulsed ultrasound wave form. In general, experimental data are difficult to obtain, mainly due to the very short time and length scales, and numerical simulation therefore offers an alternative opportunity to obtain further insight into the phenomena involved. The model simulates a full bubble–sound wave interaction in several minutes on a common PC-platform. More recently, the boundary element method was also used to simulate the collapse of microbubbles in acoustic traveling waves by Calvisi *et al.* (2007).

An often useful quantity in describing impacting jets in bubbles is the Kelvin impulse K, defined as the absolute value of the integral of the potential Φ on the bubble surface S multiplied by the fluid density and the normal vector $\mathbf{n}$ at this surface (Pearson *et al.*, 2004); i.e.,

$$K = \left| \rho \int_S \Phi \mathbf{n} dS \right| . \tag{6}$$

The Kelvin impulse is usually considered as a vector quantity, but in this axisymmetrical configuration, only the z-component of the vector, or K in Eq. (6) remains. The Kelvin impulse is somehow related to the jet speed and the broadness of the jet. For a spherically oscillating bubble, the Kelvin impulse is zero, as in that case the potential is uniform across the bubble surface and $\int_S \mathbf{n} dS = \mathbf{0}$. Thus the Kelvin impulse gives an indication of the unsymmetrical flow around the bubble. K is computed in this work to provide a comparison between the different pulsed ultrasonic wave forms (Fig. 2) interacting with different initial bubble sizes.

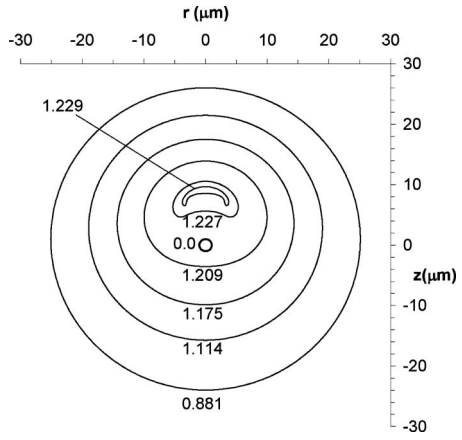


FIG. 3. The shape of a microbubble of an initial radius of 1 μm after it has been hit by pulse 1, which propagates from below. The bubble expands from its initial size at $t=0.0$ μs, indicated by the thick solid line at the center of the plot, and expands more or less spherically to a maximum radius, $R_{max}=25.9$ μm, at $t=0.881$ μs. The bubble then starts to collapse aspherically with the development of an upward-directed jet which eventually impacts on the opposite bubble surface at $t=1.229$ μs. Mainly shapes in the collapse phase are shown, the corresponding time (μs) is also indicated for each curve. The bubble moves upwards and a jet develops in the same direction as the propagation direction of the pulse. The jet attains a velocity of about 1.4 km/s upon impact (see also Fig. 5).

III. RESULTS

A. Interaction of a microbubble with pulse 1

The interaction of a single microbubble with pulse 1 is presented here. An initially stationary microbubble with initial radius 1 μm is hit by this pulse in Fig. 3. As the tensile part of the pulse moves over the bubble, the bubble expands and finally grows to approximately 26 times its original size to a maximum radius, $R_{max}=25.9$ μm, as shown by the outer most profile in Fig. 3. The bubble is still more or less spherically symmetric at this stage. Subsequently, the bubble enters its collapse phase and the bubble shapes are indicated for various times in Fig. 3. Toward the end of the collapse phase, the bottom part of the bubble deforms and a jet in the travel direction of the sound wave is formed (the sound waves travel from bottom to top).

The size distribution of microbubbles is nonuniform, and hence it is interesting to see the effect of the initial size of the bubble on the results. Therefore, bubbles with initial radii between 1 and 10 μm have been studied. The bubbles all collapse with similar profiles to the one shown in Fig. 3. They all expand to a maximum size, R_{max}, between 25 and 30 μm before collapsing with a jet in the travel direction of the pulsed ultrasonic wave. The collapse time of the microbubbles, t_c, is defined as the time difference between the instant the front of a pulse first hits the bubble and the moment the jet impacts the other side of the bubble. The collapse time is found to be between 1.2 and 1.4 μs for the range of microbubbles considered (Table I). As shown in Fig. 4, where the equivalent radius $R_{eq}=(3V/4\pi)^{1/3}$ is depicted, all the bubbles with initial radius smaller than 10 μm collapse within the first cycle of pulse 1. This observation is significant as it implies that a single cycle of ultrasound should be sufficient, if the strong jet from the collapsing

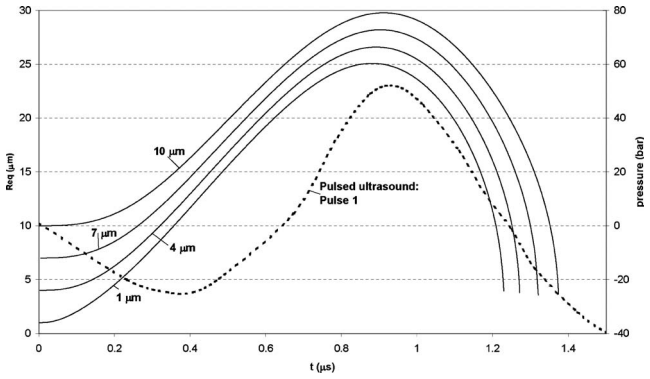


FIG. 4. Variation of equivalent bubble radius R_{eq} with time for microbubbles with radii ranging from 1 to 10 μm interacting with pulse 1. Also indicated is the pressure variation with time for pulse 1 (1000 W/cm^2, with axis on the right-hand side) at the location $z=0$. The bubbles obtain maximum radii between 25 and 30 μm and collapse between 1.23 and 1.37 μs. The collapse occurs within the first cycle of the pulsed ultrasound wave.

bubble is the main mechanism for tissue erosion or cell membrane rupture. The removal of subsequent redundant cycles might serve to reduce collateral damage to the neighboring healthy cells.

There is a slight increase in both R_{max} and t_c for increasing bubble size in Fig. 4. A larger bubble takes slightly longer to expand to its maximum size and subsequently needs more time to fully collapse. In fact all of the bubbles are still expanding when the compressive component of the sound wave has already set in (around 0.65 μs). The jet velocity, v_{jet}, attains very high values around 1.3–1.4 km/s upon impact for all cases (see also Fig. 5). The jet speed is very high when compared to those typically occurring in underwater explosions, where they attain values of only about 100 m/s as they collapse near a rigid boundary of a ship (Cole, 1948; Lauterborn and Bolle, 1975; Shima *et al.*, 1981). The speed of the opposite bubble wall (thus opposite to the jet) also attains a very high speed of 450 m/s just before the jet impacts.

B. Increasing the intensity of the pulsed ultrasound

An increase in the intensity of the sound wave yields an increase in the maximum positive and negative pressures of the pulsed ultrasound (see Fig. 2). As the intensity of the sound wave increases from 1000 to 9000 W/cm^2, the negative pressure of the first cycle of the pulse increases in magnitude from −25 bar (pulse 1) to −62 bar (pulse 4), as shown in Table I. A greater negative tensile component will result in a larger maximum bubble radius R_{max}. Thus, for a larger pulse intensity, a larger R_{max} is obtained. Also, a larger initial bubble size, R_0, will result in a slightly larger R_{max}. For example, if pulse 4 hits a 1 μm bubble, R_{max} becomes 40.5 μm: if it hits a 10 μm bubble, R_{max} increases to 44.7 μm.

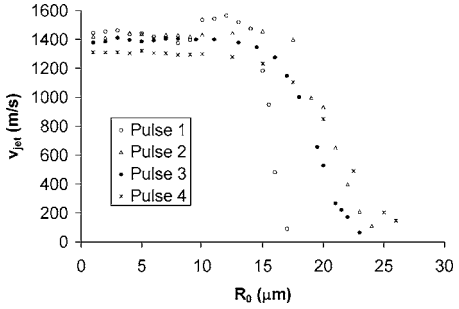


FIG. 5. Jet velocity, v_{jet}, as a function of the initial radius R_0 for the four pressure pulses, pulses 1–4. The jet velocity remains approximately constant at a value around 1.4 km/s for a large range of values of R_0 ($R_0 < 10$ μm) and does not depend significantly on pulse strength. All bubbles collapse during the first cycle of the ultrasound pulse.

For pulses 1–4 and initial bubble radii ranging from $R_0=1$ to 10 μm, the collapse times remain approximately constant, (Table I) ranging from 1.20 to 1.37 μs and are slightly larger for bubbles with larger initial size. The bubble enters its collapse phase shortly after the first compressive peak of the pulse has passed the bubble (see also Fig. 4). A very high jet speed of about 1.4 km/s (see Fig. 5) is observed for microbubbles with initial sizes smaller than 10 μm during the last phase of the collapsing bubble. The velocity of the jet only reaches a very high value just before jet impact. For bubbles with initial radii ranging from 10 to 25 μm, the jet speed gradually reduces to almost 0 km/s (Fig. 5). Even larger bubbles will reexpand again and collapse during the second positive cycle of the wave (see also Sec. III C). All bubbles collapsing during the first cycle of the wave develop a clear jet in the propagation direction of the sound wave. The intensity of the ultrasonic waves and the initial bubble radius do not seem to have a significant effect on the maximum speed obtained (at least in the range of the parameters investigated, i.e., bubbles with initial radii between 1 and 10 μm and pulses 1 to 4). For comparison, the maximum velocity in the expansion phase of the bubble is approximately 40 m/s for pulse 1 and is approximately 60 m/s for pulse 4 (for both a 1 and a 10 μm bubble radius).

Although the jet speed does not vary significantly for typical microbubbles with initial size smaller than 10 μm, the increase in intensity does broaden the jet tip. This is clearly depicted in Fig. 6, where the size of the jet just before impact for the 1 μm bubble radius has doubled from about 6 μm for pulse 1, to about 13 μm for pulse 4. An important conclusion is thus that not only the jet speed, but also the jet size, should be considered when studying the effects of a collapsing bubble. One means to do this is by investigating the Kelvin impulse. For a bubble with $R_0=1$ μm, the Kelvin impulse at the moment of jet impact increases with increasing sound wave intensity. $K=0.13$ μg m/s for pulse 1 (Fig. 7). For pulse 4, K reaches a value almost 10 times larger, i.e., $K=1.22$ μg m/s. The Kelvin impulse for a 10 μm bubble decreases slightly to $K=0.10$ μg m/s and $K=1.03$ μg m/s for pulses 1 and 4, respectively. For even larger values of R_0, the Kelvin impulse gradually tends to zero and can even become negative. It is not clear at this time how the value of the Kelvin impulse can be related directly to tissue damage. But it seems logical that a higher value of the Kelvin impulse implies a broader jet that is likely to contribute to greater damage. In Figs. 5 and 7, only the bubbles collapsing just

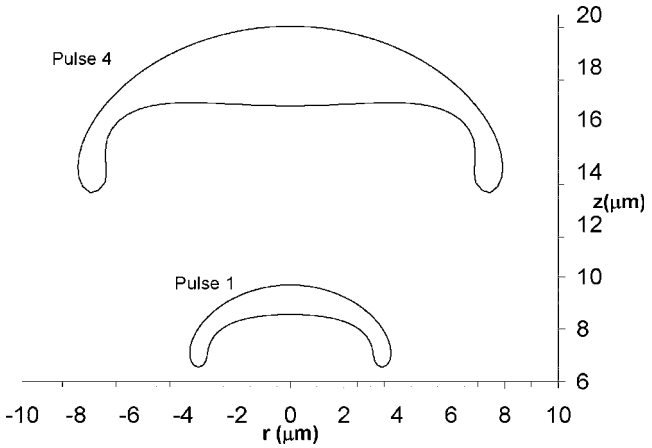


FIG. 6. Bubble shape just before jet impact for a bubble with initial radius of 1 μm interacting with pulses 1 and 4. The jet tip width is much wider for pulse 4 (13 μm) and is about twice the value of that for pulse 1 (6 μm). The vertical displacement of the bubble is also considerably larger for pulse 4 (17 μm) than that of pulse 1 (8 μm). The jet velocities are comparable, around 1.3 km/s for both pulses. The initial location of the center of the bubble for both simulations is $(r,z)=(0,0)$.

after the first compressive cycle of the wave are considered. Bubbles collapsing in the second cycle are considered in Sec. III C.

A bubble with $R_0=1$ μm is displaced 8.1 μm at the moment of jet impact in the direction of propagation of the sound wave when it interacts with pulse 1 (Table I). This is roughly 8 times R_0 (see also Figs. 3 and 6). This distance almost doubles to 15.2 μm, when this same bubble interacts with pulse 4. This trend seems to be valid for all other bubbles investigated. Larger bubbles move slightly further upwards than smaller bubbles for the same pulse intensity. This can probably be explained by the slightly larger collapse times of larger bubbles (see Fig. 4), thus the bubble has more time to move upwards.

C. Increasing the initial size of the microbubbles

The bubble behavior can change significantly, when the initial bubble size is increased beyond $R_0=10$ μm. If the bubbles collapse during or just after the first compressive cycle of the pulse, a clear upwards jet can be observed. Larger bubbles can "survive" the first compressive stage and

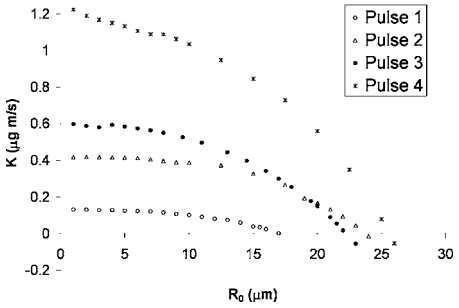


FIG. 7. Kelvin impulse, K, as a function of the initial radius R_0 for pressure pulses 1–4. The Kelvin impulse becomes lower, and even attains negative values, for larger values of R_0. The highest values of K are found for the strongest pulse, pulse 4. All bubbles in this plot collapse during the first cycle of the ultrasound pulse.

reexpand during the second tensile stage of the wave. They will only collapse during the second or third compressive stage of the wave. This can have a profound influence on the jetting behavior. Two typical examples will be shown here. Similar trends were observed for other bubble sizes.

The first example shows a bubble with $R_0=17.5$ μm, which collapses during the second compressive part of the sound wave (Fig. 8). From 0.94 to 1.68 μs, the bubble is in its first contraction phase [Fig. 8(a)]. However, due to the resistance of the internal pressure of this particular bubble, the contraction stops and the bubble rebounds [Fig. 8(b)]. The bubble has now assumed a slightly nonspherical shape, with the top and notably the bottom being flattened (this can be considered as the onset of the formation of a jet). During the second expansion, this deviation from the spherical shape continues, as the top and bottom part lag behind in their expansion. The bubble now assumes a peanut shell shape. During the second contraction [Fig. 8(c)], a ring-jet develops around the middle region of the bubble, and impacts at approximately $z=7$ μm. The jet will split the original bubble in two smaller bubbles. The velocity of this jet is again very high at 0.93 km/s, but is lower than the jet velocities observed earlier. Thus in this case, what could be considered as a shape instability, prevents the appearance of a clear upwards directed jet as was observed in the previous sections. A second example investigates the behavior of a bubble with $R_0=45$ μm, interacting with pulse 4. In Fig. 9, the positions of the bubble surface on the z-axis are indicated as a function of time. Both the upper and lower bubble surfaces, z_{up} and z_{low}, are indicated. They are not necessarily the uppermost and lowermost points on the bubble surface, respectively. The bubble exhibits very complicated behavior and undergoes three oscillations. Several shape "instabilities" develop until the jet impact occurs at $t=3.7$ μs. The shape of the bubble just before collapse is even more complex than in the previous example and is depicted in Fig. 10. More simulations were performed with other initial radii R_0 and pulse strengths, and the results are similar (not shown). In general, the more oscillations the bubble undergoes before collapsing, the more shape instabilities are developed, and a clear "clean" jet, e.g., that of a bubble collapsing in a single cycle of the wave as discussed in Section B, is rarely observed.

Even though bubbles larger than $R_0=10$ μm are unlikely to be found in microbubble applications, similar behavior is expected to occur for smaller bubbles interacting with different pressure pulses. The results presented here are thus thought to be quite general. That is, a clear jet develops if the bubble collapses within the first compressive part of the sound wave. However, multiple jets and a complicated collapse pattern can develop when the bubble collapses in the second or later compressive part. If the bubbles are initially very large (>100 μm), they are hardly perturbed by pulses 1–4.

IV. DISCUSSION

Microbubble contrast agents are becoming more readily available and research on their use in therapeutic procedures is currently being explored and developed, for example in

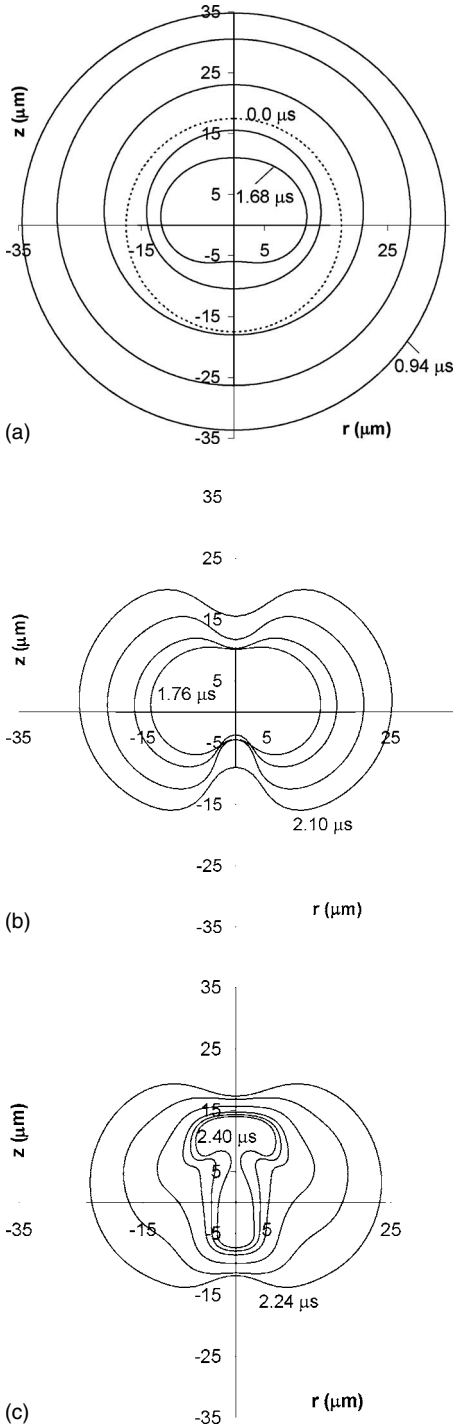


FIG. 8. Bubble shapes for a bubble with initial radius $R_0=17.5$ μm interacting with pulse 1. (a) First contraction stage starting from 0.94 (most outer curve) to 1.68 μs (most inner curve). The initial bubble shape is also indicated (dotted line). (b) Second expansion phase from 1.76 (most inner curve) to 2.10 μs (most outer curve). (c) Second contraction stage from 2.24 (most outer curve) to 2.40 μs (most inner curve). A ring jet is formed at the sides of the bubble, which impacts at 2.40 μs and causes the bubble to break up in two parts.

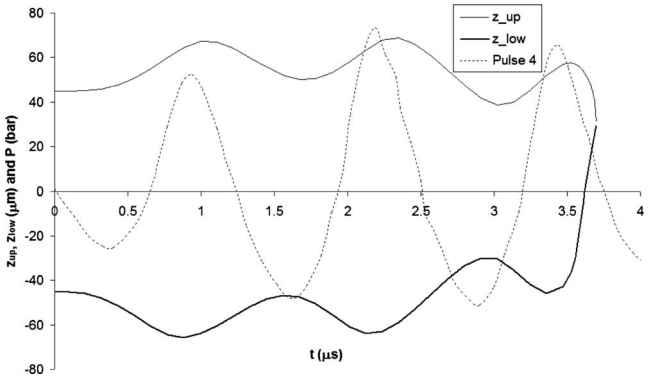


FIG. 9. Oscillation of a bubble with initial radius $R_0=45$ μm interacting with pulse 4. The positions of the bubble surface on the z-axis, z_{up} and z_{low} (upper and lower z value of the bubble on the axis of symmetry) are indicated as a function of time. The bubble undergoes three oscillations before it collapses with an upward directed jet. pulse 4 is also indicated as a function of time. At $t=3.7$ μs jet impact occurs: the upper and lower z-curves touch at this instant. The shape of the bubble just before collapse can be found in Fig. 10.

targeted drug delivery and cancer treatment. Strong ultrasound waves are utilized for cell sonoporation, where the cell membrane is temporarily perforated such that macromolecular drugs or genes can enter the cell more readily. Remote destruction of diseased tissue by ultrasound near the collapsing microbubbles after they are attached to this tissue is another potential application. These treatments with high intensity ultrasound and microbubbles have the advantages of being noninvasive, localized and potentially economical.

To understand how this concept can be utilized, the current numerical study focuses mainly on the interaction of microbubbles of typical initial radii ranging from 1 to 10 μm, with pulsed ultrasound waves as used by Xu *et al.* (2005). It is found that all the bubbles within this range collapse within the first cycle of the pulsed ultrasound (which consists of three cycles). They obtain maximum radii between 25 and 30 μm before collapsing. The bubbles expand to larger sizes before collapsing, when the intensity of the sound waves is increased from pulse 1 with 1000 W/cm^2 to pulse 4 with 9000 W/cm^2. Also, the maximum jet velocity obtained is not significantly affected by the increase in ultrasonic intensity. Nevertheless, the enhancement of tissue ero-

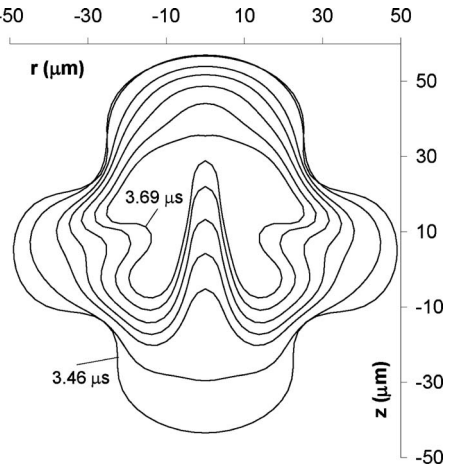


FIG. 10. Collapse of a bubble with initial radius $R_0=45$ μm interacting with pulse 4. The shape of the bubble during the phase leading to the final collapse is shown here from 3.46 (most outer curve) to 3.69 μs (most inner curve). The bubble assumes very complicated shapes and finally develops an upwards jet, whereas a ring jet is also being created. The final upwards jet has a velocity of 273 m/s prior to impact.

sion by the increased intensity ultrasound, as reported in Xu *et al.* (2005), could possibly be explained by the broadening of the jet radius from pulses 1 to 4, observed in Sec. III B.

There are a number of parameters involved in defining a pulsed ultrasound wave, the most important ones are the peak positive and negative amplitude (in pressure), the frequency, and the pulse duration. Therefore the set of possible pulsed ultrasound wave pressure profiles is very large. As the focus here is on medical applications, the pulsed ultrasound wave forms from Xu *et al.* (2005) which have direct clinical relevance (for tissue perforation in the treatment of hypoplastic left heart syndrome) are used. Moreover, these waves are found to be the optimal wave forms (Xu *et al.*, 2004) for their purpose in terms of erosion rate, i.e., thickness of tissue perforated over time needed to perforate it and energy efficiency, i.e., thickness of tissue eroded per energy input unit. These sound waves, as shown through the simulations, are capable of collapsing a microbubble and producing high speed jets in the direction of sound propagation. Other sound wave profiles may be the subject of future studies, but the trends observed in this work are expected to be quite general.

It is possible that when microbubble contrast agents break into smaller bubbles after their collapse during the first peak of the pressure pulse, the subsequent cycles of the ultrasound wave will cause these very small bubbles to expand to considerable sizes again, and bring about further unwanted collateral damage to the surrounding tissues. It is believed that one cycle of ultrasound consisting of a single negative part followed by a single compressive part would be the optimized wave form; a single clear clean jet in the direction of propagation of the wave will then be formed.

Larger bubbles (>10 μm) oscillate a few times, surviving several peaks of the pressure pulse, before collapsing. However, a clear well directed jet is seldom obtained. Instead, a very complicated flow pattern can be observed with several jets developing in the bubble. If the bubbles are large, for instance a few hundred micrometer, they are hardly perturbed even by the strongest sound wave, e.g., pulse 4.

Other microbubbles present in the vicinity might influence the behavior of the microbubble under consideration. In this work all bubbles are assumed to be at least 10 times the maximum bubble radius away from each other, and therefore mutual interactions are not taken into account. Bubbles placed closer to each other than the stipulated distance fall outside the scope of this article.

Last, it is mentioned that although microbubbles are often applied near biological tissue, the nearby biomaterials are not modeled in this study. The consistency of this material is often very close to that of water. In this case then, the tissue might have little influence in the subsequent bubble dynamics. This was confirmed by the experiments of Bourne and Field (1999), where a shock wave interacted with a jelly-like material, in which according to them: "The shock wave easily overcame the weak polymeric bonding in the gel, which flows in a manner indistinguishable from water." However, for low intensity sound waves with maximum amplitudes lower than 1 bar, the behavior of bubble–biomaterial interaction can be quite complicated (Fong *et al.*, 2006). A similar behavior was observed for an oscillating bubble, with no ultrasound, near an elastic (bio-) material (Klaseboer and Khoo, 2004; Ong *et al.*, 2005). The wave might bounce back and interact for a second time with the bubble, if there is strong reflection or if very hard materials are present near the bubble. This scenario is also outside the scope of the current article.

In our model of the microbubble, the thin membrane is not considered to be important. It is noted that this is unlike the case for the study of microbubble oscillations in a sinusoidal ultrasound field intended for bioimaging purposes. When the microbubbles oscillate gently, the surface tension due to their usually tens of nanometer thick membrane is important in influencing the bubble dynamics, e.g., the resonance frequency (Khismatullin, 2004). Yet in the case of strong sound waves, the extreme rapid expansion and subsequent more rapid collapse of the microbubbles, render the membrane effect to be of secondary importance. In fact, as shown in various experiments using strong pressure waves (Wolfrum *et al.*, 2002; Prentice *et al.*, 2005), these microbubbles collapse violently with a high speed jet, similar to that of a cavitation bubble collapsing after being hit by a shock wave (Klaseboer *et al.*, 2007). Therefore, in this study the microbubble is modeled as a free ideal gas bubble with an initial internal pressure of P_{ATM}.

As mentioned previously, the effect of surface tension can easily be incorporated in the model if desired. Now a pressure difference equal to the surface tension, σ, multiplied by the local curvature κ exists across the bubble wall, which can be taken into account in an extended version of Eq. (4) as

$$\rho\frac{D\Phi}{Dt} = P(t^*) - p_0\left(\frac{V_0}{V}\right)^{\gamma} + \frac{1}{2}\rho|\mathbf{u}|^2 + \sigma\kappa. \tag{7}$$

The initial bubble pressure must now be chosen to be $p_0 = P_{\mathrm{ATM}} + 2\sigma/R_0$, in order to ensure that the bubble stays at rest before the arrival of the sound wave. The solution procedure is equivalent to the one described in Sec. II. The surface tension is chosen to be $\sigma = 73\times10^{-3}$ N/m. It is thus assumed that the membrane immediately breaks once the sound wave hits it and that a water–air interface remains. However, in the future a more thorough study on the effect of different membrane types can be conducted. The results obtained with and without surface tension show little difference. This is indicated in Fig. 11 for example, where the jet velocity for a bubble with initial radius of 1 μm hit by pulse 1 is shown with and without the effect of surface tension. This case has been chosen, as the smallest bubble with the smallest pulse strength is most likely to be influenced by the effects of surface tension. Nevertheless, the results for the jet velocity are almost indistinguishable. Similar results were obtained for the position of the bottom z-coordinate on the axis of symmetry (Fig. 12). Other parameters, such as the Kelvin impulse and the collapse time differ by a small amount, for example the collapse time is 1.217 μs for a bubble with surface tension and 1.225 μs for a bubble without surface tension. Other bubbles with different initial radii ranging from 1 to 10 μm have also been investigated, but the results for the jet velocity and Kelvin impulse resemble closely those of Figs. 5 and 7. It can therefore be concluded

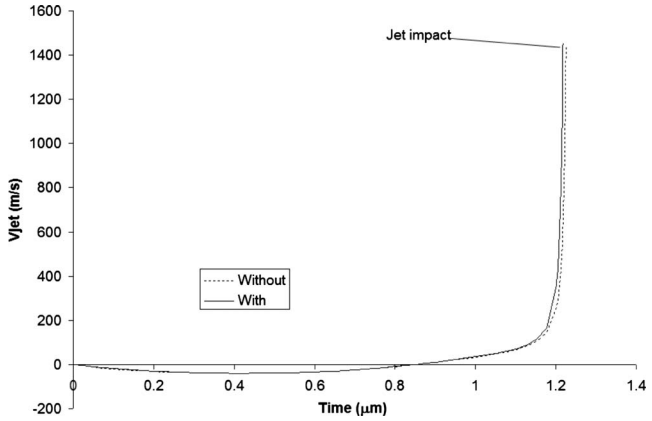


FIG. 11. Effect of surface tension: Velocity of the bottom part of the bubble (jet velocity) as a function of time for a bubble where surface tension has been taken into account (solid line) and for a bubble with surface tension neglected (dashed line). The initial radius of the bubble is 1 μm and it is hit by pulse 1 (similar to the case of Fig. 3).

that the effect of surface tension is negligible, at least for the range of parameters investigated. These results confirm the assumption that surface tension effects are negligible in Eq. (4).

The main advantage of using microbubbles combined with sound wave treatments is that the location where the bubble will occur can be fixed or predetermined. In ordinary ultrasound treatments, the bubbles can occur at any location near the focal point.

V. CONCLUSION

The shapes, jet velocities, and other parameters of interest were investigated for a microbubble interacting with a sound wave consisting of several cycles using a full BEM simulation. The dependency with time and space of the ultrasound wave is taken into account through adjusting the reference pressure for each node on the bubble surface. The first part of the wave to hit the bubble is a tensile part, which causes the microbubble to expand to several times its original size. The bubble collapses in one of the compressive cycles of the wave, depending on the initial size of the bubble and the strength of the waves. A very well defined jet in the propagation direction of the wave is observed, when the bubble collapses in the first compressive cycle. For bubbles collapsing in the second or third cycle, a very complicated bubble shape and often several jets appear in the interior of the bubble. For small bubbles, only one compressive cycle is enough to make the bubble collapse. Bubbles with larger initial sizes can collapse in subsequent cycles. More cycles do not seem beneficial if the aim is to obtain a well defined jet. The jet speed appears to not be the only parameter of interest: the size of the jet can have a determining factor on the "destructive" power of a collapsing bubble. This was investigated using the Kelvin impulse concept. As shown, for the range of parameters investigated, the inclusion of surface tension effects does not change the results noticeably.

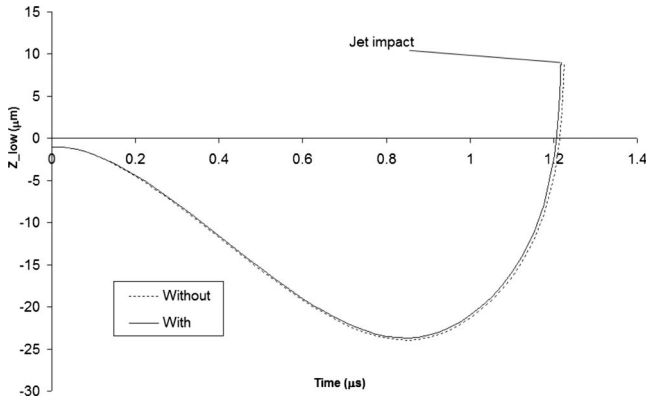


FIG. 12. The effect of surface tension on the displacement. The z-position of the bottom node on the axis of symmetry z_{low}, the position where the jet occurs, for a bubble with initial radius 1 μm and which is hit by pulse 1 as a function of time. Results with inclusion of surface tension, solid line, and without surface tension, dashed line, are shown.

Bekeredjian, R., Grayburn, P. A., and Shohet, R. V. (**2005**). "Use of ultrasound contrast agents for gene or drug delivery in cardiovascular medicine," J. Am. Coll. Cardiol. **45**, 329–335.

Bourne, N. K., and Field, J. E. (**1999**). "Shock-induced collapse and luminescence by cavities," Philos. Trans. R. Soc. London, Ser. A **357**, 295–311.

Calvisi.M. L., Lindau, O., Blake, J. R., and Szeri, A. J. (**2007**). "Shape stability and violent collapse of microbubbles in acoustic traveling waves," Phys. Fluids **19**, 047101, 1–15.

Cole, R. H. (**1948**). *Underwater Explosions* (Princeton University Press, New York).

Dayton, P. A., Pearson, D., Clark, J., Simon, S., Schumann, P. A., Zutshi, R., Matsunaga, T. O., and Ferrara, K. W. (**2004**). "Ultrasonic analysis of peptide- and antibody-targeted microbubble contrast agents for molecular imaging of $\alpha_v\beta_3$-expressing cells," Mol. Imaging **3**, 125–134.

Ding, Z., and Gracewski, S. M. (**1996**). "The behaviour of a gas cavity impacted by a weak or strong shock wave," J. Fluid Mech. **309**, 183–209.

Feinstein, S. B. (**2004**). "The powerful microbubble: from bench to bedside, from intravascular indicator to therapeutic delivery system, and beyond," Am. J. Physiol. Heart Circ. Physiol. **287**, H450–H457.

Fong, S. W., Klaseboer, E., Turangan, C. K., Khoo, B. C., and Hung, K. C. (**2006**). "Numerical analysis of a gas bubble near bio-materials in an ultrasound field," Ultrasound Med. Biol. **32**, 925–942.

Ikeda, T., Yoshizawa, S., Tosaki, M., Allen, J., Takagi, S., Ohta, N., Kitamura, T., and Matsumoto, Y. (**2006**). "Cloud cavitation control for lithotripsy using high intensity focused ultrasound," Ultrasound Med. Biol. **32**, 1383–1397.

Khismatullin, D. B. (**2004**). "Resonance frequency of microbubbles: Effect of viscosity," J. Acoust. Soc. Am. **116**, 1463–1473.

Klaseboer, E., Fong, S. W., Turangan, C. K., Khoo, B. C., Szeri, A. J., Calvisi, M. L., Sankin, G. N., and Zhong, P. (**2007**). "Interaction of lithotripter shockwaves with single inertial cavitation bubbles," J. Fluid Mech. **593**, 33–56.

Klaseboer, E., and Khoo, B. C. (**2004**). "An oscillating bubble near an elastic material," J. Appl. Phys. **96**, 5808–5818.

Klaseboer, E., Turangan, C. K., Fong, S. W., Liu, T. G., Hung, K. C., and Khoo, B. C. (**2006**). "Simulations of pressure pulse-bubble interaction using boundary element method," Comput. Methods Appl. Mech. Eng. **195**, 4287–4302.

Klibanov, A. L. (**1999**). "Targeted delivery of gas-filled microspheres, contrast agents for ultrasound imaging," Adv. Drug Delivery Rev. **37**, 139–157.

Lanza, G. M., Wallace, K. D., Scott, M. J., Cacheris, W. P., Abendschein, D. R., Christy, D. H., Sharkey, A. M., Miller, J. G., Gaffney, P. J., and Wickline, S. A. (**1996**). "A novel site-targeted ultrasonic contrast agent with broad biomedical application," Circulation **94**, 3334–3340.

Lauterborn, W., and Bolle, H. (**1975**). "Experimental investigations of cavitation bubble collapse in the neighbourhood of a solid boundary," J. Fluid Mech. **72**, 391–399.

Le Gac, S., Zwaan, E., van den Berg, A., and Ohl, C. D. (**2007**). "Sonoporation of suspension cells with a single cavitation bubble in a microfluidic confinement," Lab Chip **7**, 1666–1672.

Leighton, T. G. (**1994**). *The Acoustic Bubble* (Academic, Cambridge).
Lepper, W., Belcik, T., Wei, K., Lindner, J. R., Sklenar, J., and Kaul, S. (**2004**). "Myocardial contrast echocardiography," Circulation **109**, 3132–3135.
Li, T., Tachibana, K., Kuroki, M., and Kuroki, M. (**2003**). "Gene transfer with echo-enhanced contrast agents: comparison between Albunex, Optison and Levovist in mice-initial results," Radiology **229**, 423–428.
Lindner, J. R. (**2004**). "Microbubbles in medical imaging: current applications and future directions," Nat. Rev. Drug Discovery **3**, 527–532.
May, D. J., Allen, J. S., and Ferrara, K. W. (**2002**). "Dynamics and fragmentation of thick-shelled microbubbles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1400–1410.
Miller, D. L., Pislaru, S. V., and Greenleaf, J. F. (**2002**). "Sonoporation: mechanical DNA delivery by ultrasonic cavitation," Somatic Cell Mol. Genet. **27**, 115–134.
Ohl, C.-D., Arora, M., Ikink, R., de Jong, N., Versluis, M., Delius, M., and Lohse, D. (**2006**). "Sonoporation from jetting cavitation bubbles," Biophys. J. **91**, 4285–4295.
Ong, G. P., Khoo, B. C., Turangan, C., Klaseboer, E., and Fong, S. W. (**2005**). "Behavior of oscillating bubbles near elastic membranes: an experimental and numerical study," Mod. Phys. Lett. B **19**, 1579–1582.
Pearson, A., Blake, J. R., and Otto, S. R. (**2004**). "Jets in bubbles," J. Eng. Math. **48**, 391–412.
Pitt, W. G., Husseini, G. A., and Staples, B. J. (**2004**). "Ultrasonic drug delivery—A general review," Expert Opin. Drug Deliv. **1**, 37–56.
Prentice, P., Cuschieri, A., Dholakia, K., Prausnitz, M., and Campbell, P. (**2005**). "Membrane disruption by optically controlled microbubble cavitation," Nat. Phys. **1**, 107–110.
Shima, A., Takayama, K., Tomita, Y., and Miura, N. (**1981**). "An experimental study on effects of a solid wall on the motion of bubbles and shock waves in bubble collapse," Acustica **48**, 293–301.
Tachibana, K., and Tachibana, S. (**1995**). "Albumin microbubble echocontrast material as an enhancer for ultrasound accelerated thrombolysis," Circulation **92**, 1148–1150.
Taniyama, Y., Tachibana, K., Hiraoka, K., Namba, T., Yamasaki, K., Hashiya, N., Aoki, M., Ogihara, T., Yasufumi, K., and Morishita, R. (**2002**). "Local delivery of plasmid DNA into rat carotid artery using ultrasound," Circulation **105**, 1233–1239.
Tsutsui, J. M., Xie, F., Johanning, J., Lof, J., Cory, B., He, A., Thomas, L., Matsunaga, T., Unger, E., and Porter, T. R. (**2006**). "Treatment of deeply located acute intravascular thrombi with therapeutic ultrasound guided by diagnostic ultrasound and intravenous microbubbles," J. Ultrasound Med. **25**, 1161–1168.
Turangan, C. K., Jamaluddin, A. R., Ball, G. J., and Leighton, T. G. (**2008**). "Free-Lagrange simulations of the expansion and jetting collapse of air bubbles in water," J. Fluid Mech. **598**, 1–25.
Unger, E. C., McCreery, T. P., Sweitzer, R. H., Shen, D., and Wu, G. (**1998**). "In vitro studies of a new thrombus-specific ultrasound contrast agent," Am. J. Cardiol. **81**, 58G–61G.
Wang, Q. X., Yeo, K. S., Khoo, B. C., and Lam, K. Y. (**1996**). "Strong interaction between a buoyancy bubble and a free surface," Theor. Comput. Fluid Dyn. **8**, 73–88.
Wolfrum, B., Mettin, R., Kurz, T., and Lauterborn, W. (**2002**). "Observations of pressure-wave-excited contrast agent bubbles in the vicinity of cells," Appl. Phys. Lett. **81**, 5060–5062.
Xu, Z., Fowlkes, J. B., Rothman, E. D., Levin, A. M., and Cain, C. A. (**2005**). "Controlled ultrasound tissue erosion: the role of dynamic interaction between insonation and microbubble activity," J. Acoust. Soc. Am. **117**, 424-435.
Xu, Z., Ludomirsky, A., Eun, L. Y., Hall, T. L., Tran, B. C., Fowlkes, J. B. and Cain, C. A. (**2004**). "Controlled ultrasound tissue erosion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 426–435.

In vivo ultrasonic attenuation slope estimates for detecting cervical ripening in rats: Preliminary results

Timothy A. Bigelow[a)]
Department of Electrical Engineering, University of North Dakota, Box 7165, Grand Forks, North Dakota 58202

Barbara L. McFarlin
Department of Maternal Child Nursing, University of Illinois at Chicago, M/C 802, Room 858, 845 S. Damen Avenue, Chicago, Illinois 60612

William D. O'Brien, Jr. and Michael L. Oelze
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, Illinois 61801

(Received 11 September 2007; revised 10 December 2007; accepted 11 December 2007)

To effectively postpone preterm birth, cervical ripening needs to be detected and delayed. As the cervix ripens, the spacing between the collagen fibers increases and fills with water, hyaluronan, decorin, and enzymes suggesting that the ultrasonic attenuation of the cervix should decrease. The decrease in ultrasonic attenuation may be detectable, leading to an effective means of detecting cervical ripening. Herein, the traditional attenuation slope-estimation algorithm based on measuring the downshift in center frequency of the ultrasonic backscattered signal with propagation depth was modified and applied to the cervix of rats. The modified algorithm was verified using computer simulations and an *ex vivo* tissue sample before being evaluated in *in vivo* animal studies. Spherically-focused f/3 transducers with 33-MHz center frequencies and with 9-mm focal lengths were used in both the simulations and experiments. The accuracy was better than 15% in the simulations, and the attenuation slope of the cervix in the *ex vivo* experiment was 2.6 ± 0.6 dB/cm-MHz, which is comparable to 2.5 ± 0.4 dB/cm-MHz measured using a through-transmission insertion loss technique. For the *in vivo* experiments, a statistically significant effect of ultrasonic attenuation with gestational age was not observed. The large variances in the *in vivo* results were most likely due to the natural variation in attenuation for biological tissue between animals.
 [DOI: 10.1121/1.2832317]

PACS number(s): 43.80.Vj, 43.80.Ev, 43.80.Qf [CCC] Pages: 1794–1800

I. INTRODUCTION

Premature delivery is the leading cause of infant mortality in the United States (Callaghan *et al.*, 2006). Among the infant survivors, 23% suffer a major neurological disability (Wilson-Costello *et al.*, 2007), and it is estimated that just the initial in-hospital care after birth costs $10 billion annually (St. John *et al.*, 2000). In the United States, there has been a 20% increase in the preterm birth rate from 1990 to 2005, to 12.7% of all births (Hamilton *et al.*, 2006). Therefore, there is a medically significant need to develop new methods to prevent premature delivery. Currently, clinicians can only attempt to delay delivery once extensive uterine contractions have been initiated, but these contractions represent the final stage of the process. The cervix prepares for the delivery weeks to months before labor by a process termed preterm cervical ripening without any signs or symptoms currently detectable noninvasively. If preterm cervical ripening could be reliably detected noninvasively before the onset of contractions, then new treatments could potentially be developed to prevent premature delivery.

The role of the cervix during pregnancy before the fetus is mature is to remain strong and resist loading forces that cause it to change and allow delivery of the fetus. The strength of the cervix is due to its anatomic properties (cervical length, thickness) and its tissue properties (collagen and extracellular matrix remodeling) (House and Socrate, 2006). The loading forces are: (1) passive, the enlarged size and weight of the uterus; and (2) active, the forces of the uterine contractions (House and Socrate, 2006). Parturition involves orderly and biologically timed events of the uterus and cervix (Leppert, 1995; Mahendroo *et al.*, 1999; Straach *et al.*, 2005). As the cervix prepares for labor and birth, it first softens early in pregnancy while preserving its ability to maintain the fetus within the uterus and then undergoes a process of remodeling where the collagen concentration decreases and collagen fibrils become disorganized (cervical ripening). Once the cervix softens and then ripens, the forces allow the anatomic changes of cervical thinning, shortening, and finally dilation.

Due to the change in water concentration and tissue morphology (Clark *et al.*, 2006; Feltovich *et al.*, 2005; Leppert *et al.*, 2000), it has been hypothesized that the ultrasonic attenuation of the cervix should drastically decrease during

[a)]Electronic mail: timothybigelow@mail.und.nodak.edu

cervical ripening. This hypothesis recently received some additional support when the through-transmission insertion loss of *ex vivo* cervical samples from rats at varying gestational ages, 15 to 21 days, was observed to decrease with increasing gestational age (McFarlin *et al.*, 2006). Before the hypothesis can be validated, however, the observations need to be reproduced *in vivo*. Therefore, an algorithm for determining the attenuation from backscattered ultrasonic echoes needs to be implemented and validated for the cervix.

Over the years many algorithms have been proposed for making *in vivo* estimates of the attenuation from backscattered ultrasonic signals. The most common algorithms are based on assuming that the backscattered signals have a Gaussian power spectrum and that the ultrasonic attenuation (loss per distance; e.g., dB/cm) linearly increases with frequency, thus allowing the attenuation to be determined from the downshift in center frequency of the backscattered spectrum (Baldeweck *et al.*, 1994, 1993, 1995; Girault *et al.*, 1998; Hyungsuk and Varghese, 2007; Narayana and Ophir, 1983a, b; Oosterveld *et al.*, 1991). In this study, we followed the traditional approach of using the downshift in center frequency to determine the ultrasonic attenuation slope (attenuation per frequency; e.g., dB/cm-MHz). We implemented the algorithm with the intention of measuring the ultrasonic attenuation slopes of rat cervices *in vivo* using a 33-MHz spherically focused f/3 transducer with a focal length of 9 mm. The algorithm was implemented in the frequency domain, as opposed to using an autoregressive approach, which has been utilized by other investigators (Baldeweck *et al.*, 1994, 1993, 1995; Girault *et al.*, 1998), so that we could compensate for focusing. The accuracy and precision of our implementation of the algorithm was then validated by computer simulations and an experiment performed on an *ex vivo* tissue samples from a rat cervix. The feasibility of extending the work *in vivo* was also demonstrated by preliminary *in vivo* experiments that compared the attenuation slope of pregnant and nonpregnant rats.

II. SUMMARY OF ALGORITHM

When using the downshift in center frequency of the backscattered power spectrum with depth to estimate the *in vivo* attenuation slope of backscattered ultrasonic signals using focused sources, it is necessary to compensate for the effects of focusing (Hyungsuk and Varghese, 2007; Oosterveld *et al.*, 1991). Otherwise, windowed regions in front of the focus will result in an underestimate of the attenuation slope and windowed regions beyond the focus will overestimate the attenuation slope. In our study, we compensated for diffraction or focusing by approximating the field along the beam axis in the focal region by a Gaussian function. Earlier studies have demonstrated that this approximation is sufficiently valid when estimating spectral properties for the purpose of tissue characterization (Bigelow and O'Brien 2004a, b). Also, the width of the window used to gate the echoes was assumed to be small compared to the depth of focus for the transducer so that variations of the field within each gated region could be ignored.

Based on these approximations, the expected backscattered amplitude spectrum from the tissue, $E[|V_{\text{refl}}(f)|]$, would be proportional to

$$E[|V_{\text{refl}}(f)|] \propto f^2|V_{\text{plane}}(f)|e^{-2\alpha_{\text{eff}}z_T}(F_\gamma(f,a_{\text{eff}}))^{1/2}e^{-2\alpha f z_0}e^{-2(z_0^2/w_z^2)}, \quad (1)$$

where $|V_{\text{plane}}(f)|$ is the amplitude spectrum returned from a rigid plane placed at the focal plane in a water bath. For the other variables, α_{eff} is the effective attenuation along the propagation path from the transducer to the focus, z_T is the distance from the aperture plane to the focal plane, $F_\gamma(f,a_{\text{eff}})$ is a form factor that accounts for the frequency dependence of the ultrasonic backscatter at the focus (Insana *et al.*, 1990), z_0 is the distance of the windowed region from the focus, α is the slope of ultrasonic attenuation versus frequency in the focal region that we are attempting to estimate, and w_z is the effective Gaussian depth of focus that results from approximating the field with a Gaussian function. w_z depends linearly on wavelength and is given by $w_z = 6.01\lambda(f\#)^2$ for an ideal spherically focused source, where $f\#$ is the f-number of the source (Bigelow and O'Brien, 2004b). Assuming that the backscattered spectrum can be approximated by $\exp(-(f-\tilde{f}_0)^2/2\tilde{\sigma}_\omega^2)$, an approximation that is valid for soft tissue using standard pulse-echo imaging, Eq. (1) can be written as

$$E[|V_{\text{refl}}(f)|] \propto \exp\left(-\frac{(f-\tilde{f}_0)^2}{2\tilde{\sigma}_\omega^2} - 2\alpha f z_0 - 2\left(\frac{z_0^2}{w_z^2}\right)\right), \quad (2)$$

where $\tilde{f}_0$ and $\tilde{\sigma}_\omega$ are the spectral peak frequency and Gaussian beam width of the backscattered signal from the focus (i.e., $z_0=0$).

Therefore, the impact of focusing can be removed by multiplying the received amplitude spectrum (i.e., magnitude of the Fourier transform of windowed signal) at every window location by $\exp(2z_0^2/w_z^2)$ prior to finding the center frequency of the spectrum of the signal. The ultrasonic attenuation slope of the medium can then be calculated by finding the change in spectral peak frequency, f_{peak}, with depth into the tissue using

$$\alpha = -\frac{1}{2\tilde{\sigma}_\omega^2}\frac{\partial f_{\text{peak}}}{\partial z_0}. \quad (3)$$

III. COMPUTER SIMULATIONS

A. Simulation parameters

The algorithm was first validated using computer simulations. In the simulations, 1000 backscattered echo wave forms were generated by randomly positioning scatterers in a homogeneous attenuating half-space. The scatterers had a Gaussian correlation function [i.e., form factor $F_\gamma(f,a_{\text{eff}}) = \exp(-0.827(ka_{\text{eff}})^2)$] with an a_{eff} of 6 μm and were positioned at a density of 6000/mm^3 ($\sim$5 scatterers/resolution cell). The attenuation slope and sound speed of the half-space were 2 dB/cm-MHz and 1540 m/s, respectively. The correlation length of 6 μm and attenuation slope of

2 dB/cm-MHz correspond to previous values measured for the cervix of pregnant rats (McFarlin *et al.*, 2006). The simulated f/3 source used to obtain the echoes had a center frequency of 33 MHz and a focal length of 9 mm. The reference amplitude spectrum returned from a plane placed at the focal plane for the source was given by

$$|V_{\text{plane}}(f)| \propto \exp\left(-2\left(\frac{f-33\ \text{MHz}}{19.2\ \text{MHz}}\right)^2\right), \tag{4}$$

similar to the source used to determine the ultrasonic attenuation of the rat cervix *in vivo*.

After generating the echoes, the wave forms were combined into 20 sets of 50 wave forms per set. Each wave form was then windowed into five sections with each section having a corresponding length of 0.5 mm along the beam axis. This window length is considerably smaller than the depth of focus (2.9 mm at 33 MHz) for our transducer; therefore, field variations within each window can be neglected. There was 50% overlap between the sections resulting in a total length of 1.5 mm along the beam axis. After windowing, the Fourier transform was obtained for each section, and the amplitude spectra from all 50 wave forms corresponding to the same window location were averaged to obtain an estimate of $E[|V_{\text{refl}}(f)|]$ at that tissue depth. The estimate for $E[|V_{\text{refl}}(f)|]$ was then multiplied by $\exp(2z_0^2/w_z^2)$ before being fit with a Gaussian function to find an estimate for $\tilde{\sigma}_\omega$ and f_{peak}, the new spectral peak frequency, of the spectrum at that window location, z_0. After finding an estimate for $\tilde{\sigma}_\omega$ and f_{peak} for all five locations along the echo, a line was fit to f_{peak} vs z_0 to obtain $\partial f_{\text{peak}}/\partial z_0$ from which α can be calculated using Eq. (3). The accuracy was further improved by using the mean value of $\tilde{\sigma}_\omega$ from the five window locations for $\tilde{\sigma}_\omega$ in the equation.

B. Simulation results

When validating the algorithm, two issues were of concern. First, the algorithm should correctly compensate for the effects of focusing. Second, the algorithm should be stable even if measurement noise reduces the usable bandwidth of the received echoes (i.e., the range of frequencies not corrupted by noise). Focusing compensation was validated by moving the center of the 1.5-mm region used to obtain the estimates from 1.2 mm in front of the focus to 1.2 mm past the focus. Therefore, the total depth used in the simulations was from 1.95 mm in front of the focus to 1.95 mm past the focus for a total length of 3.9 mm. Likewise, the reduction of usable bandwidth was investigated by varying the bandwidth used in the Gaussian fit to find an estimate for $\tilde{\sigma}_\omega$ and f_{peak} from 10 to 50 MHz centered at the approximate location for the spectral peak prior to correcting for focusing.

The results for the simulations are summarized in Fig. 1. The error in the attenuation slope estimate is typically smaller than 15% except for a few window locations. Therefore, reliable estimates should be obtainable even if the usable bandwidth of the signal is significantly reduced by noise. The simulations also demonstrate a trough in the error in the attenuation slope ($\sim -10\%$) at $z_0=0$. This error is of little concern because it is on the same order as other published results that estimated the attenuation slope from the downshift in center frequency (Hyungsuk and Varghese, 2007). The reason for this trough is not clear but may be related to the number of wave forms used (i.e., 50) to obtain the spectrum because other investigators obtained smaller errors when more wave forms (i.e., 256) were used (Baldeweck *et al.*, 1995).

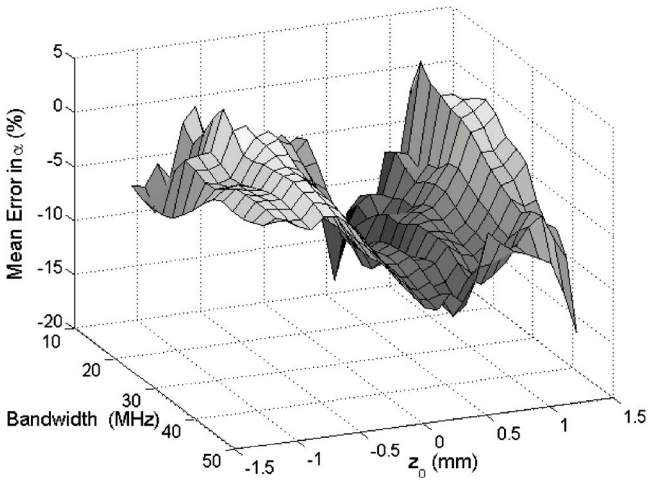


FIG. 1. (Color online) The value of the mean error in the attenuation slope for the simulations vs the usable bandwidth and the location along z_0 of the 1.5-mm section used to obtain the estimates.

Although Fig. 1 indicates that the algorithm is correctly compensating for focusing for the different values of z_0, the performance is easier to evaluate if we restrict our attention to a single frequency range. Figure 2(a) shows the error in the attenuation slope estimate in dB/cm-MHz versus location along the echo using a bandwidth of 35 MHz when performing the Gaussian fit. Figure 2(b) shows the percent error in attenuation slope both with and without focusing compensation. Prior to compensating for focusing, the attenuation slope of regions before the focus is grossly underestimated while the attenuation slope of regions beyond the focus is grossly overestimated. The attenuation slope is underestimated before the focus because $-2\alpha f z_0$ and $-2z_0^2/w_z^2$ have opposite sign (i.e., $z_0<0$) while the attenuation slope is overestimated after the focus because $-2\alpha f z_0$ and $-2z_0^2/w_z^2$ have the same sign (i.e., $z_0>0$). After compensating for focusing, the attenuation slope estimate for all of the regions is approximately the same and within $\sim 10\%$ of the true value for the attenuation.

IV. *EX VIVO* TISSUE EXPERIMENT

Once the algorithm had been validated using computer simulations, the next step was to validate the algorithm using an *ex vivo* tissue sample of a rat cervix. For the *ex vivo* study, we used a custom-fabricated 33-MHz lithium niobate f/3 transducer with a focal length of 9 mm (NIH Transducer Resource Center, University of Southern California, Los Angeles, CA). In this portion of the study, a nonpregnant Sprague-Dawley rat (Harlan, Indianapolis, IN) was euthanized by exposure to CO_2 for 5 min. The experimental protocol was approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Illinois at Urbana-

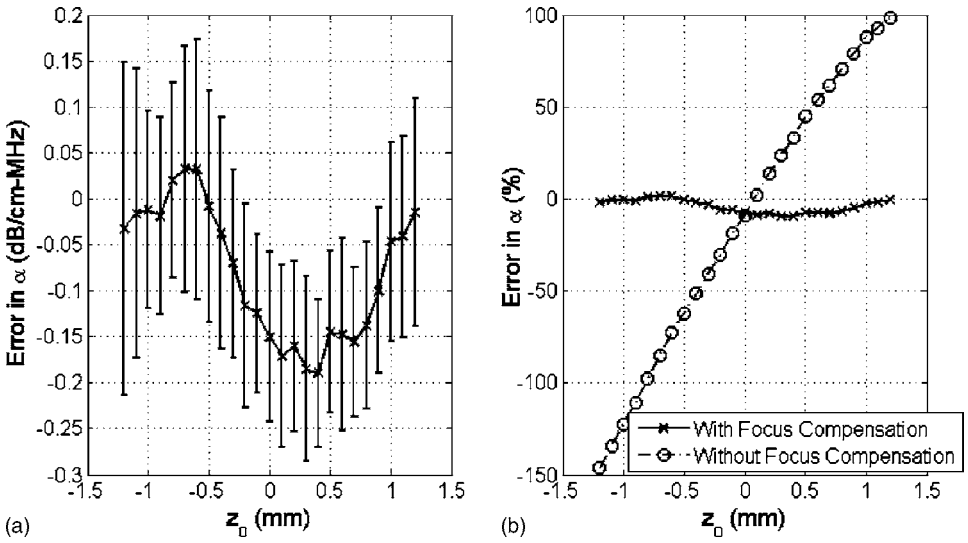


FIG. 2. (a) Difference between estimated and simulated attenuation slope value of 2 dB/cm-MHz plotted vs location along the echo in the simulations using a bandwidth of 35 MHz. The error bars represent one s.d. above and below the mean as calculated using the 20 independent sets of 50 wave forms. (b) Mean error in attenuation vs location along the echo in the simulations both with and without focusing compensation.

Champaign and satisfied all campus and NIH (National Institutes of Health) rules for the human use of laboratory animals. The rat was then weighed prior to necropsy and the cervix was dissected, trimmed, and weighed before being immediately sealed in plastic wrap (Saran Wrap™). A Saran Wrap was used to prevent degradation of the sample over time upon being immersed in degassed water. Because the tissue was surgically removed from the rat, the attenuation slope of the same rat cervix could be estimated by two separate techniques, one of which is the algorithm under evaluation.

A. Through-transmission estimate of tissue attenuation

The ultrasonic attenuation slope of the *ex vivo* cervix sample was independently estimated by a double through-transmission insertion loss technique. The backscattered ultrasonic wave forms were recorded from a Plexiglas™ block placed at the focal plane both with and without the sealed cervix sample between the transducer and the block. With the sample present, ultrasound passed through the sample twice. The frequency-dependent attenuation of the cervix could then be determined by dividing the amplitude spectra and then fitting a line versus frequency to the log of the result after compensating for the passage of the ultrasound through the Saran Wrap. Compensating for the Saran Wrap was performed by measuring the thickness of the Saran Wrap using a micrometer and applying a correction term based on the density, sound speed, and thickness of the Saran Wrap as has been done previously (Madsen *et al.*, 1999). The spectra were obtained by scanning the ultrasound beam across 3 mm of the cervix tissue sample (step size ~15 μm) and averaging every 50 lines of the received spectra from the Plexiglas block. The scanning was performed using a Panametrics 5900™ pulser/receiver (Panametrics™, Waltham, MA) and a computer-controlled micropositioning system (Daedal, Inc., Harrisburg, PA). Each set of lines had 50% overlap with adjacent line sets so a total of seven measurements of attenuation were obtained for the cervix. An example of the normalized amplitude spectra from the Plexiglas block both with and without the cervix sample is shown in Fig. 3.

Likewise, an example of the estimated ultrasonic attenuation versus frequency after comparing the spectra is shown

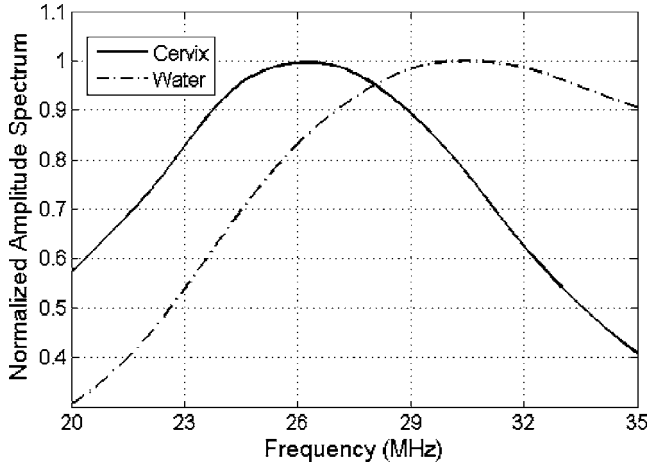


FIG. 3. An example of the normalized amplitude spectra from the Plexiglas block both with and without the cervix sample along the propagation path.

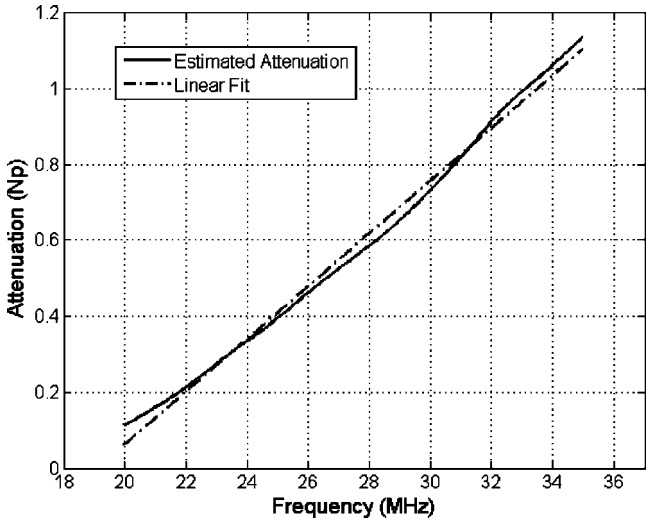


FIG. 4. An example of the estimated attenuation vs frequency after dividing the spectra shown with a linear fit vs frequency.

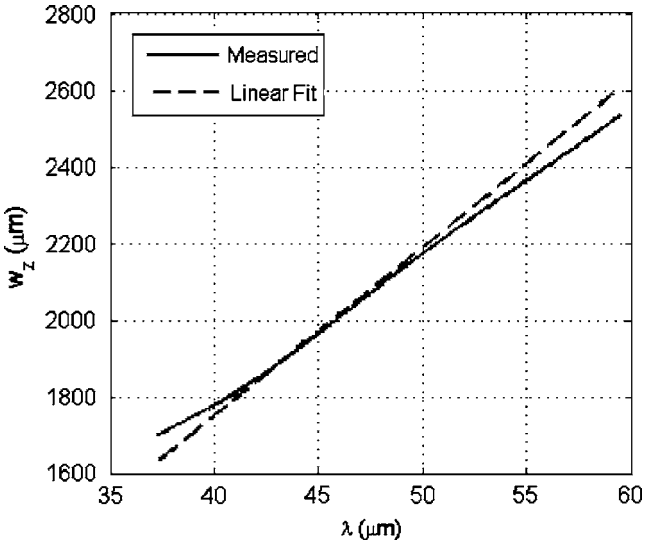


FIG. 5. Fitting a line through the measured w_z values to obtain an equation of w_z vs wavelength, λ.

in Fig. 4 along with a linear fit versus frequency. The equation for the linear fit to the attenuation values (in Np) for all seven measurements was $\alpha_{\text{true}}=(0.053\pm0.0098)f-1.25\pm0.068$ where f is in megahertz. Given that the thickness of the cervix was 1.86 ± 0.087 mm, this translates to an attenuation slope of 2.5 ± 0.4 dB/cm-MHz.

B. Determination of Gaussian depth of focus

Prior to implementing our algorithm using the *ex vivo* tissue sample, the Gaussian depth of focus, w_z, for the transducer needed to be determined as a function of wavelength. w_z was obtained for the transducer by obtaining echoes from a rigid plane located at different locations along the beam axis as controlled by a computer-controlled micropositioning system (Daedal, Inc., Harrisburg, PA) and fitting a Gaussian function versus depth to the reflected signal at each frequency. This gives a measurement of w_z at that frequency (Bigelow and O'Brien, 2004a, b). The results were generalized to other frequencies by fitting a line through the measured w_z values versus wavelength, λ, as shown in Fig. 5 for frequencies from 25 to 40 MHz. The fit was performed in terms of wavelength rather than frequency because we were anticipating a linear dependence on wavelength (Bigelow and O'Brien, 2004b). After performing the linear fit, we found that $w_z=43.8\lambda$ for the transducer used in the *ex vivo* experiment.

C. Pulse-echo estimate of attenuation using proposed algorithm

After measuring the attenuation of the cervix tissue sample by comparing the spectrum of the received signal from the Plexiglas block both with and without the tissue along the propagation path, the attenuation slope of the cervix tissue sample was estimated using our developed algorithm. The cervix tissue sample was positioned slightly in front of the focus (9 mm for this transducer) in a water bath and ultrasonic echoes from the cervix were obtained along a 3-mm length of the cervix using a step size of 15 μm. The scanning was performed using a Panametrics 5900 pulser/receiver and a computer-controlled micropositioning system (Daedal, Inc.). The echoes were windowed into five sections with each section having a depth of 0.48 mm. The sections overlapped by 50%, similar to the simulations, and spanned depths between 7.8 and 9.25 mm so that only echoes from within the cervix would be used in the estimate for attenuation. The echoes were then grouped into sets of 50 wave forms with 50% overlap between the sets and an estimate for $E[|V_{\text{refl}}(f)|]$ at each depth was obtained by averaging the amplitude spectrum of the 50 wave forms. The 50 wave forms spanned a range of approximately five beam widths, which was suggested to be optimal in terms of resolution versus estimate variance for parameters estimated from the frequency dependence of ultrasonic backscatter (Oelze and O'Brien, 2004). Then, prior to finding the frequency corresponding to the spectral peak at each depth, $E[|V_{\text{refl}}(f)|]$ was multiplied by $\exp(2z_0^2/w_z^2)$ using $w_z=43.8\lambda$ as was measured for our transducer while assuming a sound speed of water for the tissue.

The value for $\tilde{\sigma}_\omega$ and f_{peak} at each depth was then found by fitting the spectra to a Gaussian function using a frequency bandwidth of 40 MHz. Then, the value of $\tilde{\sigma}_\omega$ to be used in the calculation of attenuation slope was obtained by averaging the values of $\tilde{\sigma}_\omega$ obtained from the fit at each depth. Based on these calculations, the measured attenuation slope for the *ex vivo* tissue sample using our developed algorithm was 2.6 ± 0.6 dB/cm-MHz, which is in good agreement with the attenuation slope value of 2.5 ± 0.4 dB/cm-MHz using the basic insertion loss technique.

V. PRELIMINARY *IN VIVO* RESULTS

After completing the *ex vivo* experiment, the algorithm was used to obtain attenuation slope estimates for rat cervices *in vivo*. The experimental protocol was approved by the IACUC at the University of Illinois at Urbana-Champaign and satisfied all campus and NIH rules for the humane use of laboratory animals. Eight nonpregnant and 42 timed-pregnant (days 15, 17, 19, 20, and 21 of pregnancy) Sprague Dawley rats (Harlan) were randomly assigned to the *in vivo* experiment. After an adequate level of anesthesia was obtained [ketamine hydrochloride (87 mg/kg) and xylazine (13 mg/kg) administered intramuscularly], the rat was placed in dorsal recumbency and the ultrasonic transducer was inserted into the vaginal cavity of the rat with ultrasound coupling gel (Aquasonic 100, Parker Labs, Fairfield, NJ). The position of the transducer in the vagina was controlled by the computer-controlled micropositioning system (Daedal, Inc.). The backscattered signal from the cervix was then collected by a Panametrics 5900 pulser/receiver and monitored on an oscilloscope (Lecroy 9354 TM; Chestnut Ridge, NY). The position of the transducer was adjusted until the backscattered signal from the cervix was maximized.

Two different custom-fabricated 33-MHz lithium niobate ultrasonic transducers were used in the *in vivo* study because the original transducer used in the *ex vivo* experiment stopped working during the middle of the *in vivo* ex-

periments. The initial transducer had a focal length of 9.01 mm and a w_z value of $w_z=43.8\lambda$, while the second transducer had a focal length of 9.16 mm and a w_z value of $w_z=34.7\lambda+640$ μm. The value of w_z for the second transducer was found in the same manner as discussed in Sec. IV B. A total of 50 rats were used in the experiment including 8 nonpregnant rats, 8 rats at a gestational age of 15 days, 9 rats at a gestational age of 17 days, 8 rats at a gestational age of 19 days, 9 rats at a gestational age of 20 days, and 8 rats at a gestational age of 21 days.

After positioning, the transducer was scanned along the cervix using the micropositioning system (Daedal, Inc., Harrisburg, PA), and ultrasonic echoes were collected every 15 μm except for one scan of a nonpregnant rat where the echoes were collected every 10 μm. The usable scan width varied from 0.8 to 7 mm with a mean of 2.5 mm depending on the orientation of the cervix in the focal region and the animal's breathing. For 8 of the rats (2 at day 15, 1 at day 17, 3 at day 19, and 2 at day 21), no usable data were obtainable because the cervix was too far away from the focal region to correctly compensate for diffraction due to the bandwidth limitations of the transducer. Our method of compensating for diffraction by multiplying the spectrum by $\exp(2z_0^2/w_z^2)$ will amplify values of the spectrum at higher or lower frequencies depending on the value of z_0, the distance of the echo from the focus. Hence, for large values of z_0, the compensation term will only be amplifying noise and a reliable estimate for attenuation slope cannot be obtained.

After obtaining the usable echoes, the echoes were windowed into five sections with varying depths (0.33–0.9 mm with a mean of 0.56 mm) depending on the thickness of the cervix. The sections also overlapped in depth by 50%, similar to the simulations and *ex vivo* experiment. The echoes were then grouped into sets of 50 wave forms laterally with 50% overlap between adjacent sets and an estimate for $E[|V_{\text{refl}}(f)|]$ at each depth was obtained by averaging the amplitude spectrum of the 50 wave forms. If a cervix had fewer than 50 usable wave forms, all of the wave forms were used to obtain an estimate. The number of attenuation slope estimates (i.e., sets of 50 wave forms) obtained for a particular cervix varied from 1 to 13 depending on the usable scan width for each cervix.

The attenuation slope estimates for all of the gestational ages are summarized in Fig. 6. The bars correspond to the mean attenuation slope value for a particular cervix at each gestational age while the error bars associated with each bar represent the standard deviation for that cervix. The horizontal lines and corresponding error bars associated with each gestational age are the means and standard deviations in cervix attenuation slope for all of the animals at a particular gestational age.

While it is challenging to compare the standard deviations due to the variable number of estimates obtained for each cervix, the variance in the mean cervix attenuation slope estimates between the animals at the same gestational age is larger than the variance between the different gestational ages. Also, the nonpregnant rats have a statistically significant ($p=0.029$) higher attenuation slope (3.2 ± 1.7 dB/cm-MHz) than the pregnant rats (mean over all gestational ages of 1.5 ± 1.3 dB/cm-MHz). However, a statistically significant effect of ultrasonic attenuation slope with gestational age is not observed because the variance in the estimates is too large.

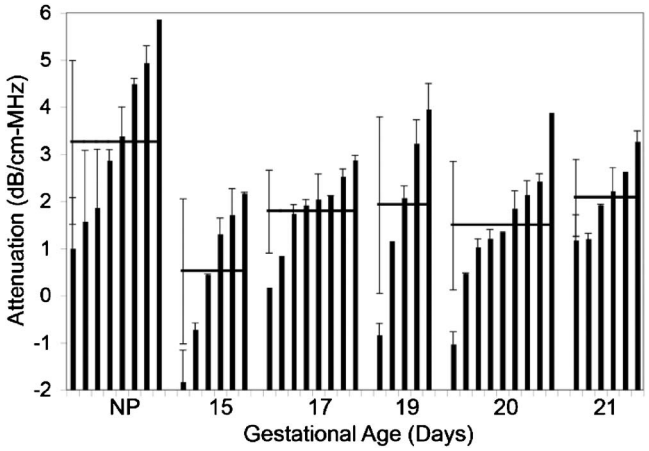


FIG. 6. *In vivo* ultrasound attenuation slope estimates for rat cervix showing the mean and standard deviation for every cervix sample (bars) vs gestational age as well as the mean and standard deviation for all of the cervices at each gestational age (horizontal lines).

VI. DISCUSSION

Our previous study suggested that cervical ripening and consequently preterm delivery could be predicted based on the ultrasonic attenuation of the cervix (McFarlin *et al.*, 2006). The goal of this study was to modify a traditional algorithm for estimating attenuation slope *in vivo*, so that it could be used to assess cervical ripening. The algorithm was modified by compensating for focusing by assuming the field along the beam axis in the focal region could be approximated by a Gaussian function and consequently multiplying the received spectra by an appropriate term. The validity of the modified algorithm was demonstrated via computer simulations and an *ex vivo* tissue experiment prior to being applied to *in vivo* experiments on rat cervices. In the simulations, the accuracy of the modified algorithm was better than 15%, and the attenuation slope of the cervix in the *ex vivo* experiment was found to be 2.6 ± 0.6 dB/cm-MHz, which is in good agreement with 2.5 ± 0.4 dB/cm-MHz estimated using an insertion loss technique.

For the *in vivo* experiments, the attenuation slope of the cervix for the nonpregnant rats was found to be approximately twice as much as the cervix for the pregnant rats on average (i.e., mean of 3.2 dB/cm-MHz versus mean of 1.5 dB/cm-MHz). However, the variance of the estimates was too large to deduce a statistically significant trend in the attenuation slope for the different gestational ages. The variance could probably be reduced by using a different animal model with a larger cervix so that more data could be used to estimate the amplitude spectrum of the backscattered signal. For example, the standard deviation for the attenuation slope for all of the rats at a gestational age of 20 days shown in Fig. 6 is 1.36 dB/cm-MHz. If we restrict our attention to only the three rats where more than 3 mm of the cervix could be scanned, the standard deviation between the different animals is 0.70 dB/cm-MHz.

The cause for the large variances is most likely the natural variation in attenuation for biological tissue between animals as is indicated by Fig. 6. Also, if we return to the three rats at a gestational age of 20 days where more than 3 mm of the cervix could be scanned, the standard deviation between the animals was 0.70 dB/cm-MHz, whereas the standard deviation for a particular animal was only 0.24 dB/cm-MHz on average. The variance in ultrasonic attenuation for tissue of the same type has been observed in other studies as well, e.g., glycogen in liver (Parker *et al.*, 1988; Tuthill *et al.*, 1989). Also, 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 dB/cm 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. Fortunately, the variance of an attenuation slope estimate of the cervix is manageable when considering a single animal with a sufficient number of backscattered echoes from the cervix. Therefore, longitudinal studies of attenuation slope estimates from individual rats throughout pregnancy are more likely to reveal statistically significant trends that could lead to prediction and detection of cervical ripening.

ACKNOWLEDGMENTS

This project was supported by Grant No. P30 NR009014 Center for Reducing Risks in Vulnerable Populations (CRRVP) from the National Institute of Nursing Research as well as the University of North Dakota School of Engineering and Mines. The project was also supported by Grant No. R01 CA111289. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Nursing Research or the National Institutes of Health.

Baldeweck, T., Herment, A., Laugier, P., and Berger, G. (**1994**). "Attenuation estimation in highly attenuating media using high frequencies: A comparison study between different mean frequency estimators," Proc.-IEEE Ultrason. Symp. **1783**, 1783–1786.

Baldeweck, T., Laugier, P., Herment, A., and Berger, G. (**1993**). "Application of autoregressive spectral analysis for ultrasound attenuation: Interest in highly attenuating medium," Proc.-IEEE Ultrason. Symp. **1182**, 1181–1186.

Baldeweck, T., Laugier, P., Herment, A., and Berger, G. (**1995**). "Application of autoregressive spectral analysis for ultrasound attenuation estimation: Interest in highly attenuating medium," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 99–110.

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.

Callaghan, W. M., MacDorman, M. F., Rasmussen, S. A., Qin, C., and Lackritz, E. M. (**2006**). "The contribution of preterm birth to infant mortality rates in the United States," Pediatrics **118**, 1566–1573.

Clark, K., Ji, H., Feltovich, H., Janowski, J., Carroll, C., and Chien, E. K. (**2006**). "Mifepristone-induced cervical ripening: Structural, biomechanical, and molecular events," Am. J. Obstet. Gynecol. **194**, 1391–1398.

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, National Institutes of Health Project No. A454, Progress Report, Washington, DC.

Feltovich, H., Ji, H., Janowski, J. W., Delance, N. C., Moran, C. C., and Chien, E. K. (**2005**). "Effects of selective and nonselective PGE2 receptor agonists on cervical tensile strength and collagen organization and microstructure in the pregnant rat at term," Am. J. Obstet. Gynecol. **192**, 753–760.

Girault, J. M., Ossant, F., Ouahabi, A., Kouame, D., and Patat, F. (**1998**). "Time-varying autoregressive spectral estimation for ultrasound attenuation in tissue characterization," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 650–659.

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.

Hamilton, B. E., Martin, J. A., and Ventura, S. J. (**2006**). "Births: Preliminary data for 2005," Natl. Vital Stat. Rep. **55**, 1–18.

House, M., and Socrate, S. (**2006**). "The cervix as a biomechanical structure," Ultrasound Obstet. Gynecol. **28**, 745–749.

Hyungsuk, K., and Varghese, T. (**2007**). "Attenuation estimation using spectral cross-correlation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 510–519.

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.

Leppert, P. C. (**1995**). "Anatomy and physiology of cervical ripening," Clin. Obstet. Gynecol. **38**, 267–279.

Leppert, P. C., Kokenyesi, R., Klemenich, C. A., and Fisher, J. (**2000**). "Further evidence of a decorin-collagen interaction in the disruption of cervical collagen fibers during rat gestation," Am. J. Obstet. Gynecol. **182**, 805–812.

Madsen, E. L., Dong, F., Frank, G. R., Garra, B. S., Wear, K. A., Wilson, T., Zagzebski, J. A., Miller, H. L., Shung, K. K., Wang, S. H., Feleppa, E. J., Liu, T., O'Brien, W. D., Jr., Topp, K. A., Sanghvi, N. T., Zaitsev, A. V., Hall, T. J., Fowlkes, J. B., Kripfgans, O. D., and Miller, J. G. (**1999**). "Interlaboratory comparison of ultrasonic backscatter, attenuation, and speed measurements," J. Ultrasound Med. **18**, 615–631.

Mahendroo, M. S., Porter, A., Russell, D. W., and Word, R. A. (**1999**). "The parturition defect steroid 5{alpha}-reductase type 1 knockout mice is due to impaired cervical ripening," Mol. Endocrinol. **13**, 981–992.

McFarlin, B. L., O'Brien, W. D., Jr., Oelze, M. L., Zachary, J. F., and White-Traut, R. C. (**2006**). "Quantitative ultrasound assessment of the rat cervix," J. Ultrasound Med. **25**, 1031–1040.

Narayana, P. A., and Ophir, J. (**1983a**). "A closed form method for the measurement of attenuation in nonlinearly dispersive media," Ultrason. Imaging **5**, 17–21.

Narayana, P. A., and Ophir, J. (**1983b**). "On the validity of the linear approximation in the parametric measurement of attenuation in tissues," Ultrasound Med. Biol. **9**, 357–361.

Oelze, M. L., and O'Brien, W. D., Jr. (**2004**). "Defining optimal axial and lateral resolution for estimating scatterer properties from volumes using ultrasound backscatter," J. Acoust. Soc. Am. **115**, 3226–3234.

Oosterveld, B. J., Thijssen, J. M., Hartman, P. C., Romijn, R. L., and Rosenbusch, G. J. E. (**1991**). "Ultrasound attenuation and texture analysis of diffuse liver disease: Methods and preliminary results," Phys. Med. Biol. **36**, 1039–1064.

Parker, K. J., Tuthill, T. A., and Baggs, R. B. (**1988**). "The role of glycogen and phosphate in ultrasonic attenuation of liver," J. Acoust. Soc. Am. **83**, 374–378.

St. John, E. B, Nelson, K. G., Cliver, S. P., Bishnoi, R. R., and Goldenberg, R. L. (**2000**). "Cost of neonatal care according to gestational age at birth and survival status," Am. J. Obstet. Gynecol. **182**, 170–175.

Straach, K. J., Shelton, J. M., Richardson, J. A., Hascall, V. C., and Mahendroo, M. S. (**2005**). "Regulation of hyaluronan expression during cervical ripening," Glycobiology **15**, 55–65.

Tuthill, T. A., Baggs, R. B., and and Parker, K. J. (**1989**). "Liver glycogen and water storage: Effect on ultrasound attenuation," Ultrasound Med. Biol. **15**, 621–627.

Wilson-Costello, D., Friedman, H., Minich, N., Siner, B., Taylor, G., Schluchter, M., and Hack, M. (**2007**). "Improved neurodevelopmental outcomes for extremely low birth weight infants in 2000-2002," Pediatrics **119**, 37–45.

Erratum: "Supporting evidence for reverse cochlear traveling waves" [J. Acoust. Soc. Am. 123 (1), 222–240 (2008)]

W. Dong and E. S. Olson
Department of Otolaryngology, Head and Neck Surgery, Columbia University, P & S 11-452, 630 West 168th Street, New York, New York 10032, USA

(Received 6 September 2007; revised 25 October 2007; accepted 27 October 2007)
[DOI: 10.1121/1.2899934]

PACS number(s): 43.64.Kc, 43.64.Jb, 43.64.Bt, 43.10.Vx

Several paragraphs of text were omitted from the original online and print publication of this article due to a publisher's error. The online version of this article has been corrected. The omitted text should have appeared in Section V. Discussion, subsection C. DP spatial variation. The heading for section V. Summary was also omitted. The entire corrected section reads as follows:

C. DP spatial variation

In Sec. IV B 2, we show and attempt to understand paradoxical results at frequencies around the BF in which, when a DP was related to a DPOAE via the phase of the two responses, the DPOAE appeared to precede the DP that is expected to be its precursor. We further showed that by looking at the DPOAE–DP phase in specific interpretable frequency regions, a reverse wave was in evidence. However, given the obvious pitfalls in relating phase responses, it is useful to be able to explore the question of reverse-traveling waves from a completely different vantage point. Spatial pressure variations provide this.

As introduced in Figs. 1(B) and 1(C), the intracochlear response to sound comprises two modes: the traveling-wave and compression wave. Theoretically, the pressure distributions of these two pressure modes within the cochlea are strikingly different. At frequencies in the region of the BF, the traveling-wave pressure, being governed by BM width and traveling-wave wavelength, is expected to vary over distances of tens of micrometers (Steele and Taber, 1979b; Andoh and Wada, 2004; Yoon *et al.*, 2006). At the same frequencies the compression pressure will vary over distances corresponding to quarter sound wavelengths—centimeters (Peterson and Bogert, 1950). The compression pressure can be thought of as a time varying, approximately space filling pressure. Experimentally, these two pressure modes have been detected in the spatial variation of pressure with single-tone stimulation, where it was found that the slow traveling wave dominated close to the BM, $\sim$100 μm from it the compression pressure dominated, producing a spatially invariant pressure (Olson, 1998, 1999, 2001; Dong and Olson, 2005b) in amplitude and phase. Therefore, a marked advantage of pressure measurements is that they can detect both the compression and slow-traveling-wave pressure modes. It has been proposed that DPs travel from their place of generation to the stapes through the cochlear fluid as a compression pressure, another possibility is that they are transmitted via a reverse cochlear traveling wave. The spatial pressure variation of the DPs is a criterion to distinguish the two possibilities: a rapid spatial variation argues for the reverse-traveling-wave mode, whereas if the DP pressure is spatially unvarying, the compression mode is supported. The results in Fig. 6 and results in Dong and Olson (2005b) showed that the DP pressure dropped sharply with distance from the BM. Therefore, the DP pressure distribution within the cochlea appears to be dominated by the traveling-wave pressure mode.

Given the observation that the DP pressure falls into the noise with distance from the BM, the DP compression pressure plateau, if one exists, must be less than 60 dB SPL in the cochlea (the sensor noise floor). Therefore, with the 40 dB loss from the middle-ear reverse transmission (Dong and Olson, 2006), the contribution to the DPOAE from the compression mode would be at most 20 dB SPL, which is much lower than the actual DPOAE we have measured in the EC. In summary, the spatial variation measurement does not rule out a contribution by a compression wave to the DPOAE, but does set the maximum limit of its contribution to DPOAE. Our results suggest that the reverse-traveling wave plays the more important role in DP reverse transmission.

 0001-4966/2008/123(3)/1801/1/$23.00